AN INVESTIGATION OF PROXIMITY CATCH DIGRAPHS IN DELAUNAY TESSELLATIONS

by

Elvan Ceyhan

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

October 2004

Abstract

In this dissertation, we introduce proximity catch digraphs (PCDs) based on two sets of points, $\mathcal{X}_n$ and $\mathcal{Y}_m$, from classes $\mathcal{X}$ and $\mathcal{Y}$, respectively, in $\mathbb{R}^d$ with $d > 1$, investigate their properties, and present examples in $\mathbb{R}^2$ for illustrative purposes. PCDs are a generalization of the class cover catch digraphs (CCCDs) introduced by Priebe et al. in [35]; a slight modification of the CCCD yields a special case of PCDs. The PCDs are constructed based on the relative positions of members of, say, class $\mathcal{X}$ points, with respect to the Delaunay tessellation of class $\mathcal{Y}$ points. Our main motivation for introducing PCDs is that a direct extension of the CCCD to multidimensional data lacks mathematical tractability of the distribution of the domination number, moments of relative density, and geometry invariance for uniform $\mathcal{X}_n$ in the convex hull of $\mathcal{Y}_m$. We investigate two major aspects of the PCDs, namely, the distribution of the domination number and the relative density. In $\mathbb{R}^2$, we analyze these concepts for $\mathcal{X}_n$ in one triangle (formed by $\mathcal{Y}_3$) and then generalize the analyses for $\mathcal{X}_n$ in multiple triangles from the Delaunay triangulation (assumed to exist) of $\mathcal{Y}_m$ with $m > 3$. Our PCDs make tractable the mathematics in multiple dimensions, thereby enhancing the applicability of the methodology to statistical hypothesis testing and pattern classification. We compute the asymptotic distribution of the domination number of one of the PCDs. Furthermore, the relative density of the PCDs is shown to be a U-statistic which avails the asymptotic normality of the relative density. The domination number and relative density are both used to test a type of complete spatial randomness against spatial point patterns of segregation and association. The power of the tests is investigated by using asymptotic efficacy methods such as Pitman asymptotic efficacy, Hodges-Lehman asymptotic efficacy, asymptotic power function analysis. Finite sample power is analyzed by Monte Carlo simulations. The methods are illustrated in the two dimensional case, but are applicable to higher dimensions as well as to other types of proximity maps.

Advisor: Carey E. Priebe

Readers: Carey E. Priebe and David J. Marchette

Acknowledgments

First of all, I would like to express my gratitute and thanks to Carey E. Priebe for his wonderful advising. His endurance, cheerful spirit, and perseverence were invaluable throughout this process. He has been supportive in every way an advisor could be, and for that I can not really thank him enough. A computational geometry problem that he has brought up to my attention has blossomed into this dissertation. I would also like to thank the second reader David J. Marchette and my other defense committee members John C. Wierman, Daniel Q. Naiman, and Shih-Ping Han whose time and suggestions improved this thesis.

Special thanks to Professors Jim Fill and John C. Wierman for their enlightening discussions and insightful comments and Prof. Goldman for his wisdom. Prof. Ed Scheinerman has been a great department head during my education at Applied Mathematics and Statistics (formerly Mathematical Sciences Department). I also need to thank the staff, especially, Mrs. Lutz, Berdan, Kirt, and Bechtel, without whom this place would not exist, for everything they have done for me.

Thanks to my fellow graduate students for making life more livable and enjoyable. Thank you, Leslie Cope, for making your experience available any time I asked.

Finally, I wish to thank my brother Hacı Ahmet for his continuous support and my family abroad for providing me emotional and spiritual shelter in good and bad times. To them I owe much.

Elvan Ceyhan

Contents

Abstract ii

Acknowledgments iii

Contents iv

List of Figures x

I Preliminaries 1

1 Introduction 2

1.1 Overview of the Proximity Maps and the Associated Digraphs 2

1.1.1 Data-random Proximity Catch Digraphs 4

1.2 Chapter Overview 7

2 Preliminary Tools and Foundation 8

2.1 Voronoi Diagrams and Delaunay Tessellations 8

2.1.1 Poisson Delaunay Triangles 10

2.2 Transformations Preserving Uniformity on Triangles in $\mathbb{R}^2$ 12

2.2.1 Transformation of T_b to T_e 13

2.3 Triangle Centers 14

2.4 The Spatial Patterns of Segregation and Association 16

2.5 U-Statistics 19

2.6 Consistency of Tests 22

2.7 Comparison of Tests 23

2.7.1 Asymptotic Power Function 23

2.7.2 Asymptotic Relative Efficiency 24

2.7.2.1 Pitman Asymptotic Efficacy 26
2.7.2.2 Hodges-Lehmann Asymptotic Efficacy 27

II Theory and Applications **29**

3 Proximity Maps and the Associated Γ_1-Regions **30**

3.1 Preliminaries and Foundation 30
3.2 Vertex and Edge Regions 33
3.2.1 Vertex Regions 33
3.2.1.1 CC-Vertex Regions 34
3.2.1.2 CM-Vertex Regions 35
3.2.1.3 M-Vertex Regions 36
3.2.2 Edge Regions 36
3.2.2.1 CM-Edge Regions 37
3.3 Proximity Maps in Delaunay Triangles 38
3.3.1 Arc-Slice Proximity Maps 39
3.3.2 r-Factor Proportional-Edge Proximity Maps 40
3.3.2.1 Extension of N_{PE}^r to Higher Dimensions 44
3.3.3 τ-Factor Central Similarity Proximity Maps 45
3.3.3.1 Extension of N_{CS}^τ to Higher Dimensions 48
3.3.4 The Behaviour of Proximity Maps 48
3.4 Γ_1-Regions and the Related Concepts for Proximity Maps 51
3.4.1 Γ_1-Regions for Arc-Slice Proximity Maps 53
3.4.2 Γ_1-Regions for r-Factor Proportional-Edge Proximity Maps 57
3.4.3 Γ_1-Regions for τ-Factor Central Similarity Proximity Maps 63
3.4.4 Characterization of Proximity Maps Using η-Values 65
3.4.5 The Behaviour of $\Gamma_1(\mathcal{X}_n, N)$ for the Proximity Maps in $T(\mathcal{Y})$ 67
3.4.6 Expected Area of Γ_1-Regions Associated with $N(\cdot)$ 70
3.4.6.1 Expected Area of $\Gamma_1(\mathcal{X}_n, N_{AS}, M)$ 71
3.4.6.2 The Limit of Expected Area of $\Gamma_1(\mathcal{X}_n, N, M)$ for N_{PE}^r and N_{CS}^τ 73
3.5 κ-Values for the Proximity Maps in $T(\mathcal{Y})$ 75

3.5.1 Characterization of Proximity Maps Using κ-Values 78

4 Distribution of the Domination Number of Proximity Catch Digraphs in Delaunay Tessellations 80

4.1 Minimum Dominating Sets and Domination Number of Digraphs 80

4.2 The Asymptotic Distribution of Domination Number of the CCCD for Uniform Data on Compact Intervals in $\mathbb{R}$ 81

4.3 The Asymptotic Distribution of $\gamma(\mathcal{X}_n, N_{AS}, M)$ 83

4.4 The Asymptotic Distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ 93

4.5 Summary of the Asymptotic Distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ 101

4.6 The Use of the Domination Number $\gamma(\mathcal{X}_n, N_{PE}^r, M_C)$ for Testing Spatial Patterns of Segregation and Association 103

4.6.1 Null Distribution of Domination Number 103

4.6.2 The Null Distribution of Mean Domination Number in the Multiple Triangle Case 104

4.6.3 Segregation and Association Alternatives 106

4.6.4 Monte Carlo Power Analysis 108

4.6.5 Asymptotic Efficacy Analysis 108

4.7 The Asymptotic Distribution of $\gamma\left(\mathcal{X}_n, N_{CS}^r, M\right)$ 109

4.7.1 Summary of Asymptotic Distribution of $\gamma(\mathcal{X}_n, N)$ 114

5 Relative Density of Proximity Catch Digraphs 116

5.1 Preliminaries and Foundation 116

5.2 Relative Density of the CCCD Based on N_S and Uniform Data on Compact Intervals in $\mathbb{R}$ 119

5.3 Relative Density of Proximity Catch Digraphs Based on Arc-Slice Proximity Maps 121

5.4 Relative Density of Random Proximity Catch Digraphs for Testing Spatial Patterns of Segregation and Association 124

5.4.1 Relative Density of Proximity Catch Digraphs Based in r-Factor Proportional-Edge Proximity Maps 124

5.4.2 Asymptotic Normality Under the Null Hypothesis 124

5.4.3 Asymptotic Normality Under the Alternatives 128
5.4.4 The Test and the Analysis 130
5.4.5 Consistency 130
5.4.6 Monte Carlo Power Analysis Under Segregation 131
5.4.7 Monte Carlo Power Analysis Under Association 136
5.4.8 Pitman Asymptotic Efficacy 143
5.4.8.1 Pitman Asymptotic Efficacy Under Segregation Alternatives . 143
5.4.8.2 Pitman Asymptotic Efficacy Under Association Alternatives . 146
5.4.9 Hodges-Lehmann Asymptotic Efficacy 149
5.4.9.1 Hodges-Lehmann Asymptotic Efficacy Under Segregation Alternatives 149
5.4.9.2 Hodges-Lehmann Asymptotic Efficacy Under Association Alternatives 154
5.4.10 Asymptotic Power Function Analysis 160
5.4.10.1 Asymptotic Power Function Analysis Under Segregation 160
5.4.10.2 Asymptotic Power Function Analysis Under Association 162
5.4.11 Multiple Triangle Case 163
5.4.11.1 Related Test Statistics in Multiple Triangle Case 167
5.4.11.2 Asymptotic Efficacy Analysis for $J > 1$ 169
5.5 Relative Density of Random τ-Factor Central Similarity Proximity Catch Digraphs 172
5.5.1 Asymptotic Normality Under the Null Hypothesis 172
5.5.2 Asymptotic Normality Under the Alternatives 176
5.5.3 The Test and Analysis 177
5.5.4 Consistency 178
5.5.5 Monte Carlo Power Analysis Under Segregation 178
5.5.6 Monte Carlo Power Analysis Under Association 183
5.5.7 Pitman Asymptotic Efficacy 191
5.5.7.1 Pitman Asymptotic Efficacy Under Segregation Alternatives . 191
5.5.7.2 Pitman Asymptotic Efficacy Under Association Alternatives . 193
5.5.8 Multiple Triangle Case 195

5.5.8.1 Asymptotic Efficacy Analysis for $J > 1$ 197

6 Conclusion **200**

III Appendices and Bibliography **205**

A Proofs of Some of the Theorems in Chapters 3 and 4 **206**

A.1 Proof of Theorem 3.4.30 206

A.1.1 Proof of Theorem 3.4.30 206

A.2 Proofs of Some of the Theorems in Chapter 4 209

A.2.1 Proof of Proposition 4.3.4 209

A.2.2 Proof of Theorem 4.3.5 211

A.2.3 Proof of Theorem 4.4.8 213

A.2.4 Proof of Theorem 4.4.9 215

A.2.5 Proof of Theorem 4.4.10 218

A.2.6 Proof of Theorem 4.4.11 223

A.2.7 Proof of Theorem 4.4.12 225

A.2.8 Proof of Theorem 4.7.4 231

A.2.9 Proof of Theorem 4.7.6 235

A.2.10 Proof of Theorem 4.7.7 237

A.2.11 Proof of Theorem 4.7.8 239

B Derivations of Mean and Variance Terms in Section 5.4.1 **242**

B.1 Derivation of $\mu(N_{PE}^r)$ in Theorem 5.4.2 242

B.2 Derivation of $\nu(N_{PE}^r)$ in Theorem 5.4.2 243

B.3 Derivation of $\mu(N_{PE}^r, \varepsilon)$ 258

B.3.1 Derivation of $\mu_S(N_{PE}^r, \varepsilon)$ and $\nu_S(N_{PE}^r, \varepsilon)$ for Segregation with $\varepsilon = \sqrt{3}/8$ 262

B.4 The Mean $\mu(N_{PE}^r, \varepsilon)$ Under Segregation and Association Alternatives 270

B.4.1 $\mu_S(N_{PE}^r, \varepsilon)$ Under Segregation Alternatives 271

B.4.2 $\mu_A\big(N_{PE}^r, \varepsilon\big)$ Under Association Alternatives 273

C Derivations of Means and Variances in Section 5.5 **276**

C.1 Derivation of $\nu\big(N_{CS}^r\big)$ 276

C.2 Derivation of $\mu(N_{CS}^{\tau}, \varepsilon)$ 279
C.3 The Mean $\mu(N_{CS}^{\tau}, \varepsilon)$ Under Segregation and Association Alternatives 282
C.3.1 $\mu_S(N_{CS}^{\tau}, \varepsilon)$ Under Segregation Alternatives 282
C.3.2 $\mu_A(N_{CS}^{\tau}, \varepsilon)$ Under Association Alternatives 285

Bibliography **288**

Vita **292**

List of Figures

2.1.1 The Voronoi diagram based on a realization of 10 Y points (left) and the associated Delaunay triangulation (right). 9

2.1.2 A realization of 200 X points and the Delaunay triangulation based on 10 Y points in Figure 2.1.1. 10

2.2.1 The description of $\phi_e(x, y)$ for $(x, y) \in T_b$ (left) and the equilateral triangle $\phi_e(T_b) = T_e$ (right). 14

2.3.1 The circumcenter, circumcircle, and circumradius of an acute triangle (top), an obtuse triangle (bottom). 17

2.3.2 The incircle, incenter, inradius of a triangle (left), and the centroid (center of mass) of a triangle (right). 18

3.2.1 The CC-vertex regions in an acute triangle (left) and in an obtuse triangle (right). 34

3.2.2 The CM-vertex regions with median lines. 35

3.2.3 CM-edge regions $R_{CM}(e_j)$, $j \in \{1, 2, 3\}$. 38

3.3.1 $N_{AS}(x, M_{CC})$ with an $x \in R_{CC}(\mathsf{y}_2)$ (left) and the superset region $\mathscr{R}_S(N_{AS}, M_C)$ in $T(\mathcal{Y})$ (right). 39

3.3.2 Construction of r-factor proximity region, $N_{PE}^2(x)$ (shaded region). 41

3.3.3 Superset region $\mathscr{R}_S\left(N_{PE}^2, M_{CC}\right)$ in an acute triangle 43

3.3.4 The triangle $\mathscr{T}^{r=\sqrt{2}}$. 44

3.3.5 Construction of τ-factor central similarity proximity region, $N_{CS}^{\tau=1/2}(x, M_C)$ (shaded region). 46

3.3.6 $N_{CS}^{\tau=1}(x, M)$ with an $x \in R_M(e_3)$ (left); $N_{CS}^{\tau=1}(x, M_C)$ with an $x \in R_{M_C}(e_3)$ (right) 47

3.3.7 The figure for x, $y \in R_M(e_3)$ described in Theorem 3.3.8. 50

3.4.1 $\Gamma_1(x, N_{AS}, M_{CC})$ and $x \in R_{CC}(\mathsf{y}_2)$ (left); the ball $B(M_2, |e_2|/2)$ that divides the region $R_{CC}(\mathsf{y}_2)$ for which Γ_1-region is a hexagon or a pentagon (right). . . . 54

3.4.2 An empirical Γ_1-region, $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC})$ with $n = 10$ 55

3.4.3 Figure for $X_{\mathrm{y}_1}(n) = X_{\mathrm{y}_2}(n) = (x, y)$ (left) and B_j for $j \in \{1, 2, 3\}$ for distinct vertex extrema (right). 56

3.4.4 Construction of the Γ_1-region, $\Gamma_1(x, N_{PE}^{r=2}, M_C)$ (shaded region). 58

3.4.5 The figure for $X_{e_1}(n) = X_{e_2}(n) = (x, y)$ 61

3.4.6 Examples of the four cases of the Γ_1-region, $\Gamma_1\left(x, N_{CS}^{\tau=1/2}, M_C\right)$ with four distinct $x \in R_{CM}(e_3)$ (shaded regions). 64

3.4.7 The Γ_1-region $\Gamma_1\left(x, N_{CS}^{\tau}, M_C\right)$ with $x \in R_M(e_2)$ (left) and $\Gamma_1\left(X_n, N_{CS}^{\tau}, M_C\right)$ with $n > 1$ (right). 65

3.4.8 The smallest arc-circle $S_C(T(\mathcal{Y}))$ that contains $T(\mathcal{Y})$ 72

3.5.1 A vertex with inner angle equal to $\pi/3$ (left), the domain of α and β for which $\kappa(N_{AS}) = 4$ and $\kappa(N_{AS}) = 5$ (right). 76

3.5.2 The figure for $\kappa\left(N_{CS}^{\tau}\right) = n$. 79

4.3.1 A realization of $H(\mathcal{X}_n, \mathcal{Y})$ based on a given set of vertex extrema $X_{\mathrm{y}_j}(n) = x_{\mathrm{y}_j}$ from $\mathcal{X}_n$. 85

4.3.2 A figure for the description of the event $E_n^{\varepsilon}(\mathcal{Y})$ (left) and the pdf of $X_{\mathrm{y}}(n)$ (right) given three distinct vertex extrema $X_{\mathrm{y}_j}(n) = x_{\mathrm{y}_j}$ for $j \in \{1, 2, 3\}$. 87

4.3.3 The empirical estimates of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ versus various n values for T_e (left) and for $T_b = \big((0,0), (1,0), (1/2, 1/2)\big)$ (right). 93

4.3.4 The empirical estimates of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ versus various n values for $T_b = \big((0,0), (1,0), (1/2, 1/4)\big)$. 94

4.4.1 The empirical estimates of $P\left(\gamma_n = k\right)$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right)$ versus various n values with $r = 2$ and $M = M_C$. 96

4.4.2 The empirical estimates of $P\left(\gamma_n = k\right)$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right)$ versus various n values with $r = 5/4$ and $M = M_C$. 98

4.4.3 The empirical estimates of $P\left(\gamma_n = k\right)$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right)$ versus various n values with $r = 5/4$ and $M = \big(3/5, \sqrt{3}/10\big)$. 99

4.4.4 Plotted is the probability $p_r = \lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right) = 2\right)$ as a function of r for $r \in [1, 3/2)$ and $M \in \{t_1, t_2, t_3\}$. 100

4.4.5 The empirical estimates of $P\left(\gamma_n = k\right)$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right)$ versus various n values with $r = 5/4$ and $M = \big(7/10, \sqrt{3}/10\big)$. 101

4.4.6 The empirical values of $P\left(\gamma_n = k\right)$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^{r}, M_C\right)$ for various n. . 102

4.5.1 The empirical values of $\gamma_n = k$ versus various n values where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^{4/3}, M_C\right)$. 103

4.6.1 Realizations of segregation (left), null case (middle), and association (right) for $|\mathcal{Y}| = 10$, $J = 13$, and $n = 1000$. 104

4.6.2 Depicted are $\overline{G}_J \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu \approx 2.2587, \sigma^2/J \approx .1917/J\right)$ for $J = 13$, $n = 100$ (left); $J = 13$, $n = 1000$ (middle); and $J = 30$, $n = 1000$ (right). Histograms are based on 1000 Monte Carlo replicates and the curves are the associated asymptotic normal curves. 105

4.6.3 Depicted are $\overline{G}_J \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu \approx 2.2587, \sigma^2/J \approx .1917/J\right)$ for $J = 13$, $n = 1000$ (left); $J = 30$, $n = 3000$ (middle); and $J = 30$, $n = 5000$ (right). Histograms are based on 1000 Monte Carlo replicates and the curves are the associated asymptotic normal curves. 105

4.6.4 Two Monte Carlo experiments against the segregation alternatives $H^S_{\sqrt{3}/8}$ i.e., $\delta = 1/16$. Depicted are kernel density estimates of $\overline{G}_J$ for $J = 13$ and $n = 1000$ with 1000 Monte Carlo replicates (left) and $J = 30$ and $n = 5000$ with 1000 Monte Carlo replicates (right) under the null (solid) and the segregation alternative (dashed). 109

4.6.5 Two Monte Carlo experiments against the association alternatives $H^A_{\sqrt{3}/21}$, i.e., $\delta = 16/49$. Depicted are kernel density estimates of $\overline{G}_J$ for $J = 13$ and $n = 1000$ with 1000 Monte Carlo replicates (left) and $J = 30$ and $n = 5000$ with 100 Monte Carlo replicates (right) under the null (solid) and the association alternative (dashed). 110

4.7.1 The empirical estimates of $P(\gamma_n = k)$ for $k = 1, 2$ and $k > 2$ versus various n values with $\tau = 1$ and $M = M_C$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)$. 113

4.7.2 The empirical estimates of $P(\gamma_n = k)$ versus various n values with $\tau = 1$ and $M = M_C$ where $\gamma_n = \gamma_{ind}\left(\mathcal{X}_n \cap T_s, N_{CS}^{\tau}, M\right)$. 114

4.7.3 The empirical estimates of $P(\gamma_n = k)$ versus various n values with $\tau = 1$ and $M = M_C$ where $\gamma_n = \gamma_{ind}\left(\mathcal{X}_n^{''}, N_{CS}^{\tau}, M\right)$. 115

5.3.1 The two regions in T_s for which $\Gamma_1(x, N_{AS})$ is a pentagon or a hexagon. 122

5.3.2 A sample figure for pentagonal $\Gamma_1(x, N_{AS})$ (left) and a hexagonal $\Gamma_1(x, N_{AS})$ (right). 122

5.4.1 Asymptotic null mean $\mu(N_{PE}^r)$ (left) and variance $\nu(N_{PE}^r)$ (right) from Theorem 5.4.2 for $r \in [1, 5]$. The vertical lines indicate the endpoints of the intervals in the piecewise definition of the functions. Notice that the vertical axes are differently scaled. 126

5.4.2 The plot of $\omega(N_{PE}^r) = \mathbf{Var}\left[h_{12}(N_{PE}^r)\right]$ as a function of r for $r \in [1, 5]$. 127

5.4.3 Depicted are the distributions of $\rho_n(N_{PE}^2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{5}{8}, \frac{25}{192\,n}\right)$ for $10, 20, 100$ (left to right). Histograms are based on 1000 Monte Carlo replicates. Solid curves represent the approximating normal densities given in Theorem 5.4.2. Note that the vertical axes are differently scaled. 128

5.4.4 Depicted are the histograms for 10000 Monte Carlo replicates of $\rho_{10}(N_{PE}^1)$ (left) and $\rho_{10}(N_{PE}^5)$ (right) indicating severe small sample skewness for extreme values of r. 129

5.4.5 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/8}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2,\ 3\}$ (left-to-right). 132

5.4.6 Two Monte Carlo experiments against the segregation alternative $H^S_{\sqrt{3}/8}$. Depicted are kernel density estimates for $\rho_n(N_{PE}^{11/10})$ for $n = 10$ (left) and $n = 100$ (right) under the null (solid) and alternative (dashed). 133

5.4.7 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/4}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2\}$ (left-to-right). 134

5.4.8 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{2\sqrt{3}/7}$ (dashed) for $r \in \{1,\ 21/20,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2}\}$ (left-to-right). 135

5.4.9 Monte Carlo power using the empirical critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of r for $n = 10$. 136

5.4.10 Monte Carlo power using the asymptotic critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of r for $n = 10$. The circles represent the empirical significance levels while triangles represent the empirical power values. 136

5.4.11 Kernel density estimates for the null (solid) and the association alternative $H^A_{5\sqrt{3}/24}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2,\ 3,\ 5,\ 10\}$ (left-to-right). 138

5.4.12 Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/12}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2,\ 3,\ 5,\ 10\}$ (left-to-right). 139

5.4.13 Two Monte Carlo experiments against the association alternative $H^A_{\sqrt{3}/12}$. Depicted are kernel density estimates for $\rho_n(N_{PE}^{11/10})$ for $n = 10$ (left) and $n = 100$ (right) under the null (solid) and alternative (dashed). 140

5.4.14 Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/21}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2,\ 3,\ 5,\ 10\}$ (left-to-right). 141

5.4.15 Monte Carlo power using the empirical critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of r for $n = 10$. 142

5.4.16 Monte Carlo power using the asymptotic critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of r for $n = 10$. The circles represent the empirical significance levels while triangles represent the empirical power values. 142

5.4.17 Pitman asymptotic efficacy against segregation (left) and against association (right) as a function of r. 146

5.4.18 Hodges-Lehmann asymptotic efficacy against segregation alternative H^S_ε as a function of r for $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$ (left to right). 154

5.4.19 The mean $\mu_S(N^r_{PE}, \varepsilon)$ (left) and asymptotic variance $\nu_S(N^r_{PE}, \varepsilon)$ (right) as a function of r under segregation with $\varepsilon \in \{0, \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$. 154

5.4.20 Hodges-Lehmann asymptotic efficacy against association alternative H^A_ε as a function of r for $\varepsilon \in \{\sqrt{3}/21,\ \sqrt{3}/12,\ 5\sqrt{3}/24\}$ (left to right). 159

5.4.21 The mean $\mu_A(N^r_{PE}, \varepsilon)$ (left) and asymptotic variance $\nu_A(N^r_{PE}, \varepsilon)$ (right) as a function of r under association with $\varepsilon \in \{0, \sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24\}$. 159

5.4.22 Asymptotic power function against segregation alternative $H^S_{\sqrt{3}/8}$ as a function of r for $n \in \{5, 10, 15, 20, 50, 100\}$. 161

5.4.23 Asymptotic power function against segregation alternative $H^S_{\sqrt{3}/4}$ as a function of r for $n = 3$ (left) and $n = 5$ (right). 161

5.4.24 Asymptotic power function against segregation alternative $H^S_{\sqrt{3}/7}$ as a function of r for $n = 3$ (left) and $n = 5$ (right). 162

5.4.25 Asymptotic power function against association alternative $H^A_{\sqrt{3}/21}$ as a function of r for $n = 5, 10, 100$ 163

5.4.26 Asymptotic power function against association alternative $H^A_{\sqrt{3}/12}$ as a function of r for $n \in \{5, 10, 100\}$ 163

5.4.27 Asymptotic power function against association alternative $H^A_{5\sqrt{3}/24}$ as a function of r for $n \in \{5, 10\}$. 164

5.4.28 Realization of segregation (left), H_0 (middle), and association (right) for $|\mathcal{Y}| = 10$, $J = 13$, and $n = 100$. 164

5.4.29 Monte Carlo power using the asymptotic critical value against $H^S_{\sqrt{3}/8}$, as a function of r, for $n = 100$ (left), $n = 200$ (middle), and $n = 500$ (right) conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. The circles represent the empirical significance levels while triangles represent the empirical power values. 168

5.4.30 Monte Carlo power using the asymptotic critical value against $H^A_{\sqrt{3}/12}$ as a function of r, for $n = 100$ (left), $n = 200$ (middle), and $n = 500$ (right) conditional on the realization of $\mathcal{Y}$ in Figure 4.6.1. The circles represent the empirical significance levels while triangles represent the empirical power values. 168

5.4.31 Pitman asymptotic efficacy against segregation (left) and association (right) as a function of r with the realization of $\mathcal{Y}$ in Figure 5.4.28. Notice that vertical axes are differently scaled. 170

5.4.32 Hodges-Lehmann asymptotic efficacy against segregation alternative H^S_ε as a function of r for $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$ (left to right) conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. 171

5.4.33 Hodges-Lehmann asymptotic efficacy against association alternative H^A_ε as a function of r for $\varepsilon \in \{\sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24\}$ (left to right) conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. 171

5.5.1 The asymptotic null mean $\mu(N^\tau_{CS})$ for $N^\tau_{CS}(\cdot, M)$ (left) and asymptotic null variance $\nu(N^\tau_{CS})$ from Equation 5.5.2 in Theorem 5.5.3 (right). 174

5.5.2 The graph of $\omega(N^\tau_{CS}) = \mathbf{Var}\left[h_{12}(N^\tau_{CS}, M_C)\right]$ as a function of τ. 175

5.5.3 Depicted are $\rho_n\left(N_{CS}^{1/2}\right) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{1}{24}, \frac{19}{2880\,n}\right)$ for 10, 20, 100 (left to right). Histograms are based on 1000 Monte Carlo replicates. Solid curves represent the approximating normal densities given in Theorem 5.4.2. Note that the vertical axes are differently scaled. 176

5.5.4 Depicted are the histograms for 10000 Monte Carlo replicates of $\rho_{10}\left(N_{CS}^{1/4}\right)$ (left), $\rho_{10}\left(N_{CS}^{3/4}\right)$ (middle), and $\rho_{10}\left(N_{CS}^{1}\right)$ (right) indicating severe small sample skewness for small values of τ. 176

5.5.5 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/8}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$, and $N = 10000$ (left-to-right). 179

5.5.6 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/4}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$, and $N = 10000$ (left-to-right). 180

5.5.7 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{2\sqrt{3}/7}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$, and $N = 10000$ (left-to-right). 180

5.5.8 Monte Carlo power using the empirical critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of τ for $n = 10$. 181

5.5.9 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/8}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 100$, and $N = 1000$ (left-to-right). 181

5.5.10 Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/4}$ (dashed) for $\tau = .5$ with $n = 10$ and $N = 10000$ (left) and $n = 100$, $N = 1000$ (right). 182

5.5.11 Monte Carlo power using the empirical critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of τ for $n = 100$. 182

5.5.12 Monte Carlo power using the asymptotic critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of τ for $n = 10$ and $N = 10000$. The circles represent the empirical significance levels while triangles represent the empirical power values. 183

5.5.13 Monte Carlo power using the asymptotic critical value against segregation alternative $H^S_{\sqrt{3}/8}$, as a function of τ for $n = 10$ and $N = 10000$ (left), $n = 20$ and $N = 10000$ (middle), and $n = 100$ and $N = 1000$ (right). The circles represent the empirical significance levels while triangles represent the empirical power values. 184

5.5.14 Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/21}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ with $n = 10$ and $N = 10000$ (left-to-right). 185

5.5.15 Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/12}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ (left-to-right). 186

5.5.16 Kernel density estimates for the null (solid) and the association alternative $H^A_{5\sqrt{3}/24}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ (left-to-right). 186

5.5.17 Monte Carlo power using the empirical critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of τ with $n = 10$ and $N = 10000$. 187

5.5.18 Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/21}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ with $n = 100$, $N = 1000$ (left-to-right). 187

5.5.19 Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/12}$ (dashed) for $\tau = .5$ with $n = 10$ and $N = 10000$ (left) and $n = 100$, $N = 1000$ (right). 188

5.5.20 Monte Carlo power using the empirical critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of τ with $n = 100$ and $N = 1000$. 188

5.5.21 Monte Carlo power using the asymptotic critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of τ for $n = 10$. The circles represent the empirical significance levels while triangles represent the empirical power values. 189

5.5.22 Monte Carlo power using the asymptotic critical value against segregation alternative $H^A_{\sqrt{3}/12}$, as a function of τ for $n = 10$ and $N = 10000$ (left), $n = 20$ and $N = 10000$ (middle), and $n = 100$, $N = 1000$ (right). The circles represent the empirical significance levels while triangles represent the empirical power values. 190

5.5.23 Pitman asymptotic efficacy against segregation (left) and against association (right) as a function of τ. 192

5.5.24 Monte Carlo power using the asymptotic critical value against $H^S_{\sqrt{3}/8}$ (left), $H^A_{\sqrt{3}/12}$ (right) as a function of τ for $n = 100$ conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. The circles represent the empirical significance levels while triangles represent the empirical power values. 197

5.5.25 Monte Carlo power using the asymptotic critical value against $H^S_{\sqrt{3}/8}$ (left), $H^A_{\sqrt{3}/12}$ (right) as a function of τ for $n = 500$ conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. The circles represent the empirical significance levels while triangles represent the empirical power values. 198

5.5.26 Pitman asymptotic efficacy against segregation (left) and association (right) as a function of τ with $J = 13$. Notice that vertical axes are differently scaled. . . . 199

A.1.1 The shaded regions are the triangular $\Gamma_1\left(x_{e_j}, N_{PE}^{3/2}, M_C\right) \cap R_{CM}(\mathsf{y}_j)$ regions for $j \in \{1,2,3\}$ (left); the figure for the description of the pdf of $X_{e_j}(n)$, the shaded region is $T\left(\vec{\zeta}\right)$ (right) given $X_{e_j}(n) = x_{e_j} = (x_j, y_j)$ for $j \in \{1,2,3\}$. 207

A.2.1 A figure for the description of the ball $B(M_{CC}, \varepsilon)$. 210

A.2.2 The vertex regions $R_{CC}(\mathsf{y})$ in an obtuse triangle. 211

A.2.3 A figure for the description of the pdf of $X_3^f(n)$ with a given $X_3^f(n) = x_3^f$. 212

A.2.4 The explanatory figure for $N_1((x_1, y_1))$ for $(x_1, y_1) \in R_{CC}(\mathsf{y}_1)$ (left); the graph of $f(x)$ in Equation A.2.1 multiplied by $A(R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2))$ (right). 212

A.2.5 A figure for the description of the pdf of $X_{e_3}(n)$ (left) and $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M)$ (right) given $X_{e_3}(n) = x_{e_3} = (x, y)$. 214

A.2.6 A figure for the description of the pdf of $\hat{Q}_1(n)$ and $\hat{Q}_3(n)$ (left) and the unshaded region is $N_{PE}^r(\hat{q}_1, M) \cup N_{PE}^r(\hat{q}_3, M)$ (right). 215

A.2.7 A figure for the description of the pdf of $\hat{Q}_1(n)$ and $\hat{Q}_3(n)$ (left) and the unshaded region is $N_{PE}^r(\hat{q}_1, M) \cup N_{PE}^r(\hat{q}_3, M)$ (right) given $\hat{Q}_j(n) = \hat{q}_j$ for $j \in \{1, 3\}$. . . . 218

A.2.8 A figure for the description of the pdf of $\hat{Q}_2(n)$ and $\hat{Q}_3(n)$ (left) and unshaded region is $N_{PE}^r(\hat{q}_2) \cup N_{PE}^r(\hat{q}_3)$ (right) given $\hat{Q}_j(n) = \hat{q}_j$ for $j \in \{2, 3\}$. 221

A.2.9 A figure for the description of the pdf of $\tilde{Q}_2(n)$ given $\tilde{Q}_2(n) = \tilde{q}_2$. The shaded region is $T(\tilde{q}_2)$ (left). A figure for the description of the joint pdf of $\tilde{Q}_1(n), \tilde{Q}_2(n)$ given $\tilde{Q}_j(n) = \tilde{q}_j$ for $j \in \{1, 2\}$, and the event $E_n^{1,2}$. The unshaded region is $N_{PE}^{3/2}(\tilde{q}_1) \cup N_{PE}^{3/2}(\tilde{q}_2) \setminus [T(\tilde{q}_1) \cup T(\tilde{q}_2)]$ (right). 224

A.2.10 A figure for the description of the pdf of $\tilde{Q}_1(n), \tilde{Q}_2(n), \tilde{Q}_3(n)$ and the event $E_n^{1,2} \cap E_n^{1,3}$. The unshaded region is $N_{PE}^{3/2}(\tilde{q}_1) \cup [N_{PE}^{3/2}(\tilde{q}_2) \cap N_{PE}^{3/2}(\tilde{q}_2)] \setminus [T(\tilde{q}_1) \cup T(\tilde{q}_2)]$ (left). A figure for the description of the event $E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}$. The unshaded region is $([N_{PE}^{3/2}(\tilde{q}_1) \cup N_{PE}^{3/2}(\tilde{q}_2)] \cap [N_{PE}^{3/2}(\tilde{q}_1) \cup N_{PE}^{3/2}(\tilde{q}_3)] \cap [N_{PE}^{3/2}(\tilde{q}_2) \cup N_{PE}^{3/2}(\tilde{q}_3)]) \setminus [T(\tilde{q}_1) \cup T(\tilde{q}_2) \cup T(\tilde{q}_3)]$ (right). 228

A.2.11 A figure for the description of the event $E_1(n, \varepsilon)$ (left) and $E_2(n, \varepsilon)$ (right). The unshaded regions are the corresponding ε-strips around the edges given $X_{e_j}(n) = x_{e_j} = (x_j, y_j)$ for $j \in \{1, 2, 3\}$. 232

A.2.12 The Γ_1-region for $N_{CS}^{\tau=1}(\cdot, M)$(left) and $N_{CS}^{\tau=1}(\cdot, M_C)$ (right) in $T(\mathcal{Y})$. 233

A.2.13 A figure for the description of the pdf of $X_{e_3}^f(n)$ (left) and $N_{CS}^{\tau}\left(X_{e_3}^f(n), M_C\right)$ (right) given $X_{e_3}^f(n) = x_{e_3}^f = (x_3, y_3)$. 235

A.2.14 A figure for the description of the joint pdf of $X_{e_1}^f(n), X_{e_3}^f(n)$ (left), and $N_{CS}^{\tau}\left(X_{e_1}^f(n), M_C\right) \cup N_{CS}^{\tau}\left(X_{e_3}^f(n), M_C\right)$ (right) given $X_{e_j}^f(n) = x_{e_1}^f = (x_j, y_j)$ for $j \in \{1, 3\}$. 238

A.2.15 The triangle T_s (left) and a realization of the special extremum Z_M^1 in T_s (right). 240

A.2.16 The explanatory figure for $\widetilde{N_{CS}^{\tau=1}}(Z_M^1)$ (left) and the asymptotically accurate support for the pdf of Z_M^1 (right) given $Z_M^1 = z_M^1 = (x_1, y_1)$. 240

B.1.1 The cases for relative position of $\ell_s(r, x)$ with various r values. 243

B.2.1 The prototypes of the six cases of $\Gamma_1(x_1, N_{PE}^r)$ for $x_1 \in T_s$ for $r \in [1, 4/3)$. . . . 245

B.2.2 The regions corresponding to the prototypes of the six cases for $r \in [1, 4/3)$ with $r = 1.25$. 246

B.2.3 The prototype of the new case for $\Gamma_1(x_1, N_{PE}^r)$ for $x_1 \in T_s$ for $r \in [4/3, 3/2)$. . . 250

B.2.4 The regions corresponding to the six cases for $r \in [4/3, 3/2)$ 250
B.2.5 The regions corresponding to the three cases for $r \in [3/2, 2)$ with $r = 1.65$ 253
B.2.6 The regions corresponding to the two cases for $N^r_{PE}(x_1)$ for $r \in [2, \infty)$ with $r = 2.5$ 257
B.3.1 The support under H^S_ε for $\varepsilon \in (0, \sqrt{3}/4)$ and the two types $\ell_r(x_1, x)$ for $r_1 < r_2$. 258
B.3.2 The partition of T_s for different types of $N^r_{PE}(\cdot, \varepsilon)$ under H^S_ε with $r \in [2, \sqrt{3}/(2\,\varepsilon))$. 259
B.3.3 The regions corresponding to the seven cases for $r \in [1, 4)$ with $r = 1.9$ 265

C.1.1 The regions corresponding to the prototypes of the four cases with $\tau = 1/2$. . . . 277
C.2.1 The prototypes of the four cases of $N^\tau_{CS}(x)$ for four distinct $x \in R_{CM}(e_3)$ (shaded regions). 279
C.2.2 The regions corresponding to the prototypes of the four cases shown in Figure C.2.1. 280

The formulation of a problem is often more essential than its solution which may be merely a matter of mathematical or experimental skill.

Albert Einstein

Part I

Preliminaries

CHAPTER 1

Introduction

The proximity catch digraphs (PCDs) are a special type of proximity graphs. In this chapter, we provide the development of proximity graphs in the literature, with various examples. Then we define the PCDs together with their closest relative in literature, namely, class cover catch digraphs (CCCDs).

1.1 Overview of the Proximity Maps and the Associated Digraphs

Proximity maps and the associated (di)graphs are used in disciplines for which shape and structure are crucial. Examples include computer vision (dot patterns), image analysis, pattern recognition (prototype selection), geography and cartography, visual perception, biology, etc. *Proximity graphs* were first introduced by Touissaint, who called them *relative neighborhood graphs* in [41]. The notion of relative neighborhood graph has been generalized in several directions and all of these graphs are now called proximity graphs. From a mathematical and algorithmic point of view proximity graphs fall under the category of *computational geometry.*

A general definition of proximity graphs is as follows:

Definition 1.1.1. Let V be any finite or infinite set of points in $\mathbb{R}^d$. Each (unordered) pair of points $(p, q) \in V \times V$ is associated with a neighborhood $\mathfrak{N}(p, q) \subseteq \mathbb{R}^d$. Let $\mathfrak{P}$ be a property defined on $\mathfrak{N} = \{\mathfrak{N}(p, q) : (p, q) \in V \times V\}$. A *proximity (or neighborhood) graph* $G_{\mathfrak{N},\mathfrak{P}}(V, E)$ defined by the property $\mathfrak{P}$ is a graph with the set of vertices V and the set of edges E such that $(p, q) \in E$ iff $\mathfrak{N}(p, q)$ satisfies property $\mathfrak{P}$.

Examples of most commonly used proximity graphs are the Delaunay tessellation, the boundary of the convex hull, the Gabriel graph, relative neighborhood graph, Euclidean minimum spanning tree, and sphere of influence graph of a finite data set. See, e.g., [19].

The *Delaunay tessellation* of a finite set of points V, $\mathcal{D}(V)$, is the dual of the Voronoi diagram generated by V. See Section 2.1 for further details.

The *convex hull* of a set $V \subset \mathbb{R}^d$, denoted $C_H(V)$, encircles V as if by a rubber band so that the region inside is $C_H(V)$. More formally, $C_H(V)$ is the intersection of all convex sets (there exists infinitely many of them) that contain V. The boundary of $C_H(V)$ can be viewed as a proximity graph which is also a subgraph of $\mathcal{D}(V)$.

The *Gabriel graph* of V, denoted $GG(V)$, is defined as the graph in which (p, q) is an edge of $GG(V)$ iff the circle centered at the midpoint of the line segment $\overline{pq}$ and with diameter $d(p, q)$, the distance between p and q, does not contain any other points from V.

The *relative neighborhood graph* of V is a prominent representative of the family of graphs which are defined by some sort of neighborliness. For a set of points $V \subset \mathbb{R}^d$, the relative neighborhood graph of V, denoted $RNG(V)$, is a graph with vertex set V and edge set which are exactly the pairs (p, q) of points for which $d(p, q) \leq \min_{v \in V} \max(d(p, v), d(q, v))$. That is, (p, q) is an edge of $RNG(V)$ iff Lune(p, q) does not contain any other points of V, where Lune(p, q) is defined as the intersection of two discs centered at p, q each with radius $d(p, q)$ (see, e.g., [19]).

The *Euclidean minimum spanning tree* of V, denoted $EMST(V)$, is defined as the spanning tree in which the sum of the Euclidean lengths of the edges yield the minimum over all spanning trees with vertex set V.

Note that $EMST(V)$ is a subgraph of $RNG(V)$ which is a subgraph of $GG(V)$ which is subgraph of $\mathcal{D}(V)$ (see [30]).

The *sphere of influence graph* on V, denoted $SIG(V)$, has vertex set V and (p, q) as an edge iff the circles centered at p and q with radii $\min_{v \in V \setminus \{p\}} d(p, v)$ and $\min_{v \in V \setminus \{q\}} d(q, v)$, respectively, have nonempty intersection.

In the examples above, $d(x, y)$, can be any distance in $\mathbb{R}^d$. Furthermore, the distance between a point x and a set A is defined as $d(x, A) := \inf_{y \in A} d(x, y)$, and the distance between two sets A and B is defined as $d(A, B) := \inf_{(x,y) \in A \times B} d(x, y)$.

A *digraph* is a directed graph, i.e., a graph with directed edges from one vertex to another based on a binary relation. Then the pair $(p, q) \in V \times V$ is an ordered pair and (p, q) is an arc (directed edge) denoted pq to reflect the difference between an arc and an edge. For example, the nearest neigbor (di)graph in [32] is a proximity digraph. The nearest neighbor digraph,

denoted $NND(V)$, has the vertex set V and pq as an arc iff $d(p,q) = \min_{v \in V \setminus \{p\}} d(p,v)$. That is, pq is an arc of $NND(V)$ iff q is a nearest neighbor of p. Note that if pq is an arc in $NND(V)$, then (p,q) is an edge in $RNG(V)$.

Our proximity catch digraphs are based on the property $\mathfrak{P}$ that is determined by the following mapping which is defined in a more general space than $\mathbb{R}^d$.

Definition 1.1.2. Let $(\Omega, \mathcal{M})$ be a measurable space. The *proximity map* $N(\cdot)$ is given by $N : \Omega \to \wp(\Omega)$, where $\wp(\cdot)$ is the power set functional, and the *proximity region* of $x \in \Omega$, denoted $N(x)$, is the image of $x \in \Omega$ under $N(\cdot)$.

The points in $N(x)$ are thought of as being "closer" to $x \in \Omega$ than are the points in $\Omega \setminus N(x)$. Proximity maps are the building blocks of the *proximity graphs* of Toussaint in [41]; an extensive survey is available in [19].

Definition 1.1.3. The *proximity catch digraph* D has the vertex set $\mathcal{V} = \{p_1, \ldots, p_n\}$ and the arc set $\mathcal{A}$ is defined by $p_i p_j \in \mathcal{A}$ iff $p_j \in N(p_i)$ for $i \neq j$.

Notice that the proximity catch digraph D depends on the *proximity* map $N(\cdot)$, and if $p_j \in N(p_i)$, then we call $N(p_i)$ *catches* p_j. Hence the name *proximity catch digraph*.

If arcs of the form $p_i p_i$ (i.e., loops) were allowed, D would have been called a *pseudodigraph* according to some authors (see, e.g., [4]).

1.1.1 Data-random Proximity Catch Digraphs

Classification and clustering have received considerable attention in the statistical literature. In recent years, a new classification approach has been developed which is based on the relative positions of the data points from various classes. Priebe et al. [35] introduced the class cover catch digraphs (CCCDs) and gave the exact and the asymptotic distribution of the domination number of the CCCD based on two classes $\mathcal{X}_n$ and $\mathcal{Y}_m$ both of which are random samples from uniform distribution on a compact interval in $\mathbb{R}$. DeVinney and Priebe [9], DeVinney et al. [10], Marchette and Priebe [29], Priebe et al. [37], [36] applied the concept in higher dimensions and demonstrated relatively good performance of CCCD in classification. The methods employed involve data reduction (condensing) by using approximate minimum dominating sets as prototype sets (since finding the exact minimum dominating set is in general an NP-hard problem — In particular, for CCCD — (see [8]). Furthermore, the exact and the asymptotic distribution of

the domination number of the CCCDs are not analytically tractable in dimensions greater than 1.

Let $(\Omega, \mathcal{M})$ be a measurable space and $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ and $\mathcal{Y}_m = \{Y_1, \ldots, Y_m\}$ be two classes of Ω-valued random variables whose joint pdf is $F_{X,Y}$. For simplicity in notation, we will use $\mathcal{Y}$ instead of $\mathcal{Y}_m$ henceforth. Let $d(\cdot,\cdot) : \Omega \times \Omega \to [0, \infty)$ be a distance function. The class cover problem for a target class, say $\mathcal{X}_n$, refers to finding a collection of neighborhoods, $N(X_i)$ around $X_i \in \mathcal{X}_n$ such that (i) $\mathcal{X}_n \subset \left(\cup_i N(X_i)\right)$ and (ii) $\mathcal{Y} \cap \left(\cup_i N(X_i)\right) = \emptyset$. A collection of neighborhoods satisfying both conditions is called a *class cover*. A cover satisfying condition (i) is a *proper cover* of class $\mathcal{X}_n$ while a collection satisfying condition (ii) is a *pure cover* relative to class $\mathcal{Y}$. From a practical point of view, for example for classification, of particular interest are the class covers satisfying both (i) and (ii) with the smallest collection of neighborhoods, i.e., minimum cardinality cover.

This class cover problem is a generalization of the set cover problem in [14] that emerged in statistical pattern recognition and machine learning, where an edited or condensed set (prototype set) is selected from $\mathcal{X}_n$ (see, e.g., Devroye et al. [11]).

In particular, we construct the proximity regions using two classes of data sets. Given $\mathcal{Y} \subseteq \Omega$, the *proximity map* $N_{\mathcal{Y}}(\cdot) : \Omega \to \wp(\Omega)$ associates a *proximity region* $N_{\mathcal{Y}}(x) \subseteq \Omega$ with each point $x \in \Omega$. The region $N_{\mathcal{Y}}(x)$ is defined in terms of the distance between x and $\mathcal{Y}$. More specifically, our proximity maps will be based on the relative position of points from class $\mathcal{X}_n$ with respect to the Delaunay tessellation of the class $\mathcal{Y}$. See Section 2.1 for more on Delaunay tessellation.

If $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ is a set of Ω-valued random variables then $N_{\mathcal{Y}}(X_i)$ are random sets. If X_i are independent identically distributed then so are the random sets $N_{\mathcal{Y}}(X_i)$. We define the data-random proximity catch digraph D — associated with $N_{\mathcal{Y}}(\cdot)$ — with vertex set $\mathcal{X}_n = \{X_1, \cdots, X_n\}$ and arc set $\mathcal{A}$ by $X_i X_j \in \mathcal{A} \iff X_j \in N_{\mathcal{Y}}(X_i)$. Since this relationship is not symmetric, a digraph is needed rather than a graph. The random digraph D depends on the (joint) distribution of the X_i and on the map $N_{\mathcal{Y}}(\cdot)$.

The PCDs are closely related to the *proximity graphs* of Toussaint and Jaromczyk [19] and might be considered as a special case of *covering sets* of Tuza [42] and *intersection digraphs* of Sen et al. [39]. This data random proximity digraph is a *vertex-random proximity digraph* which is not of standard type (see, e.g., Janson et al. [18]). The randomness of the PCDs lies in the fact that the vertices are random with joint pdf $F_{X,Y}$, but arcs $X_i X_j$ are deterministic

functions of the random variable X_j and the set $N_{\mathcal{Y}}(X_i)$.

For example, the CCCD of Priebe et al. [35] can be viewed as an example of PCD with $N_{\mathcal{Y}}(x) = B(x, r(x))$, where $r(x) := \min_{\mathrm{y} \in \mathcal{Y}} d(x, \mathrm{y})$. The CCCD is the digraph of order n with vertex set $\mathcal{X}_n$ and an arc from X_i to X_j iff $X_j \in B(X_i, r(X_i))$. That is, there is an arc from X_i to X_j iff there exists an open ball centered at X_i which is "pure" (or contains no elements) of $\mathcal{Y}$, and simultaneously contains (or "catches") point X_j.

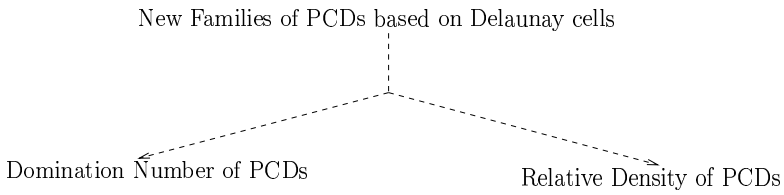
Notice that the CCCDs are defined with (open) balls only, whereas PCDs are not based on a particular geometric shape or a functional form; that is, PCDs admit $N_{\mathcal{Y}}(\cdot)$ to be any type of region, e.g., circle (ball), arc slice, triangle, a convex or nonconvex polygon, etc. In this sense, the PCDs are defined in a more general setting compared to CCCD. On the other hand, the types of PCDs we introduce in this dissertation are well-defined for points restricted to the convex hull of $\mathcal{Y}$, $C_H(\mathcal{Y})$. Moreover, the three families of proximity maps we introduce will yield closed regions.

Furthermore, the CCCDs based on balls use proximity regions are defined by the obvious metric, while the PCDs do not suggest an obvious metric. In particular, our PCDs will be based on some sort of dissimilarity measure, but no metric underlying this measure will exist.

In this dissertation, we have three major result categories:

- Introduction of three families of PCDs, where one is a modification of CCCD based on the balls, the other two are parametrized families.
  - Investigation of the domination number of the PCDs.
  - Investigation of the relative density of the PCDs.

Schematically, the major result categories are depicted as



The domination number and the relative density of PCDs are first investigated for data in one Delaunay cell (triangle in $\mathbb{R}^2$) and the analyses are generalized to data in multiple Delaunay cells.

1.2 Chapter Overview

In this thesis, we provide the preliminary tools and the foundation in Chapter 2. A proximity catch digraph which is a modification of the CCCD, and two parametrized families of PCDs in $\mathbb{R}^2$ and the related concepts are introduced in Chapter 3. In Chapter 4, the asymptotic distribution of the domination number of one family of PCDs is computed. The domination number is then used in testing for the spatial patterns of segregation and association. Furthermore, lower and upper bounds are provided for other two families. In Chapter 5, the asymptotic (normal) distribution of the relative density of the parametrized PCDs is computed by using the central limit theory of U-statistics. Furthermore, relative density is also used in testing for the spatial patterns of segregation and association. Asymptotic relative efficacy methods such as Pitman asymptotic efficacy, Hodges-Lehmann asymptotic efficacy, asymptotic power function analysis, and Monte Carlo simulations are employed in the selection of optimal proximity parameters.

In Chapter 6, we compare the proximity maps and the associated PCDs in terms of some appealing properties, asymptotic behaviour and tractability of the distributions of the domination number and relative density of the PCDs, as well as their performance in testing for the spatial patterns. Also provided are open problems and directions for prospective research.

In Appendices A, B, and C we provide details of lengthy proofs and expressions.

CHAPTER 2

Preliminary Tools and Foundation

2.1 Voronoi Diagrams and Delaunay Tessellations

Our proximity catch digraphs will be based on the relative location of points from one class with respect to the Delaunay tessellation of points from another class. In this section, we provide a brief description of Delaunay tessellations and related concepts.

A *tessellation* is a partition of a space into convex polytopes; tessellation of the plane (into convex polygons) is the most frequently studied case.

Given $2 \leq n < \infty$ distinct points in $\mathbb{R}^2$, we associate all points in the space with the closest member(s) of the point set with respect to the Euclidean distance. The result is a tessellation of the plane into a set of regions associated with the n points. We call this tessellation the *planar ordinary Voronoi diagram* generated by the point set and the regions *ordinary Voronoi polygons*. See Figure 2.1.1 (left) for an example with $m = |\mathcal{Y}| = 10$ Y points iid from $\mathcal{U}\big((0,1) \times (0,1)\big)$.

In general, let $P = \{p_1, p_2, \ldots, p_n\}$ be n points in $\mathbb{R}^d$ where $2 \leq n < \infty$ and $p_i \neq p_j$ for $i \neq j$, $i, j \in [n] := \{1, 2, \ldots, n\}$ and let $||\cdot||$ denote the norm functional. We call the region $\mathcal{V}_C(p_i) = \{x \in \mathbb{R}^d : ||x - p_i|| \leq ||x - p_j|| \text{ for } j \neq i,\ j \in [n]\}$ the (ordinary) *Voronoi polygon* or *cell* associated with p_i and the set $\mathfrak{V} = \{\mathcal{V}_C(p_1), \ldots, \mathcal{V}_C(p_n)\}$ the *Voronoi diagram* or *Dirichlet tessellation* generated by P. We call p_i the *generator* of $\mathcal{V}_C(p_i)$. The Voronoi diagram partitions the space into disjoint regions (which are also called *tiles* or *Thiessen polygons* in $\mathbb{R}^2$). Notice that we still say $\mathfrak{V}$ *partitions* the space $\mathbb{R}^d$, although $\mathcal{V}_C(p_j)$ are not necessarily disjoint, but if nonempty the intersection lies in a lower dimension, or equivalently, has zero $\mathbb{R}^{d-1}$-Lebesgue measure. We stick to this convention throughout this dissertation. The intersection of two Voronoi cells, if nonempty, i.e., for $i \neq j$, $\mathcal{V}_C(p_i) \cap \mathcal{V}_C(p_j) \neq \emptyset$, is called a *Voronoi edge*. If a Voronoi edge is not a point, then $\mathcal{V}_C(p_i)$ and $\mathcal{V}_C(p_j)$ are said to be *adjacent*. An end point of a Voronoi edge is called a *Voronoi vertex*.

If at a Voronoi vertex, more than three Voronoi polygons intersect, we say $\mathfrak{V}$ is *degenerate*;

otherwise, $\mathfrak{V}$ is *non-degenerate*. A detailed discussion including the history of Voronoi diagrams is available in [30].

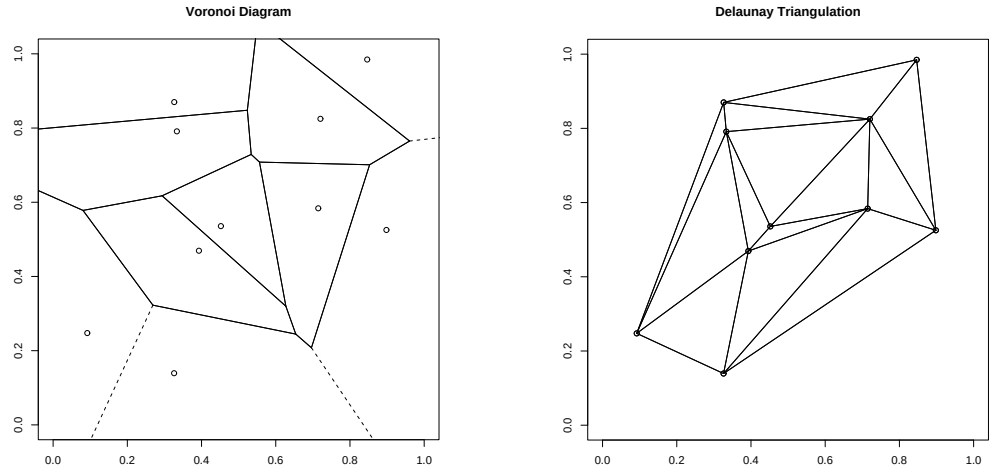


Figure 2.1.1: The Voronoi diagram based on a realization of 10 Y points (left) and the associated Delaunay triangulation (right).

Given a Voronoi diagram with $n \geq d+1$ non-coplanar (i.e., not all the points lie on a $(d-1)$-dimensional hyperplane) generators, P, in $\mathbb{R}^d$, we join all pairs of generators whose Voronoi cells have a common Voronoi edge. The resulting tessellation is called the *Delaunay tessellation* of P. See Figure 2.1.1 (right) for the Delaunay triangulation associated with the Voronoi diagram in Figure 2.1.1 (left) based on 10 points iid from $\mathcal{U}\big((0,1) \times (0,1)\big)$. By definition a Delaunay tessellation of a finite set, P, is the dual of the *Voronoi diagram* constructed by using the same set. The tessellation yields a (unique) polytopization provided that no more than $(d+1)$ points in $\mathbb{R}^d$ are cospherical (i.e., no more than $(d+1)$ points lie on the boundary of a (hyper)sphere in $\mathbb{R}^d$). Moreover, the circumsphere of each Delaunay polytope (i.e., the sphere that contains the vertices of the Delaunay polytope on its boundary) is pure from the set P, i.e., the interior of the circumsphere of the Delaunay polytope does not contain any points from P. The Delaunay tessellation partitions the convex hull, $\mathcal{C}_H(P)$, of P. In particular, in $\mathbb{R}^2$, the tessellation is a *triangulation* that yields triangles T_j (including the interior), $j = 1, \ldots, J$ (see, e.g., [30]) provided that no more than three points are cocircular (i.e., no more than three points lie on the boundary of some circle in $\mathbb{R}^2$). In this dissertation we adopt the convention that a triangle refers to the closed region bounded by its edges. See Figure 2.1.2 for an example with $n = 200$

X points $\overset{iid}{\sim} \mathcal{U}\big((0,1)\times(0,1)\big)$, the uniform distribution on the unit square and the Delaunay triangulation is based on Y points in Figure 2.1.1.

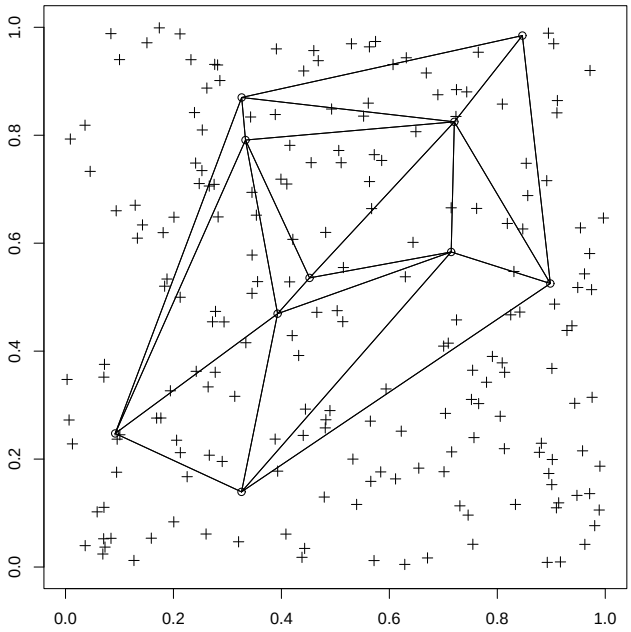


Figure 2.1.2: A realization of 200 X points and the Delaunay triangulation based on 10 Y points in Figure 2.1.1.

2.1.1 Poisson Delaunay Triangles

The Delaunay triangles are based on a given set of points $\mathcal{Y}$. The set $\mathcal{Y}$ can be assumed to come from a Poission point process on a finite region, and in the application of PCDs, to remove the conditioning on $\mathcal{Y}$, it is suggested that $\mathcal{Y}$ comes from a Poisson point process for prospective research directions. We briefly describe the Poisson point processes and Poisson Delaunay triangles.

A *stochastic point process* on $\mathbb{R}^d$ is defined to be a process in which points are generated according to a probability distribution $P\big(|\mathcal{Y}\cap B| = k\big)$, $k = 0, 1, 2, \ldots$, over any $B \subseteq \mathbb{R}^d$. For example, a binomial point process is a stochastic point process in which n points are generated over a bounded set $S \subsetneq \mathbb{R}^d$ according to the uniform distribution.

In particular, if $d = 2$ the process is called a *planar stochastic point process.* If two points

coincide with probability zero, then it is a *simple stochastic point process.* A stochastic point process is said to be *locally finite* if any finite region $B \subset \mathbb{R}^d$ contains a finite number of points with probability 1 under the process (see [30]).

We have built the Delaunay tessellation using $\mathcal{Y}$ with finite sample size. Suppose $\mathcal{Y}$ is from a stochastic point process. One of the most fundamental locally finite stochastic point processes is the *Poisson point process*, which is defined as the process that satisfies

$$P\big(|\mathcal{Y} \cap B| = k\big) = \frac{\lambda \cdot \mathrm{V}(B) \cdot e^{-\lambda \cdot \mathrm{V}(B) \cdot k}}{k!}, \quad k = 0, 1, 2, \ldots$$

for any $B \subsetneq \mathbb{R}^d$, where $\mathrm{V}(\cdot)$ denotes the d-dimensional volume functional and $\lambda > 0$ is the intensity of the process. We can also define the Poisson point process as the limit of the binomial point process in the sense of expanding the finite region S to an infinite region while keeping $\lambda = \frac{n}{\mathrm{V}(S)}$ constant. We call the Delaunay tessellation based on a finite data set from a Poisson point process *Poisson Delaunay tessellation* and denote it $\mathcal{D}_P$. The associated Voronoi diagram is called the *Poisson Voronoi diagram* and denoted $\mathcal{V}_P$. For more details on the properties of $\mathcal{V}_P$ (see, for example, [30]).

A *simplex* in $\mathbb{R}^d$ is the convex hull of any $(d+1)$ points in general position, i.e., no $(d+1)$ of the points lie in a $(d-1)$-dimensional hyperplane in $\mathbb{R}^d$. The simplex is the point itself for $d = 0$, the line segment joining the two points for $d = 1$, a triangle for $d = 2$, a tetrahedron for $d = 3$, and so on. Each Poisson Delaunay cell of $\mathcal{D}_P$ is a $(d+1)$-dimensional simplex whose vertices are $x_0, x_1, \ldots, x_m$ from the Poisson point process. Any s-face of $\mathcal{D}_P$ is an $(s+1)$-dimensional simplex with vertices $x_0, \ldots, x_s$, also points from the Poisson point process. There are $\binom{d+s}{s+1}$ many s-faces contained in a Poisson Delaunay cell for $0 \leq s \leq d$.

Let c and r be the circumcenter and circumradius, respectively, of a $(d+1)$-dimensional Poisson Delaunay cell in $\mathbb{R}^d$. Then the $(d+1)$ vertices of the cell are the points $\{c + r\, u_i\}$ where $\{u_i\}$ are the unit vectors for $i = 0, 1, \ldots, d$. The ergodic joint probability density function of $\mathcal{D}_P$, the probability density function of r, and k^{th} moment of the volume of a typical Poisson Delaunay cell are provided in [30].

In $\mathbb{R}^2$, the probability density function (pdf) of a pair of inner angles arbitrarily selected from an arbitrary triangle in $\mathcal{D}_P$ is given by

$$f(x, y) = \frac{8}{3\pi}(\sin x)(\sin y)\sin(x + y), \text{ for } x, y > 0 \text{ and } x + y < \pi.$$

Notice that the mode of this density is at $x = y = \pi/3$, which implies that the most frequent triangles in a $\mathcal{D}_P$ are nearly equilateral triangles.

By integrating over y, we obtain the pdf of a randomly selected inner angle of an arbitrary triangle from $\mathcal{D}_P$:

$$f(x) = \left[\frac{4}{3\pi}\left((\pi - x)\cos x + \sin x\right)\sin x\right] \mathbf{I}(0 < x < \pi)$$

where $\mathbf{I}(\cdot)$ is the indicator function. Then the expected value of X is $\mathbf{E}[X] = \pi/3$ and $\mathbf{E}[X^2] = 2\pi^2/9 - 5/6$.

The pdf of the minimum angle and the pdf of the maximum angle, and the distribution of the length of an arbitray edge of an arbitrary triangle from $\mathcal{D}_P$, are also provided in [30] with relevant references.

2.2 Transformations Preserving Uniformity on Triangles in $\mathbb{R}^2$

We will assume $\mathcal{X}_n$ is a set of iid uniform random variables on the convex hull of $\mathcal{Y}$; i.e., a ramdom sample from $\mathcal{C}_H(Y)$. In particular, conditional on $|\mathcal{X}_n \cap T_j| > 0$ being fixed, $\mathcal{X}_n \cap T_j$ will also be a set of iid uniform random variables on T_j for $j \in \{1, 2\}, \ldots, J$, where T_j is the j^{th} Delaunay triangle and J is the total number of Delaunay triangles. Reducing the triangle T_j as much as possible while preserving uniformity and the probabilities related to PCDs will simplify the notation and calculations. Below, we present such a transformation that reduces a single triangle.

Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\} \subset \mathbb{R}^2$ be three non-collinear points and $T(\mathcal{Y})$ be the triangle (including the interior) with vertices $\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3$. Let $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}))$, the uniform distribution on $T(\mathcal{Y})$, for $i = 1, \ldots, n$. The probability density function (pdf) of $\mathcal{U}(T(\mathcal{Y}))$ is

$$f(u) = \frac{1}{A(T(\mathcal{Y}))}\mathbf{I}(u \in T(\mathcal{Y})),$$

where $A(\cdot)$ is the area functional.

The triangle $T(\mathcal{Y})$ can be carried into the first quadrant by a composition of transformations in such a way that the largest edge has unit length and lies on the x-axis, and the x-coordinate of the vertex nonadjacent to largest edge is less than $1/2$. We call the resultant triangle the *basic triangle* and denote it as T_b.

Although such transformations are simple, we will describe them in prose only, due to the complexity of notation.

Let e_j be the edge opposite vertex y_j for $j \in \{1,2,3\}$. To transform $T(\mathcal{Y})$ to the corresponding basic triangle T_b, first find the lengths of the edges; suppose e_3 is of maximum length. Then scale the triangle so that e_3 is of unit length. Next translate y_1 to $(0,0)$, and rotate (if necessary) the triangle so that $\mathsf{y}_2 = (1,0)$. If the y-coordinate of y_3 is negative reflect around the x-axis, then if x-coordinate of y_3 is greater than $1/2$, reflect around $x = 1/2$, then the associated basic triangle T_b is obtained. That is, the basic triangle T_b can be obtained by a composition of rigid motion transformations: scaling, translation, rotation, and reflection. Hence if $T(\mathcal{Y})$ is transformed into T_b, then $T(\mathcal{Y})$ is similar to T_b. Thus the random variables $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}))$ transformed along with $T(\mathcal{Y})$ in the described fashion are $\overset{iid}{\sim} \mathcal{U}(T_b)$. So, without loss of generality, we can assume $T(\mathcal{Y})$ to be the basic triangle, $T_b = \big((0,0),(1,0),(c_1,c_2)\big)$ where $0 < c_1 \leq 1/2$, and $c_2 > 0$ and $(1-c_1)^2 + c_2^2 \leq 1$. The functional form of T_b is

$$T_b = \left\{(x,y) \in \mathbb{R}^2 : y \geq 0;\ y \leq (c_2\, x)/c_1;\ y \leq c_2\,(1-x)/(1-c_1)\right\}.$$

The basic triangle T_b is an equilateral triangle, if $c_1 = 1/2$ and $c_2 = \sqrt{3}/2$; an obtuse triangle, if $c_2 < \sqrt{c_1 - c_1^2}$; a right triangle, if $c_2 = \sqrt{c_1 - c_1^2}$; is an acute triangle, if $c_2 > \sqrt{c_1 - c_1^2}$. If $c_2 = 0$, then the T_b reduces to the unit interval $(0,1)$.

Note that there are also transformations that preserve uniformity of the random variable, but not similarity of the triangles. We only describe the transformation that maps $T(\mathcal{Y})$ to the standard equilateral triangle, $T_e = T\left((0,0),(1,0),\left(1/2,\sqrt{3}/2\right)\right)$ for exploiting the symmetry in calculations using T_e.

2.2.1 Transformation of T_b to T_e

Let $\phi_e : (x,y) \to (u,v)$, where $u(x,y) = x + \frac{1-2\,c_1}{\sqrt{3}}\, y$ and $v(x,y) = \frac{\sqrt{3}}{2\,c_2}\, y$. Then y_1 is mapped to $(0,0)$, y_2 is mapped to $(1,0)$, and y_3 is mapped to $\left(1/2,\sqrt{3}/2\right)$. See also Figure 2.2.1.

Note that the inverse transformation is $\phi_e^{-1}(u,v) = \big(x(u,v), y(u,v)\big)$ where $x(u,v) = u - \frac{(1-2\,c_1)}{\sqrt{3}}\, v$ and $y(u,v) = \frac{2\,c_2}{\sqrt{3}}\, u$. Then the Jacobian is given by

$$J(x,y) \quad = \quad \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & \frac{2\,c_1 - 1}{\sqrt{3}} \\ 0 & \frac{2\,c_2}{\sqrt{3}} \end{vmatrix} = \frac{2\,c_2}{\sqrt{3}}.$$

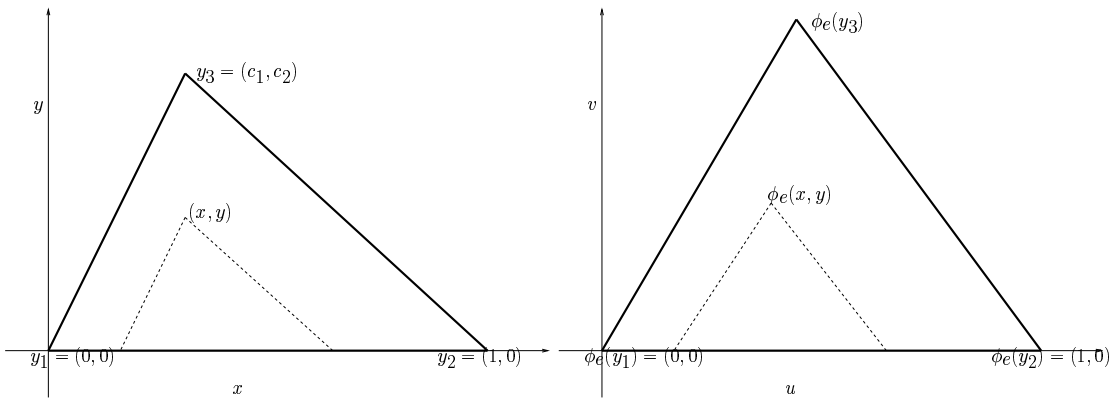


Figure 2.2.1: The description of $\phi_e(x, y)$ for $(x, y) \in T_b$ (left) and the equilateral triangle $\phi_e(T_b) = T_e$ (right).

So $f_{U,V}(u, v) = f_{X,Y}(\phi_e^{-1}(u, v))\,|J| = \frac{4}{\sqrt{3}}\,\mathbf{I}\big((u, v) \in T_e\big)$. Hence uniformity is preserved.

2.3 Triangle Centers

Our PCDs will be defined using the vertex and edge regions, which will be constructed using a point, preferably, in the interior of the triangle, e.g., a triangle center.

Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\} \subset \mathbb{R}^2$ be non-collinear and $T(\mathcal{Y})$ be the corresponding triangle. The *trilinear coordinates* of a point P with respect to $T(\mathcal{Y})$ are an ordered triple of numbers, each of which is proportional to the distance from P to one of the edges. Trilinear coordinates are denoted as $(\alpha : \beta : \gamma)$ and also are known as *homogeneous coordinates* or *trilinears*. Trilinear coordinates were introduced by Plücker in 1835 (see [43]). The triplet of trilinear coordinates obtained by multiplying a given triplet by any nonzero constant k describes the same point, i.e., $(\alpha : \beta : \gamma) = (k\alpha : k\beta : k\gamma)$, for any $k > 0$.

By convention, the three vertices y_1, y_2, and y_3 of $T(\mathcal{Y})$ are commonly written as $(1 : 0 : 0)$, $(0 : 1 : 0)$, and $(0 : 0 : 1)$, respectively (see [43]).

Definition 2.3.1. A *triangle center* is a point whose trilinear coordinates are defined in terms of the edge lengths and (inner) angles of a triangle. The function giving the coordinates $(\alpha : \beta : \gamma)$ is called the *triangle center function*.

Kimberling enumerates 360 triangle centers in [23], among which four have been widely known since the ancient times; namely, circumcenter (CC), incenter (IC), center of mass or

centroid (CM), and orthocenter (OC). The point where the center is located in $T(\mathcal{Y})$ will be labeled accordingly; e.g., M_{CC} will denote the circumcenter of $T(\mathcal{Y})$.

The *circumcircle* is a triangle's circumscribed circle; i.e., the unique circle that passes through each of the triangle's three vertices y_1, y_2, y_3. The center of the circumcircle is called the *circumcenter*, denoted M_{CC}, and the circle's radius is called the *circumradius*, denoted r_{cc}. By construction, the distances from circumcenter to the vertices are equal (to r_{cc}). Furthermore, the triangle's three edge bisectors perpendicular to edges e_j at M_j for $j \in \{1,2,3\}$ intersect at M_{CC}. See Figure 2.3.1. The trilinear coordinates of M_{CC} are $(\cos\theta_1 : \cos\theta_2 : \cos\theta_3)$ where θ_j is the inner angle of $T(\mathcal{Y})$ at vertex y_j for $j \in \{1,2,3\}$ and the trilinears for M_{cc} can also be written as $(r_{cc}\cos\theta_1 : r_{cc}\cos\theta_2 : r_{cc}\cos\theta_3)$.

The circumcenter of a triangle is in the interior, at the midpoint of the hypotenuse, or in the exterior of the triangle, if the triangle is acute, right, or obtuse, respectively. See Figure 2.3.1 where an acute and an obtuse triangle are depicted. Using the pdf of an arbitrary angle of a triangle T_j from Poisson Delaunay triangulation $\mathcal{D}_P$, we see that,

$$P(T_j \text{ is a right triangle}) = P(\theta = \pi/2) = 0,$$

hence $P(M_{CC}$ is the midpoint of the hypothenuse$) = 0$. Furthermore,

$$\begin{aligned} P(T_j \text{ is an obtuse triangle}) &= P(M_{CC} \notin T_j) = P(\theta_{\max} > \pi/2) = \int_{\pi/2}^{\pi} f_3(x)dx \\ &= \frac{\left(3\, f_S\left(\sqrt{2\,\pi}\right) - f_C\left(\sqrt{2\,\pi}\right) - 3\, f_S\left(\sqrt{\frac{\pi}{2}}\right) + f_C\left(\sqrt{\frac{\pi}{2}}\right)\right)}{\sqrt{2\,\pi}} \\ &\approx .03726 \end{aligned} \tag{2.3.1}$$

where

$$f_3(x) = \left[\frac{2}{\pi}(3x(\sin 2x) - \cos 2x + \cos 4x - \pi \sin 2x)\right] \mathbf{I}(\pi/3 < x < \pi/2) + \left[\frac{1}{\pi}(4\pi(\cos x)(\sin x) + 3\sin x^2 - \cos x^2 - 4x(\cos x)(\sin x) + 1)\right] \mathbf{I}(\pi/2 < x < \pi).$$

is the pdf of the maximum angle, $f_C(x) = \int_0^x \cos(\pi\, t^2/2)dt$, and $f_S(x) = \int_0^x \sin(\pi\, t^2/2)dt$ are the Fresnel cosine and sine functions, respectively. The coordintates of M_{CC} in the basic triangle

T_b are $\left(\frac{1}{2}, \frac{c_1^2 - c_1 + c_2^2}{2\,c_2}\right)$.

The *incircle* is the inscribed circle of a triangle, i.e., the unique circle that is tangent to the edges e_j at P_j for $j \in \{1,2,3\}$. The center of the incircle is called the *incenter*, denoted M_I, and the radius of the incircle is called the *inradius*, denoted r_{ic}. Incenter has trilinear coordinates $(1:1:1)$. The incenter is the point where the triangle's inner angle bisectors intersect. See Figure 2.3.2 (left).

The coordinates of M_I for the basic triangle T_b are (x_{ic}, y_{ic}), where

$$x_{ic} = \frac{c_1 - \sqrt{c_1^2 + c_2^2}}{1 + \sqrt{c_1^2 + c_2^2} + \sqrt{(1-c_1)^2 + c_2^2}}, \quad y_{ic} = \frac{c_2}{1 + \sqrt{c_1^2 + c_2^2} + \sqrt{(1-c_1)^2 + c_2^2}}.$$

The distance between M_{CC} and M_I is $d(M_{CC}, M_I) = \sqrt{r_{cc}(r_{cc} - 2\,r_{ic})}$. Unlike the circumcenter, the incenter is guaranteed to be inside the triangle.

The *median line* of a triangle is the line from one of its vertices to the midpoint of the opposite edge. The three median lines of any triangle intersect at the triangle's *centroid*, denoted M_C. The centroid is the *center of mass* of the vertices of a triangle. Since M_C is also the intersection of the triangle's three median lines, it is sometimes called the *median point*. It has trilinear coordinates $(1/|e_1| : 1/|e_2| : 1/|e_3|)$ or $(\csc\theta_1 : \csc\theta_1 : \csc\theta_1)$ where e_j denotes the edge opposite vertex y_j for $j \in \{1,2,3\}$. The centroid is also guaranteed to be in the interior of the triangle. See Figure 2.3.2 (right). The coordinates of M_C for the basic triangle are $\big((1+c_1)/3, c_2/3\big)$.

The intersection of the three altitudes of a triangle is called the *orthocenter*, M_O, which has trilinear coordinates $(\cos\theta_2\cos\theta_3 : \cos\theta_1\cos\theta_3 : \cos\theta_1\cos\theta_2)$. The orthocenter of a triangle is in the interior, at vertex y_3, or in the exterior of the basic triangle, T_b, if T_b is acute, right, or obtuse, respectively. The functional form of M_O in the basic triangle is $(c_1, c_1\,(1-c_1)/c_2)$.

Note that in an equilateral triangle, $M_I = M_{CC} = M_O = M_C$ (i.e., all four centers we have described coincide).

2.4 The Spatial Patterns of Segregation and Association

The PCDs are used in testing spatial patterns of segregation and association. We present a brief overview of these patterns and the related tests in the literature.

The spatial relationships between two or more classes have important consequences in many areas of science, especially, ecology and species population biology. Most of the early work

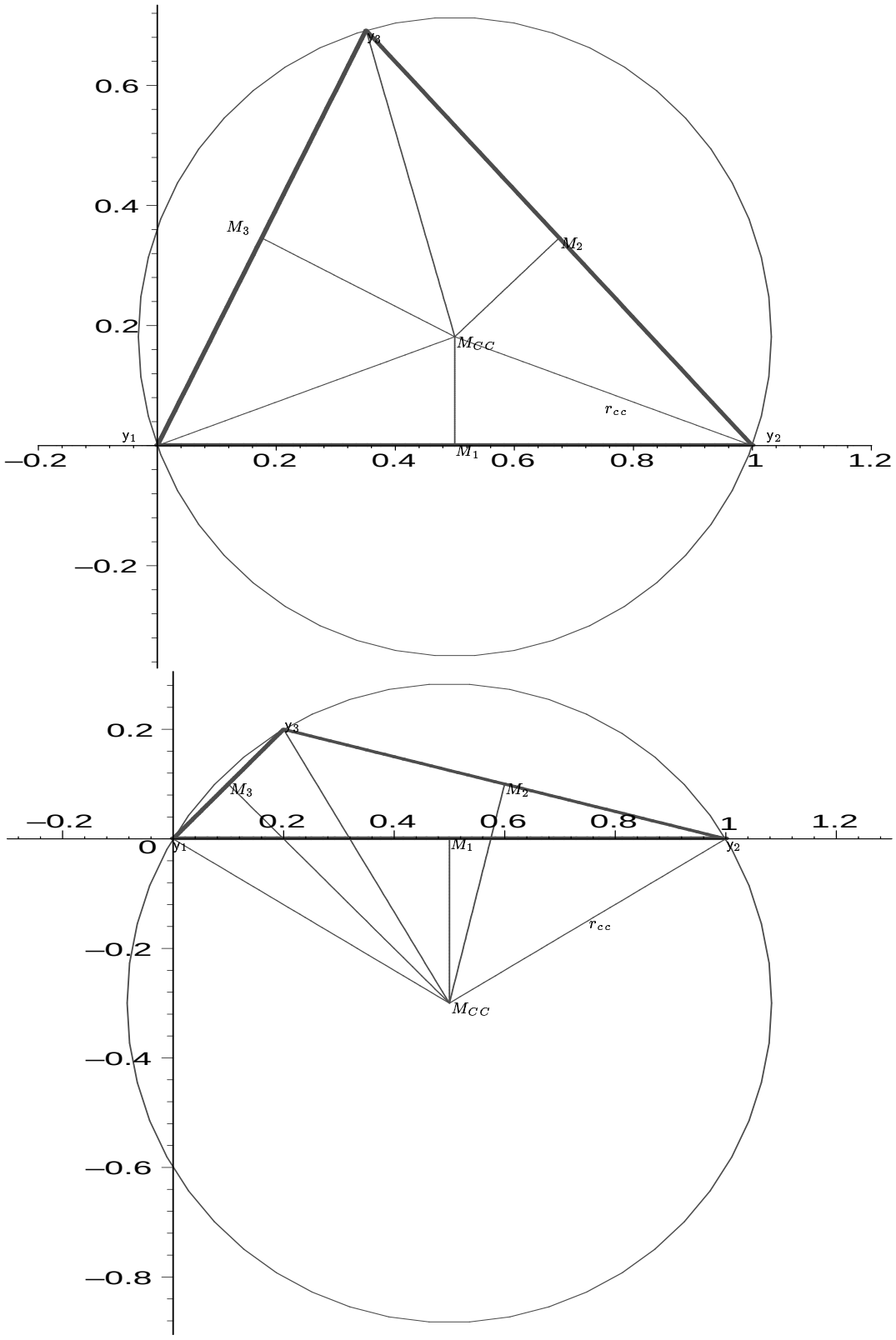


Figure 2.3.1: The circumcenter, circumcircle, and circumradius of an acute triangle (top), an obtuse triangle (bottom).

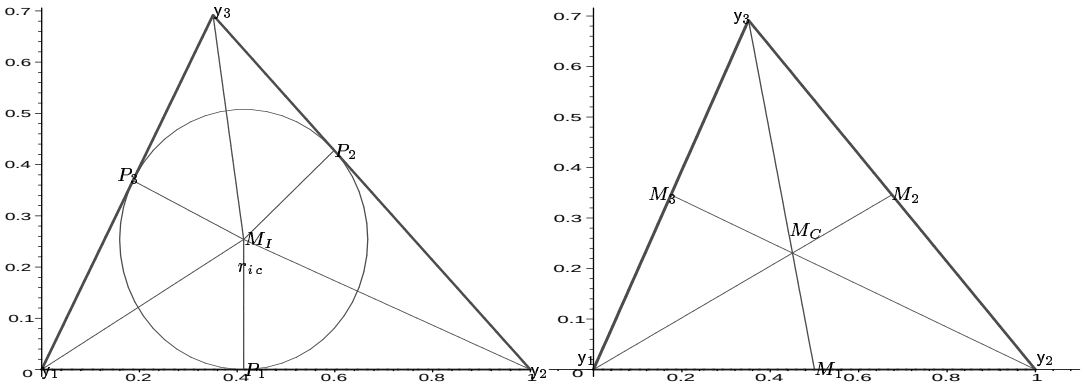


Figure 2.3.2: The incircle, incenter, inradius of a triangle (left), and the centroid (center of mass) of a triangle (right).

in spatial patterns of observations (e.g., species of plants) was done one class (or species) at a time only. In 1961, Pielou [33] considered the relative position of the species with respect to each other by describing the spatial patterns in two categories: two species of similar size are "unsegregated" if a member of a species is equally likely to be the nearest neighbor of a member of same species or a member of the opposite species; two species are "segregated" if the species tend to exist in clumps or clusters, or it is more likely for a member to be located near a member of its own kind. Over the years, the relative spatial patterns of two classes, have been divided into three categories: segregation, association, or neither (e.g., complete spatial randomness). These three categories are also called positive segregation, negative segregation, or unsegregation by some authors (see, e.g., [33]). The phenomenon known as *segregation* involves members from different classes having a tendency to repel each other, which in return implies that members of a species tend to be found near conspecifics (i.e., members of the same species). *Association* involves members from different classes having a tendency to attract one another. See for instance, Coomes et al. [6].

One straightforward approach to testing spatial patterns is to partition the spatial region into rectangular or square quadrats and compare the intensity of each species in quadrats to the chance expectation and decide for the existence or non-existence of segregation. However, this procedure is inevitably dependent on the quadrat size, which usually is arbitrary. There are many possible patterns for segregation. Rather than testing against these particular patterns, Pielou's test is based on the investigation of the nearest neighbor relationships of the populations.

In her approach, a 2×2 nearest neighbor contingency table is constructed by recording the class of each observation and the class of its nearest neighbor; that is, a 2×2 contingency table $C = [c_{i,j}]$ is constructed, where $c_{i,j}$ is the number of individuals from class j that are nearest neighbors of class i. The resulting test is a χ^2-test of independence with 1 degree of freedom, which compares the observed cell counts to the expected cell counts. Pielou's method, which has been frequently used in ecological studies, tests whether two species have the expected proportion of the same species as the nearest neighbor under independence.

Many other tests have been proposed to test segregation in spatial data. These include the comparison of nearest neighbors (Diggle, 1983; Cuzick and Edwards, 1990), comparisons of Ripley's $K(t)$ or $L(t)$ functions (Ripley 1981), formal testing of the difference between two $K(t)$ curves (Diggle and Chetwynd, 1991). See Dixon [12] for a broader discussion and the references mentioned in this paragraph.

A test of segregation depends on the distribution of the cell counts under a null hypothesis of non-existence of segregation. The null hypothesis here can be formulated in two slightly different forms:

(i) Complete spatial randomness occurs when each class is randomly distributed throughout the area of interest. Complete spatial randomness not only describes the association between species but also arrangement of locations.

(ii) Random labeling, on the other hand, is less restrictive in the sense that locations may have any particular arrangement (perhaps clustered or spread in a particular pattern) but labels, e.g., species names, are randomly assigned to locations (see [12]).

2.5 U-Statistics

The relative density of the PCDs is shown to be a U-statistic. Here we describe U-statistics, and the related asymptotic results.

U-statistics are a class of unbiased estimators of characteristics (or parameters) of one or more populations. The basic theory of U-statistics was developed by W. Hoeffding in 1948 (see [17]). Let $\mathcal{F}$ be a family of probability distributions on an arbitrary measurable space and $\theta(F)$ be a real-valued function depending on $F \in \mathcal{F}$. We start with one-sample U-statistics. A parameter $\theta(F)$ is said to be *estimable of degree m* for the family of distributions $\mathcal{F}$, if m is the minimum

sample size for which there exists a function $h(x_1, \ldots, x_m)$ such that $\mathbf{E}\left[h(X_1, \ldots, X_m)\right] = \theta(F)$ for every distribution $F \in \mathcal{F}$, where $\{X_1, \ldots, X_m\}$ is a random sample (i.e., set of iid random variables) from F and $h(X_1, \ldots, X_m)$ is a statistic that does not depend on F. The function $h(\cdot)$ is called the *kernel* of the parameter θ. A random sample from a distribution refers to a set of iid random variables from the same distribution throughout the dissertation.

The function $h(\cdot)$ can be assumed to be symmetric in its arguments, since for any given kernel $g(x_1, \ldots, x_m)$, we can always construct one that is symmetric in its arguments as

$$h(x_1, \ldots, x_m) = \frac{1}{m!} \sum_{\pi \in \Pi_m} g(x_{\pi(1)}, \ldots, x_{\pi(m)})$$

where the summation is over $\Pi_m = \{\pi : \pi \text{ is a permutation of integers } 1, \ldots, m\}$. Obviously, $h(\cdot)$ is symmetric in its arguments since $h(x_1, \ldots, x_m) = h(x_{\pi(1)}, \ldots, x_{\pi(m)})$ for every permutation $(\pi(1), \ldots, \pi(m))$ of the integers $[m]$ and $h(x_1, \ldots, x_m)$ is an unbiased estimator of $\theta(F)$ for any $F \in \mathcal{F}$.

For a real-valued kernel function $h(x_1, \ldots, x_m)$ and for a sample $X_1, \ldots, X_n$ of size $n \geq m$ from a distribution F, a one-sample U-statistic for the estimable parameter $\theta = \theta(F)$ of degree m constructed by $h(\cdot)$ is given by

$$U_n = U(X_1, \ldots, X_n) = \frac{(n-m)!}{n!} \sum_{P_{m,n}} h\left(X_{i(1)}, \ldots, X_{i(m)}\right),$$

where the summation is over the set $P_{m,n}$ of all $n!/(n-m)!$ permutations of $(i(1), \ldots, i(m))$ of size m chosen from $[n]$. If the kernel $h(\cdot)$ is symmetric in its arguments, then U_n has the equivalent form

$$U_n = U(X_1, \ldots, X_n) = \frac{1}{\binom{n}{m}} \sum_{C_{m,n}} h\left(X_{i(1)}, \ldots, X_{i(m)}\right),$$

where $C_{m,n}$ is the set of all $\binom{n}{m}$ combinations of m integers $i(1), i(2), \ldots, i(m)$ chosen from $[n]$. Note that a U-statistic is an unbiased estimator of θ for every $F \in \mathcal{F}$ and is symmetric in its arguments. In fact, when $\mathcal{F}$ includes all continuous distributions, it can be shown that a U-statistic is the unique minimum variance unbiased estimator of θ (see [40]).

The asymptotic distribution of U_n is given below.

Lemma 2.5.1. *(Theorem 3.3.13 in [38]) Let $\{X_1, \ldots, X_n\}$ denote a random sample from some*

population. Let θ be an estimable parameter of degree m with symmetric kernel $h(x_1, \ldots, x_m)$. If $\mathbf{E}\left[h^2(X_1, \ldots, X_m)\right] < \infty$ and if

$$U(X_1, \ldots, X_m) = \frac{1}{\binom{n}{m}} \sum_{\beta \in C_{m,n}} h\left(X_{\beta(1)}, \ldots, X_{\beta(m)}\right)$$

where $C_{m,n}$ consists of the subsets of m integers chosen without replacement from $[n]$, then $\sqrt{n}\left[U_n - \theta\right]$ has a limiting normal distribution with mean 0 and variance $\xi_1 \cdot m^2$, provided $\xi_1 = \mathbf{E}\left[h(X_1, \ldots, X_m)\, h(X_1, X_m, X_{m+1} \ldots, X_{2\,m-1})\right] - \theta^2$ is positive.

The sharpest rate of convergence in the asymptotic normality of U-statistics in the literature is provided in [2] as follows.

Lemma 2.5.2. *If $v = E|h|^3 < \infty$ and $\xi_1 > 0$ then*

$$\sup_{t \in \mathbb{R}} \left| P\left(\frac{\sqrt{n}(U_n - \theta)}{m\,\sqrt{\xi_1}} \leq t \right) - \Phi(t) \right| \leq C \cdot v \cdot \left(m^2\,\xi_1\right)^{-3/2} \cdot n^{-1/2}$$

where C is a constant and $\Phi(\cdot)$ is the standard normal distribution function.

A parameter $\theta(F_1, F_2)$ is said to be *estimable of degree* (r, s) for the family of distributions $\mathcal{F}$, if $r+s$ is the minimum sample size for which there exists a function $h(x_1, \ldots, x_r; y_1, \ldots, y_s)$ such that $\mathbf{E}\left[h(X_1, \ldots, X_r; Y_1, \ldots, Y_s)\right] = \theta(F_1, F_2)$ for distributions $F_1, F_2 \in \mathcal{F}$, where $\left\{X_1, \ldots, X_r\right\}$ and $\left\{Y_1, \ldots, Y_s\right\}$ are random samples from F_1 and F_2, respectively, and $h(X_1, \ldots, X_r; Y_1, \ldots, Y_s)$ is a statistic that does not depend on F_1, F_2.

For an estimable parameter θ of degree (r, s) with a symmetric kernel $h(\cdot)$, a *two-sample U-statistic* has, for $1 \leq r \leq n$ and $1 \leq s \leq m$, the form

$$U(X_1, \ldots, X_n; Y_1, \ldots, Y_m) = \frac{1}{\binom{n}{r}\,\binom{m}{s}} \sum_{\alpha \in C_{r,n}} \sum_{\beta \in C_{s,m}} h(X_{\alpha(1)}, \ldots, X_{\alpha(r)}; Y_{\beta(1)}, \ldots, Y_{\beta(s)})$$

where $C_{r,n}$ ($C_{s,m}$) is the collection of all subsets of r (s) integers chosen without replacement from integers $[n]$ ($[m]$). The variance expression for the two-sample U-statistics is often considerably more complicated than that of one-sample case. For the limiting distribution see Theorem 3.4.13 of Randles and Wolfe [38] and for the rate of convergence see [15]. For more theory of U-statistics, see [25] and [28].

2.6 Consistency of Tests

The statistics based on our PCDs will be used in testing spatial point patterns. Consistency is a reasonable property a test should enjoy.

Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a random sample from a distribution F. As the sample size n increases, we obtain more information about the precise value of the parameter $\theta = \theta(F)$. Reasonable procedures should reflect this fact. In estimation, if a sequence of estimators, $T_n = T_n(\mathcal{X}_n)$ is a "reasonable" estimator of θ, then T_n should be close to θ with higher probability as n increases. More precisely, T_n is called a *weakly consistent estimator* of θ iff $P\big(|T_n - \theta| \geq \varepsilon\big) \to 0$ as $n \to \infty$ for all $\varepsilon > 0$ (i.e., $T_n \to \theta$ in probability). If $T_n \to \theta$ a.s., then T_n is called a *strongly consistent estimator* of θ.

In a hypothesis testing format, a roughly corresponding property may be defined for α-level tests for $\alpha \in (0,1)$. Suppose the joint distribution of $\mathcal{X}_n$ is $F = F_n(\vec{x}, \theta)$ for $\theta \in \Theta$, a general parameter space, and the hypothesis of interest is

$$H_0 : \theta \in \Theta_0 \subsetneq \Theta \text{ versus } H_a : \theta \in \Theta_a = \Theta \setminus \Theta_0.$$

If $\{\delta(\mathcal{X}_n)\}_{n \geq 1}$ is a sequence of α-level tests for testing the above hypothesis, then $\delta(\mathcal{X}_n)$ is said to be *consistent* against $\theta_1 \in \Theta_a$ if $\beta(\theta_1, \delta(\mathcal{X}_n)) \to 1$ as $n \to \infty$, where $\beta(\theta, \delta) = P(\text{reject } H_0 | \theta)$. In particular, if $\theta \in \Theta_0$, $\beta(\theta, \delta)$ is the probability of Type-I error; if $\theta \in \Theta_a$, then $\beta(\theta, \delta)$ is the power against θ. For further details on consistency of tests, see [1] and [22].

For example, for a one-sided test with $\Theta \subseteq \mathbb{R}$, $H_0 : \theta \leq \theta_0$ versus $H_a : \theta > \theta_0$, where $\theta, \theta_0 \in \Theta$, if $\theta_a > \theta_0$, the sequence of tests T_n is consistent if $P_{\theta_a}(T_n > c_n(\alpha, \theta_0)) \to 1$ where $c_n(\alpha, \theta_0)$ is the α-level critical value based on n observations. For $H_0 : \theta \geq \theta_0$ versus $H_a : \theta < \theta_0$, where $\theta, \theta_0 \in \Theta$, and if $\theta_a < \theta_0$, the sequence of tests T_n is consistent if $P_{\theta_a}(T_n < c_n(\alpha, \theta_0)) \to 1$ where $c_n(\alpha, \theta_0)$ is the α-level critical value based on n observations.

Lemma 2.6.1. *Suppose T_n is a sequence of test statistics for testing $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$ with mean $\mu_n = \mathbf{E}[T_n]$ and variance $\sigma_n^2 = \mathbf{Var}(T_n)$ (both of which are assumed to exist) where $\mu_n \to \mu$ and $\sqrt{n}\,\sigma_n \to \sigma$ in probability. Suppose also that $\sqrt{n}\,(T_n - \mu_n)/\sigma_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, the standard normal distribution. Then $\{T_n\}$ is consistent for all $\mu > \mu_0$.*

Proof: Under both the null and alternative hypotheses $\sigma_n^2 \to 0$ as $n \to \infty$. Let $c_n(\alpha, \mu_0)$ be the α-level critical value. Then $c_n(\alpha, \mu_0) \to \mu_0$ in probability as $n \to \infty$. If $\mu > \mu_0$,

$P_\mu\big(T_n > c_n(\alpha, \mu_0)\big) \to 1$ since $\sigma_n^2 \to 0$; consistency follows. ■

2.7 Comparison of Tests

The test statistics based on the relative density of the PCDs are parametrized. The optimal parameter for the hypothesis testing is selected by using asymptotic relative efficacy analysis of the tests. Note that throughout this dissertation relative efficacy and relative efficiency are used interchangably.

Consider the problem of testing $H_0 : \theta \in \Theta_0 \subsetneq \Theta$ versus $H_a : \theta \in \Theta_a = \Theta \setminus \Theta_0$ with two sequences of test statistics $\{T_n\}$ and $\{S_n\}$. Let θ_a be an arbitrary element in the alternative space Θ_a, and $c_n(\alpha, S)$ and $c_n(\alpha, T)$ denote the α-level critical regions for S_n and T_n, respectively. That is,

$$P_\theta\big(S_n \in c_n(\alpha, S)\big) \leq \alpha, \quad \text{and } P_\theta\big(T_n \in c_n(\alpha, T)\big) \leq \alpha$$

for all $\theta \in \Theta_0$ and all n. Let β be fixed in $(\alpha, 1)$ and define n' and m' to be the minimum sample sizes for which S_n and T_n attain a power of at least β against the alternative θ_a. So

$$P_{\theta_a}\big(S_{n'} \in c_{n'}(\alpha, S)\big) \geq \beta, \quad \text{and } P_{\theta_a}\big(T_{m'} \in c_{m'}(\alpha, T)\big) \geq \beta.$$

Note that $n' = n'(\alpha, \beta, \theta_a)$ and $m' = m'(\alpha, \beta, \theta_a)$ depend on α, β, θ_a and the underlying distribution.

The *relative efficiency* of $\{S_n\}$ with respect to $\{T_n\}$ for $\theta_a \in \Theta_a$ is defined to be

$$e_{S,T} := e_{S,T}(\alpha, \beta, \theta_a) := \frac{m'(\alpha, \beta, \theta_a)}{n'(\alpha, \beta, \theta_a)}.$$

By the nature of the definition, $e_{S,T} > 1$ means that S_n should be preferred over T_n, because it requires fewer observations to get the same power compared to T_n at level α.

2.7.1 Asymptotic Power Function

The relative efficiency of two tests depends on three arguments — the size of the test α, the "distance" (in terms of some parameter θ) between the null and alternative hypotheses, and the sample size n required by the "efficient" (i.e., the most powerful test in the class considered) test. Even if we consider only a few typical values of α, we still need a table of two entries for

the comparison of tests by relative efficiency. To achieve a single quantity to assess efficiency, we must resort to asymptotics.

One straightforward approach involves letting the sample sizes of the two tests considered go to infinity and then take the ratio of the power functions as a measure of relative efficiency. However, if the tests are α-level consistent tests (which is a reasonable requirement) then by definition the power converges to 1, hence the ratio of powers tends to 1 also, as the sample sizes tend to infinity. Thus, for any fixed alternative value θ_a, the relative efficiency will always be 1 as $n \to \infty$. Hence the asymptotics in this sense is quite useless. Thus statisticians consider limiting results of other parameters (together with letting the sample size tend to infinity) which give rise to the *asymptotic relative efficacy* (ARE) measures discussed in section 2.7.2.

However, we can still make use of asymptotic critical values for finite sample size tests. Let T_1 and T_2 be two consistent test statistics for the hypotheses $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$. Suppose T_1 and T_2 are asymptotically normal for all values of θ. Let $\mu_{j0} = \mathbf{E}[T_j|H_0]$, $\sigma_{j0}^2 = \mathbf{Var}[T_j|H_0]$, $\mu_{ja} = \mathbf{E}[T_j|H_a]$, and $\sigma_{ja}^2 = \mathbf{Var}[T_j|H_a]$ for $j \in \{1, 2\}$. Then large sample α-level tests have the critical region $c_j^\alpha = \{t \in \mathbb{R} : t > \mu_{j0} + z_\alpha \cdot \sigma_{j0}/\sqrt{n}\}$, where $z_\alpha = \Phi^{-1}(1-\alpha)$. Therefore the asymptotic power function for T_j is

$$P_{n,j}(\theta) = \Phi\left(\sqrt{n}\left[\mu_{j1} - (\mu_{j0} + z_\alpha \cdot \sigma_{j0}/\sqrt{n})\right]/\sigma_{j1}\right), \text{for } j \in \{1, 2\}.$$

(See, e.g., Kendall and Stuart, [22].) Note that this power expression is calculated with an appeal to asymptotic normality using the asymptotic critical value, but is still a function of the sample size.

2.7.2 Asymptotic Relative Efficiency

To evaluate the relative efficiency of two test procedures for large samples, a variety of approaches exist in the statistical literature. The earliest investigation of asymptotic relative efficiency is due to Pitman [34]. These approaches differ depending on the constraints on Type-I and Type-II errors for large samples and/or alternatives. Many statisticians worked in this area, owing to the variety of these constraints; some of these statisticians who introduced new techniques include Pitman, Chernoff, Bahadur, Hodges, Lehmann, Hoeffding, Rubin, and Sethuraman. We summarize their work below.

Under a specified distribution, $e_{S,T}(\alpha, \beta, \theta_a)$ has three arguments, so the optimum efficiency

on the entire space of $(\alpha, \beta, \theta_a)$ is usually too difficult, if not impossible, to calculate. Hence, limiting values of $e_{S,T}(\alpha, \beta, \theta_a)$ are proposed to compare the efficiency of test statistics. Following are five widely known asymptotic (relative) efficiency measures.

1. *Bahadur Asymptotic Relative Efficiency* (ARE) of $\{S_n\}$ with respect to $\{T_n\}$ is defined to be

$$e^B_{S,T}(\beta, \theta_a) := \lim_{\alpha \to 0} e_{S,T}(\alpha, \beta, \theta_a),$$

   if the limit exists for $\beta \in (0,1)$ and $\theta_a \in \Theta_a$.

2. *Hodges-Lehmann ARE* of $\{S_n\}$ with respect to $\{T_n\}$ is defined to be

$$e^{HL}_{S,T}(\alpha, \theta_a) := \lim_{\beta \to 1} e_{S,T}(\alpha, \beta, \theta_a),$$

   if the limit exists for $\alpha \in (0,1)$ and $\theta_a \in \Theta_a$.

3. *Pitman ARE* of $\{S_n\}$ with respect to $\{T_n\}$ is defined to be

$$e^P_{S,T}(\alpha, \beta, \theta_0) := \lim_{\theta_a \to \theta_0} e_{S,T}(\alpha, \beta, \theta_a),$$

   if the limit exists for $0 < \alpha < \beta < 1$ and $\theta_a \to \theta_0 \in \partial(\Theta_0)$, the boundary of the null space Θ_0.

4. *Chernoff ARE* of $\{S_n\}$ with respect to $\{T_n\}$ is defined to be

$$e^C_{S,T}(\theta_a) := \lim_{\substack{\alpha \to 0, \\ \beta \to 1}} e_{S,T}(\alpha, \beta, \theta_a),$$

   if the limit exists for $0 < \alpha < \beta < 1$ and $\theta_a \in \Theta_a$.

5. *Rubin-Sethuraman ARE* of $\{S_n\}$ with respect to $\{T_n\}$ is defined to be

$$e^{RS}_{S,T}(\theta_0) := \lim_{\substack{\alpha \to 0, \\ \beta \to 1, \\ \theta_a \to \theta_0}} e_{S,T}(\alpha, \beta, \theta_a),$$

   if the limit exists for $0 < \alpha < \beta < 1$ and $\theta_a \to \theta_0 \in \partial(\Theta_0)$.

Bahadur ARE usually does not depend on β, Hodges-Lehmann ARE does not depend on α, and Pitman ARE usually does not depend on α or β. Chernoff ARE falls somewhere intermediate

among the above ARE's. Rubin-Sethuraman ARE uses the concept of Bayes risk. (See [45] for the details.) Each type of asymptotic relative efficiency has appeal and motivation of its own.

From a practical point of view, tests with small levels, high power, and alternatives close to the null are important, so Pitman ARE is widely used in the literature. Hodges-Lehmann ARE is also desirable, because it is of practical interest to compare the power under fixed alternatives in some contexts.

2.7.2.1 Pitman Asymptotic Efficacy

Suppose that the distribution F under consideration may be indexed by a set $\Theta \subseteq \mathbb{R}$ and consider $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$.

Consider the test sequence $S = \{S_n\}$ satisfying the following conditions for θ in a neighborhood $[\theta_0, \theta_0 + \delta]$ of the null parameter for some $\delta > 0$.

Pitman's Conditions:

(PC1) For some functions $\mu_n(\theta)$ and $\sigma_n(\theta)$, the distribution F_θ of $[S_n - \mu_n(\theta)]/\sigma_n(\theta)$ converges to $\mathcal{N}(0,1)$ uniformly on $[\theta_0, \theta_0 + \delta]$; i.e.,

$$\sup_{\theta_0 \le \theta \le \theta_0 + \delta} \sup_{t \in \mathbb{R}} \left| P\left(\frac{S_n - \mu_n(\theta)}{\sigma_n(\theta)} \le t \right) - \Phi(t) \right| \to 0 \text{ as } n \to \infty.$$

(PC2) For $\theta \in [\theta_0, \theta_0 + \delta]$, $\mu_n(\theta)$ is differentiable with $\mu_n'(\theta_0) > 0$.

(PC3) For $\theta_n = \theta_0 + O\left(n^{-1/2}\right)$, $\lim_{n\to\infty} \frac{\mu_n'(\theta_n)}{\mu_n'(\theta_0)} = 1$.

(PC4) For $\theta_n = \theta_0 + O\left(n^{-1/2}\right)$, $\lim_{n\to\infty} \frac{\sigma_n(\theta_n)}{\sigma_n(\theta_0)} = 1$.

(PC5) For some constant $C_S > 0$,

$$\lim_{n\to\infty} \frac{\mu_n'(\theta_0)}{\sqrt{n}\,\sigma_n(\theta_0)} = C_S.$$

Condition (PC1) is equivalent to

(PC1)′ For some functions $\mu_n(\theta)$ and $\sigma_n(\theta)$ and $\theta_n = \theta_0 + O\left(n^{-1/2}\right)$, the distribution F_θ of $[S_n - \mu_n(\theta_n)]/\sigma_n(\theta_n)$ converges to a standard normal distribution uniformly (see [13]).

Note that if $\mu_n^{(k)}(\theta_0) > 0$ and $\mu_n^{(l)}(\theta_0) = 0$, for all $l = 1, 2, \ldots, k-1$, then $\mu_n'(\theta_0) > 0$ in (PC2) can be replaced by $\mu_n^{(k)}(\theta_0) > 0$ and $\mu_n'(\theta_n)$ in (PC3) and (PC5) can be replaced by $\mu_n^{(k)}(\theta_n)$ (see [22]).

Lemma 2.7.1. *(Pitman-Noether)*

(i) Let $S = \{S_n\}$ satisfy (PC1)-(PC5). Consider testing $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ by the critical regions $S_n > c_{\alpha_n}$ with $\alpha_n = P_{\theta_0}(S_n > c_{\alpha_n}) \to \alpha$ as $n \to \infty$ where $\alpha \in (0,1)$. For $\beta \in (0,1)$ and $\theta_n = \theta_0 + O(n^{-1/2})$, we have

$$\beta_n(\theta_n) = P_{\theta_n}(S_n > c_{\alpha_n}) \to \beta \text{ iff } C_S\sqrt{n}(\theta_n - \theta) \to \Phi^{-1}(1-\alpha) - \Phi^{-1}(\beta)$$

where C_S is defined in (PC5).

(ii) Let $S = \{S_n\}$ and $Q = \{Q_n\}$ each satisfy satisfy (PC1)–(PC5). *Then the asymptotic relative efficiency of S relative to Q is given by $ARE(S,Q) = (C_S/C_Q)^2$.*

Thus, to evaluate $ARE(S,Q)$ under the conditions (PC1)–(PC5), we need only calculate the quantities C_S and C_Q, $PAE(S) = C_S^2$ is called the *Pitman Asymptotic Efficacy* (PAE) of the test based on S_n. Using similar notation and terminology for Q_n,

$$ARE(S,Q) = \frac{\text{PAE}(S)}{\text{PAE}(Q)}.$$

Let $\{X_1, \ldots, X_n\}$ be a random sample (i.e. a set of iid random variables) from a distribution $F(x-\theta)$, which is continuous and symmetric around θ with the additional assumption that $F(x) + F(-x) = 1$ for all x. The hypothesis test of interest is $H_0 : \theta = 0$ versus $H_a : \theta > 0$. The Pitman alternatives are $\theta_i = \theta_0 + c/\sqrt{n_i}$, which implies $F(x-\theta_i) = F(x - \theta_0 - (c/\sqrt{n_i}))$ for $i = 1, 2, 3, \ldots$ where c is a constant. We compare the Student's t-test $T_n = \overline{X}/s$ with $s^2 = \frac{1}{(n-1)}\sum_{i=1}^n (X_i - \overline{X})^2$ to the sign test statistic, $S_n = \frac{1}{n}\sum_{i=1}^n \mathbf{I}(X_i > 0)$. The $ARE(T_n, S_n) = \frac{\text{PAE}(T_n)}{\text{PAE}(S_n)} = \frac{\pi}{2}$. See [13] for the details.

2.7.2.2 Hodges-Lehmann Asymptotic Efficacy

In 1956, Hodges and Lehmann proposed a comparison of the speed at which the power tends to 1 under a fixed alternative (see [16]). With the same test setting as above, consider the comparison of test sequences $S = \{S_n\}$ and $T = \{T_n\}$ for a fixed $\theta_a > \theta_0$. Let $\beta_S(n, \alpha, \theta_a)$ and $\beta_T(n, \alpha, \theta_a)$ be the power at level α for test S and T, respectively. Suppose $(1 - \beta_S(n, \alpha, \theta_a))^{1/n} \to A_S(\alpha, \theta_a)$ and $(1 - \beta_T(n, \alpha, \theta_a))^{1/n} \to A_T(\alpha, \theta_a)$ as $n \to \infty$. Then it is easy to see that $N_T/N_S =$

$\log(A_T(\alpha,\theta_a))/\log(A_S(\alpha,\theta_a))$, where N_T and N_S are the sample sizes required by the tests T and S to attain the same power.

Example 2.7.2. For testing $H_0 : \theta = 0$ versus $H_a : \theta > 0$ for $X_i \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ using $S_n := \overline{X}$, reject H_0 when $\sqrt{n} \cdot S_n > z_\alpha$ where $\Phi(z_\alpha) = 1 - \alpha$, and power is $\beta(n, \alpha, \theta_a) = 1 - \Phi(z_\alpha - \sqrt{n}\,\theta_a)$ where $\theta_a > 0$ is a fixed alternative. Now, let $n \to \infty$ when α and θ_a are fixed; $1 - \beta = P_{II}$ (probability of Type II error) is equivalent to

$$\frac{1}{\sqrt{n}\,\theta_a} \cdot \frac{1}{\sqrt{2\,\pi}} \exp\left(-\theta_a^2/2\right).$$

Then $\lim_{n\to\infty} \sqrt[n]{P_{II}} = \exp\left(-\theta_a^2/2\right)$.

With the sign test, $T_n := \sum_{i=1}^n \mathbf{I}(X_i > 0)$, reject H_0 for high number of positive X_i. The test statistic T_n is distributed as a binomial with $p = 1/2$ under H_0 and $p = \Phi(\theta_a)$ under H_a. Therefore, $(1 - \beta_T(n, \alpha, \theta_a))^{1/n} \to 2\sqrt{\Phi(\theta_a)\,(1 - \Phi(\theta_a))}$. Hence $e_{S,T}^{HL} = \frac{-\theta_a^2}{2\log 2 + \log \Phi(\theta_a) + \log(1-\Phi(\theta_a))}$. Notice that this quantity is independent of α, but not θ_a. As $\theta_a \to 0$, $e_{S,T}^{HL} \to \pi/2$, which is the Pitman ARE. See [16] for further details. □

Notice that the idea behind this example can be generalized as follows. Let S_n be a sequence of test statistics used to test $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$. Given $A_S(\alpha, \theta_a) = \lim_{n\to\infty} \sqrt[n]{1 - \beta_S(n, \alpha, \theta_a)}$, we call $2\,|\log(A_S(\alpha, \theta_a)|$ the Hodges-Lehmann Asymptotic Efficacy (HLAE) of S_n, and let $c_{n,\alpha}$ be the critical value for finite sample size n at level α.

Proposition 2.7.3. *Suppose S_n is a sequence of test statistics that converges in law to a normal distribution, i.e., $\sqrt{n}\,S_n \xrightarrow{\mathcal{L}} \mathcal{N}(\theta, \sigma^2)$ for $\theta \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$. Let $\beta_S(n, \alpha, \theta) := P_\theta(S_n > c_{n,\alpha})$ be the power of S_n at the given level α. Then $A_S(\alpha, \theta_a) = \exp\left(-\left(\frac{(\theta_a - \theta_0)^2}{2\,\sigma^2}\right)\right)$; i.e., HLAE of S_n is $(\theta_a - \theta_0)^2/\sigma^2$.*

Proof: Let $\widetilde{S}_n := (S_n - \theta_0)/\sigma$. Then $\widetilde{S}_n \xrightarrow{\mathcal{L}} \mathcal{N}\left((\theta - \theta_0)/\sigma, 1\right)$, and $\lim_{n\to\infty} c_{n,\alpha} = z_\alpha$. So the hypotheses become $H_0 : \mathbf{E}\left[\widetilde{S}_n\right] = 0$ versus $H_a : \mathbf{E}\left[\widetilde{S}_n\right] > 0$. Let $\theta = (\theta_a - \theta_0)/\sigma$ for some $\theta_a > \theta_0$. Then for sufficiently large n, $\beta(n, \alpha, \theta) \approx 1 - \Phi(z_\alpha - \sqrt{n}\,\theta)$. Now, letting $n \to \infty$ for fixed α and θ_a, $P_{II} = 1 - \beta$ in asymptotics as in Example 2.7.2. Then $\lim_{n\to\infty} \sqrt[n]{P_{II}} = \exp\left(-\theta^2/2\right)$. Hence HLAE of S_n is $\theta^2 = (\theta_a - \theta_0)^2/\sigma^2$. ■

Part II

Theory and Applications

CHAPTER 3

Proximity Maps and the Associated Γ_1-Regions

In this chapter, we note some appealing properties of the proximity map associated with CCCD for data in a compact interval in $\mathbb{R}$ and use them as guidelines for defining new proximity maps in higher dimensions. Furthermore, we introduce the auxiliary tools used for the construction of the new proximity maps, as well as some related concepts that will be used in the investigation and comparison of the proximity maps.

3.1 Preliminaries and Foundation

Let $\mathcal{Y} = \{\mathsf{y}_1, \ldots, \mathsf{y}_m\} \subset \mathbb{R}$. Then the proximity map associated with CCCD is defined as the open ball $N_S(x) := B(x, r(x))$ for all $x \in \mathbb{R} \setminus \mathcal{Y}$, where $r(x) = \min_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$ (see Section 1.1.1 and [35]) with $d(x, y)$ being the Euclidean distance between x and y. For $x \in \mathcal{Y}$, define $N_S(x) = \{x\}$. Notice that a ball is a sphere in higher dimensions, hence the notation N_S. Furthermore, dependence on $\mathcal{Y}$ is through $r(x)$. Note that, this proximity map is based on the intervals $I_j = \left(\mathsf{y}_{(j-1):m}, \mathsf{y}_{j:m}\right)$ for $j = 0, \ldots, m+1$ with $\mathsf{y}_{0:m} = -\infty$ and $\mathsf{y}_{(m+1):m} = \infty$ where $y_{j:m}$ is the j^{th} order statistic in $\mathcal{Y}$. Our next goal is to extend this concept to higher dimensions and investigate the properties of the associated PCDs.

A natural extension of the proximity region $N_S(x)$ to multiple dimensions (i.e., to $\mathbb{R}^d$ with $d > 1$) is obtained by the same definition as above; that is, $N_S(x) := B(x, r(x))$ where $r(x) := \min_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$. We will call it *spherical proximity map*. The spherical proximity map $N_S(x)$ is well-defined for all $x \in \mathbb{R}^d$ provided that $\mathcal{Y} \neq \emptyset$. Extensions to $\mathbb{R}^2$ and higher dimensions with the spherical proximity map — with applications in classification — are investigated in [9], [10], [29], [36], [37]. However, finding the minimum dominating set of the PCD associated with $N_S(\cdot)$ is an NP-hard problem and the distribution of the domination number is not analytically tractable for $d > 1$. This drawback has motivated us to define new types of proximity maps in higher dimensions. Note that for $d = 1$, such problems do not occur. Hence we state some

appealing properties of the proximity map $N_S(x) = B(x, r(x))$ in $\mathbb{R}$ and use them as guidelines for our definition of new proximity maps.

Some appealing properties of the proximity map $N_S(x)$ in $\mathbb{R}$ are noted below.

P1 $N_S(x)$ is well-defined for all $x \in \mathcal{C}_H(\mathcal{Y}) = [\mathsf{y}_{1:m}, \mathsf{y}_{m:m}]$.

P2 $x \in N_S(x)$ for all $x \in \mathcal{C}_H(\mathcal{Y})$.

P3 The point x is at the *center* of $N_S(x)$ for all $x \in \mathcal{C}_H(\mathcal{Y})$.

P4 For $x \in I_j \subseteq \mathcal{C}_H(\mathcal{Y})$, $N_S(x)$ and I_j are of the *same type*; they are both intervals.

P5 For $x \in I_j \subseteq \mathcal{C}_H(\mathcal{Y})$, $N_S(x)$ mimics the shape of I_j; i.e., it is *similar* to I_j.

P6 For $x \in I_j$, $N_S(x)$ is a proper subset of I_j for all $x \in I_j \setminus \{(\mathsf{y}_{(j-1):m} + \mathsf{y}_{j:m})/2\}$ (or almost everywhere in I_j).

P7 For $x \in I_j$ and $y \in I_k$ with $j \neq k$, $N_S(x)$ and $N_S(y)$ are disjoint regions.

P8 The size of $N_S(x)$ is continuous in x; that is, for each $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that $\big||N_S(y)| - |N_S(x)|\big| < \varepsilon$ whenever $|d(x, y)| < \delta(\varepsilon)$.

Notice that properties **P1**, **P2**, and **P3** also hold for all $x \in \mathbb{R}$.

Partitioning of the convex hull of $\mathcal{Y}$, $\mathcal{C}_H(\mathcal{Y})$, in $\mathbb{R}$ to intervals can be viewed as a tessellation. For $d > 1$, a natural tessellation that partitions $\mathcal{C}_H(\mathcal{Y})$ is the Delaunay tessellation (see Section 2.1), where each Delaunay cell is a $(d+1)$-simplex. Let $\mathcal{T}_j$ be the j^{th} Delaunay cell in the Delaunay tessellation of $\mathcal{Y}$ for $j = 1, \ldots, J$, where J is the total number of Delaunay cells. In $\mathbb{R}$, we implicitly use the Delaunay cell that contains x to define the proximity map for $x \in \mathcal{C}_H(\mathcal{Y})$.

Note that **P5** and **P4** are equivalent when $d = 1$ for $x \in \mathcal{C}_H(\mathcal{Y})$, since any two (compact) intervals are (geometrically) similar in $\mathbb{R}$. For $d > 1$, **P5** implies **P4** only, since, for example, for $d = 2$, any two triangles are not necessarily similar, but similar triangles are always of the same type; they are triangles.

Notice that $N_S(\cdot)$ satisfies only **P1**, **P2**, **P3**, and **P8** in $\mathbb{R}^d$ with $d > 1$. For any $x \in \mathcal{T}_j \subset \mathbb{R}^d$, $B(x, r(x)) \not\subset \mathcal{T}_j$, which implies that two proximity regions $N_S(x)$ and $N_S(y)$ might overlap for x, y from two distinct cells, hence **P7** is violated. Such an overlap of the regions make the distribution of the domination number of the PCD associated with $N_S(\cdot)$, if not impossible, hard to calculate. In order to avoid the overlap of regions $B(x, r(x))$ and $B(y, r(y))$ for x, y in different

Delaunay cells, we restrict the balls to the corresponding cells, which leads to *arc-slice proximity regions*, $N_{AS}(x) := \overline{B}(x, r(x)) \cap \mathcal{T}_j$, where $\overline{B}(x, r(x))$ is the closure of the ball $B(x, r(x))$. We chose the closed ball in the definition of the arc-slice proximity map for consistency with the other proximity maps that we will define on Delaunay cells. The arc-slice proximity map $N_{AS}(x)$ is well-defined only in $\mathcal{C}_H(\mathcal{Y})$, provided that $\mathcal{Y}$ is in general position and $|\mathcal{Y}| \geq d+1$ in $\mathbb{R}^d$. So **P1** holds. $N_{AS}(\cdot)$ also satisfies **P2**, **P6**, **P7**, and **P8**. We define new types of proximity maps that satisfy more of the properties **P1**-**P8** and introduce new explanatory concepts for investigating the properties of these proximity maps. We focus on possible extensions to multiple dimensions based on the Delaunay tessellation of $\mathcal{Y}$, hence the extensions will also be well-defined only in $\mathcal{C}_H(\mathcal{Y})$.

The appealing properties mentioned above can be extended to more general measurable spaces.

Property **P6** suggests a new concept. For $x \in I_j$, $N_S(x) = I_j$ iff $x = \left(\mathsf{y}_{(j-1):m} + \mathsf{y}_{j:m}\right)/2$. We define an associated region for such points in the general context.

Definition 3.1.1. The *superset region* for any proximity map $N(\cdot)$ in Ω is defined to be $\mathscr{R}_S(N) := \{x \in \Omega : N(x) = \Omega\}$.

For example, for $\Omega = I_j \subsetneq \mathbb{R}$, $\mathscr{R}_S(N_S) := \{x \in I_j : N_S(x) = I_j\} = \left\{\left(\mathsf{y}_{(j-1):m} + \mathsf{y}_{j:m}\right)/2\right\}$, and for $\Omega = \mathcal{T}_j \subsetneq \mathbb{R}^d$, $\mathscr{R}_S(N_S) := \{x \in \mathcal{T}_j : N_S(x) = \mathcal{T}_j\}$. Note that for $x \in I_j$, $\lambda(N_S(x)) \leq \lambda(I_j)$ and $\lambda(N_S(x)) = \lambda(I_j)$ iff $x \in \mathscr{R}_S(N_S)$ where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}$ (also called as $\mathbb{R}$-Lebesgue measure). So the proximity region of a point in $\mathscr{R}_S(N_S)$ has the largest $\mathbb{R}$-Lebesgue measure. Note also that given $\mathcal{Y}$, $\mathscr{R}_S(N_S)$ is not a random set, but $\mathbf{I}(X \in \mathscr{R}_S(N_S))$ is a random variable. Furthermore, **P6** is equivalent to $\mathscr{R}_S(N_S)$ having zero $\mathbb{R}$-Lebesgue measure. On the other hand, for $x \in \partial(I_j) = \{\mathsf{y}_{(j-1):m}, \mathsf{y}_j\}$, the proximity region $N_S(x) = \{x\}$ has zero $\mathbb{R}$-Lebesgue measure. This suggests the following concept.

Definition 3.1.2. Let (Ω, μ) be a measurable space. The Λ_0*-region* for any proximity map $N(\cdot)$ is defined to be $\Lambda_0(N) := \{x \in \Omega : \mu(N(x)) = 0\}$.

For $\Omega = \mathbb{R}^d$, $\Lambda_0(N_S) := \{x \in \mathbb{R}^d : \lambda(N_S(x)) = 0\}$. For example, for $\Omega = \mathcal{C}_H(\mathcal{Y}) \subsetneq \mathbb{R}$, $\Lambda_0(N_S) = \mathcal{Y}$, since $\lambda(N_S(x)) = 0$ iff $x \in \mathcal{Y}$.

Furthermore, given a set B of size n in $[\mathsf{y}_{1:m}, \mathsf{y}_{m:m}] \setminus \mathcal{Y}$, **P7** implies that the number of disconnected components in the PCD based on $N_S(\cdot)$ is at least the cardinality of $\{j \in [m] :$

$B \cap I_j \neq \emptyset\}$, which is the set of indices of the intervals that contain some point(s) from B.

3.2 Vertex and Edge Regions

Our new proximity maps will be based on the Delaunay cell $\mathcal{T}_j$ that contains x. The region $N_{\mathcal{Y}}(x)$ will also depend on the location of x in $\mathcal{T}_j$ with respect to the vertices or faces (edges in $\mathbb{R}^2$) of $\mathcal{T}_j$. Hence for $N_{\mathcal{Y}}(x)$ to be well-defined, the vertex or face of $\mathcal{T}_j$ associated with x should be uniquely determined. This will give rise to two new concepts: vertex regions and face regions (edge regions in $\mathbb{R}^2$).

3.2.1 Vertex Regions

Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ be the triangle with vertices $\mathcal{Y}$. Then for $x \in T(\mathcal{Y})$, $N_{AS}(x) = \overline{B}(x, r(x)) \cap T(\mathcal{Y})$ where $r(x) = \min_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$. That is, $r(x) = d(x, \mathsf{y}_j)$ iff $x \in \mathcal{V}_C(\mathsf{y}_j) \cap T(\mathcal{Y})$ for $j \in \{1, 2, 3\}$, where $\mathcal{V}_C(\mathsf{y}_j)$ is the Voronoi cell generated by y_j in the Voronoi tessellation based on $\mathcal{Y}$. Notice that these cells partition the triangle $T(\mathcal{Y})$ (in the sense that the intersection of the cells have zero $\mathbb{R}^2$-Lebesgue measure) and each $\mathcal{V}_C(\mathsf{y}_j) \cap T(\mathcal{Y})$ is adjacent only to vertex y_j and if $T(\mathcal{Y})$ is non-obtuse, their intersection is the point M which is equidistant to the vertices, so M is in fact the circumcenter, M_{CC}, of $T(\mathcal{Y})$. To define new proximity regions based on some sort of distance or dissimilarity relative to the vertices $\mathcal{Y}$, we associate each point in $T(\mathcal{Y})$ to a vertex of $T(\mathcal{Y})$ as in the arc-slice case. This gives rise to the concept of *vertex regions*. Note that $N_{AS}(x)$ is constructed using the vertex region based on the closest vertex, $\operatorname{argmin}_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$. If two vertices were equidistant from x (i.e. $\operatorname{argmin}_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$ were not unique), we arbitrarily assign x to a vertex region. In fact, for N_{AS}, by construction, it would not matter which vertex to pick when the vertices are equidistant to x, the region $N_{AS}(x)$ will be the same.

Definition 3.2.1. The connected regions that partition the triangle, $T(\mathcal{Y})$, (in the sense that the intersections of the regions have zero $\mathbb{R}^2$-Lebesgue measure) such that each region has one and only one vertex of $T(\mathcal{Y})$ on its boundary are called *vertex regions*.

This definition implies that we have three regions. In fact, we can describe the vertex regions starting with a point $M \in \mathbb{R}^2 \setminus \mathcal{Y}$. Join the point M to a point on each edge by a curve such that the resultant regions satisfy the above definition. We call such regions *M-vertex regions*

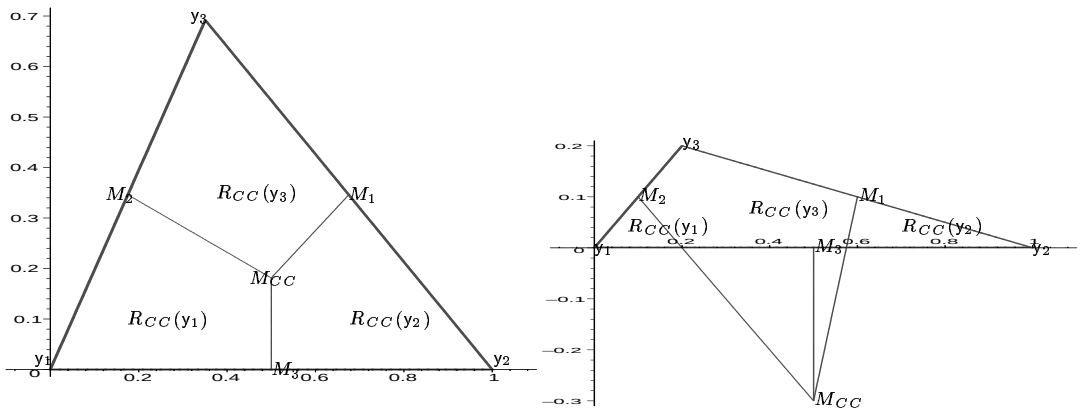


Figure 3.2.1: The CC-vertex regions in an acute triangle (left) and in an obtuse triangle (right).

and denote the vertex region associated with vertex y as $R_M(\mathsf{y})$ for $\mathsf{y} \in \mathcal{Y}$. In particular, we use a *center* of $T(\mathcal{Y})$ as the starting point M for vertex regions. See the discussion of triangle centers in Section 2.3. We think of the points in $R_M(\mathsf{y})$ as being "closer" to y than to the other vertices.

It is reasonable to require that the area of the region $R_M(\mathsf{y})$ gets larger as $d(M, \mathsf{y})$ increases. Usually the curves will be taken to be linear or even the orthogonal projections to the edges. But these lines do not necessarily yield three vertex regions for M in the exterior of $T(\mathcal{Y})$. Unless stated otherwise, M-vertex regions will refer to regions constructed by joining M to the edges with *straight line segments.*

3.2.1.1 CC-Vertex Regions

The region $\mathcal{V}_C(\mathsf{y}) \cap T(\mathcal{Y})$ is a special type of vertex regions, which can also be obtained geometrically by starting at M_{CC} and drawing the orthogonal projections to the edges. Hence we will call these regions *CC-vertex regions.* We can also construct CC-vertex regions by drawing the perpendicular (mid)edge bisectors. See Figure 3.2.1, where M_j are the midpoints of the edges.

The functional forms of $R_{CC}(\mathsf{y}_j)$ for $j \in \{1, 2, 3\}$, in the basic triangle $T_b = \big((0,0), (1,0), (c_1, c_2)\big)$

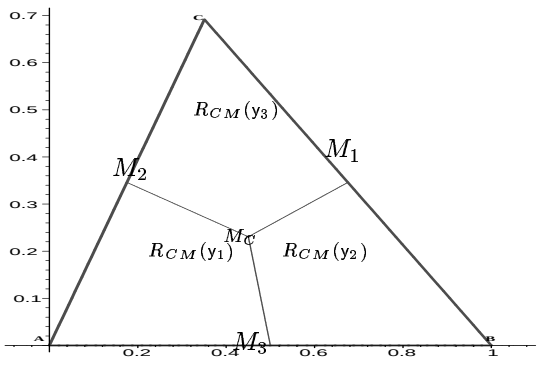


Figure 3.2.2: The CM-vertex regions with median lines.

(see Section 2.2) are given by

$$R_{CC}(y_1) = \left\{(x,y) \in T(\mathcal{Y}) : x \leq \frac{1}{2};\ y \leq \frac{c_1^2 + c_2^2 - 2\,c_1\,x}{2\,c_2}\right\},$$
$$R_{CC}(y_2) = \left\{(x,y) \in T(\mathcal{Y}) : x \geq \frac{1}{2};\ y \leq \frac{c_1^2 + c_2^2 - 1 + 2\,(1+c_1)\,x}{2\,c_2}\right\},$$
$$R_{CC}(y_3) = \left\{(x,y) \in T(\mathcal{Y}) : y \geq \frac{c_1^2 + c_2^2 - 2\,c_1\,x}{2\,c_2};\ y \geq \frac{c_1^2 + c_2^2 - 1 + 2\,(1+c_1)\,x}{2\,c_2}\right\}.$$

If x falls on the boundary of two CC-vertex regions, then x is arbitrarily assigned to one of the CC-vertex regions.

3.2.1.2 CM-Vertex Regions

The motivation behind CM-vertex regions is that unlike the circumcenter, center of mass is guaranteed to be inside the triangle. We define the CM-vertex regions by using the median lines and denote the regions as $R_{CM}(y_j)$ for $j \in \{1,2,3\}$ (see Figure 3.2.2).

The functional forms of $R_{CM}(y_j)$ for $j \in \{1,2,3\}$ in the basic triangle T_b are given by

$$R_{CM}(y_1) = \left\{(x,y) \in T(\mathcal{Y}) : y \leq \frac{c_2\,(2\,x-1)}{2\,c_1 - 1};\ y \leq \frac{c_2\,(x-1)}{c_1 - 2}\right\},$$
$$R_{CM}(y_2) = \left\{(x,y) \in T(\mathcal{Y}) : y \leq \frac{c_2\,(2\,x-1)}{2\,c_1 - 1};\ y \leq \frac{c_2\,x}{1+c_1}\right\},$$
$$R_{CM}(y_3) = \left\{(x,y) \in T(\mathcal{Y}) : y \geq \frac{c_2\,(x-1)}{c_1 - 2};\ y \geq \frac{c_2\,x}{1+c_1}\right\}.$$

If x falls on the boundary of two CM-vertex regions, then x is arbitrarily assigned to one of

the CM-vertex regions.

We can also define the incenter vertex regions by using the inner angle bisectors.

3.2.1.3 M-Vertex Regions

We can also construct M-vertex regions with $M \in T(\mathcal{Y})^o$ by using the extensions of the line segments joining y to M for all $\mathsf{y} \in \mathcal{Y}$. See Figure 3.2.2 with $M = M_C$. The functional forms of $R_M(\mathsf{y}_j)$ for $j \in \{1,2,3\}$ with $M = (m_1, m_2)$ and $m_1 > c_1$ in the basic triangle, T_b, are given by

$$
\begin{aligned}
R_M(\mathsf{y}_1) &= \left\{ (x,y) \in T(\mathcal{Y}) : y \le \frac{m_2\,(x-1)}{m_1 - 1};\ y \le \frac{m_2(c_1 - x) + c_2(x - m_1)}{c_1 - m_1} \right\}, \\
R_M(\mathsf{y}_2) &= \left\{ (x,y) \in T(\mathcal{Y}) : y \le \frac{m_2\,x}{m_1};\ y \ge \frac{m_2(c_1 - x) + c_2(x - m_1)}{c_1 - m_1} \right\}, \text{ and} \\
R_M(\mathsf{y}_3) &= \left\{ (x,y) \in T(\mathcal{Y}) : y \ge \frac{m_2\,x}{m_1};\ y \ge \frac{m_2\,(x-1)}{m_1 - 1} \right\}.
\end{aligned}
$$

For $m_1 < c_1$, $R_M(\mathsf{y}_j)$ for $j \in \{1,2,3\}$ are defined similarly.

If x falls on the boundary of two M-vertex regions, then x is arbitrarily assigned to one of the M-vertex regions.

3.2.2 Edge Regions

The arc-slice proximity region seen earlier is constructed by using the vertex region based on the closest vertex, $\text{argmin}_{\mathsf{y} \in \mathcal{Y}}\, d(x, \mathsf{y})$. We can also use the closest edge, $\text{argmin}_{j \in \{1,2,3\}}\, d(x, e_j)$, in defining a proximity region, which suggests the concept of *edge regions*.

While using the edge $\text{argmin}_{j \in \{1,2,3\}}\, d(x, e_j)$, we again partition the triangle into three regions whose intersection is some point M with Euclidean distance to the edges $d(M, e_1) = d(M, e_2) = d(M, e_3)$, so M is in fact the incenter of $T(\mathcal{Y})$ and $d(M, e) = r_{ic}$ is the inradius (see Section 2.3 for incenter and inradius).

Definition 3.2.2. The connected regions that partition the triangle, $T(\mathcal{Y})$, in such a way that each region has one and only one edge of $T(\mathcal{Y})$ on its boundary, are called *edge regions*.

This definition implies that we have exactly three regions which intersect at only one point, M in $T(\mathcal{Y})^o$, the interior of $T(\mathcal{Y})$. In fact, we can describe the edge regions starting with M. Join the point M to the vertices by curves such that the resultant regions satisfy the above definition. We call such regions *M-edge regions* and denote the edge region for edge e as $R_M(e)$

for $e \in \{e_1, e_2, e_3\}$. Unless stated otherwise, M-edge regions will refer to the regions constructed by joining M to the vertices by straight lines. In particular, we use a *center* of $T(\mathcal{Y})$ for the starting point M as the edge regions. We can also consider the points in $R_M(e)$ to be "closer" to e than to the other edges. Furthermore, it is reasonable to require that the area of the region $R_M(e)$ get larger as $d(M, e)$ increases. Moreover, in higher dimensions, the corresponding regions are called "face regions".

The functional forms of $R_M(e_j)$ for $j \in \{1, 2, 3\}$, for $M = (m_1, m_2) \in T(\mathcal{Y})^o$ and $m_1 > c_1$ in the basic triangle are given by

$$R_M(e_1) = \left\{(x, y) \in T(\mathcal{Y}) : y \geq \frac{m_2\,(x-1)}{m_1 - 1};\ y \geq \frac{c_2\,(x - m_1) - m_2\,(x - c_1)}{c_1 - m_1}\right\},$$
$$R_M(e_2) = \left\{(x, y) \in T(\mathcal{Y}) : y \geq \frac{m_2\,x}{m_1};\ y \leq \frac{c_2\,(x - m_1) - m_2\,(x - c_1)}{c_1 - m_1}\right\}, \text{ and}$$
$$R_M(e_3) = \left\{(x, y) \in T(\mathcal{Y}) : \frac{m_2\,x}{m_1};\ y \leq \frac{m_2\,(x-1)}{m_1 - 1}\right\}.$$

If x falls on the boundary of two M-edge regions, then x is arbitrarily assigned to one of the M-edge regions.

We describe the center of mass edge regions (CM-edge regions) in detail, as we will use them in defining a new class of proximity maps.

3.2.2.1 CM-Edge Regions

We can divide $T(\mathcal{Y})$ into three regions by using the median lines which intersect at the centroid (see Section 2.3), or equivalently, joining the centroid M_C to the vertices by straight lines will yield the CM-edge regions. Let $R_{CM}(e)$ be the region for edge $e \in \{e_1, e_2, e_3\}$. See Figure 3.2.3.

The functional forms of $R_{CM}(e_j)$ for $j \in \{1, 2, 3\}$, in the basic triangle, T_b, are given by

$$R_{CM}(e_1) = \left\{(x, y) \in T(\mathcal{Y}) : y \leq \frac{c_2\,(1 - 2\,x)}{1 - 2\,c_1};\ y \geq \frac{c_2\,(1 - x)}{2 - c_1}\right\},$$
$$R_{CM}(e_2) = \left\{(x, y) \in T(\mathcal{Y}) : y \geq \frac{c_2\,(1 - 2\,x)}{1 - 2\,c_1};\ y \geq \frac{c_2\,x}{1 + c_1}\right\},$$
$$R_{CM}(e_3) = \left\{(x, y) \in T(\mathcal{Y}) : y \leq \frac{c_2\,x}{1 + c_1};\ y \leq \frac{c_2\,(1 - x)}{2 - c_1}\right\}.$$

If x falls on the boundary of two CM-edge regions, then x is arbitrarily assigned to one the CM-edge regions.

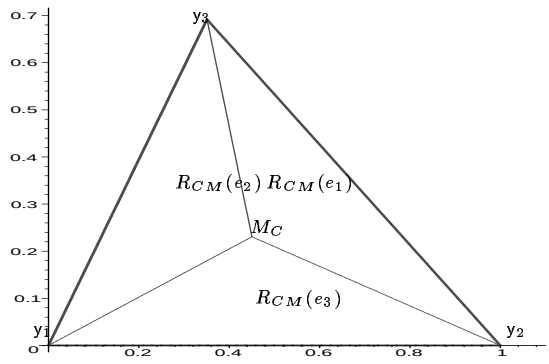


Figure 3.2.3: CM-edge regions $R_{CM}(e_j),\ j \in \{1,2,3\}$.

Remark 3.2.3. We can also divide $T(\mathcal{Y})$ into three regions by using the inner angle bisectors which intersect at the incenter, yielding the incenter edge regions (IC-edge regions). Let $R_{IC}(e)$ be the region for edge $e \in \{e_1, e_2, e_3\}$. Notice that the closest edge to any point in $R_{IC}(e)$ is edge e, i.e., $x \in R_M(e)$ iff $\text{argmin}_{u \in \{e_1,e_2,e_3\}}\, d(x,u) = e$. If two edges are equidistant from x, we arbitrarily assign x to an edge region. □

Remark 3.2.4. In $\mathbb{R}$, we can view the end points of $[0,1]$, $\{0,1\}$, as vertices or edges. So $[0,1/2]$ and $[1/2,1]$ can be viewed as either vertex regions or edge regions. □

3.3 Proximity Maps in Delaunay Triangles

Let $\mathcal{Y} = \{\mathsf{y}_1, \ldots, \mathsf{y}_m\}$ be m points in general position in $\mathbb{R}^d$ and $\mathcal{T}_j$ be the Delaunay cell for $j = 1, \ldots, J$, where J is the number of Delaunay cells. Let $\mathcal{X}_n$ be a random sample (i.e., set of iid random variables) from distribution F in $\mathbb{R}^d$ with support $\mathcal{S}(F) \subseteq \mathcal{C}_H(\mathcal{Y})$. Then the appealing properties for proximity regions in Section 3.1 can be adapted with I_j being replaced by $\mathcal{T}_j$.

In particular, for illustrative purposes, we focus on $\mathbb{R}^2$, where a Delaunay tessellation is a triangulation, provided that no more than three points in $\mathcal{Y}$ are cocircular.

Furthermore, for simplicity, let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ be the triangle with vertices $\mathcal{Y}$. Let $\mathcal{X}_n$ be a random sample from F with support $\mathcal{S}(F) \subseteq T(\mathcal{Y})$. The spherical proximity map is the only type of proximity map defined in literature (see [9], [10], [29], [36], [37]) where CC-vertex regions are implicitly used for points in $\mathcal{C}_H(\mathcal{Y})$. In this section, we will describe arc-slice proximity maps $N_{AS}(\cdot)$ in more detail and

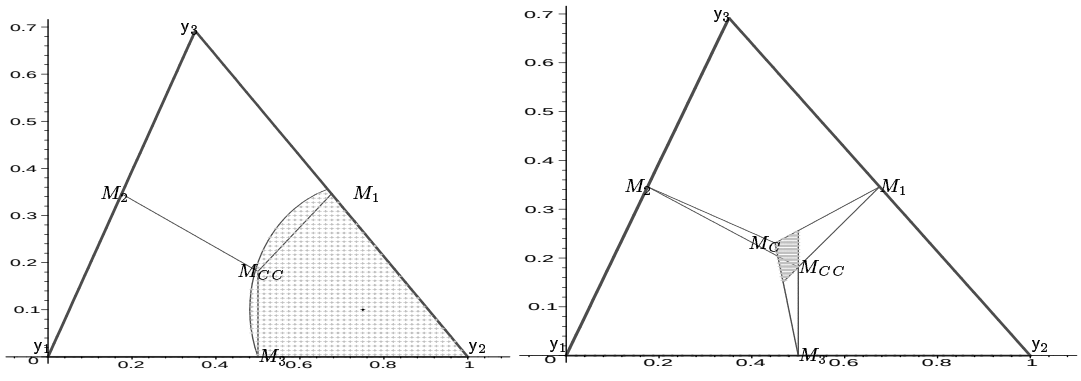


Figure 3.3.1: $N_{AS}(x, M_{CC})$ with an $x \in R_{CC}(\mathsf{y}_2)$ (left) and the superset region $\mathscr{R}_S(N_{AS}, M_C)$ in $T(\mathcal{Y})$ (right).

define two families of triangular proximity regions for which **P4** and **P5** will automatically hold.

3.3.1 Arc-Slice Proximity Maps

Recall that the arc-slice proximity region of $x \in T(\mathcal{Y})$ is given by $N_{AS}(x) := \overline{B}(x, r(x)) \cap T(\mathcal{Y})$ where $\overline{B}(x, r(x))$ is the closed ball centered at x with radius $r(x) := \min_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$. By construction, we implicitly use the CC-vertex regions, since $x \in R_{CC}(\mathsf{y})$ iff $\mathsf{y} = \operatorname{argmin}_{u \in \mathcal{Y}} d(x, u)$. To make this dependence explicit, we use $N_{AS}(\cdot, M_{CC})$. See Figure 3.3.1 (left) for $N_{AS}(x, M_{CC})$ with $x \in R_{CC}(\mathsf{y}_2)$.

The functional form of $N_{AS}(x, M_{CC})$ for an $x \in R_{CC}(\mathsf{y})$ is given by

$$N_{AS}(x, M_{CC}) := \left\{ z \in T(\mathcal{Y}) : d(z, x) \leq r(x) = d(x, \mathsf{y}) \right\}.$$

Notice that, the region $N_{AS}(x, M_{CC})$ is a closed region, unlike $N_S(x) = B(x, r(x))$. The properties **P1**, **P2**, **P7** hold by definition. Notice that $N_{AS}(x, M_{CC}) \subseteq T(\mathcal{Y})$ for all $x \in T(\mathcal{Y})$ and $N_{AS}(x, M_{CC}) = T(\mathcal{Y})$ iff $x = M_{CC}$, since $\overline{B}(x, r(x)) \supset T(\mathcal{Y})$ only when $x = M_{CC}$. Hence the superset region for arc-slice proximity maps with CC-vertex regions is $\mathscr{R}_S(N_{AS}, M_{CC}) = \{M_{CC}\}$. So **P6** follows. Furthermore, **P8** holds since the area $A\left(N_{AS}(x, M_{CC})\right)$ is a continuous function of $r(x) = \min_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$ which is a continuous function of x.

We can define arc-slice proximity regions with any type of M-vertex regions as

$$N_{AS}(x, M) := \overline{B}(x, r(x)) \cap T(\mathcal{Y}) \text{ where } r(x) := d(x, \mathsf{y}) \text{ for } x \in R_M(\mathsf{y}).$$

But for $M \neq M_{CC}$, $N_{AS}(\cdot, M)$ satisfies only **P1**, **P2**, and **P7**. **P6** fails to hold since $\mathscr{R}_S(N_{AS}, M)$ has positive area and **P8** fails since the size of $N_{AS}(x, M)$ is not continuous in x. See, for example, Figure 3.3.1 (right) for $\mathscr{R}_S(N_{AS}, M_C)$. In terms of the properties in Section 3.1, $N_{AS}(\cdot, M_{CC})$ is the most appealing map among $\mathscr{N}_{AS} := \{N_{AS}(\cdot, M) : M \in \mathbb{R}^2 \setminus \mathcal{Y}\}$.

Moreover, $\Lambda_0(N_{AS}, M) = \mathcal{Y}$ for all $M \in \mathbb{R}^2 \setminus \mathcal{Y}$ since $\lambda(N_{AS}(x, M)) = 0$ iff $x \in \mathcal{Y}$.

Next, we define *triangular proximity regions*, which, by definition, will satisfy properties **P4** and **P5**. These proximity regions are the building blocks of the PCDs for which more rigorous mathematical analysis — compared to the PCDs based on spherical and arc-slice proximity maps — will be possible.

3.3.2 r-Factor Proportional-Edge Proximity Maps

The first type of triangular proximity map we introduce is the r-factor proportional-edge proximity map. For this proximity map, the asymptotic distribution of domination number and the relative density of the corresponding PCD will have mathematical tractability.

For $r \in [1, \infty]$, define $N^r_{PE}(\cdot, M) := N(\cdot, M; r, \mathcal{Y})$ to be the *r-factor proportional-edge proximity map* with M-vertex regions as follows (see also Figure 3.3.2 with $M = M_C$ and $r = 2$). For $x \in T(\mathcal{Y}) \setminus \mathcal{Y}$, let $v(x) \in \mathcal{Y}$ be the vertex whose region contains x; i.e., $x \in R_M(v(x))$. If x falls on the boundary of two M-vertex regions, we assign $v(x)$ arbitrarily. Let $e(x)$ be the edge of $T(\mathcal{Y})$ opposite $v(x)$. Let $\ell(v(x), x)$ be the line parallel to $e(x)$ through x. Let $d(v(x), \ell(v(x), x))$ be the Euclidean (perpendicular) distance from $v(x)$ to $\ell(v(x), x)$. For $r \in [1, \infty)$, let $\ell_r(v(x), x)$ be the line parallel to $e(x)$ such that

$$d(v(x), \ell_r(v(x), x)) = r\, d(v(x), \ell(v(x), x))$$

and

$$d(\ell(v(x), x), \ell_r(v(x), x)) < d(v(x), \ell_r(v(x), x)).$$

Let $T_r(x)$ be the triangle similar to and with the same orientation as $T(\mathcal{Y})$ having $v(x)$ as a vertex and $\ell_r(v(x), x)$ as the opposite edge. Then the r-factor proportional-edge proximity region $N^r_{PE}(x, M)$ is defined to be $T_r(x) \cap T(\mathcal{Y})$. Notice that $\ell(v(x), x)$ divides the edges of $T_r(x)$ (other than $\ell_r(v(x), x)$) proportionally with the factor r. Hence the name *r-factor proportional edge proximity region.*

Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M)$. Furthermore, $\lim_{r\to\infty} N_{PE}^r(x, M) = T(\mathcal{Y})$ for all $x \in T(\mathcal{Y}) \setminus \mathcal{Y}$, so we define $N_{PE}^\infty(x, M) = T(\mathcal{Y})$ for all such x. For $x \in \mathcal{Y}$, we define $N_{PE}^r(x, M) = \{x\}$ for all $r \in [1, \infty]$.

Notice that $X_i \overset{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional probability density function f exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y})$, implies that the special case in the construction of N_{PE}^r — X falls on the boundary of two vertex regions — occurs with probability zero. Note that for such an F, $N_{PE}^r(X)$ is a triangle a.s.

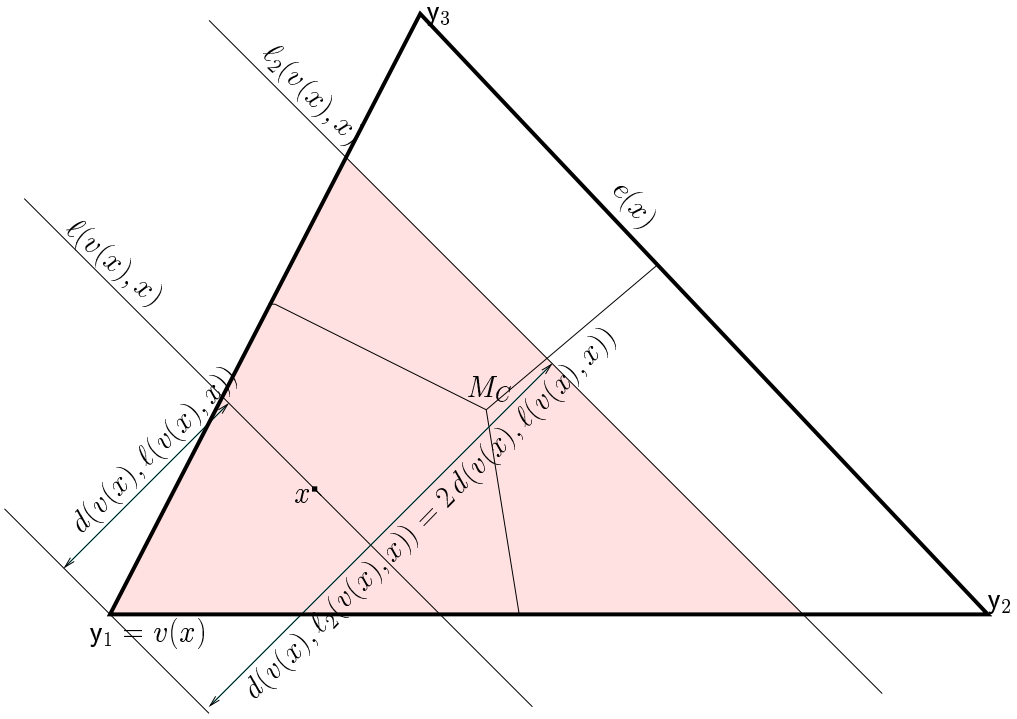


Figure 3.3.2: Construction of r-factor proximity region, $N_{PE}^2(x)$ (shaded region).

The functional form of $N_{PE}^r(x, M)$ for $x = (x_0, y_0) \in T_b$ is given by

$$N_{PE}^r(x, M) = \left\{(x, y) \in T_b : y \leq \tfrac{r\,(y_0\,(1-c_1)+c_2\,x_0)-c_2\,x}{1-c_1}\right\} \text{ for } x \in R_M(\mathsf{y}_1),$$

$$N_{PE}^r(x, M) = \left\{(x, y) \in T_b : y \leq \tfrac{r\,(y_0\,c_1-c_2\,(x_0+1))-c_2+c_2\,x}{c_1}\right\} \text{ for } x \in R_M(\mathsf{y}_2),$$

$$N_{PE}^r(x, M) = \{(x, y) \in T_b : y \geq 2\,y_0 - c_2(r-1)\} \text{ for } x \in R_M(\mathsf{y}_3).$$

Of particular interest is $N_{PE}^r(x, M)$ with any M and $r \in \{\sqrt{2},\, 3/2, 2\}$. For $r = \sqrt{2}$, $\ell(v(x), x)$ divides $T_{\sqrt{2}}(x)$ into two regions of equal area, hence $N_{PE}^{\sqrt{2}}(x, M)$ is also referred to as *double-area proximity region.* For $r = 2$, $\ell(v(x), x)$ divides the edges of $T_2(x)$ —other than $\ell_r(v(x), x)$ —

into two segments of equal length, hence $N_{PE}^2(x, M)$ is also referred to as *double-edge proximity region*. For $r < 3/2$, $\mathscr{R}_S(N_{PE}^r, M_C) = \emptyset$, and for $r > 3/2$, $\mathscr{R}_S(N_{PE}^r, M_C)$ has positive area; for $r = 3/2$, $\mathscr{R}_S(N_{PE}^r, M_C) = \{M_C\}$. Therefore, $r = 3/2$ is the threshold for $N_{PE}^r(\cdot, M_C)$ to satisfy **P6**. Furthermore, we will see that $r = 3/2$ will be the value at which the asymptotic distribution of the domination number of the PCD based on $N_{PE}^r(\cdot, M_C)$ will be nondegenerate (see Section 4.4 and Equation 4.4.2).

The properties **P1**, **P2**, **P4**, **P5**, and **P7** follow by definition for all M and r. Property **P5** holds with similarity ratio of $N_{PE}^r(x, M)$ to $T(\mathcal{Y})$: $\frac{\min(d(v(x),e(x)), r\, d(v(x),\ell(v(x),x)))}{d(v(x),e(x))}$, that is, $N_{PE}^r(x, M)$ is similar to $T(\mathcal{Y})$ with the given ratio. **P6** holds depending on the pair M and r. That is, there exists an $r_0(M)$ so that $N_{PE}^r(x, M)$ satisfies **P6** for all $r \leq r_0(M)$, and fails to satisfy otherwise. **P6** fails for all M when $r = \infty$. **P8** holds only when $M = M_C$. With CM-vertex regions, for all $r \in [1, \infty]$, the area $A\left(N_{PE}^r(x, M_C)\right)$ is a continuous function of $d(\ell_r(v(x), x), v(x))$ which is a continuous function of $d(\ell(v(x), x), v(x))$ which is a continuous function of x.

Moreover, $\Lambda_0(N_{PE}^r, M) = \mathcal{Y}$ for all $r \in [1, \infty]$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}$, since the $\mathbb{R}^2$-Lebesgue measure $\lambda(N_{PE}^r(x, M)) = 0$ iff $x \in \mathcal{Y}$.

As for **P3**, for $T_2(x) \subseteq T(\mathcal{Y})$ we can loosen the concept of center by treating the line $\ell(v(x), x)$ as the *edge-wise central line*, so **P3** is satisfied in this loose sense. Notice that x is not the unique center in this sense but a point on a central line. Let M_j, $j \in \{1, 2, 3\}$, be the midpoints of the edges of $T(\mathcal{Y})$, then for any $x \in T(M_1, M_2, M_3)$, $N_{PE}^2(x, M) = T(\mathcal{Y})$, so $T(M_1, M_2, M_3) \subseteq \mathscr{R}_S(N_{PE}^2, M)$ where equality holds for $M = M_C$ for all triangles and for $M = M_{CC}$ in non-obtuse triangles (see Figure 3.3.3).

In obtuse triangles, $\mathscr{R}_S(N_{PE}^2, M_{CC}) \supsetneq T(M_1, M_2, M_3)$ and is a quadrilateral.

The functional form of the superset region, $\mathscr{R}_S(N_{PE}^r, M)$, in T_b is given by

$$\mathscr{R}_S(N_{PE}^r, M) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \geq \frac{c_2\,(1 - r\,x)}{r\,(1 - c_1)}\right\} \cup \left\{(x, y) \in R_M(\mathsf{y}_2) : y \geq \frac{c_2\,(r\,(x - 1) + 1)}{r\,c_1}\right\} \cup \left\{(x, y) \in R_M(\mathsf{y}_3) : y \leq c_2 \frac{r - 1}{r}\right\},$$

and the functional form of $T(M_1, M_2, M_3)$ in T_b is given by

$$T(M_1, M_2, M_3) = \left\{(x, y) \in T_b : y \leq \frac{c_2}{2};\ y \geq \frac{c_2\,(-1 + 2\,x)}{2\,c_1};\ y \geq \frac{c_2\,(1 - 2\,x)}{2\,(1 - c_1)}\right\}.$$

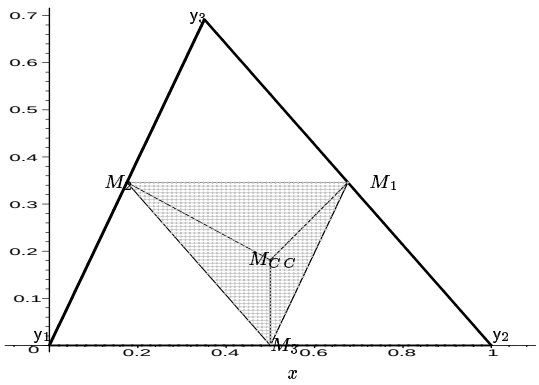


Figure 3.3.3: Superset region $\mathscr{R}_S\left(N_{PE}^2, M_{CC}\right)$ in an acute triangle

For $r = \sqrt{2}$, we can loosen the concept of center by treating the line $\ell(v(x), x)$ as the *area-wise central line* in $N_{PE}^{\sqrt{2}}(x, M)$, so **P3** is satisfied in this loose sense. Note that if x is close enough to M, we might have $N_{PE}^{\sqrt{2}}(x, M) = T(\mathcal{Y})$.

In $T(\mathcal{Y})$, drawing the lines $q_j(r, x)$ such that $d(\mathsf{y}_j, e_j) = r\, d(q_j(r, x), \mathsf{y}_j)$ for $j \in \{1, 2, 3\}$ yields a triangle, $\mathscr{T}^r$, for $r < 3/2$. See Figure 3.3.4 for $\mathscr{T}^r$ with $r = \sqrt{2}$.

The functional form of $\mathscr{T}^r$ in T_b is

$$\mathscr{T}^r = \left\{(x, y) \in T_b : y \geq \frac{c_2\,(r-1)}{r};\ y \leq \frac{c_2\,(1 - r\,x)}{r\,(1-c_1)};\ y \leq \frac{c_2\,(r\,(x-1)+1)}{r\,c_1}\right\} \tag{3.3.1}$$
$$= T\left(\left(\frac{(r-1)\,(1+c_1)}{r}, \frac{c_2\,(r-1)}{r}\right), \left(\frac{2-r+c_1\,(r-1)}{r}, \frac{c_2\,(r-1)}{r}\right), \left(\frac{c_1\,(2-r)+r-1}{r}, \frac{c_2\,(r-2)}{r}\right)\right).$$

There is a crucial difference between $\mathscr{T}^r$ and $T(M_1, M_2, M_3)$: $T(M_1, M_2, M_3) \subseteq \mathscr{R}_S(N_{PE}^r, M)$ for all M and $r \geq 2$, but $(\mathscr{T}^r)^o$ and $\mathscr{R}_S(N_{PE}^r, M)$ are disjoint (i.e., $(\mathscr{T}^r)^o \cap \mathscr{R}_S(N_{PE}^r, M) = \emptyset$) for all M and r.

So if $M \in (\mathscr{T}^r)^o$, then $\mathscr{R}_S(N_{PE}^r, M) = \emptyset$; if $M \in \partial(\mathscr{T}^r)$, then $\mathscr{R}_S(N_{PE}^r, M) = \{M\}$; and if $M \notin \mathscr{T}^r$, then $\mathscr{R}_S(N_{PE}^r, M)$ has positive area. Thus $N_{PE}^r(\cdot, M)$ fails to satisfy **P6** if $M \notin \mathscr{T}^r$.

The triangle $\mathscr{T}^r$ defined above plays a crucial role in the analysis of the distribution of the domination number of the PCD based on r-factor proximity maps. The superset region $\mathscr{R}_S(N_{PE}^r, M)$ will be important for both the domination number and the relative density of the corresponding PCDs.

Remark 3.3.1. In terms of the properties stated in Section 3.1, $N_{PE}^{3/2}(\cdot, M_C)$ is the most appealing amongst $\mathscr{N}_{\mathcal{Y}}^r := \left\{N_{PE}^r(\cdot, M) : r \in [1, \infty],\ M \in \mathbb{R}^2 \setminus \mathcal{Y}\right\}$. It is also noteworthy that the asymp-

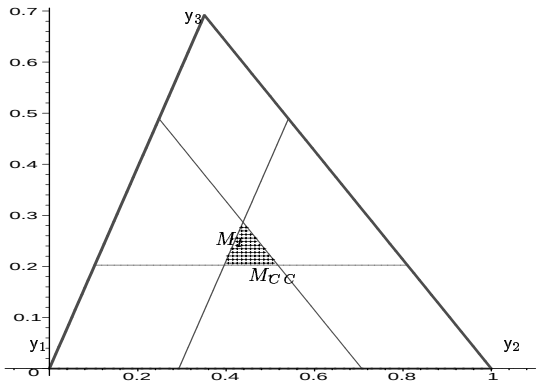


Figure 3.3.4: The triangle $\mathscr{T}^{r=\sqrt{2}}$.

totic distribution of the domination number of the PCD based on $N_{PE}^{3/2}(\cdot, M_C)$ is nondegenerate. □

Remark 3.3.2.

- For $r_1 \leq r_2$, $N_{PE}^{r_1}(x, M) \subseteq N_{PE}^{r_2}(x, M)$ for all $x \in T(\mathcal{Y})$. For $r_1 < r_2$, $N_{PE}^{r_1}(x, M) \subseteq N_{PE}^{r_2}(x, M)$ with equality holding for only $x \in \mathcal{Y}$ or $x \in \mathscr{R}_S(N_{PE}^{\min(r_1, r_2)}, M)$.
- For $3/2 < r_1 < r_2$, $\mathscr{R}_S(N_{PE}^{r_1}, M) \subset \mathscr{R}_S(N_{PE}^{r_2}, M)$ and $\mathscr{R}_S(N_{PE}^{r}, M) = \emptyset$ for $r < 3/2$.
- For $r_1 < r_2$, $A(N_{PE}^{r_1}(X, M)) \leq^{ST} A(N_{PE}^{r_2}(X, M))$ for X from a continous distribution on $T(\mathcal{Y})$. □

3.3.2.1 Extension of N_{PE}^{r} to Higher Dimensions

The extension to $\mathbb{R}^d$ for $d > 2$ is straightforward. We give the extension with $M = M_C$, but the extension for general M is similar. Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \cdots, \mathsf{y}_{d+1}\}$ be $d+1$ points that do not lie on the same hyperplane. Denote the simplex formed by these $d+1$ points as $\mathfrak{S}(\mathcal{Y})$. A simplex is the simplest polytope in $\mathbb{R}^d$ having $d+1$ vertices, $d(d+1)/2$ edges and $d+1$ faces of dimension $(d-1)$. For $r \in [1, \infty]$, define the r-factor proximity map as follows. Given a point x in $\mathfrak{S}(\mathcal{Y})$, let $v := \text{argmin}_{\mathsf{y} \in \mathcal{Y}} \mathrm{V}(Q_{\mathsf{y}}(x))$ where $Q_{\mathsf{y}}(x)$ is the polytope with vertices being the $d(d+1)/2$ midpoints of the edges, the vertex v and x and $\mathrm{V}(\cdot)$ is the d-dimensional volume functional. That is, the vertex region for vertex v is the polytope with vertices given by v and the midpoints of the edges. Let $v(x)$ be the vertex in whose region x falls. If x falls on the boundary of two vertex regions, we assign $v(x)$ arbitrarily. Let $\varphi(x)$ be the face opposite to vertex $v(x)$, and

$\Upsilon(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ which contains x. Let $d(v(x), \Upsilon(v(x), x))$ be the (perpendicular) Euclidean distance from $v(x)$ to $\Upsilon(v(x), x)$. For $r \in [1, \infty)$, let $\Upsilon_r(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ such that

$$d(v(x), \Upsilon_r(v(x), x)) = r\, d(v(x), \Upsilon(v(x), x))$$

and

$$d(\Upsilon(v(x), x), \Upsilon_r(v(x), x)) < d(v(x), \Upsilon_r(v(x), x)).$$

Let $\mathfrak{S}_r(x)$ be the polytope similar to and with the same orientation as $\mathfrak{S}$ having $v(x)$ as a vertex and $\Upsilon_r(v(x), x)$ as the opposite face. Then the r-factor proximity region $N_{PE}^r(x, M_C) := \mathfrak{S}_r(x) \cap \mathfrak{S}(\mathcal{Y})$. Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M_C)$.

3.3.3 τ-Factor Central Similarity Proximity Maps

The other type of triangular proximity map we introduce is the τ-factor central similarity proximity map. This will turn out to be the most appealing in terms of the properties in Section 3.1. Furthermore, the relative density of the corresponding PCD will have mathematical tractability. Alas, the distribution of the domination number of the associated PCD is still an open problem.

For $\tau \in [0, 1]$, define $N_{CS}^\tau(\cdot, M) := N(\cdot, M; \tau, \mathcal{Y})$ to be the *τ-factor central similarity proximity map* with M-edge regions as follows; see also Figure 3.3.5 with $M = M_C$. For $x \in T(\mathcal{Y}) \setminus \mathcal{Y}$, let $e(x)$ be the edge in whose region x falls; i.e., $x \in R_M(e(x))$. If x falls on the boundary of two edge regions, we assign $e(x)$ arbitrarily. For $\tau \in (0, 1]$, the *τ-factor* central similarity proximity region $N_{CS}^\tau(x, M)$ is defined to be the triangle $T_\tau(x)$ with the following properties:

(i) $T_\tau(x)$ has edges $e_j^\tau(x)$ parallel to e_j for each $j \in \{1, 2, 3\}$, and for $x \in R_M(e(x))$, $d(x, e^\tau(x)) = \tau\, d(x, e(x))$ and $d(e^\tau(x), e(x)) \leq d(x, e(x))$ where $d(x, e(x))$ is the Euclidean (perpendicular) distance from x to $e(x)$;

(ii) $T_\tau(x)$ has the same orientation as and is similar to $T(\mathcal{Y})$;

(iii) x is the same type of center of $T_\tau(x)$ as M is of $T(\mathcal{Y})$.

Note that (i) explains the "τ-factor", (ii) explains "similarity", and (iii) explains "central" in the name, *τ-factor central similarity proximity map.* For $\tau = 0$, we define $N_{CS}^{\tau=0}(x, M) = \{x\}$ for all $x \in T(\mathcal{Y})$. For $x \in \partial(T(\mathcal{Y}))$, we define $N_{CS}^{\tau}(x, M) = \{x\}$ for all $\tau \in [0, 1]$.

Notice that by definition $x \in N_{CS}^{\tau}(x)$ for all $x \in T(\mathcal{Y})$. Furthermore, $\tau \leq 1$ implies that $N_{CS}^{\tau}(x, M) \subseteq T(\mathcal{Y})$ for all $x \in T(\mathcal{Y})$ and $M \in T(\mathcal{Y})^o$. For all $x \in T(\mathcal{Y})^o \cap R_M(e(x))$, the edges $e^{\tau}(x)$ and $e(x)$ are coincident iff $\tau = 1$.

Notice that $X_i \overset{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional probability density function f exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y})$, implies that the special case in the construction of $N_{CS}^{\tau}(\cdot)$ — X falls on the boundary of two edge regions — occurs with probability zero. Note that for such an F, $N_{CS}^{\tau}(X, M)$ is a triangle a.s.

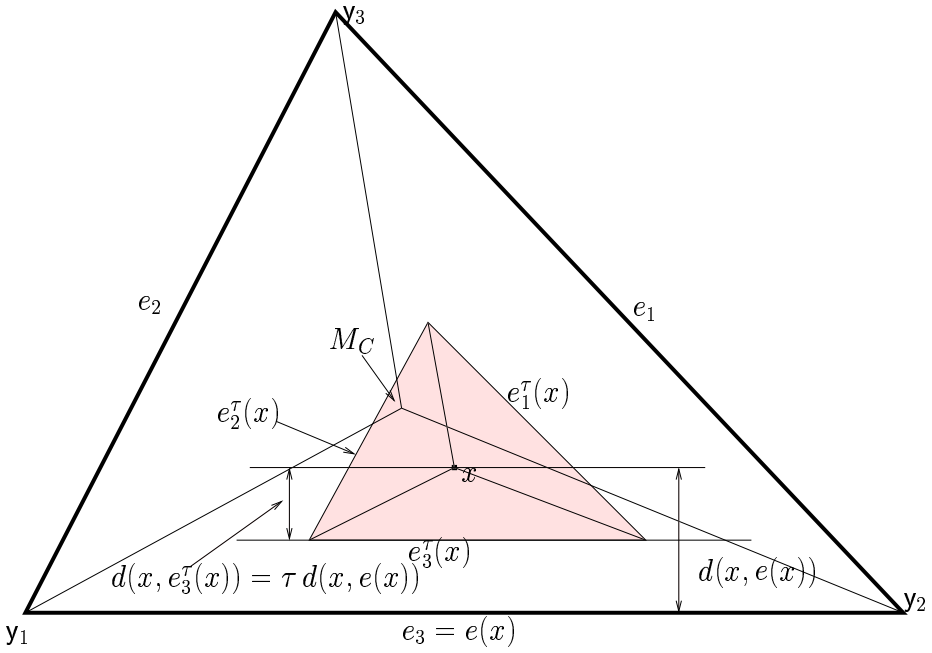


Figure 3.3.5: Construction of τ-factor central similarity proximity region, $N_{CS}^{\tau=1/2}(x, M_C)$ (shaded region).

Notice that τ-factor central similarity proximity maps are defined with M-edge regions for $M \in T(\mathcal{Y})^o$. Among the four centers we consider in Section 2.3, M_C and M_I are inside the triangle, so they can be used in construction of the τ-factor central similarity proximity map.

In general, for τ-factor central similarity proximity regions with M-edge regions, the similarity ratio of $N_{CS}^{\tau}(x, M)$ to $T(\mathcal{Y})$ is $d(x, e^{\tau}(x))/d(M, e(x))$. See Figure 3.3.6 (right) for $N_{CS}^{\tau=1}(x, M)$ with $e(x) = e_3$. Notice that $N_{CS}^{\tau}(\cdot, M)$ satisfies properties **P1**-**P8**.

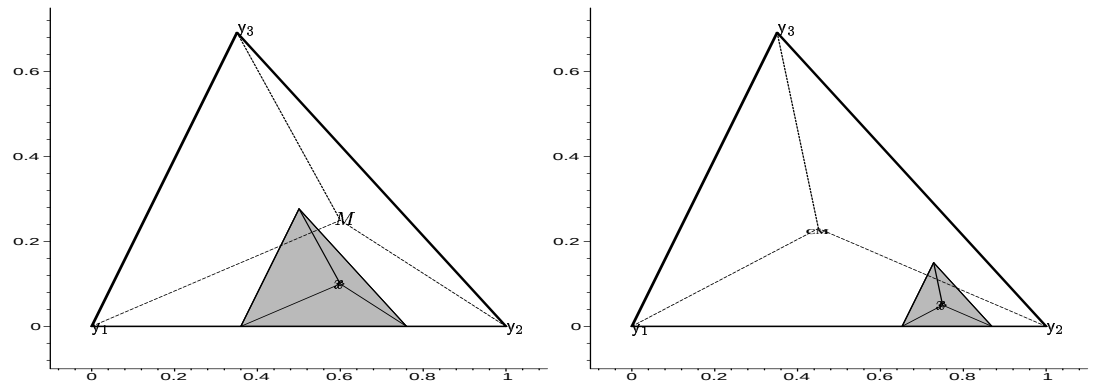


Figure 3.3.6: $N_{CS}^{\tau=1}(x, M)$ with an $x \in R_M(e_3)$ (left); $N_{CS}^{\tau=1}(x, M_C)$ with an $x \in R_{M_C}(e_3)$ (right)

See Figure 3.3.5 for $N_{CS}^{\tau=1/2}(x, M_C)$ with $e = e_3$ and Figure 3.3.6 (left) for $N_{CS}^{\tau=1}(x, M_C)$ with $e = e_3$. The functional form of $N_{CS}^{\tau}(x, M_C)$ for an $x = (x_0, y_0) \in R_{CM}(e)$ is as follows:

$$\text{For } x \in R_{CM}(e_1),\ N_{CS}^{\tau}(x, M_C) = \Big\{(x,y) \in T_b : y \geq y_0 + \tau\,(y_0\,(1 - c_1)),$$
$$y \leq \tfrac{(1-\tau)\,(y_0\,(1-c_1)-x_0\,c_2)+c_2\,(\tau-x)}{1-c_1};\ y \leq \tfrac{y_0\,(c_1\,(1+\tau)-\tau)-x_0\,c_2\,(1+\tau)+c_2\,(\tau+x)}{c_1}\Big\}.$$

$$\text{For } x \in R_{CM}(e_2),\ N_{CS}^{\tau}(x, M_C) = \Big\{(x,y) \in T_b : y \geq y_0 + \tau\,(y_0\,c_1 - c_2\,x_0),$$
$$y \leq \tfrac{(1-\tau)\,(y_0\,c_1-x_0\,c_2)+c_2\,x}{c_1};\ y \leq \tfrac{y_0\,(1-c_1\,(1+\tau))+x_0\,c_2\,(1+\tau)}{1-c_1}\Big\}.$$

$$\text{For } x \in R_{CM}(e_3),\ N_{CS}^{\tau}(x, M_C) = \Big\{(x,y) \in T_b : y \geq y_0\,(1-\tau);\ y \leq \tfrac{(y_0\,(\tau+c_1)+c_2\,(x-x_0))}{c_1};$$
$$y \leq \tfrac{y_0\,(1-c_1+\tau)-c_2\,(x-x_0)}{1-c_1}\Big\}.$$

Furthermore, $\Lambda_0\left(N_{CS}^{\tau}(\cdot, M)\right) = \partial(T(\mathcal{Y}))$ for all $\tau \in (0, 1]$ and $\Lambda_0\left(N_{CS}^{\tau=0}(\cdot, M)\right) = T(\mathcal{Y})$, since $\lambda\left(N_{CS}^{\tau}(x)\right) = 0$ iff $x \in e_j$ for $j \in \{1, 2, 3\}$ or $\tau = 0$.

Remark 3.3.3. Among $\mathscr{N}_{CS}^{\tau} := \left\{N_{CS}^{\tau}(\cdot, M) :\ \tau \in [0, 1],\ M \in T(\mathcal{Y})^o\right\}$, every $N_{CS}^{\tau}(\cdot, M)$ with $\tau \in (0, 1]$ satisfies all the properties in Section 3.1. □

Remark 3.3.4.

- For $\tau_1 \leq \tau_2$, $N_{CS}^{\tau_1}(x, M) \subseteq N_{CS}^{\tau_2}(x, M)$ for all $x \in T(\mathcal{Y})$. For $\tau_1 < \tau_2$, $N_{CS}^{\tau_1}(x, M) \subseteq N_{CS}^{\tau_2}(x, M)$ with equality holding only for $x \in \partial(T(\mathcal{Y}))$.
- The superset region $\mathscr{R}_S\left(N_{CS}^{\tau}, M\right) = \emptyset$ for $\tau \in [0, 1)1$ and $\mathscr{R}_S\left(N_{CS}^{\tau=1}, M\right) = \{M\}$.
- For $\tau_1 < \tau_2$, $A\left(N_{CS}^{\tau_1}(X, M)\right) \leq^{ST} A\left(N_{CS}^{\tau_2}(X, M)\right)$ for X from a continuous distribution on $T(\mathcal{Y})$. □

3.3.3.1 Extension of N_{CS}^{τ} to Higher Dimensions

The extension of N_{CS}^{τ} to $\mathbb{R}^d$ for $d > 2$ is straightforward. We describe the extension for $M = M_C$, the extension for general M is similar. Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \cdots, \mathsf{y}_{d+1}\}$ be $d+1$ points that do not lie on the same hyperplane. Denote the simplex formed by these $d+1$ points as $\mathfrak{S}(\mathcal{Y})$. For $\tau \in (0,1]$, define the τ-factor central similarity proximity map as follows. Let φ_j be the face opposite vertex y_j for $j \in \{1,2\}, \ldots, d+1$, and "face regions" $R_{CM}(\varphi_1), \ldots, R_{CM}(\varphi_{d+1})$ partition $\mathfrak{S}(\mathcal{Y})$ into $d+1$ regions, namely the $d+1$ polytopes with vertices being the center of mass together with d vertices chosen from $d+1$ vertices. For $x \in \mathfrak{S}(\mathcal{Y}) \setminus \mathcal{Y}$, let $\varphi(x)$ be the face in whose region x falls; $x \in R(\varphi(x))$. If x falls on the boundary of two face regions, we assign $\varphi(x)$ arbitrarily. For $\tau \in (0,1]$, the *τ-factor* central similarity proximity region $N_{CS}^{\tau}(x, M_C) = N_{PE}^{\tau}(x, M_C)$ is defined to be the simplex $\mathfrak{S}_\tau(x)$ with the following properties:

(i) $\mathfrak{S}_\tau(x)$ has faces $\varphi_j^\tau(x)$ parallel to $\varphi_j(x)$ for $j \in \{1,2\}, \ldots, d+1$, and for $x \in R_{CM}(\varphi(x))$, $\tau\, d(x, \varphi(x)) = d(\varphi^\tau(x), x)$ where $d(x, \varphi(x))$ is the Euclidean (perpendicular) distance from x to $\varphi(x)$;

(ii) $\mathfrak{S}_\tau(x)$ has the same orientation as and similar to $\mathfrak{S}(\mathcal{Y})$;

(iii) x is the center of mass of $\mathfrak{S}_\tau(x)$, as M_C is of $\mathfrak{S}(\mathcal{Y})$.

3.3.4 The Behaviour of Proximity Maps

In this section, we provide the conditions for x, which, if satisfied, will imply some sort of increase in the size of the proximity regions we have defined.

Let $N(\cdot)$ be any proximity map defined on the measurable space Ω with measure μ, and let $\{x_n\}_{n=1}^{\infty}$ be a sequence of points in Ω. We say $N(x_n)$ *gets larger* if $N(x_n) \subseteq N(x_m)$ for all $m \geq n$, and $N(x_n)$ *gets strictly larger* if $N(x_n) \subsetneq N(x_m)$ for all $m > n$.

In the following theorems we will assume $\Omega = \mathbb{R}^2$ with μ being the $\mathbb{R}^2$-Lebesgue measure λ. M-vertex regions are defined with points $M \in \mathbb{R}^2 \setminus \mathcal{Y}$.

Theorem 3.3.5. *For arc-slice proximity regions with M-vertex regions for an $M \in \mathbb{R}^2 \setminus \mathcal{Y}$, as $d(x, \mathsf{y})$ (strictly) increases for x lying on a ray from y in $R_M(\mathsf{y}) \setminus \mathscr{R}_S(N_{AS}, M)$, $N_{AS}(x, M)$ gets (strictly) larger.*

Proof: For x, y lying on a ray from y in $R_M(\mathsf{y}) \setminus \mathscr{R}_S(N_{AS}, M)$, if $d(x, \mathsf{y}) \leq d(y, \mathsf{y})$, then $B(x, r(x)) \subseteq B(y, r(y))$, which implies $N_{AS}(x, M) \subseteq N_{AS}(y, M)$, hence $N_{AS}(x, M)$ gets larger

as $d(x, \mathsf{y})$ increases for x lying on a ray from y in $R_M(\mathsf{y}) \setminus \mathscr{R}_S(N_{AS}, M)$. The strict version follows similarly. If $x,\ y \in R_M(\mathsf{y}) \cap \mathscr{R}_S(N_{AS}, M)$, then $N_{AS}(x, M) = N_{AS}(y, M) = T(\mathcal{Y})$. ■

Let $\ell(\mathsf{y}, x)$ be the line at x parallel to $e(x)$ for $x \in R_M(\mathsf{y})$ where $e(x)$ is the edge opposite vertex y.

Theorem 3.3.6. *For the r-factor proportional-edge proximity maps with M-vertex regions for an $M \in \mathbb{R}^2 \setminus \mathcal{Y}$, as $d(\ell(\mathsf{y}, x), \mathsf{y})$ (strictly) increases for $x \in R_M(\mathsf{y}) \setminus \mathscr{R}_S(N_{PE}^r, M)$, $N_{PE}^r(x, M)$ gets (strictly) larger for $r < \infty$.*

Proof: For $x,\ y \in R_M(\mathsf{y}) \setminus \mathscr{R}_S(N_{PE}^r, M)$, if $d(\ell(\mathsf{y}, x), \mathsf{y}) \leq d(\ell(\mathsf{y}, y), \mathsf{y})$, then by definition $N_{PE}^r(x, M) \subseteq N_{PE}^r(y, M)$, hence the result follows. The strict version follows similarly. If $x,\ y \in R_M(\mathsf{y}) \cap \mathscr{R}_S(N_{PE}^r, M)$, then $N_{PE}^r(x, M) = N_{PE}^r(y, M) = T(\mathcal{Y})$, and if $r = \infty$ and $x, y \in T(\mathcal{Y}) \setminus \mathcal{Y}$, $N_{PE}^r(x, M) = N_{PE}^r(y, M) = T(\mathcal{Y})$.■

Note that as $d(\ell(\mathsf{y}, x), \mathsf{y})$ increases for $x \in R_M(\mathsf{y})$, $d(\ell(\mathsf{y}, x), M)$ decreases, provided that $M \in T(\mathcal{Y})^o$ and M-vertex regions are convex.

We define the M-edge regions, $R_M(e)$, with points $M \in T(\mathcal{Y})^o$.

Theorem 3.3.7. *For τ-factor central similarity proximity regions with M-edge regions for an $M \in T(\mathcal{Y})^o$, as $d(x, e)$ (strictly) increases for $x \in R_M(e)$, the area $A\left(N_{CS}^\tau(x, M)\right)$ (strictly) increases for $\tau \in (0, 1]$.*

Proof: For $x,\ y \in R_M(e)$ and $\tau \in (0, 1]$, if $d(x, e) \leq d(y, e)$ then the similarity ratio of $N_{CS}^\tau(y, M)$ to $T(\mathcal{Y})$ is larger than or equal to that of $N_{CS}^\tau(x, M)$, which in turn implies that $A\left(N_{CS}^\tau(x, M)\right) \leq A\left(N_{CS}^\tau(y, M)\right)$. The strict version follows similarly. ■

Observe that the statement of Theorem 3.3.7 is about the area $A\left(N_{CS}^\tau(x, M)\right)$. We need further conditions for $N_{CS}^\tau(x, M)$ to get larger.

Theorem 3.3.8. *Let $\ell_M(\mathsf{y})$ be the line joining M and vertex $\mathsf{y} \in \mathcal{Y}$. As $d(x, \ell_M(\mathsf{y}_j))$ and $d(x, \ell_M(\mathsf{y}_k))$ both (strictly) decrease for $x \in R_M(e_l)$ where j, k, l are distinct, $N_{CS}^\tau(x, M)$ (strictly) increases for $\tau \in (0, 1]$.*

Proof: Suppose, without loss of generality, that $x,\ y \in R_M(e_3)$. Consider the set

$$S(e_3, x) := \{y \in R_M(e_3) : d(y, \ell_M(\mathsf{y}_1)) \leq d(x, \ell_M(\mathsf{y}_1)) \text{ and } d(y, \ell_M(\mathsf{y}_2)) \leq d(x, \ell_M(\mathsf{y}_2))\},$$

which a parallelogram. See Figure 3.3.7 for an example of $S(e_3, x)$ with $M = M_C$ and $e = e_3$. Given x, for $y \in S(e_3, x)$, by construction, $N^r_{CS}(x, M) \subseteq N^r_{CS}(y, M)$. Then the desired result follows for $\tau \in (0, 1]$. (Observe that if x_{n+1} is in $S(e_3, x_n)$, then $d(x_n, \ell_M(\mathsf{y}_1))$ and $d(x_n, \ell_M(\mathsf{y}_2))$ both decrease.) The strict version follows similarly. ■

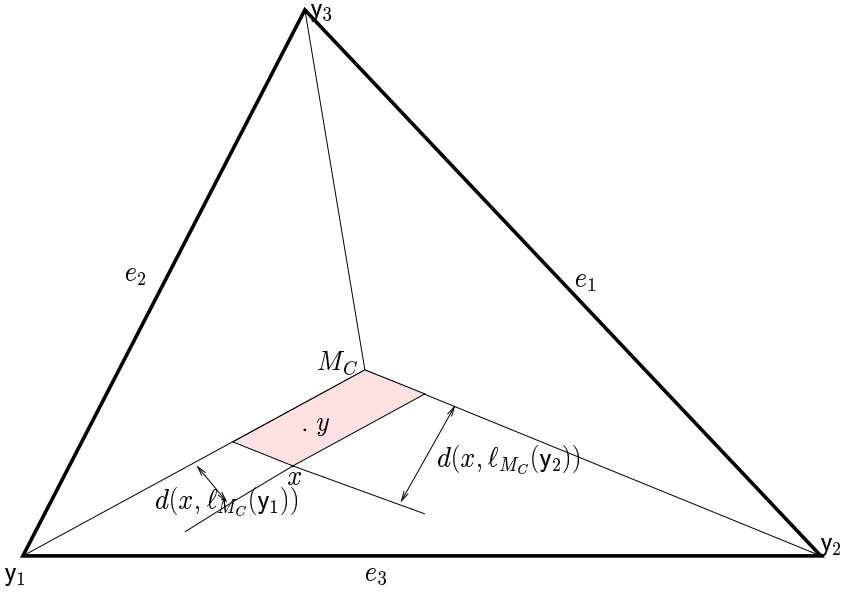


Figure 3.3.7: The figure for x, $y \in R_M(e_3)$ described in Theorem 3.3.8.

Remark 3.3.9. For $\mathscr{R}_S(N_{\mathcal{Y}})$ with positive area, by definition, as $x \to y \in \mathscr{R}_S(N_{\mathcal{Y}})$, $N_{\mathcal{Y}}(x) \to T(\mathcal{Y})$ and hence $\operatorname{argsup}_{x \in T(\mathcal{Y})} A(N_{\mathcal{Y}}(x)) \in \mathscr{R}_S(N_{\mathcal{Y}})$ with $\sup_{x \in T(\mathcal{Y})} A(N_{\mathcal{Y}}(x)) = A(T(\mathcal{Y}))$. Furthermore, the following also hold.

1. As $x \to M_{CC}$ in a non-obtuse triangle $T(\mathcal{Y})$, $N_{AS}(x, M_{CC}) \to T(\mathcal{Y})$ and $\operatorname{argsup}_{x \in T(\mathcal{Y})} A\left(N_{AS}(x, M_{CC})\right) = M_{CC}$ with $\sup_{x \in T(\mathcal{Y})} A\left(N_{AS}(x, M_{CC})\right) = A(T(\mathcal{Y}))$.

2. For $r > 3/2$, as $x \to y \in (\mathscr{R}_S(N^r_{PE}, M))^o$, $N^r_{PE}(x, M) \to T(\mathcal{Y})$ hence $\operatorname{argsup}_{x \in T(\mathcal{Y})} A(N^r_{PE}(x, M)) \in \mathscr{R}_S(N^r_{PE}, M)$ with $\sup_{x \in T(\mathcal{Y})} A(N^r_{PE}(x, M)) = A(T(\mathcal{Y}))$.

3. For $r < 3/2$, if $M \notin \mathscr{T}^r$, then as $x \to M$, $N^r_{PE}(x, M) \to T(\mathcal{Y})$ and $\operatorname{argsup}_{x \in T(\mathcal{Y})} A(N^r_{PE}(x, M)) = M$ with $\sup_{x \in T(\mathcal{Y})} A(N^r_{PE}(x, M)) = A(T(\mathcal{Y}))$.

   If $M \in (\mathscr{T}^r)^o$, then as $x \to M$, $N^r_{PE}(x, M) \to N^r_{PE}(M, M) \subsetneq T(\mathcal{Y})$, but still $\operatorname{argsup}_{x \in T(\mathcal{Y})} A(N^r_{PE}(x, M)) = M$ with $\sup_{x \in T(\mathcal{Y})} A(N^r_{PE}(x, M)) = A(N^r_{PE}(M, M))$.

If $M \in \partial(\mathscr{T}^r)$, then as $x \to M$, $N_{PE}^r(x, M) \to N_{PE}^r(M, M) \subseteq T(\mathcal{Y})$, and $\text{argsup}_{x \in T(\mathcal{Y})}$ $A(N_{PE}^r(x, M)) = M$ with $\sup_{x \in T(\mathcal{Y})} A(N_{PE}^r(x, M)) = A(N_{PE}^r(M, M))$ which might be $T(\mathcal{Y})$ or a proper subset of $T(\mathcal{Y})$.

4. As $x \to M_C$, $N_{PE}^{3/2}(x, M_C) \to T(\mathcal{Y})$ and $\text{argsup}_{x \in T(\mathcal{Y})} A\left(N_{PE}^{3/2}(x, M_C)\right) = M_C$ with $\sup_{x \in T(\mathcal{Y})} A\left(N_{PE}^{3/2}(x, M_C)\right) = A(T(\mathcal{Y}))$.

5. As $x \to M$, $N_{CS}^{\tau=1}(x, M) \to T(\mathcal{Y})$ and $\text{argsup}_{x \in T(\mathcal{Y})} A(N_{CS}^{\tau=1}(x, M)) = M$ with $\sup_{x \in T(\mathcal{Y})} A(N_{CS}^{\tau=1}(x, M)) = A(T(\mathcal{Y}))$. For $\tau < 1$, as $x \to M$, $N_{CS}^{\tau}(x, M) \to N_{CS}^{\tau}(M, M)$ and $\text{argsup}_{x \in T(\mathcal{Y})} A\left(N_{CS}^{\tau}(x, M)\right) = M$ with $\sup_{x \in T(\mathcal{Y})} A\left(N_{CS}^{\tau}(x, M)\right) = A(N_{CS}^{\tau}(M, M))$. □

Although the comments in the above remark follow by definition, they will be indicative of whether the asymptotic distribution of the domination number of the associated PCD is degenerate or not.

3.4 Γ_1-Regions and the Related Concepts for Proximity Maps

Let $(\Omega, \mathcal{M})$ be a measurable space and consider the proximity map $N : \Omega \to \wp(\Omega)$, where $\wp(\cdot)$ is the power set functional. For any set $B \subseteq \Omega$, the Γ_1-*region* of B associated with $N(\cdot)$, is defined to be the region $\Gamma_1(B, N) := \{z \in \Omega : B \subseteq N(z)\}$. For $x \in \Omega$, we denote $\Gamma_1\big(\{x\}, N\big)$ as $\Gamma_1(x, N)$.

If $\mathcal{X}_n = \big\{X_1, X_2, \cdots, X_n\big\}$ is a set of Ω-valued random variables, then $\Gamma_1(X_i, N), i = 1, \cdots, n$ are random sets. If the X_i are independent and identically distributed, then so are the random sets $\Gamma_1(X_i, N)$. Furthermore, $\Gamma_1(\mathcal{X}_n, N)$ is a random set.

For $X_1, \cdots, X_n \overset{iid}{\sim} F$ the domination number of the associated data-random proximity catch digraph D, denoted $\gamma(\mathcal{X}_n, N)$, is the minimum number of points that dominate all points in $\mathcal{X}_n$. Note that, $\gamma(\mathcal{X}_n, N) = 1$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N) \neq \emptyset$. Hence the name Γ_1-*region*. Suppose μ is a measure on Ω. Following are some general results about $\Gamma_1(\cdot, N)$ for general N.

Proposition 3.4.1. *For any proximity map N and set $B \subseteq \Omega$, $\mathscr{R}_S(N) \subseteq \Gamma_1(B, N)$.*

Proof: For $x \in \mathscr{R}_S(N)$, $N(x) = \Omega$, so $B \subseteq N(x)$, then $x \in \Gamma_1(B, N)$, hence $\mathscr{R}_S(N) \subseteq \Gamma_1(B, N)$. ■

Lemma 3.4.2. *For any proximity map N and $B \subseteq \Omega$, $\Gamma_1(B, N) = \cap_{x \in B} \Gamma_1(x, N)$.*

Proof: Given a particular type of proximity map N and subset $B \subseteq \Omega$, $y \in \Gamma_1(B, N)$ iff $B \subseteq N(y)$ iff $x \in N(y)$ for all $x \in B$ iff $y \in \Gamma_1(x, N)$ for all $x \in B$ iff $y \in \cap_{x \in B} \Gamma_1(x, N)$. Hence the result follows. ■

A problem of interest is finding, if possible, a (proper) subset of B, say $G \subset B$, such that $\Gamma_1(B, N) = \cap_{x \in G} \Gamma_1(x, N)$. This implies that only the points in G are *active* in determining $\Gamma_1(B, N)$.

Definition 3.4.3. An *active set* of points $S_A(B) \subseteq \Omega$ for determining $\Gamma_1(B, N)$ is defined to be a subset of B such that $\Gamma_1(B, N) = \cap_{x \in S_A(B)} \Gamma_1(x, N)$.

This definition allows B to be an active set, which always holds by Lemma 3.4.2. If B is a set of finitely many points, so is the associated active set. Among the active sets, we seek an active set of minimum cardinality.

Definition 3.4.4. Let B be a set of finitely many points. An active subset is called a *minimal active subset*, denoted $S_\mu(B)$, if there is no other active subset S_A of B such that $S_A(B) \subsetneq S_\mu(B)$. The minimum cardinality among the active subsets of B is called the *η-value* and denoted as $\eta(B, N)$. An active subset of cardinality $\eta(B, N)$ is called a *minimum active subset* denoted as $S_M(B)$, that is, $\eta(B, N) := |S_M(B)|$.

We will suppress the dependence on B for $S_A(B)$, $S_\mu(B)$, and $S_M(B)$ if it is obvious. In particular, if $B = \mathcal{X}_n$ is a set of Ω-valued random variables, then $\eta(\mathcal{X}_n, N)$ is a random quantity.

For example, in $\mathbb{R}$ with $\mathcal{Y} = \{0, 1\}$, and $\mathcal{X}_n$ a random sample of size $n > 1$ from F in $(0, 1)$, $\Gamma_1(\mathcal{X}_n, N_S) = \left(\frac{X_{n:n}}{2}, \frac{1+X_{1:n}}{2}\right)$, where $X_{j:n}$ is the j^{th} largest value in $\mathcal{X}_n$. So the extrema (minimum and maximum) of the set $\mathcal{X}_n$ are sufficient to determine the Γ_1-region; i.e., $S_M = \{X_{1:n}, X_{n:n}\}$. Then $\eta(\mathcal{X}_n, N_S) = 1 + \mathbf{I}(n > 1)$ a.s. for $\mathcal{X}_n$ a random sample from a continuous distribution on $(0, 1)$.

In the multidimensional case there is no natural extension of ordering that yields natural extrema such as minimum or maximum. To get the minimum active sets associated with our proximity maps, we will resort to some other sort of extrema, such as, the closest points to edges or vertices in $T(\mathcal{Y})$.

For any proximity map N and $\mathcal{X}_n$, $\eta(\mathcal{X}_n, N) \leq n$ follows trivially.

Lemma 3.4.5. *Given a sequence of Ω-valued random variables $X_1, X_2, X_3, \ldots$ from distribution F, let $\mathcal{X}(n) := \mathcal{X}(n-1) \cup \{X_n\}$ for $n = 0, 1, 2, 3, \ldots$ with $\mathcal{X}(0) := \emptyset$. Then $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing in n in the sense that $\Gamma_1(\mathcal{X}(n), N) \supseteq \Gamma_1(\mathcal{X}(n+1), N)$.*

Proof: Given a particular type of proximity map N and a data set $\mathcal{X}(n) = \{X_1, \ldots, X_n\}$, by Lemma 3.4.2, $\Gamma_1(\mathcal{X}(n), N) = \cap_{j=1}^{n} \Gamma_1(X_j, N)$ and by definition, $\mathcal{X}(n+1) = \mathcal{X}(n) \cup \{X_{n+1}\}$. So, $\Gamma_1(\mathcal{X}(n+1), N) = \cap_{j=1}^{n+1} \Gamma_1(X_j, N) = [\cap_{j=1}^{n} \Gamma_1(X_j, N)] \cap \Gamma_1(X_{n+1}, N) = \Gamma_1(\mathcal{X}(n), N) \cap \Gamma_1(X_{n+1}, N) \subseteq \Gamma_1(\mathcal{X}(n), N)$.

Thus we have shown that $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing in n, i.e., $\Gamma_1(\mathcal{X}(n), N) \supseteq \Gamma_1(\mathcal{X}(n+1), N)$. ■

Remark 3.4.6. By monotone sequential continuity from above, $\{\Gamma_1(\mathcal{X}(n), N)\}_{n=1}^{\infty}$ has a limit

$$\begin{aligned} G_1 &:= \cap_{k=1}^{\infty} \Gamma_1(\mathcal{X}(k), N) = \lim_{m \to \infty} \cap_{k=1}^{m} \Gamma_1(\mathcal{X}(k), N) = \lim_{m \to \infty} \Gamma_1(\mathcal{X}(m), N) \\ &= \lim_{m \to \infty} \cap_{j=1}^{m} \Gamma_1(X_j, N) = \cap_{j=1}^{\infty} \Gamma_1(X_j, N). \ \square \end{aligned}$$

Note however that $\Gamma_1(\mathcal{X}_n, N)$ is neither strictly decreasing (in the sense that $\Gamma_1(\mathcal{X}_n, N) \supsetneq \Gamma_1(\mathcal{X}_{n+1}, N)$) nor non-increasing (in the sense that $\Gamma_1(\mathcal{X}_n, N) \supseteq \Gamma_1(\mathcal{X}_{n+1}, N)$) provided that $\mathscr{R}_S(N) \neq \Omega$ for all $\mathcal{X}_n$, because we might have $\Gamma_1(\mathcal{X}_n, N) \subsetneq \Gamma_1(\mathcal{X}_m, N)$ for some $m > n$.

Since $\Gamma_1(\mathcal{X}_n, N) = \cap_{j=1}^{n} \Gamma_1(X_j, N)$ for a given realization of the data set $\mathcal{X}_n$, first we describe the region $\Gamma_1(X, N)$ for $X \in \mathcal{X}_n$, and then describe the region $\Gamma_1(\mathcal{X}_n, N)$.

Recall that the domination number of the PCD is $\gamma(\mathcal{X}_n, N) = 1$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N) \neq \emptyset$. Then the distribution of the domination number of the PCD will depend on the Γ_1-region. Furthermore, we will see in Chapter 5 that, the Γ_1-regions are also used in finding the distribution of the relative density of the PCDs. Hence we describe the Γ_1-regions and the associated concepts for the proximity maps $\{N_{AS}, N_{PE}^{r}, N_{CS}^{\tau}\}$.

3.4.1 Γ_1-Regions for Arc-Slice Proximity Maps

Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\} \subset \mathbb{R}^2$ be three non-collinear points and $T(\mathcal{Y})$ be the triangle with vertices $\mathcal{Y}$. For $x \in T(\mathcal{Y})$, draw perpendicular lines at the midpoints of the line segments joining x with vertices $\mathcal{Y}$. The Γ_1-region is the region bounded by these perpendicular lines. Notice that $\Gamma_1(x, N_{AS}, M_{CC}) = \mathcal{V}_C(x) \cap T(\mathcal{Y})$, where M_{CC} is included in the notation to make the dependence on the CC-vertex regions explicit, and $\mathcal{V}_C(x)$ is the Voronoi cell generated by x

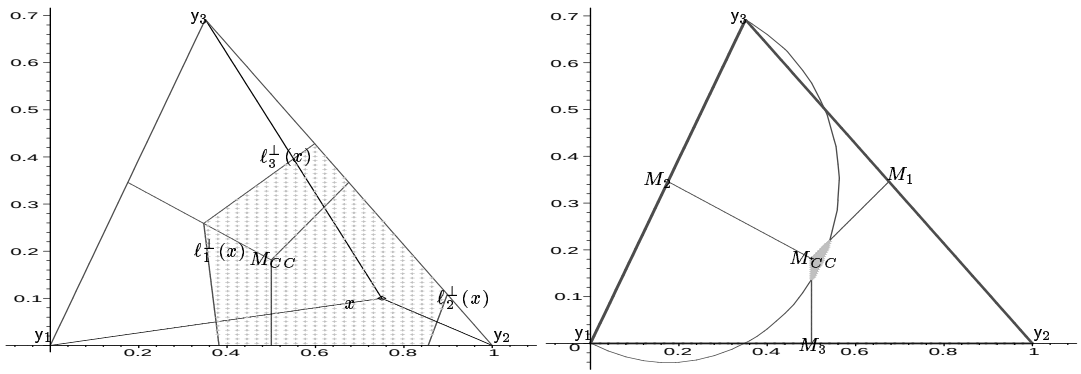


Figure 3.4.1: $\Gamma_1(x, N_{AS}, M_{CC})$ and $x \in R_{CC}(\mathsf{y}_2)$ (left); the ball $B(M_2, |e_2|/2)$ that divides the region $R_{CC}(\mathsf{y}_2)$ for which Γ_1-region is a hexagon or a pentagon (right).

in the Voronoi tessellation of $\mathcal{Y} \cup \{x\}$. See Figure 3.4.1 (left) where the Γ_1-region is a convex pentagon. $\Gamma_1(x, N_{AS}, M_{CC})$ can be a hexagon for x close enough to M_{CC}. Let $\ell_j^\perp(x)$ be the perpendicular line at the midpoint between x and vertex y_j for $j \in \{1, 2, 3\}$. For example, for $x \in R_{CC}(\mathsf{y}_2) \setminus \{\mathsf{y}_2\}$, $\Gamma_1(x, N_{AS}, M_{CC})$ will be a pentagon that has M_2 as a vertex iff $\ell_1^\perp(x) \cap \ell_3^\perp(x) = \{M_2\}$. The locus of such points in the basic triangle, T_b, in the implicit functional form, is given by

$$x^2 - c_1\, x - c_2\, y + y^2 = (x - c_1/2)^2 + (y - c_2/2)^2 = \left(c_1^2 + c_2^2\right)/4,$$

which is the equation of the circle centered at M_2 with radius $\left(\sqrt{c_1^2 + c_2^2}\right)\Big/2$. See Figure 3.4.1 (right). Therefore for $x \in R_{CC}(\mathsf{y}_2)$, $\Gamma_1(x, N_{AS}, M_{CC})$ is a hexagon iff $d(x, M_2) < \left(\sqrt{c_1^2 + c_2^2}\right)\Big/2$; i.e., $x \in (B(M_2, |e_2|/2))^o$ (see the shaded region in Figure 3.4.1 (right)), otherwise it will be a pentagon.

For any $x = (x_0, y_0) \in T(\mathcal{Y}) = T_b$, the functional form of the Γ_1-region is given by

$$\Gamma_1(x = (x_0, y_0), N_{AS}, M_{CC}) = \left\{(x, y) \in T(\mathcal{Y}) : y \geq \ell_1^\perp(x);\ y \geq \ell_2^\perp(x);\ y \leq \ell_3^\perp(x)\right\}$$

where $\ell_1^\perp(x) = \frac{y_0^2 - 2\,x_0\,x + x_0^2}{2\,y_0}$; $\ell_2^\perp(x) = \frac{y_0^2 - 2\,x_0\,x + x_0^2 + 2\,x + 1}{2\,y_0}$; $\ell_3^\perp(x) = \frac{-y_0^2 + c_2^2 + 2\,x_0\,x - x_0^2 - 2\,c_1\,x + c_1^2}{2\,(c_2 - y_0)}$.

We have found and expressed the form of Γ_1-region for a point in $T(\mathcal{Y})$. For a data set of size $n > 1$, we wish to do the same. However, the region $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC})$ is not a k-gon for some fixed $k \in \{k_1, \ldots, k_j\} \subsetneq [n]$ where $j < n$. Furthermore, the η-value, $\eta(\mathcal{X}_n, N_{AS})$, (the

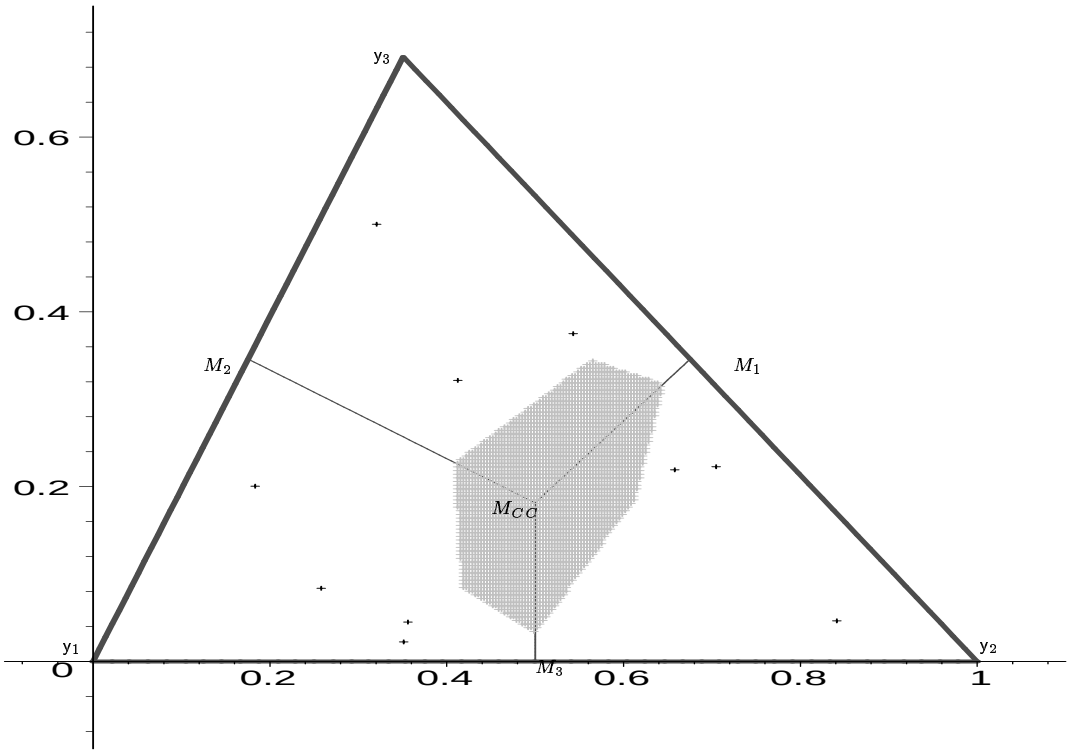


Figure 3.4.2: An empirical Γ_1-region, $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC})$ with $n = 10$

minimum cardinality of an active set to describe the Γ_1-region) is not a fixed constant for all n. We conjecture the following on the upper bound of η-value.

Conjecture 3.4.7. Let $\mathcal{X}_n$ be a random sample from F with continuous density f on $T(\mathcal{Y})$. Then for arc-slice proximity maps, $\eta(\mathcal{X}_n, N_{AS}) = n$ with positive probability for each $n < \infty$.

We can grid the smallest rectangle that contains $T(\mathcal{Y})$ and estimate the Γ_1-region empirically. The empirical Γ_1-region (the union of the grid squares whose centers are verified to be in the actual Γ_1-region) for a particular sample of size $n = 10$ is in Figure 3.4.2. Notice that the empirical region suggests that the actual Γ_1-region is a convex polygon for this data set and $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) = \emptyset$, which implies $\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) > 1$.

However, we can find regions slightly larger than the Γ_1-region for arc-slice proximity regions by picking $k < n$ points which have some particular property and find the intersection of their Γ_1-regions. For example:

(i) Pick the closest points to the vertices; i.e., pick the vertex extrema $X_{\mathsf{y}}(n) \in \text{argmin}_{X \in \mathcal{X}_n} d(X, \mathsf{y})$ for each $\mathsf{y} \in \mathcal{Y}$. Break the distance ties arbitrarily whenever they occur.

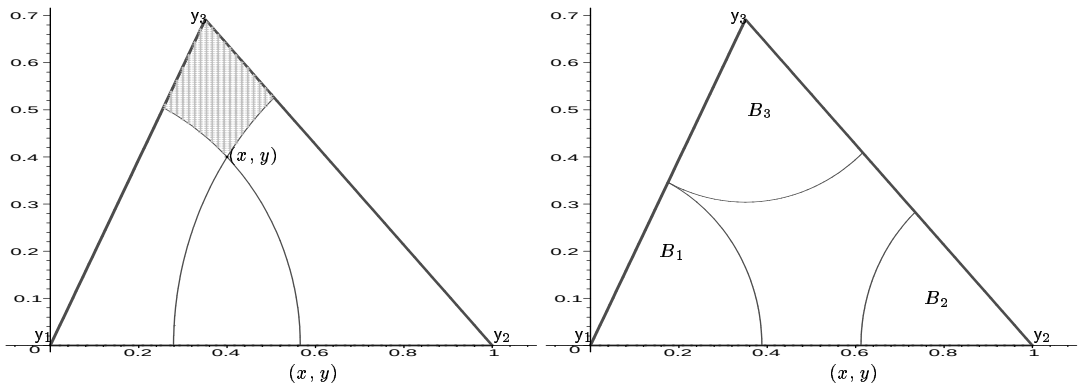


Figure 3.4.3: Figure for $X_{\mathsf{y}_1}(n) = X_{\mathsf{y}_2}(n) = (x, y)$ (left) and B_j for $j \in \{1, 2, 3\}$ for distinct vertex extrema (right).

(ii) Pick the furthest points from M_{CC} in the CC-vertex regions; that is, $\text{argmax}_{X \in \mathcal{X}_n \cap R_{CC}(\mathsf{y})}$ $d(X, M_{CC})$ for each $\mathsf{y} \in \mathcal{Y}$. If $\text{argmax}_{X \in \mathcal{X}_n \cap R_{CC}(\mathsf{y})}$ is not unique, pick any one of the furthest points from M_{CC} arbitrarily. If there are no points in $R_{CC}(\mathsf{y})$ for some y, then don't pick any point in that vertex region.

The methods (i) and (ii) above give regions slightly larger than the exact Γ_1-region for large n, because these points are more "restrictive" than the subsets of $\mathcal{X}_n$ of same size in some way. However, for each method, we can find examples so that Γ_1-region is a proper subset of the region obtained by the method.

The most straightforward choice is picking the vertex extrema, denoted as $X_{\mathsf{y}_j}(n)$ for $j \in \{1, 2, 3\}$. It is possible to have two or three distinct vertex extrema, but $P(X_{\mathsf{y}_1}(n) = X_{\mathsf{y}_2}(n) = X_{\mathsf{y}_3}(n)) = \mathbf{I}(n = 1)$.

If the cardinality of the set of the vertex extrema is two, then two of the vertex extrema should coincide, say $X_{\mathsf{y}_1}(n) = X_{\mathsf{y}_2}(n)$, then the rest of the points should fall, for example, in the shaded region in Figure 3.4.3 (left) given $X_{\mathsf{y}_1}(n) = X_{\mathsf{y}_2}(n) = (x, y)$.

Proposition 3.4.8. *Let $\mathcal{X}_n$ be a random sample from F whose support is in $T(\mathcal{Y})$ and density f is nonzero around y_j, for $j \in \{1, 2, 3\}$. Let $E_{\mathcal{Y}}(n)$ be the event that vertex extrema of $\mathcal{X}_n$ are distinct. Then $P(E_{\mathcal{Y}}(n)) \to 1$ as $n \to \infty$.*

Proof: Let $\rho := d(\mathsf{y}_1, \mathsf{y}_2)/2$ in the basic triangle, T_b. So $\rho = |e_2|/2 = \left(\sqrt{c_1^2 + c_2^2}\right)/2$. Let $B_j := B(\mathsf{y}_j, \rho) \cap T(\mathcal{Y})$ be the ball slices at y_j for $j \in \{1, 2, 3\}$. See Figure 3.4.3 (right).

Let D_j be the event that $\mathcal{X}_n \cap B_j \neq \emptyset$, for $j \in \{1,2,3\}$. If $\mathcal{X}_n \cap B_j \neq \emptyset$, for all $j \in \{1,2,3\}$, then three distinct vertex extrema will exist. Then $P\left(\cap_{j=1}^3 D_j\right) \leq P(E_{\mathcal{Y}}(n))$ and $P\left(\cap_{j=1}^3 D_j\right) = 1 - P\left(\cup_{j=1}^3 D_j^c\}\right)$, where D_j^c is the complement of D_j. Thus

$$P\left(\cup_{j=1}^3 D_j^c\right) = P(D_1^c)+P(D_2^c)+P(D_3^c)-P(D_1^c\cap D_2^c)-P(D_1^c\cap D_3^c)-P(D_2^c\cap D_3^c)+P(D_1^c\cap D_2^c\cap D_3^c).$$

Now,

$$P(D_j^c) = (1 - P(X \in B_j))^n \to 0 \text{ as } n \to \infty,$$

since by hypothesis $P(X \in B_j) > 0$ for $j \in \{1,2,3\}$. Then $P(D_j^c \cap D_k^c) \leq P(D_j^c) \to 0$ as $n \to \infty$ for $j \neq k$ and $j,k \in \{1,2,3\}$, and $P(D_1^c \cap D_2^c \cap D_3^c) \leq P(D_j^c) \to 0$ as $n \to \infty$. Hence $P\left(\cup_{j=1}^3 D_j^c\right) \to 0$ as $n \to \infty$, so $P\left(\cap_{j=1}^3 D_j\right) = 1 - P\left(\cup_{j=1}^3 D_j^c\right) \to 1$ as $n \to \infty$. Thus $P(E_{\mathcal{Y}}(n)) \to 1$ as $n \to \infty$. ■

The region $\cap_{j=1}^3 \Gamma_1(X_{\mathsf{y}_j}(n), N_{AS}, M_{CC})$ which is slightly larger than the actual Γ_1-region will be used in finding the asymptotic bounds for the domination number of the corresponding PCD.

3.4.2 Γ_1-Regions for r-Factor Proportional-Edge Proximity Maps

For $r \in [1,\infty]$, let N_{PE}^r be the *r-factor central similarity proximity map* (see Section 3.3.2). For $N_{PE}^r(\cdot, M)$, the Γ_1-region is constructed as follows; see also Figure 3.4.4. Let $\xi_j(r,x)$ be the line parallel to e_j such that $\xi_j(r,x) \cap T(\mathcal{Y}) \neq \emptyset$ and $r\, d(\mathsf{y}_j, \xi_j(r,x)) = d(\mathsf{y}_j, \ell(\mathsf{y}_j, x))$ for $j \in \{1,2,3\}$. Then

$$\Gamma_1\left(x, N_{PE}^r, M\right) = \cup_{j=1}^3 \left[\Gamma_1\left(x, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_j)\right]$$

where

$$\Gamma_1\left(x, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_j) = \{z \in R_M(\mathsf{y}_j) : d(\mathsf{y}_j, \ell(\mathsf{y}_j, z)) \geq d(\mathsf{y}_j, \xi_j(r,x)\} \text{ for } j \in \{1,2,3\}.$$

Notice that $r \geq 1$ implies $x \in \Gamma_1\left(x, N_{PE}^r, M\right)$. Furthermore, $\lim_{r\to\infty} \Gamma_1\left(x, N_{PE}^r, M\right) = T(\mathcal{Y})$ for all $x \in T(\mathcal{Y}) \setminus \mathcal{Y}$ and so we define $\Gamma_1(x, N_{PE}^{r=\infty}, M) = T(\mathcal{Y})$ for all such x. For $x \in \mathcal{Y}$, $\Gamma_1\left(x, N_{PE}^r, M\right) = \{x\}$ for all $r \in [1,\infty]$.

The functional form of $\Gamma_1(x = (x_0, y_0), N_{PE}^r, M)$ in the basic triangle T_b is given by

$$\Gamma_1\left(x = (x_0,y_0), N_{PE}^r, M\right) = \cup_{j=1}^3 \left[\Gamma_1(x = (x_0,y_0), N_{PE}^r, M) \cap R_M(y_j)\right]$$

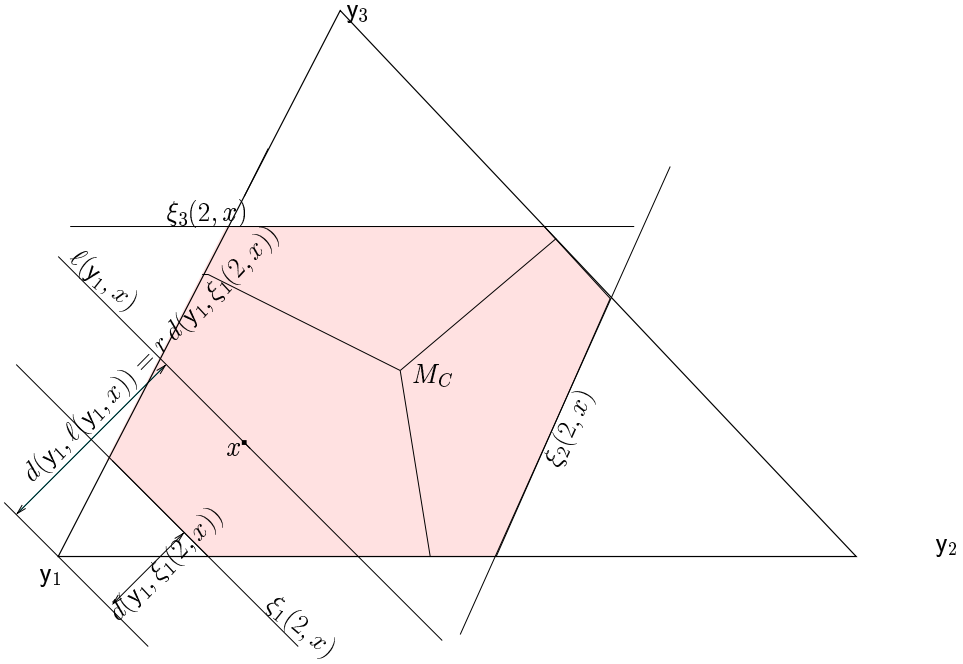


Figure 3.4.4: Construction of the Γ_1-region, $\Gamma_1(x, N_{PE}^{r=2}, M_C)$ (shaded region).

where

$$\Gamma_1(x=(x_0,y_0),N_{PE}^r,M)\cap R_M(\mathsf{y}_1)=\left\{(x,y)\in R_M(\mathsf{y}_1):y\geq\frac{y_0}{r}-\frac{c_2\,(r\,x-x_0)}{(1-c_1)\,r}\right\},$$
$$\Gamma_1(x=(x_0,y_0),N_{PE}^r,M)\cap R_M(\mathsf{y}_2)=\left\{(x,y)\in R_M(\mathsf{y}_1):y\geq\frac{y_0}{r}-\frac{c_2\,(r\,(x-1)+1-x_0}{c_1\,r}\right\},$$
$$\Gamma_1(x=(x_0,y_0),N_{PE}^r,M)\cap R_M(\mathsf{y}_3)=\left\{(x,y)\in R_M(\mathsf{y}_1):y\leq\frac{y_0-c_2\,(1-r)}{r}\right\}.$$

Notice that $\Gamma_1\,(x, N_{PE}^r, M_C)$ is a convex hexagon for all $r \geq 2$ and $x \in T(\mathcal{Y}) \setminus \mathcal{Y}$, (since for such an x, $\Gamma_1\,(x, N_{PE}^r, M_C)$ is bounded by $\xi_j(r,x)$ and e_j for all $j \in \{1,2,3\}$, see also Figure 3.4.4.) else it is either a convex hexagon or a non-convex polygon depending on the location of x and the value of r.

So far, we have described the Γ_1-region for a point in $x \in T(\mathcal{Y})$. For a set $\mathcal{X}_n$ of size n in $T(\mathcal{Y})$, the region $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M)$ can be exactly described by the edge extrema.

Definition 3.4.9. The (closest) *edge extrema* of a set B in $T(\mathcal{Y})$ are the points closest to the edges of $T(\mathcal{Y})$, denoted x_{e_j} for $j \in \{1,2,3\}$; that is, $x_{e_j} \in \operatorname{arginf}_{x\in B} d(x, e_j)$.

Note that if $B = \mathcal{X}_n$ is a random sample of size n from F then the edge extrema, denoted $X_{e_j}(n)$, are random variables.

Proposition 3.4.10. *Let B be any set of n distinct points in $T(\mathcal{Y})$. For r-factor proportional-edge proximity maps with M-vertex regions, $\Gamma_1(B, N_{PE}^r) = \cap_{k=1}^3 \Gamma_1(x_{e_k}, N_{PE}^r)$.*

Proof: Given $B = \{x_1, \ldots, x_n\}$ in $T(\mathcal{Y})$. Note that

$$\Gamma_1(B, N_{PE}^r, M) \cap R_M(\mathsf{y}_j) = \left[\cap_{i=1}^n \Gamma_1(x_i, N_{PE}^r, M)\right] \cap R_M(\mathsf{y}_j),$$

but by definition $x_{e_j} \in \text{argmax}_{x \in B}\, d(\mathsf{y}_j, \xi_j(r, x))$, so

$$\Gamma_1(B, N_{PE}^r, M) \cap R_M(\mathsf{y}_j) = \Gamma_1(x_{e_j}, N_{PE}^r, M) \cap R_M(\mathsf{y}_j) \text{ for } j \in \{1, 2, 3\}.$$

Furthermore, $\Gamma_1(B, N_{PE}^r, M) = \cup_{j=1}^3 \left[\Gamma_1(x_{e_j}, N_{PE}^r, M) \cap R_M(\mathsf{y}_j)\right]$, and

$$\Gamma_1(x_{e_j}, N_{PE}^r, M) \cap R_M(\mathsf{y}_j) = \cap_{k=1}^3 \left[\Gamma_1(x_{e_k}, N_{PE}^r, M) \cap R_M(\mathsf{y}_j)\right] \text{ for } j \in \{1, 2, 3\}.$$

Combining these two results, we obtain $\Gamma_1(B, N_{PE}^r, M) = \cap_{k=1}^3 \Gamma_1(x_{e_k}, N_{PE}^r, M)$. ■

From the above proposition, we see that the Γ_1-region for B as in proposition can also be written as the union of three regions of the form

$$\Gamma_1(B, N_{PE}^r, M) \cap R_M(\mathsf{y}_j) = \{z \in R_M(\mathsf{y}_j) : d(\mathsf{y}_j, \ell(\mathsf{y}_j, z)) \geq d(\mathsf{y}_j, \xi_j(r, x_{e_j}))\} \text{ for } j \in \{1, 2, 3\}.$$

Corollary 3.4.11. *Let $\mathcal{X}_n$ be a random sample from a continuous distribution F on $T(\mathcal{Y})$. For r-factor proportional-edge proximity maps with M-vertex regions, $\eta(\mathcal{X}_n, N_{PE}^r) \leq 3$ with equality holding with positive probability for $n \geq 3$.*

Proof: From Proposition 3.4.10, $\eta(\mathcal{X}_n, N_{PE}^r) \leq 3$. Furthermore, $X_e(n)$ is unique for each edge e a.s. since F is continuous, and there are three distinct edge extrema with positive probability. Hence $P(\eta(\mathcal{X}_n, N_{PE}^r) = 3) > 0$ for $n \geq 3$. ■

Note that $P(\eta(\mathcal{X}_n, N_{PE}^r) = 3) \to 1$ as $n \to \infty$ for $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}))$, since edge extrema are distinct with probability 1 as $n \to \infty$ as shown in the following theorem.

Theorem 3.4.12. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and let $E_{c,3}(n)$ be the event that (closest) edge extrema are distinct. Then $P(E_{c,3}(n)) \to 1$ as $n \to \infty$.*

Proof: Using the uniformity preserving transformation $\phi_e : (x, y) \to (u, v)$, where $u(x, y) = x + \frac{1-2c_1}{\sqrt{3}} y$ and $v(x, y) = \frac{\sqrt{3}}{2c_2} y$ from Section 2.2.1, we can, without loss of generality, assume

$\mathcal{X}_n$ is a random sample from $\mathcal{U}(T_e)$. Observe also that the edge extrema in T_b are mapped into the edge extrema in T_e. Note that the probability of edge extrema all being equal to each other is $P(X_{e_1}(n) = X_{e_2}(n) = X_{e_3}(n)) = \mathbf{I}(n = 1)$. Let $E_{c,2}(n)$ be the event that there are only two distinct (closest) edge extrema. Then for $n > 1$,

$$P(E_{c,2}(n)) = P(X_{e_1}(n) = X_{e_2}(n)) + P(X_{e_1}(n) = X_{e_3}(n)) + P(X_{e_2}(n) = X_{e_3}(n))$$

since the intersection of events $X_{e_i}(n) = X_{e_j}(n)$ and $X_{e_i}(n) = X_{e_k}(n)$ is equivalent to the event $X_{e_1}(n) = X_{e_2}(n) = X_{e_3}(n)$. Notice also that $P(E_{c,2}(n = 2)) = 1$. So, for $n > 2$, there are two or three distinct edge extrema with probability 1. Hence $P(E_{c,3}(n)) + P(E_{c,2}(n)) = 1$ for $n > 2$.

We will show that $P(E_{c,2}(n)) \to 0$ as $n \to \infty$, which will imply the desired result.

First consider $P(X_{e_1}(n) = X_{e_2}(n))$. The event $X_{e_1}(n) = X_{e_2}(n) = X_e$ is equivalent to the event that $\mathcal{X}_n \subset \{U \in T_e : \ d(\mathsf{y}_1, \ell(\mathsf{y}_1, U)) \leq d(\mathsf{y}_1, \ell(\mathsf{y}_1, X_e)),\ d(\mathsf{y}_2, \ell(\mathsf{y}_2, U)) \leq d(\mathsf{y}_2, \ell(\mathsf{y}_2, X_e))\}$. For example, if given $X_{e_1}(n) = X_{e_2}(n) = (x, y)$ the remaining $n - 1$ points will lie in the shaded region in Figure 3.4.5.

The pdf of such (X, Y) is $f(x, y) = n\ \left(4/\sqrt{3}\right)\ \left(y^2/\sqrt{3}\right)^{n-1}$. Let $\varepsilon > 0$, by Markov's inequality, $P\left(\sqrt{3}/2 - Y > \varepsilon\right) \leq \mathbf{E}\left[\sqrt{3}/2 - Y\right]/\varepsilon$. But,

$$\begin{aligned}
\mathbf{E}\left[\sqrt{3}/2 - Y\right] &= \int_0^{1/2} \int_0^{\sqrt{3}\,x} \left(\sqrt{3}/2 - y\right)\ n\ \left(y^2/\sqrt{3}\right)^{n-1}\ \left(4/\sqrt{3}\right) dy\, dx \\
&\quad + \int_{1/2}^{1} \int_0^{\sqrt{3}\,(1-x)} \left(\sqrt{3}/2 - y\right)\ n\ \left(y^2/\sqrt{3}\right)^{n-1}\ \left(4/\sqrt{3}\right) dy\, dx \\
&= 4\left(\sqrt{3}/4\right)^n \frac{1}{n\,(4\,n^2 - 1)},
\end{aligned}$$

which converges to 0 as $n \to \infty$. Hence for each $\varepsilon > 0$, $P\left(\sqrt{3}/2 - Y > \varepsilon\right) \to 0$ as $n \to \infty$. Hence $P(X_{e_1}(n) = X_{e_2}(n) \neq \mathsf{y}_3) \to 0$ as $n \to \infty$. Furthermore, $P(X_{e_1}(n) = X_{e_2}(n) = \mathsf{y}_3) \leq P(X_{e_2}(n) \in e_2) = 0$ for all $n \geq 1$. So $P(X_{e_1}(n) = X_{e_2}(n)) \to 0$ as $n \to \infty$.

Likewise, by symmetry, it follows that $\lim_{n\to\infty} P(X_{e_1}(n) = X_{e_3}(n)) = \lim_{n\to\infty} P(X_{e_2}(n) = X_{e_3}(n)) = 0$. Hence $P(E_{c,2}(n)) \to 0$ as $n \to \infty$. Thus $P(E_{c,3}(n)) \to 1$ as $n \to \infty$. ■

The above theorem implies that the asymptotic distribution of $\eta\left(\mathcal{X}_n, N_{PE}^r\right)$ is degenerate with $P(\eta\left(\mathcal{X}_n, N_{PE}^r\right) = 3) \to 1$ as $n \to \infty$. But for finite n, $\eta\left(\mathcal{X}_n, N_{PE}^r\right)$ for $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}))$ has

the following non-degenerate distribution.

$$\eta\left(\mathcal{X}_n, N_{PE}^r\right) \;=\; \begin{cases} 2 & \text{wp } \pi_2(n) \\ 3 & \text{wp } \pi_3(n) = 1 - \pi_2(n), \end{cases}$$

where $\pi_2(n) \in (0,1)$ is the probability of edge extrema for any two distinct edges being concurrent.

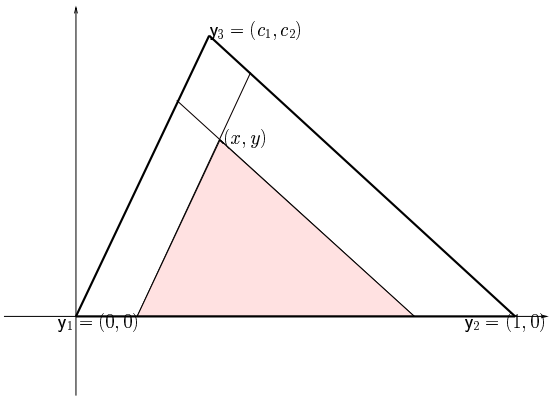


Figure 3.4.5: The figure for $X_{e_1}(n) = X_{e_2}(n) = (x, y)$.

Remark 3.4.13. If $\mathcal{X}_n$ is a random sample from F such that $\mathcal{S}(F) \cap \{x \in T(\mathcal{Y}) : d(x, e_j) \leq \varepsilon_1\}$ has positive measure and $\mathcal{S}(F) \cap B(\mathsf{y}_j, \varepsilon_2) = \emptyset$ for some $\varepsilon_1, \varepsilon_2 > 0$, then $P(E_{c,3}(n)) \to 1$ as $n \to \infty$ follows trivially. However, the case that F has positive density around the vertices $\mathcal{Y}$ needs more work. □

For $r \geq 3/2$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}$, $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M) \neq \emptyset$ a.s., since $\mathscr{R}_S(N_{PE}^r, M) \neq \emptyset$ and $\mathscr{R}_S(N_{PE}^r, M) \subseteq \Gamma_1(\mathcal{X}_n, N_{PE}^r, M)$.

Now, for $n > 1$, let $X_{e_j}(n) = x_{e_j} = (u_j, w_j)$ be given for $j \in \{1, 2, 3\}$, be the edge extrema in a given realization of $\mathcal{X}_n$. Then the functional form of Γ_1-region in T_b is given by

$$\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) = \cup_{j=1}^3 \left[\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_j)\right]$$

where

$$\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_1) = \left\{(x,y) \in R_M(\mathsf{y}_1) : y \geq \frac{w_1}{r} - \frac{c_2\left(r\,x - u_1\right)}{(1-c_1)\,r}\right\},$$
$$\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_2) = \left\{(x,y) \in R_M(\mathsf{y}_1) : y \geq \frac{w_2}{r} - \frac{c_2\left(r\left(x-1\right)+1-u_2\right)}{c_1\,r}\right\},$$
$$\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_3) = \left\{(x,y) \in R_M(\mathsf{y}_1) : y \leq \frac{w_3 - c_2\left(1-r\right)}{r}\right\}.$$

Note that, for $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}))$, $P(\eta\left(\mathcal{X}_n, N_{PE}^r\right) = 3) \to 1$ as $n \to \infty$, since the edge extrema are distinct with probability 1 as $n \to \infty$. However, for $r < 3/2$, the region $\Gamma_1\left(x, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_j)$ might be empty for some $j \in \{1,2,3\}$. Furthermore, if $M \in (\mathscr{T}^r)^o$ (see Equation 3.3.1 for $\mathscr{T}^r$) with $r < 3/2$, then $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right)$ will be empty with probability 1 as $n \to \infty$. In such a case, there is no Γ_1-region to construct. But the definition of the η-value still works in the sense that $\Gamma_1(X_n, N_{PE}^r, M) = \emptyset = \cap_{x \in S_M}\Gamma_1\left(x, N_{PE}^r, M\right)$ (see Definition 3.4.3 for S_M) and $\Gamma_1\left(x, N_{PE}^r, M\right) \neq \emptyset$ for all $x \in \mathcal{X}_n$ since $x \in \Gamma_1\left(x, N_{PE}^r, M\right)$. To determine whether the Γ_1-region is empty or not, it suffices to check the intersection of the Γ_1-regions of the edge extrema. If $M \notin (\mathscr{T}^r)^o$, the Γ_1-region is guaranteed to be nonempty.

Remark 3.4.14.

- For $r_1 \leq r_2$, $\Gamma_1\left(x, N_{PE}^{r_1}, M\right) \subseteq \Gamma_1\left(x, N_{PE}^{r_2}, M\right)$ for all $x \in T(\mathcal{Y})$.
- For $3/2 < r_1 < r_2$, $\Gamma_1\left(x, N_{PE}^{r_1}, M\right) \subsetneq \Gamma_1\left(x, N_{PE}^{r_2}, M\right)$.
- Let X be a random variable in $\mathcal{X}_n$ which is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for $r_1 < r_2$, $A\left(\Gamma_1\left(X, N_{PE}^{r_1}, M\right)\right) \leq^{ST} A\left(\Gamma_1\left(X, N_{PE}^{r_2}, M\right)\right)$ and $A(\Gamma_1(\mathcal{X}_n, N_{PE}^{r_1}, M)) \leq^{ST} A(\Gamma_1(X_n, N_{PE}^{r_2}, M))$. □

Remark 3.4.15. In $\mathbb{R}^d$ with $d > 2$, recall $\mathfrak{S}(\mathcal{Y})$, the simplex based on $d+1$ points that do not lie on the same hyperplane. Furthermore, let $\varrho_j(r,x)$ be the hyperplane such that $\varrho_j(x) \cap \mathfrak{S}(\mathcal{Y}) \neq \emptyset$ and $r\,d(\mathsf{y}_j, \varrho_j(r,x)) = d(\mathsf{y}_j, \Upsilon(\mathsf{y}_j, x))$ for $j \in \{1,2\}, \ldots, d+1$. Then

$$\Gamma_1(x, N_{PE}^r, M_C) \cap R_{CM}(\mathsf{y}_j) = \{z \in R_{CM}(\mathsf{y}_j) : d(\mathsf{y}_j, \Upsilon(\mathsf{y}_j, z)) \geq d(\mathsf{y}_j, \varrho_j(r,x)\} \text{ for } j \in \{1,2,3\}.$$

Hence $\Gamma_1(x, N_{PE}^r, M_C) = \cup_{j=1}^{d+1}(\Gamma_1(x, N_{PE}^r, M_C) \cap R_{CM}(\mathsf{y}_j))$. Furthermore, it is easy to see that $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M_C) = \cap_{j=1}^{d+1}\Gamma_1(X_{\varphi_j}(n), N_{PE}^r, M_C)$, where $X_{\varphi_j}(n)$ is one of the closest points in $\mathcal{X}_n$ to face φ_j. □

3.4.3 Γ_1-Regions for τ-Factor Central Similarity Proximity Maps

For $\tau \in [0,1]$, let N_{CS}^{τ} be the *τ-factor central similarity proximity map* (see Section 3.3.3).

For N_{CS}^{τ}, the Γ_1-region is constructed as follows. Let $e_j^{\tau}(x)$ be the edge of $T_{\tau}(x)$ parallel to edge e_j for $j \in \{1,2,3\}$. Now, suppose $u \in R_M(e_3)$ and let $q_j(\tau,x)$ for $j \in \{1,2\},\ldots,7$ be the lines such that

$$\begin{aligned}
&v \in q_1(\tau,u) \cap R_M(e_3) \Longrightarrow u \in e_1^{\tau}(v), \qquad && v \in q_2(\tau,u) \cap R_M(e_3) \Longrightarrow u \in e_2^{\tau}(v),\\
&v \in q_3(\tau,u) \cap R_M(e_1) \Longrightarrow u \in e_2^{\tau}(v), && v \in q_4(\tau,u) \cap R_M(e_1) \Longrightarrow u \in e_3^{\tau}(v),\\
&v \in q_5(\tau,u) \cap R_M(e_2) \Longrightarrow u \in e_3^{\tau}(v), && v \in q_6(\tau,u) \cap R_M(e_2) \Longrightarrow u \in e_2^{\tau}(v),\\
&v \in q_7(\tau,u) \cap R_M(e_3) \Longrightarrow u \in e_3^{\tau}(v).
\end{aligned} \tag{3.4.1}$$

Then $\Gamma_1(x, N_{CS}^{\tau}, M)$ is the region bounded by these lines. See also Figure 3.4.6. $\Gamma_1(x, N_{CS}^{\tau}, M)$ for $x \in R_M(e_j)$ for $j \in \{1,2\}$ can be described similarly.

Notice that $\tau > 0$ implies that $x \in \Gamma_1(x, N_{CS}^{\tau}, M)$. Furthermore, $\Gamma_1(x, N_{CS}^{\tau}, M) = \{x\}$ iff (i) $\tau = 0$ or (ii) $\tau < 1$ and $x \in \partial(T(\mathcal{Y}))$.

The Γ_1-region $\Gamma_1(x, N_{CS}^{\tau}, M)$ is a convex k-gon with $3 \le k \le 7$ vertices. In particular, for $\tau = 1$, $\Gamma_1(x, N_{CS}^{\tau=1}, M)$ is a convex hexagon. See Figure 3.4.7 (right).

Proposition 3.4.16. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$. For τ-factor central similarity proximity maps with M-edge regions (by definition $M \in T(\mathcal{Y})^o$) and $\tau > 0$, $\eta(\mathcal{X}_n, N_{CS}^{\tau}) \le 3$ with equality holding with positive probability for $n \ge 3$.*

Proof: Let $M \in T(\mathcal{Y})^o$ and $\tau > 0$. Then given $\mathcal{X}_n$, for $j \in \{1,2,3\}$, we have

$$\begin{aligned}
\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau}, M)) \cap R_M(e_j) &= \cap_{i=1}^{n} [\Gamma_1(X_i, N_{CS}^{\tau}, M) \cap R_M(e_j)]\\
&= \Gamma_1(X_{e_k}(n), N_{CS}^{\tau}, M) \cap \Gamma_1(X_{e_l}(n), N_{CS}^{\tau}, M) \cap R_M(e_j)
\end{aligned}$$

where $k, l \neq j$, since for $x \in R_M(e_j)$, if $\{X_{e_k}(n), X_{e_l}(n)\} \subset N_{CS}^{\tau}(x)$, then $\mathcal{X}_n \subset N_{CS}^{\tau}(x)$. Hence for each edge we need the edge extrema with respect to the other edges, then the minimum active set is $S_M = \{X_{e_1}(n), X_{e_2}(n), X_{e_3}(n)\}$, hence $\eta(\mathcal{X}_n, N_{CS}^{\tau}) \le 3$. Furthermore, for the random sample $\mathcal{X}_n$, $X_e(n)$ is unique for each edge e with probability 1 and there are three distinct edge extrema with positive probability. Hence $P(\eta(\mathcal{X}_n, N_{CS}^{\tau}) = 3) > 0$ for $n \ge 3$. ■

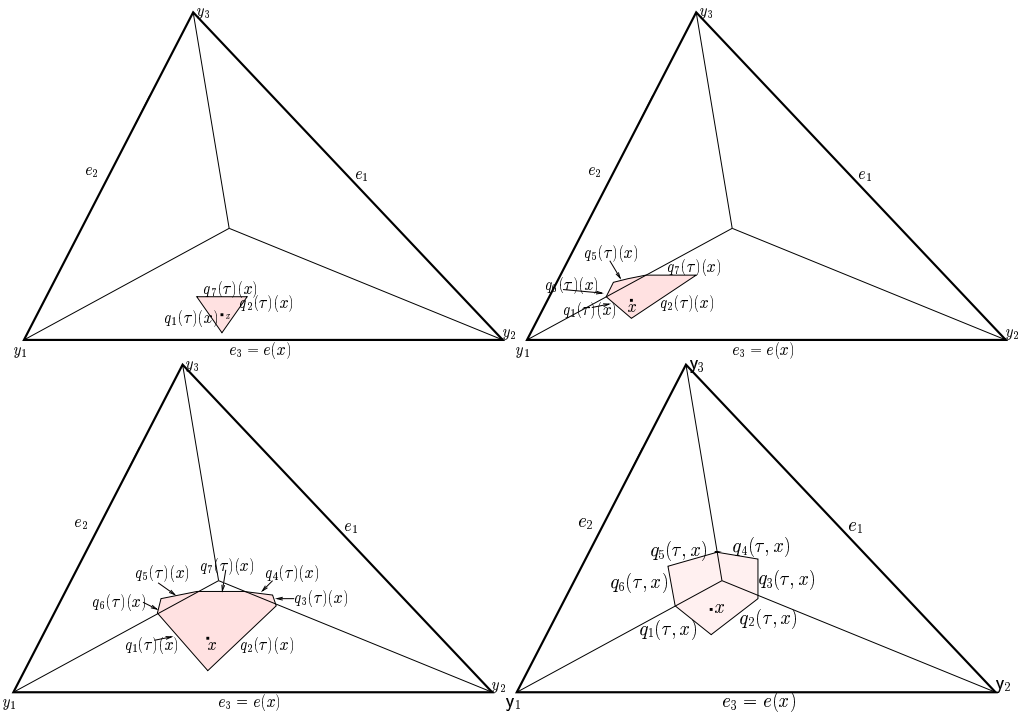


Figure 3.4.6: Examples of the four cases of the Γ_1-region, $\Gamma_1\left(x, N_{CS}^{\tau=1/2}, M_C\right)$ with four distinct $x \in R_{CM}(e_3)$ (shaded regions).

Note that for $\tau > 0$ and $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}))$, $P(\eta(\mathcal{X}_n, N_{CS}^{\tau}) = 3) \to 1$ as $n \to \infty$, since the edge extrema are distinct with probability 1 as $n \to \infty$. Here $\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau=1}, M) \neq \emptyset$ for all n, because by construction $M \in \Gamma_1(\mathcal{X}_n, N_{CS}^{\tau=1}, M)$ since $N_{CS}^{\tau=1}(M) = T(\mathcal{Y})$. Furthermore, $\eta\left(\mathcal{X}_n, N_{CS}^{\tau}\right) \stackrel{d}{=} \eta\left(\mathcal{X}_n, N_{PE}^{r}\right)$ for all $(r, \tau) \in [1, \infty) \times (0, 1]$ where $\stackrel{d}{=}$ denotes the equality in distribution.

Proposition 3.4.17. *The Γ_1-region $\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau=1}, M)$ is a convex hexagon for any set B of size n in $T(\mathcal{Y})^o$.*

Proof: This follows from the fact that each $q_j(x)$ for $j \in \{1, \ldots, 6\}$ is parallel to a line joining y_j to M for $\tau = 1$ ($q_7(x)$ is not used in construction of $\Gamma_1(B, N_{CS}^{\tau=1}, M)$.) See Figure 3.4.7 (right) for an example. ■

For $n > 1$ with $M = M_C$, let $X_{e_j}(n) = x_{e_j} = (u_j, w_j)$ be given for $j \in \{1, 2, 3\}$. Then the functional forms of the lines that determine the boundary of $\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau}, M_C\right)$ are given by

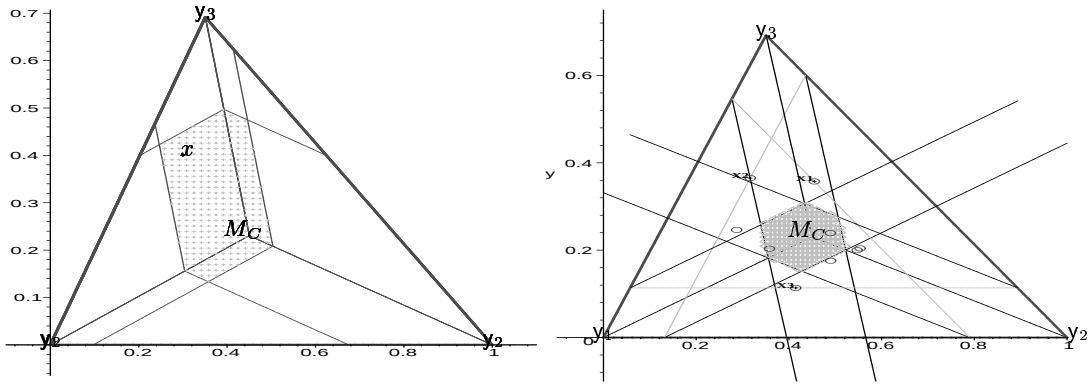


Figure 3.4.7: The Γ_1-region $\Gamma_1(x, N_{CS}^{\tau}, M_C)$ with $x \in R_M(e_2)$ (left) and $\Gamma_1(X_n, N_{CS}^{\tau}, M_C)$ with $n > 1$ (right).

$$
\begin{aligned}
q_1(x) &= \tfrac{w_2\, c_1 + c_2\,(x-u_2)}{c_1+\tau}, & q_3(x) &= \tfrac{c_2\,\tau\,(1-x)+y_0}{1+\tau\,(1-c_1)},\\
q_2(x) &= \tfrac{c_2\,(u_1-x)+w_1\,(1-c_1)}{\tau+1-c_1}, & q_5(x) &= \tfrac{c_2\,x+w_3}{1+\tau\,c_1},\\
q_4(x) &= \tfrac{\tau\,c_2\,(1-x)+c_2\,(u_3-x)-w_3\,c_1}{\tau\,(1-c_1)-c_1}, & q_7(x) &= \tfrac{w_3}{1-\tau},\\
q_6(x) &= \tfrac{c_1\,(1-c_1)\,w_1+c_1\,c_2\,u_1-c_2\,(2\,\tau\,(1-c_1)+c_1\,(1-\tau))\,x+\tau\,c_2\,(1-2\,c_1)}{c_1\,(1-c_1)\,(1-\tau)}.
\end{aligned}
$$

The functional forms of the lines that determine the boundary of $\Gamma_1((x_0, y_0), N_{CS}^{\tau})$ can be obtained by replacing u_j with x_0 and w_j with y_0 for $j \in \{1, 2, 3\}$.

See Figure 3.4.7 for $\Gamma_1(x, N_{CS}^{\tau=1}, M_C)$ and $\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C)$.

Remark 3.4.18.

- For $\tau_1 \leq \tau_2$, $\Gamma_1(x, N_{CS}^{\tau_1}, M) \subseteq \Gamma_1(x, N_{CS}^{\tau_2}, M)$ for all $x \in T(\mathcal{Y})$ and for $\tau_1 < \tau_2$, $\Gamma_1(x, N_{CS}^{\tau_1}, M) \subsetneq \Gamma_1(x, N_{CS}^{\tau_2}, M)$ for all $x \in T(\mathcal{Y})^o$.

- Let X be a random variable in $\mathcal{X}_n$ which is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for $\tau_1 < \tau_2$, $A(\Gamma_1(X, N_{CS}^{\tau_1}, M)) \leq^{ST} A(\Gamma_1(X, N_{CS}^{\tau_2}, M))$ and $A(\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau_1}, M)) \leq^{ST} A(\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau_2}, M))$. $\square$

3.4.4 Characterization of Proximity Maps Using η-Values

By definition, it is immediate that the minimum number of points to describe the Γ_1-region, $\eta(\mathcal{X}_n, N) \leq n$ for any proximity map. We have improved the upper bound for N_{PE}^{τ} and

N_{CS}^{τ}: $\eta(\mathcal{X}_n, N_{PE}^{\tau}) \leq 3$ and $\eta(\mathcal{X}_n, N_{CS}^{\tau}) \leq 3$. However, such an improvement does not hold for $\eta(\mathcal{X}_n, N_{AS})$; that is, there is no $k < n$ such that $\eta(\mathcal{X}_n, N_{AS}) \leq k$ for all $\mathcal{X}_n$.

Below we state a condition for $N(\cdot, M)$ defined with M-vertex regions to have $\eta(\mathcal{X}_n, N) \leq 3$.

Theorem 3.4.19. *Suppose $N(\cdot, M)$ is a proximity region defined with M-vertex regions and B is any set of n distinct points in $T(\mathcal{Y})$. Then $\eta(B, N) \leq 3$ if for each $\mathsf{y} \in \mathcal{Y}$ there exists a point $x(\mathsf{y}) \in B$ (i.e., related to y) such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}) = \Gamma_1(x(\mathsf{y}), N) \cap R_M(\mathsf{y})$.*

Proof: Let $B = \{x_1, \ldots, x_n\} \subset T(\mathcal{Y})$. Suppose there exists a point $x(\mathsf{y}_j) \in B$ such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_j) = \Gamma_1(x(\mathsf{y}_j), N) \cap R_M(\mathsf{y}_j)$ for each $j \in \{1, 2, 3\}$. Then

$$\Gamma_1(x(\mathsf{y}_j), N) \cap R_M(\mathsf{y}_j) = \cap_{i=1}^{n}[\Gamma_1(x_i, N) \cap R_M(\mathsf{y}_j)] = \cap_{k=1}^{3}[\Gamma_1(x(\mathsf{y}_j), N) \cap R_M(\mathsf{y}_j)]$$

and

$$\Gamma_1(B, N) = \cup_{j=1}^{3}[\Gamma_1(x(\mathsf{y}_j), N) \cap R_M(\mathsf{y}_j)].$$

Combining these two results, we get

$$\Gamma_1(B, N) = \cap_{j=1}^{3}\Gamma_1(x(\mathsf{y}_j), N).$$

Hence, the minimum active set $S_M \subseteq \{x(\mathsf{y}_1), x(\mathsf{y}_2), x(\mathsf{y}_3)\}$, which implies $\eta(B, N) \leq 3$. The η-value $\eta(B, N) < 3$ will hold if $x(\mathsf{y}_j)$ are not all distinct. ■

Notice that N_{PE}^{r} satisfies the hypothesis of the Theorem 3.4.19.

Below we state some conditions for $N(\cdot, M)$ defined with M-edge regions to have η-value less than equal to 3.

Theorem 3.4.20. *Suppose $N(\cdot, M)$ is a proximity region defined with M-edge regions and B is set of n distinct points in $T(\mathcal{Y})$. Then $\eta(B, N) \leq 3$ if*

(i) for each $e \in \{e_1, e_2, e_3\}$, there exists a point $x(e) \in B$ such that $\Gamma_1(B, N) \cap R_M(e) = \Gamma_1(x(e), N) \cap R_M(e)$,

or

(ii) there exist points $x(e_k)$, $x(e_l) \in B$ such that $\Gamma_1(B, N) \cap R_M(e_j) = \Gamma_1(x(e_k), N) \cap \Gamma_1(x(e_l), N) \cap R_M(e_j)$ for $k, l \neq j$ with $j \in \{1, 2, 3\}$ and $(k, l) \in \{(1, 2), (1, 3), (2, 3)\}$.

Proof:

(i) Suppose there exists a point $x(e_j) \in B$ such that $\Gamma_1(B,N) \cap R_M(e_j) = \Gamma_1(x(e_j),N) \cap R_M(e_j)$ for each $j \in \{1,2,3\}$. Then

$$\Gamma_1(x(e_j),N) \cap R_M(e_j) = \cap_{i=1}^{n}[\Gamma_1(x_i,N) \cap R_M(e_j)] = \cap_{k=1}^{3}[\Gamma_1(x(e_k),N) \cap R_M(e_j)]$$

and

$$\Gamma_1(B,N) = \cup_{j=1}^{3}[\Gamma_1(x(e_j),N) \cap R_M(e_j)].$$

Combining these two equalities, we get

$$\Gamma_1(\mathcal{X}_n,N) = \cap_{k=1}^{3}\Gamma_1(x(e_k),N).$$

Hence, the minimum active set $S_M \subseteq \{x(e_1),x(e_2),x(e_3)\}$ which implies $\eta(\mathcal{X}_n,N) \leq 3$.

(ii) Suppose there exist points $x(e_k)$ and $x(e_l)$ such that $\Gamma_1(\mathcal{X}_n,N) \cap R_M(e_j) = \Gamma_1(x(e_k),N) \cap \Gamma_1(x(e_l),N) \cap R_M(e_j)$ for $k,l \neq j$. Then

$$\Gamma_1(x(e_k),N) \cap \Gamma_1(x(e_l),N) \cap R_M(e_j) = \cap_{i=1}^{n}[\Gamma_1(x_i,N) \cap R_M(e_j)] = \cap_{q=1}^{3}[\Gamma_1(x(e_q),N) \cap R_M(e_j)]$$

and

$$\Gamma_1(B,N) = \cup_{k,l\neq j}[\Gamma_1(x(e_k),N) \cap \Gamma_1(x(e_l),N) \cap R_M(e_j)].$$

Combining these two results, we get $\Gamma_1(B,N) = \cap_{j=1}^{3}\Gamma_1(x(e_j),N)$. Hence, the minimum active set $S_M \subseteq \{x(e_1),x(e_2),x(e_3)\}$ which implies $\eta(B,N) \leq 3$. ■

Notice that N_{CS}^{τ} satisfies condition (ii) in Theorem 3.4.20.

3.4.5 The Behaviour of $\Gamma_1(\mathcal{X}_n,N)$ for the Proximity Maps in $T(\mathcal{Y})$

In Section 3.4, we have investigated the behaviour of $\Gamma_1(\mathcal{X}_n,N)$ for general proximity maps in Ω. The assertions made about Γ_1-regions will be stronger for the proximity regions we have defined, i.e., for N_{AS}, N_{PE}^{r}, and N_{CS}^{τ}, compared to the general assertions in Section 3.4. One property enjoyed by these proximity maps is that the region $N(x)$ gets larger as x moves along a line from $\Lambda_0(N)$ to $\mathscr{R}_S(N)$ in a region with positive $\mathbb{R}^2$-Lebesgue measure. So the modifications

of the assertions in 3.4 in fact hold not only for $\{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$ but all proximity regions with the same property (mentioned above).

Proposition 3.4.21. *For each type of proximity map $N \in \{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$ and any random sample $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ from a continuous distribution F on $T(\mathcal{Y})$, if $\mathscr{R}_S(N) \neq \emptyset$, then $\mathscr{R}_S(N) \subsetneq \Gamma_1(\mathcal{X}_n, N)$ a.s. for each $n < \infty$.*

Proof: We have shown that $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}_n, N)$ (see Proposition 3.4.1). Moreover, $\mathscr{R}_S(N) \neq \emptyset$ for N_{AS}, N_{PE}^r with $r > 3/2$, and $N_{CS}^{\tau=1}$. For these proximity regions, $\mathscr{R}_S(N) = \Gamma_1(\mathcal{X}_n, N)$ with probability 0 for each finite n since

(i) for N_{AS}, $\Gamma_1(\mathcal{X}_n, N_{AS}, M) = \mathscr{R}_S(N_{AS}, M)$ iff $X_{\mathsf{y}}(n) = \mathsf{y}$ for each $\mathsf{y} \in \mathcal{Y}$ which happens with probability 0,

(ii) for $N \in \{N_{CS}^{\tau=1}, N_{PE}^{r>3/2}\}$, $\Gamma_1(\mathcal{X}_n, N, M)) = \mathscr{R}_S(N, M)$ iff $X_{e_j}(n) \in e_j$ for each $j \in \{1, 2, 3\}$ which happens with probability 0.

Furthermore, for N_{PE}^r with $r < 3/2$, $\mathscr{R}_S(N_{PE}^r, M) \neq \emptyset$ iff $M \notin (\mathscr{T}^r)^o$ (see Equation 3.3.1 for $\mathscr{T}^r$), say M is such that $d(M, e_2) < d(\mathsf{y}_2, e_2)/r$. Then $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M) = \mathscr{R}_S(N_{PE}^r, M)$ iff $X_{e_2}(n) \in e_2$ which happens with probability 0. Similarly the same result also holds for edges e_1 and e_3. ■

Note that, if $\mathscr{R}_S(N) = \emptyset$ and $\mathcal{X}_n$ is a random sample from a continuous distribution on $T(\mathcal{Y})$, then $\Gamma_1(\mathcal{X}_n, N) = \emptyset$ a.s. as $n \to \infty$. In particular, this holds for $N_{PE}^r(\cdot, M)$ with $r < 3/2$ and $M \in (\mathscr{T}^r)^o$. Lemma 3.4.5 holds as stated. In Lemma 3.4.5, we have shown that $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing. Furthermore, for the proximity regions $\{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$, we can state that $\Gamma_1(\mathcal{X}(n), N) \supsetneq \Gamma_1(\mathcal{X}(n+1), N)$ with positive probability, since the new point in $\mathcal{X}(n+1)$ has positive probability to fall closer to the subset of $T(\mathcal{Y})$ that defines $\mathscr{R}_S(N)$ (e.g., $\partial(T(\mathcal{Y}))$). We restate Lemma 3.4.5 and prove it for the proximity maps $\{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$.

Theorem 3.4.22. *Given a sequence of random variables $X_1, X_2, X_3, \ldots \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}))$, let $\mathcal{X}(n) := \mathcal{X}(n-1) \cup \{X_n\}$ for $n = 0, 1, 2, \ldots$ with $\mathcal{X}(0) := \emptyset$. Then for each $N \in \{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$, $\Gamma_1(\mathcal{X}(n), N) \downarrow \mathscr{R}_S(N)$ as $n \to \infty$ a.s., in the sense that $\Gamma_1(\mathcal{X}(n), N) \supseteq \Gamma_1(\mathcal{X}(n+1), N)$ and $A(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)) \downarrow 0$ a.s.*

Proof: From Lemma 3.4.5, we know that $\Gamma_1(\mathcal{X}(n), N) \supseteq \Gamma_1(\mathcal{X}(n+1), N)$. Hence $G_1 := \cap_{k=1}^\infty \Gamma_1(\mathcal{X}(k), N) = \cap_{j=1}^\infty \Gamma_1(X_j, N)$, and $\mathscr{R}_S(N) \subseteq G_1$. Furthermore, $A(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)) >$

0 for each n a.s. From Proposition 5.6. in Karr [21],

$$Y_n \xrightarrow{a.s.} Y \text{ iff } \forall \varepsilon > 0, \quad \lim_{n\to\infty} P\left(\sup_{k\geq n} |Y_k - Y| > \varepsilon\right) = 0.$$

Now, let $\varepsilon > 0$. Thus we need to show that

$$\lim_{n\to\infty} P\left(\sup_{k\geq n} A(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)) > \varepsilon\right) = 0.$$

Note that $\sup_{k\geq n} A(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N) \stackrel{d}{=} A(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)$, so

$$P\left(\sup_{k\geq n} A\left(\cap_{j=1}^{k} \Gamma_1(X_j, N) \setminus \mathscr{R}_S(N)\right) \leq \varepsilon\right) = P\left(A\left(\cap_{j=1}^{n} \Gamma_1(X_j, N) \setminus \mathscr{R}_S(N)\right) \leq \varepsilon\right).$$

Hence it will suffice to show that

$$P\left(A\left(\cap_{j=1}^{n} \Gamma_1(X_j, N) \setminus \mathscr{R}_S(N)\right) \leq \varepsilon\right) \to 1 \text{ as } n \to \infty,$$

for $N \in \{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$ (with $r \in [1, \infty]$ and $\tau \in [0, 1]$). With $r = \infty$ and $\tau = 0$, the result follows trivially.

For $N \in \{N_{PE}^r, N_{CS}^\tau\}$, $\mathscr{R}_S(N)$ is determined by the edge extrema $X_e(n)$. Then for sufficiently large n, $X_e(n)$ will be close enough to e w.p. 1 for each $e \in \{e_1, e_2, e_3\}$; that is, for $\varepsilon > 0$, there exist $\delta_1(\varepsilon)$, $\delta_2(\varepsilon) > 0$ such that $P(d(X_e(n), e) \leq \delta_1(\varepsilon)) = 1$ and $d(X_e(n), e) \leq \delta_1(\varepsilon)$ implies that $\sup_{(x,y)\in\partial(\Gamma_1(\mathcal{X}(n),N)))\times\partial(\mathscr{R}_S(N))} d(x, y) \leq \delta_2(\varepsilon)$, which implies that $A(\Gamma_1(\mathcal{X}(n), N, M) \setminus \mathscr{R}_S(N, M)) \leq \varepsilon$.

For N_{AS}, $\mathscr{R}_S(N_{AS})$ is determined by $\mathcal{Y}$ and for sufficiently large n, $X_{\mathsf{y}}(n)$ will be close enough to $\mathsf{y} \in \mathcal{Y}$ w.p. 1 that $H(\mathcal{X}(n), \mathcal{Y})$ hence $\Gamma_1(\mathcal{X}(n), N_{AS}, M)$ will be slightly larger than $\mathscr{R}_S(N_{AS}, M)$; that is, $A(\Gamma_1(\mathcal{X}(n), N_{AS}, M) \setminus \mathscr{R}_S(N_{AS}, M)) \leq \varepsilon$ will follow. Then $P(A(\Gamma_1(\mathcal{X}(n), N_{AS}, M) \setminus \mathscr{R}_S(N_{AS}, M)) \leq \varepsilon) \to 1$ as $n \to \infty$ which can be shown by arguments similar to the one above. ■

Proposition 3.4.23. *For positive integers $m > n$, let $\mathcal{X}_m$ and $\mathcal{X}_n$ be two random samples from $\mathcal{U}(T(\mathcal{Y}))$. Then $A(\Gamma_1(\mathcal{X}_m, N, M)) \leq^{ST} A(\Gamma_1(\mathcal{X}_n, N, M))$ for proximity maps $N \in \{N_{PE}^r, N_{CS}^\tau\}$, for $r \in [1, \infty)$ and $\tau \in (0, 1]$.*

Proof: Let $m > n$ be two positive integers. Let $X_e(k)$ be a member of the set $\text{argmin}_{X \in \mathcal{X}_k} d(X, e)$ (note that $X_e(k)$ is unique a.s. since X is from $\mathcal{U}(T(\mathcal{Y}))$), then $d(X_e(m),\, e) \leq^{ST} d(X_e(n),\, e)$ for each edge e and $A(\Gamma_1(\mathcal{X}_n, N, M))$ is an increasing function of $d(X_e,\, e)$ for $N \in \{N_{PE}^r,\, N_{CS}^r\}$; the result follows. ■

For N_{AS} the stochastic ordering of $A\left(\Gamma_1(\mathcal{X}_n, N_{AS}, M)\right)$ is still an open problem, although we conjecture that $A\left(\Gamma_1(\mathcal{X}_m, N_{AS}, M)\right) \leq^{ST} A\left(\Gamma_1(\mathcal{X}_n, N_{AS}, M)\right.$ for $\mathcal{X}_m$ and $\mathcal{X}_n$ two random samples from $\mathcal{U}(T(\mathcal{Y}))$ with $m > n$.

Theorem 3.4.24. *Let $\{\mathcal{X}_n\}_{n=1}^{\infty}$ be a sequence of data sets which are iid $\mathcal{U}(T(\mathcal{Y}))$. Then $\Gamma_1(\mathcal{X}_n, N) \xrightarrow{a.s.} \mathscr{R}_S(N)$ as $n \to \infty$ for $N \in \{N_{AS}, N_{PE}^r, N_{CS}^r\}$.*

Proof: Given a sequence $\{\mathcal{X}_n\}_{n=1}^{\infty}$ as in the theorem, $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}_n, N)$ for all n. Let $\varepsilon > 0$. By Proposition 3.4.23, for $N \in \{N_{PE}^r, N_{CS}^r\}$, $\sup_{k \geq n} A\left(\Gamma_1(\mathcal{X}_n, N, M) \setminus \mathscr{R}_S(N, M)\right) \stackrel{d}{=} A\left(\Gamma_1(\mathcal{X}_n, N, M) \setminus \mathscr{R}_S(N, M)\right)$.

For $N \in \{N_{PE}^r, N_{CS}^r\}$, $\mathscr{R}_S(N, M)$ is determined by the edges and $\Gamma_1(\mathcal{X}_n, N, M)$ by the edge extrema $X_e(n)$. As $n \to \infty$, $X_e(n)$ is arbitrarily close to e with probability 1; that is, there exist $\delta_1(\varepsilon) > 0$ and $\delta_2(\varepsilon) > 0$ such that $P(d(X_e(n), e) \leq \varepsilon) = 1$ and $d(X_e(n), e) \leq \delta_1(\varepsilon)$ will imply $\sup_{(x,y) \in \mathscr{R}_S(N) \times \Gamma_1(\mathcal{X}_n, N)} d(x, y) \leq \delta_2(\varepsilon)$ which in turn will imply $A\left(\Gamma_1(\mathcal{X}_n, N, M) \setminus \mathscr{R}_S(N, M)\right) \leq \varepsilon$ for each $e \in \{e_1, e_2, e_3\}$. Hence

$$P\left(\sup_{k \geq n} A\left(\Gamma_1(\mathcal{X}_n, N, M) \setminus \mathscr{R}_S(N, M)\right) \leq \varepsilon\right) =$$

$$P\left(A\left(\Gamma_1(\mathcal{X}_n, N, M) \setminus \mathscr{R}_S(N, M)\right) \leq \varepsilon\right) \to 1 \text{ as } n \to \infty.$$

For N_{AS}, let $H(\mathcal{X}_n, \mathcal{Y}) := \cap_{\mathsf{y} \in \mathcal{Y}} \Gamma_1(X_{\mathsf{y}}(n), N_{AS}, M)$. Then $\Gamma_1(\mathcal{X}_n, N_{AS}, M) \subseteq H(\mathcal{X}_n, \mathcal{Y})$ and $\mathscr{R}_S(N_{AS}, M) \subseteq \Gamma_1(\mathcal{X}_n, N_{AS}, M)$. So, if $A(H(\mathcal{X}_n, \mathcal{Y})) \xrightarrow{a.s.} A(\mathscr{R}_S(N_{AS}, M))$, then the desired result will follow for N_{AS}. Note that $H(\mathcal{X}_n, \mathcal{Y})$ is determined by $\mathcal{Y}$. As $n \to \infty$, $X_{\mathsf{y}}(n)$ will be arbitrarily close to $\mathsf{y} \in \mathcal{Y}$ with probability 1, using arguments similar to above, the result follows. ■

3.4.6 Expected Area of Γ_1-Regions Associated with $N(\cdot)$

Let $\lambda(\cdot)$ be the $\mathbb{R}^d$-Lebesgue measure on $\mathbb{R}^d$ with $d \geq 1$. In $\mathbb{R}$, $\lambda(\cdot)$ is the length $|\cdot|$, in $\mathbb{R}^2$, $\lambda(\cdot)$ is the area $A(\cdot)$, and in $\mathbb{R}^d$, $\lambda(\cdot)$ is the d-dimensional volume $V(\cdot)$. In $\mathbb{R}$ with $\mathcal{Y} = \{0, 1\}$, let $\mathcal{X}_n$

be a random sample from $\mathcal{U}(0,1)$, and $N_S(x) = B(x, r(x))$ where $r(x) = \min(x, 1-x)$. Then,

$$\Gamma_1(\mathcal{X}_n, N_S) = [X_{n:n}/2, (1+X_{1:n})/2] \Longrightarrow \lambda(\Gamma_1(\mathcal{X}_n, N_S)) = |\Gamma_1(\mathcal{X}_n, N_S)| = (1 + X_{1:n} - X_{n:n})/2.$$

Hence the expected length of the Γ_1-region is

$$\begin{aligned} \mathbf{E}[\lambda(\Gamma_1(\mathcal{X}_n, N_S)] &= \mathbf{E}\left[\frac{1+X_{1:n}-X_{n:n}}{2}\right] = \frac{1+\mathbf{E}[X_{1:n}]-\mathbf{E}[X_{n:n}]}{2} \\ &= \frac{1+\frac{1}{n+1}-\frac{n}{n+1}}{2} = \frac{1}{n+1} \to 0 \text{ as } n \to \infty. \end{aligned}$$

In $\mathbb{R}^2$, with three non-collinear points $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$, let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $A(\Gamma_1(\mathcal{X}_n, N_{PE}^r, M)) > 0$ a.s. for all $n < \infty$, $r > 3/2$, $M \in \mathbb{R}^2 \setminus \mathcal{Y}$. Furthermore, $A(\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau=1}, M)) > 0$ a.s. for all $n < \infty$ and $M \in (T(\mathcal{Y}))^o$; $A(\Gamma_1(\mathcal{X}_n, N_{AS}, M)) > 0$ a.s. for all $n < \infty$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}$.

The Γ_1-region, $\Gamma_1(\mathcal{X}_n, N)$, is closely related to the distribution of the domination number of the PCD associated with N. Hence we study the asymptotic behaviour of the expected area $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N, M)]$, as $n \to \infty$, for $N \in \{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$.

3.4.6.1 Expected Area of $\Gamma_1(\mathcal{X}_n, N_{AS}, M)$

Recall that for N_{AS}, the minimum number of "active" observations for a general representation of $\Gamma_1(\mathcal{X}_n, N_{AS}, M)$; that is, $\eta(\mathcal{X}_n, N_{AS})$ has no fixed upper bound that works for all n. Hence, unlike $\Gamma_1(\mathcal{X}_n, N, M)$ with $N \in \{N_{PE}^r, N_{CS}^\tau\}$, there is not a general short form involving some sort of extrema for $\Gamma_1(\mathcal{X}_n, N_{AS}, M)$. However, we can find regions slightly larger than $\Gamma_1(\mathcal{X}_n, N_{AS}, M)$ using the vertex extrema $X_{\mathsf{y}}(n)$ for y; i.e., $X_{\mathsf{y}}(n) \in \operatorname{argmin}_{X \in \mathcal{X}_n} d(X, \mathsf{y})$. Then $\Gamma_1(\mathcal{X}_n, N_{AS}, M) \subseteq H(\mathcal{X}_n, \mathcal{Y}) := \cap_{\mathsf{y} \in \mathcal{Y}} \Gamma_1(X_{\mathsf{y}}(n), N_{AS}, M)$. Hence $A(\Gamma_1(\mathcal{X}_n, N_{AS}, M)) \leq A(H(\mathcal{X}_n, \mathcal{Y}))$, and same inequality holds for the expected values. The expected area of $H(\mathcal{X}_n, \mathcal{Y})$ is determined by the expected locus of $\mathbf{E}[X_{\mathsf{y}}(n)]$; i.e., the expected value of the distance $d(X_{\mathsf{y}}(n), \mathsf{y})$. For example, if $d(X_{\mathsf{y}}(n), \mathsf{y}) = u$, then the locus of $X_{\mathsf{y}}(n)$ is on the arc of the circle in $T(\mathcal{Y})$ with center at y and radius u.

Lemma 3.4.25. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and $R_{\mathsf{y}}(n) := d(X_{\mathsf{y}}(n), \mathsf{y})$ for each $\mathsf{y} \in \mathcal{Y}$. Then $\mathbf{E}[R_{\mathsf{y}}(n)] \to 0$ as $n \to \infty$ for each $\mathsf{y} \in \mathcal{Y}$.*

Proof: Notice that $\mathbf{E}[R_{\mathsf{y}}(n)] \to 0$ as $n \to \infty$ implies that expected locus of $X_{\mathsf{y}}(n)$ is on y.

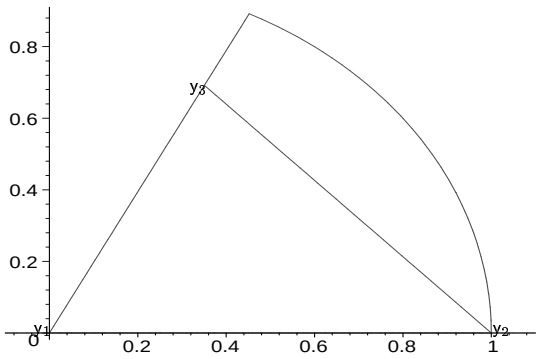


Figure 3.4.8: The smallest arc-circle $S_C(T(\mathcal{Y}))$ that contains $T(\mathcal{Y})$

For $\mathsf{y} = \mathsf{y}_1$, $R_{\mathsf{y}_1}(n) := d(X_{\mathsf{y}_1}(n), \mathsf{y}_1)$. Let $S_C(T(\mathcal{Y}))$ be the smallest slice of the circle centered at y_1 with radius 1, that contains $T(\mathcal{Y})$ (see Figure 3.4.8). Given $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}))$, and let $Y_i \overset{iid}{\sim} \mathcal{U}(S_C(T(\mathcal{Y})))$ for $i = 1, \ldots, n$. Furthermore, let $R_j(\mathsf{y}_1) := d(Y_j, \mathsf{y}_1)$ and $R_{k:n}$ be the k^{th} largest $R_j(\mathsf{y}_1)$ value. Then clearly, $P(R_{\mathsf{y}_1}(n) \leq u) \geq P(R_{1:n} \leq u)$ for all $u \in [0,1]$ with inequality holding for all $u \in [0,1]$. So $R_{\mathsf{y}_1} \leq^{ST} R_{1:n}$, which implies $\mathbf{E}[R_{\mathsf{y}_1}] \leq \mathbf{E}[R_{1:n}]$.

Next $P(R_j(\mathsf{y}_1) \leq u) = \frac{\pi u^2 \hat{\mathsf{y}}_1}{\pi \hat{\mathsf{y}}_1} = u^2$ for all $u \in [0,1]$ where $\hat{\mathsf{y}}_1$ is the inner angle at vertex y_1 in radians, then cumulative distribution function (cdf) and probability density function (pdf) of R_j are given by

$$F_R(u) = u^2\, \mathbf{I}(0 \leq u < 1) + \mathbf{I}(u \geq 1) \text{ and } f_R(u) = 2\,u\, \mathbf{I}(0 \leq u \leq 1),$$

respectively. Thus pdf of $R_{1:n}$ is $f_{R_{1:n}}(u) = 2\,n\,u\,(1-u^2)^{n-1}\mathbf{I}(0 \leq u \leq 1)$. Then

$$\begin{aligned} \mathbf{E}[R_{1:n}] &= \int_0^1 2\,n\,u^2\,(1-u^2)^{n-1} du = \frac{\pi}{2}\,\frac{\Gamma(3/2)\Gamma(n+1)}{\Gamma(n+3/2)} \\ &\approx \frac{\pi}{2}\,\frac{\sqrt{2\,\pi\,n}\,n^n\,e^{-n}}{\sqrt{2\,\pi\,(n+1/2)}(n+1/2)^{n+1/2}e^{-n-1/2}} \approx \frac{\pi}{2}\frac{1}{\sqrt{n}} \to 0, \text{ as } n \to \infty, \end{aligned}$$

where for the first approximation we used the Stirling's approximation of $\Gamma(k+1) \approx k^k\,\sqrt{2\pi k}\,e^{-k}$ and the second approximation is straightforward. Hence $\mathbf{E}[R_{1:n}] \to 0$ which implies $\mathbf{E}[R_{\mathsf{y}_1}(n)] \to 0$ as $n \to \infty$. Similarly, $\mathbf{E}[R_{\mathsf{y}_j}(n)] \to 0$ for $j \in \{2,3\}$ as $n \to \infty$. ■

Theorem 3.4.26. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N_{AS}, M))] \to A(\mathscr{R}_S(N_{AS}, M))$ as $n \to \infty$.*

Proof: Recall that $H(\mathcal{X}_n, \mathcal{Y}) = \cap_{\mathsf{y}\in\mathcal{Y}}\Gamma_1(X_{\mathsf{y}}(n), N_{AS}, M)$ where $H(\mathcal{X}_n, \mathcal{Y})$ is bounded by the perpendicular lines at the midpoints between $X_{\mathsf{y}}(n)$ and y for all $\mathsf{y} \in \mathcal{Y}$. Furthermore, $\Gamma_1(\mathcal{X}_n, N_{AS}, M) \subseteq H(\mathcal{X}_n, \mathcal{Y})$, so $A(\Gamma_1(\mathcal{X}_n, N_{AS}, M)) \leq^{ST} A(H(\mathcal{X}_n, \mathcal{Y}))$, which implies $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N_{AS}, M))] \leq \mathbf{E}[A(H(\mathcal{X}_n, \mathcal{Y}))]$. Additionally, $H(\mathcal{X}_n, \mathcal{Y}) = \mathscr{R}_S(N_{AS}, M)$ iff $X_{\mathsf{y}}(n) = \mathsf{y}$ for all $\mathsf{y} \in \mathcal{Y}$. Moreover, $\mathscr{R}_S(N_{AS}, M) \subseteq \Gamma_1(\mathcal{X}_n, N_{AS}, M)$, which implies $A(\mathscr{R}_S(N_{AS}, M)) \leq \mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N_{AS}, M))]$. In Lemma 3.4.25, we have shown that $\mathbf{E}[R_{\mathsf{y}}(n)] \to 0$ as $n \to \infty$. Hence the expected locus of the perpendicular lines (that determine the boundary of $H(\mathcal{X}_n, \mathcal{Y})$) converge to the midedge bisectors (the expected value of the angle between midedge bisector and the perpendicular lines converge to zero); i.e., the boundary of the superset region $\partial(\mathscr{R}_S(N_{AS}, M))$. Then $\mathbf{E}[A(H(\mathcal{X}_n, \mathcal{Y}))] \to A(\mathscr{R}_S(N_{AS}, M))$ as $n \to \infty$.

So, combining the results we obtain

$$A(\mathscr{R}_S(N_{AS}, M)) \leq \mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N_{AS}, M)] \leq \mathbf{E}[A(H(\mathcal{X}_n, \mathcal{Y}))] \to A(\mathscr{R}_S(N_{AS}, M)) \text{ as } n \to \infty.$$

Hence $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N_{AS}, M)] \to A(\mathscr{R}_S(N_{AS}, M))$ as $n \to \infty$. ■

In particular, if $M = M_{CC}$, then $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}))] \to 0$ as $n \to \infty$, since $\mathscr{R}_S(N_{AS}, M_{CC}) = \{M_{CC}\}$.

3.4.6.2 The Limit of Expected Area of $\Gamma_1(\mathcal{X}_n, N, M)$ for N_{PE}^{τ} and N_{CS}^{τ}

Recall that for $N \in \{N_{PE}^{\tau}, N_{CS}^{\tau}\}$, $\Gamma_1(\mathcal{X}_n, N, M)$ is determined by the (closest) edge extrema $X_e(n) \in \text{argmin}_{X\in\mathcal{X}_n} d(X, e)$, for $e \in \{e_1, e_2, e_3\}$. So, to find the expected area of $\Gamma_1(\mathcal{X}_n, N, M)$, it suffices to find the expected locus of $X_e(n)$; i.e., the expected distance $\mathbf{E}[d(X_e(n), e)]$. For example, for $\mathcal{X}_n$ a random sample from a continuous distribution F, $\text{argmin}_{X\in\mathcal{X}_n} d(X, e)$ is unique a.s., and if $d(X_e(n), e) = u$, then $X_e(n)$ falls on a line parallel to e whose distance to e is u a.s.

Lemma 3.4.27. *Let $D_{e_j}(n) := d(X_{e_j}(n), e_j)$ for $j \in \{1, 2, 3\}$ and $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $\mathbf{E}[D_{e_j}(n)] \to 0$ (i.e., the expected locus of $X_{e_j}(n)$ is on e_j) for each $j \in \{1, 2, 3\}$, as $n \to \infty$.*

Proof: Given $Z_i = (X_i, Y_i) \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}))$. Then for $e = e_3$, $D_{e_3}(n) = Y_{1:n}$ (the minimum y-coordinate of $Z_j \in \mathcal{X}_n$). First observe that $P(Y_j \leq y) = \frac{y\,(2\,c_2 - y)}{c_2^2}$, hence

$$F_Y(y) = \frac{y\,(2\,c_2 - y)}{c_2^2}\mathbf{I}(0 \leq y < c_2) + \mathbf{I}(y \geq c_2).$$

So the pdf of Y_j is $f_Y(y) = 2\,\frac{c_2 - y}{c_2^2}\mathbf{I}(0 \leq y \leq c_2)$. Then the pdf of $Y_{1:n}$ is

$$f_{1:n}(y) = 2\,n(c_2 - y)\left(1 - \frac{y\,(2\,c_2 - y)}{c_2^2}\right)^{n-1} {c_2}^{-2}.$$

Therefore,

$$\mathbf{E}\,[Y_{1:n}] = \int_0^{c_2} 2\,y\,n(c_2 - y)\left(1 - \frac{y\,(2\,c_2 - y)}{c_2^2}\right)^{n-1} {c_2}^{-2} dy = \frac{c_2}{2\,n+1} \to 0, \text{ as } n \to \infty.$$

Hence $\mathbf{E}\,[Y_{1:n}] = \mathbf{E}\,[D_{e_3}(n)] \to 0$. Similarly, $\mathbf{E}\,[D_{e_j}(n)] \to 0$ for $j \in \{1, 2\}$, as $n \to \infty$. ■

Notice that, $D_e(n) = 0$ iff $X_e(n) \in e$, so $\mathbf{E}\,[D_e(n)] \to 0$ iff expected locus of $X_e(n)$ is on the edge e, for each $e \in \{e_1, e_2, e_3\}$.

Theorem 3.4.28. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and $M \in T(\mathcal{Y})^o$. For $N \in \{N_{PE}^r,\, N_{CS}^r\}$, $\mathbf{E}\,[A(\Gamma_1(\mathcal{X}_n, N, M))] \to A(\mathscr{R}_S(N, M))$ as $n \to \infty$.*

Proof: Recall that for $N \in \{N_{PE}^r,\, N_{CS}^r\}$, $\Gamma_1(\mathcal{X}_n, N, M) = \cap_{j=1}^3 \Gamma_1(X_{e_j}(n), N, M)$. Moreover, $\Gamma_1(\mathcal{X}_n, N, M) = \mathscr{R}_S(N, M)$ iff $X_{e_j}(n) \in e_j$ for $j \in \{1, 2, 3\}$. In Lemma 3.4.27, we have shown that expected locus of $X_e(n)$ converges to edge e as $n \to \infty$. Hence the expected locus of $\partial(\Gamma_1(\mathcal{X}_n, N, M))$
$\cap\, R_M(e)$ converges to the $\partial(\mathscr{R}_S(N, M)) \cap R_M(e_j)$ for each $j \in \{1, 2, 3\}$. Hence

$$\mathbf{E}\,[A(\Gamma_1(\mathcal{X}_n, N, M)] \to A(\mathscr{R}_S(N, M)) \text{ as } n \to \infty. \blacksquare$$

Remark 3.4.29. In particular,

i- $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^2, M_C\right)\right)\right] \to 1/4$ as $n \to \infty$, since $\mathscr{R}_S\left(N_{PE}^2, M_C\right) = T(M_1, M_2, M_3)$

ii- $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right)\right)\right] \to 0$ as $n \to \infty$ if $M \in \mathscr{T}^r$, since $\mathscr{R}_S\left(N_{PE}^r, M\right) = \emptyset$ for $M \in \mathscr{T}^r$.

iii- Furthermore, $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right)\right] \to 0$ since $\mathscr{R}_S\left(N_{PE}^{3/2}, M_C\right) = \{M_C\}$,

iv- for any $M \in T(\mathcal{Y})^o$, $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right)\right)\right] \to 0$ as $n \to \infty$, since $\mathscr{R}_S\left(N_{CS}^{\tau}, M\right) = \{M\}$.

v- We also have $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{r}, M_C\right)\right)\right] \to 0$ for $r \in [1, 3/2]$ as $n \to \infty$.

vi- Furthermore, by careful geometric calculations, we get $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{r}, M_C\right)\right)\right] \to (1 - 3/(2\,r))^2\,\sqrt{3}$ for $r \in (3/2, 2]$,

vii- $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{r}, M_C\right)\right)\right] \to \sqrt{3}/4\,(1 - 3/r^2)$ for $r \in (2, \infty]$, as $n \to \infty$.

We also derive the rate of convergence of $\mathbf{E}\left[A(\Gamma_1(\mathcal{X}_n, N_{PE}^{r}, M_C))\right]$ for $r = 3/2$.

Theorem 3.4.30. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$. For $r = 3/2$, the expected area of the the Γ_1-region, $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{r}, M_C\right)\right)\right]$, converges to zero, at rate $O\left(n^{-2}\right)$.*

Proof: See Appendix A.1.1 for the proof. ■

Remark 3.4.31. We can also define the regions associated with $\gamma(\mathcal{X}_n, N) = k$ for $k \leq n$. Let B^k be the Cartesian product of k copies of the set B. For example, the Γ_2-*region* for a proximity map $N(\cdot)$ and set $B \subseteq \Omega$ is

$$\Gamma_2(B, N) = \{(x, y) \in [\Omega \setminus \Gamma_1(B, N)]^2 : \; B \subseteq N_{\mathcal{Y}}(x) \cup N_{\mathcal{Y}}(y)\}.$$

In general, Γ_k-*region* for proximity map $N_{\mathcal{Y}}(\cdot)$ and set $B \subseteq \Omega$ for $k = 1, \ldots, n$ is

$$\Gamma_k(B, N) = \big\{(x_1, x_2, \ldots, x_k) \in \Omega^k : B \subseteq \cup_{j=1}^{k} N_{\mathcal{Y}}(x_j) \text{ and all } m\text{-permutations } (u_1, u_2, \ldots, u_m) \text{ of } (x_1, x_2, \ldots, x_k) \text{ satisfy } (u_1, u_2, \ldots, u_m) \notin \Gamma_m(B, N) \text{ for each } m \in \{1, 2, \ldots, k-1\}\big\}.$$

Note that Γ_k-regions are defined for $k \leq n$ and a Γ_k-region might be empty. Moreover, Γ_k-regions are not in Ω but in Ω^k. Furthermore, $P(\gamma(\mathcal{X}_n, N) = k) = 1$ iff $P(\gamma(\mathcal{X}_n, N) < k) = 0$ and $\mathcal{X}_n^k \cap \Gamma_k(\mathcal{X}_n, N) \neq \emptyset$, where $\mathcal{X}_n^k = \{(u_1, u_2, \ldots, u_k) : u_j \in \mathcal{X}_n,\; j \in \{1, \ldots, k\}\}$. □

3.5 κ-Values for the Proximity Maps in $T(\mathcal{Y})$

Recall that the domination number, $\gamma(\mathcal{X}_n, N)$ is the cardinality of a minimum dominating set of the PCD based on N. So by definition, $\gamma(\mathcal{X}_n, N) \leq n$. We will seek an a.s. least upper bound for $\gamma(\mathcal{X}_n, N)$ which suggests the following concept.

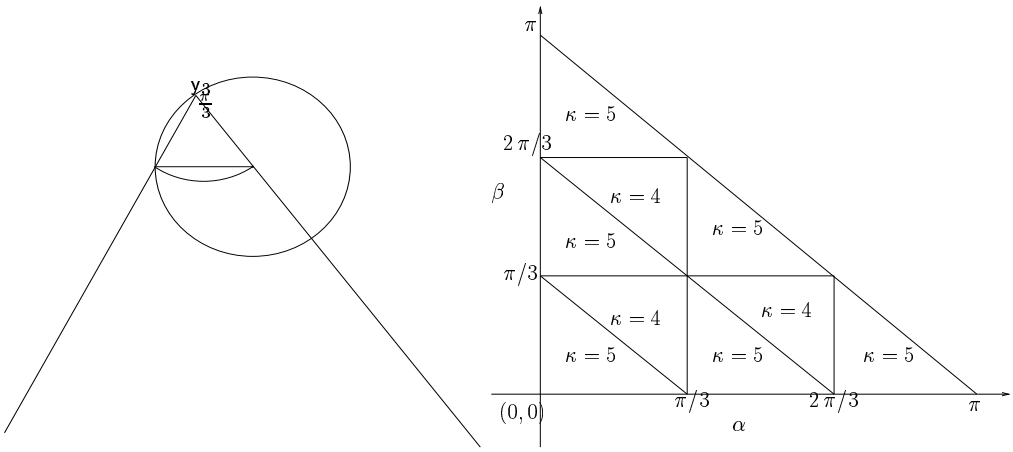


Figure 3.5.1: A vertex with inner angle equal to $\pi/3$ (left), the domain of α and β for which $\kappa(N_{AS}) = 4$ and $\kappa(N_{AS}) = 5$ (right).

Definition 3.5.1. Let $\mathcal{X}_n$ be a random sample from F on $T(\mathcal{Y})$ and let $\gamma(\mathcal{X}_n, N)$ be the domination number for the PCD based on a proximity map N. The general a.s. least upper bound for $\gamma(\mathcal{X}_n, N)$ that works for all $n \geq 1$ and is independent of n is called the *κ-value*; i.e., $\kappa(N) := \min\{k : \gamma(\mathcal{X}_n, N) \leq k \text{ a.s. for all } n \geq 1\}$.

In $\mathbb{R}$ with $\mathcal{Y} = \{0, 1\}$, for $\mathcal{X}_n$ a random sample from $\mathcal{U}(0, 1)$, $\gamma(\mathcal{X}_n, N_S) \leq 2$ with equality holding with positive probability. Hence $\kappa(N_S) = 2$. Next, we investigate the κ-values for $\{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$ in $\mathbb{R}^2$.

Theorem 3.5.2. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$. For arc-slice proximity maps with CC-vertex regions, $\kappa(N_{AS}) \leq 5$.*

Proof: The value of $\kappa(N_{AS})$ depends on the inner angles of the triangle.We need one point to cover $\mathcal{X}_n \cap R_{CC}(\mathsf{y})$ if the inner angle at y is less than or equal to $\pi/3$. See Figure 3.5.1 (left). For an equilateral triangle, $\kappa(N_{AS}) = 3$, with one angle between $\pi/3$ and $2\pi/3$ degrees, two less than $\pi/3$, $\kappa(N_{AS}) = 4$, with two angles between $\pi/3$ and $2\pi/3$ degrees, one less than $\pi/3$, $\kappa(N_{AS}) = 5$, and, with one angle larger than $2\pi/3$ degrees, two less than $\pi/3$, again $\kappa(N_{AS}) = 5$. ■

Notice that the $\kappa(N_{AS})$ depends on the geometry of $T(\mathcal{Y})$, which is caused by the lack of geometry invariance for N_{AS} (when, e.g., $T(\mathcal{Y})$ is transformed to T_e).

Since the κ-value depends on the inner angle, we can find its distribution for an arbitrary triangle from $\mathcal{D}_P$, the Delaunay Poisson triangles. See Section 2.1.1. Recall that the joint

density of any two arbitrary angles of a triangle from $\mathcal{D}_P$ is $f(\alpha,\beta) = \frac{8}{3\pi}\sin\alpha\sin\beta\sin(\alpha+\beta)$, for $\alpha,\beta > 0$ and $\alpha+\beta < \pi$ (see [30]). Using the above density, we can find the distribution of $\kappa(N_{AS})$ for $\mathcal{D}_P$.

$P(\kappa(N_{AS}) = 3) = P(\alpha = \beta = \pi/3) = 0$. A careful investigation shows that

$$\begin{aligned} P(\kappa(N_{AS}) = 4) =& P(\alpha < \pi/3, \beta < \pi/3, \alpha+\beta > \pi/3) \\ &+ P(\pi/3 < \alpha < 2\pi/3, 0 < \beta < \pi/3, 2\pi/3 < \alpha+\beta < \pi) \\ &+ P(\pi/3 < \beta < 2\pi/3, 0 < \alpha < \pi/3, 2\pi/3 < \alpha+\beta < \pi). \end{aligned}$$

See Figure 3.5.1 (right) for the domain of α and β for which $\kappa(N_{AS}) = 4$ and $\kappa(N_{AS}) = 5$. So

$$P(\alpha < \pi/3, \beta < \pi/3, \alpha+\beta > \pi/3) = \int_0^{\pi/3}\int_{\pi/3-\alpha}^{\pi/3} f(\alpha,\beta)d\beta d\alpha = 1/6.$$

and

$$P(\pi/3 < \alpha < 2\pi/3, 0 < \beta < \pi/3, 2\pi/3 < \alpha+\beta < \pi) = \int_{\pi/3}^{2\pi/3}\int_{2\pi/3-\alpha}^{\pi/3} f(\alpha,\beta)d\beta d\alpha = 1/6,$$

and by symmetry third piece is also $1/6$. Hence $P(\kappa(N_{AS}) = 4) = 1/2$, which implies $P(\kappa(N_{AS}) = 5) = 1/2$, also. Therefore,

$$\kappa(N_{AS}) = \begin{cases} 4 & \text{w.p. } 1/2 \\ 5 & \text{w.p. } 1/2. \end{cases}$$

Theorem 3.5.3. *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$, and $M \in \mathbb{R}^2 \setminus \mathcal{Y}$. For $N_{PE}^r(\cdot, M)$, $\kappa(N_{PE}^r) = 3$.*

Proof: For $N_{PE}^r(\cdot, M)$, pick the point closest to edge e_j in vertex region $R_M(\mathsf{y}_j)$; that is, pick $U_j \in \operatorname{argmin}_{X\in\mathcal{X}_n\cap R_M(\mathsf{y}_j)} d(X, e_j) = \operatorname{argmax}_{X\in\mathcal{X}_n\cap R_M(\mathsf{y}_j)} d(\ell(\mathsf{y}, X), \mathsf{y}_j)$ in the vertex region for which $X_n \cap R_M(\mathsf{y}_j) \neq \emptyset$ for $j \in \{1,2,3\}$ (note that as $n \to \infty$, U_j is unique a.s. for each j, since X is from $\mathcal{U}(T(\mathcal{Y}))$). Then $\mathcal{X}_n \cap R_M(\mathsf{y}_j) \subset N_{PE}^r(U_j, M)$. Hence $\mathcal{X}_n \subset \cup_{j=1}^3 N_{PE}^r(U_j, M)$. So $\gamma(\mathcal{X}_n, N_{PE}^r, M_C) \leq 3$ with equality holding with positive probability. Thus $\kappa(N_{PE}^r) = 3$. ■

There is no least upper bound for $\gamma(\mathcal{X}_n, N_{CS}^\tau, M)$ that works for all $n > 0$; i.e., $\kappa(N_{CS}^\tau)$ is not defined, as shown below.

Theorem 3.5.4. *Let $\mathcal{X}_n$ be random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $\gamma(\mathcal{X}_n, N_{CS}^{\tau}, M) = n$ holds with positive probability for all $\tau \in [0,1]$.*

Proof: For $\tau = 0$, the result follows trivially. For $\tau = 1$, we will prove the theorem by showing that there is a union of n regions of positive area in $T(\mathcal{Y})$, so that $u \in N_{CS}^{\tau=1}(v, M)$ iff $u = v$, for any $u, v \in \mathcal{X}_n$. Let $M = (m_1, m_2) \in T(\mathcal{Y})^o$. In $R_M(e_3)$ locate n triangles evenly on e_3 with base length $1/n$ and similar to $T(\mathcal{Y})$ (with similarity ratio $1/n$). See also Figure 3.5.2. Then locate n points in each triangle at $z_i = (x_i, y_i)$ such that (x_i, y_i) is the same type of center of T_i as M is of $T(\mathcal{Y})$. Then using the similarity ratio of $N_{CS}^{\tau=1}(z_i, M)$ to $T(\mathcal{Y})$, namely, $y_i/m_2 = 1/n$, we get $y_i = m_2/n$ for all $i = 1, \ldots, n$. Moreover, $x_i - x_{i-1} = 1/n$ for $i = 2, \ldots, n$ with $x_1 = m_1/n$ and $x_n = 1 - (1 - m_1)/n$. Then $(x_i, y_i) \in N_{CS}^{\tau=1}((x_j, y_j), M)$ iff $i = j$. Furthermore, for sufficiently small $\varepsilon > 0$, the same holds for the ε neighborhood of each $z_i = (x_i, y_i)$. That is, $N_{CS}^{\tau=1}(x, M) \cap B(z_j, \varepsilon) = \emptyset$ for all $x \in B(z_i, \varepsilon)$ for any distinct pair $i, j \in [n]$, and probability of $\mathcal{X}_n$ being composed of n points one from each $B(z_i, \varepsilon)$ is positive. Then $\gamma(\mathcal{X}_n, N_{CS}^{\tau}, M) = n$ holds with positive probability. The result for $\tau \in (0,1)$ follows similarly. ■

3.5.1 Characterization of Proximity Maps Using κ-Values

For the proximity maps we have considered, $\kappa(N_{AS}) \leq 5$, $\kappa(N_{PE}^{r}) = 3$ and $\kappa(N_{CS}^{\tau})$ is not defined.

The common property for the proximity maps for which $\kappa(N)$ is defined is that probability of having an $\mathcal{X}_n$ for which $N(X) \cap \mathcal{X}_n = \{X\}$ for all $X \in \mathcal{X}_n$ is zero. Note also that N_{AS} and N_{PE}^{r} are defined with M-vertex regions, while N_{CS}^{τ} is defined with M-edge regions.

Below we state a condition for $\kappa(N(\cdot, M)) = 3$ for $N(\cdot, M)$ defined with M-vertex regions.

Theorem 3.5.5. *Suppose $N(\cdot, M)$ is defined with M-vertex regions with $M \in \mathbb{R}^2 \setminus \mathcal{Y}$ and $N(x, M)$ gets larger as $d(\ell(\mathsf{y}, x), \mathsf{y})$ increases for $x \in R_M(\mathsf{y})$ in the sense that $N(x, M) \subseteq N(z, M)$ for all $x, z \in R_M(\mathsf{y})$ when $d(\ell(\mathsf{y}, x), \mathsf{y}) \leq d(\ell(\mathsf{y}, z), \mathsf{y})$. Furthermore, $N(X, M) \subsetneq N(Z, M)$ for all $X, Z \in R_M(\mathsf{y})$ occurs with positive probability when $d(\ell(\mathsf{y}, X), \mathsf{y}) < d(\ell(\mathsf{y}, Z), \mathsf{y})$ occurs with positive probability for X, Z from F. Then $\kappa(N) = 3$.*

Proof: When $\mathcal{X}_n \cap R_M(\mathsf{y}_j) \neq \emptyset$, pick one of the points $U_j(n) \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_j)} d(\ell(\mathsf{y}_j, X), \mathsf{y}_j)$, then $\mathcal{X}_n \cap R_M(\mathsf{y}_j) \subset N(U_j(n))$ for each $j \in \{1, 2, 3\}$. So $\gamma(\mathcal{X}_n, N, M) \leq 3$, and the condition for

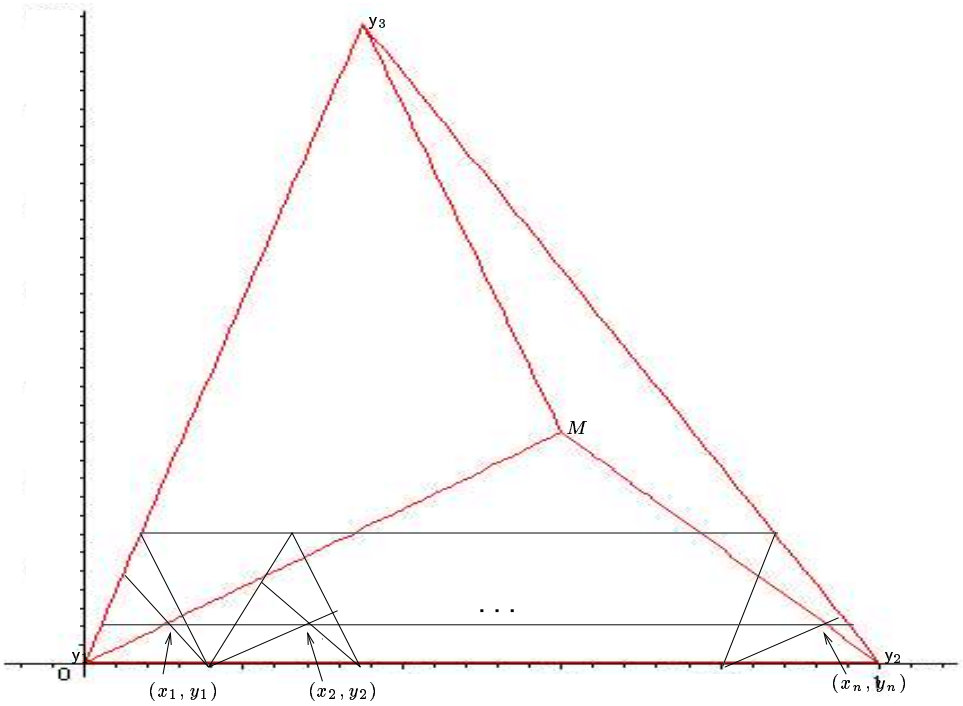


Figure 3.5.2: The figure for $\kappa\left(N_{CS}^{\tau}\right) = n$.

strict inclusion for strict inequality implies that $\gamma(\mathcal{X}_n, N, M) = 3$ holds with positive probability. Hence $\kappa(N) = 3$. ■

Notice that N_{PE}^{τ} satisfies the conditions of Theorem 3.5.5.

Theorem 3.5.6. *Suppose $N(\cdot, M)$ is defined with M-edge regions and $N(x, M)$ gets larger as $d(x, e)$ increases for $x \in R_M(e)$ in the sense that $N(x, M) \subseteq N(y, M)$ for all $x, y \in R_M(e)$ when $d(x, e) \leq d(y, e)$. Suppose also that for X, Y from F, $N(X, M) \subsetneq N(Y, M)$ for all $X, Y \in R_M(e)$ occurs with positive probability when $d(X, e) < d(Y, e)$ occurs with positive probability. Then $\kappa(N) = 3$.*

Proof: When $\mathcal{X}_n \cap R_M(e_j) \neq \emptyset$, pick one of the points $U_{e_j}(n) \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(e_j)} d(X, e_j)$. Then $\mathcal{X}_n \cap R_M(e_j) \subset N(U_e(n))$ for each $j \in \{1, 2, 3\}$. So $\gamma(\mathcal{X}_n, N, M) \leq 3$, and the condition for strict inclusion for strict inequality implies that $\gamma(\mathcal{X}_n, N, M) = 3$ holds with positive probability. Hence $\kappa(N) = 3$. ■

CHAPTER 4

Distribution of the Domination Number of Proximity Catch Digraphs in Delaunay Tessellations

In this chapter, first, we give the formal definitions of the domination number and some related concepts and provide some background on the domination number of general digraphs. Then we investigate the distribution of the domination number of the PCDs based on $N \in \{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$.

4.1 Minimum Dominating Sets and Domination Number of Digraphs

Definition 4.1.1. In a digraph $D = (\mathcal{V}, \mathcal{A})$ a vertex $v \in \mathcal{V}$ *dominates* itself and all vertices of the form $\{u : vu \in \mathcal{A}\}$. A *dominating set* S_D for the digraph D is a subset of $\mathcal{V}$ such that each vertex $v \in \mathcal{V}$ is dominated by a vertex in S_D. A *minimum dominating set* S_D^* is a dominating set of minimum cardinality and the *domination number* $\gamma(D)$ is defined as $\gamma(D) := |S_D^*|$ (see [26]) where $|\cdot|$ denotes the set cardinality functional. See Chartrand and Lesniak [4] and West [44] for more on graphs and digraphs. If a minimum dominating set is of size one, we call it a *dominating point*.

Note that since $\mathcal{V}$ is always a dominating set, $1 \leq \gamma(D) \leq n$.

In a digraph $D(\mathcal{V}, \mathcal{A})$, the *open neighborhood* of a vertex $v \in \mathcal{V}$, denoted $\mathbf{N}(v)$, is the set of vertices $u \in \mathcal{V}$ such that $vu \in \mathcal{A}$. The *closed neighborhood* is defined by $\mathbf{N}[v] = \mathbf{N}(v) \cup \{v\}$. In general finding a minimum dominating set is an NP-Hard optimization problem (see [20] and [5]). However an approximately minimum dominating set, $\widehat{S}_D$, can be obtained in $O(|\mathcal{V}|^2)$ by a well-known greedy algorithm (see [5] and [31]). The greedy algorithm begins by selecting the vertex that has the most neighbors: $S_D^1 = \{v^1\}$, where $v^1 \in \text{argmax}_{v \in \mathcal{V}} |\mathbf{N}[v]|$. Cardinality ties are broken arbitrarily whenever they occur. Then for iteration $t \geq 2$ the algorithm picks $v^t \in \text{argmax}_{v \in \mathcal{V} \setminus S_D^{t-1}} \left| \mathbf{N}[v] \setminus \cup_{u \in S_D^{t-1}} \mathbf{N}[u] \right|$ and set $S_D^t = S_D^{t-1} \cup \{v^t\}$ until $\mathcal{V} = \cup_{v \in S_D^{t-1}} \mathbf{N}[v]$. This

algorithm is guaranteed to terminate after at most $|\mathcal{V}|$ iterations. If the algorithm terminates after t^* iterations, the set $\widehat{S}_D = S_D^{t^*}$ is a dominating set. The approximation to the domination number $\gamma(D)$ of the digraph D is $\widehat{\gamma}(D) = \left|\widehat{S}_D\right|$; and $\widehat{\gamma}(D) \geq \gamma(D)$.

For $\mathcal{X}_n = \{X_1, \cdots, X_n\}$ a random sample from F, the domination number of the associated data-random proximity catch digraph based on the proximity map $N(\cdot)$, denoted $\gamma(\mathcal{X}_n, N)$, is the minimum number of points that dominate all points in $\mathcal{X}_n$. Recall that $\gamma(\mathcal{X}_n, N) = 1$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N) \neq \emptyset$.

The random variable $\gamma(\mathcal{X}_n, N)$ depends on $\mathcal{X}_n$ and $N(\cdot)$ explicitly, and on F implicitly. Furthermore, in general, the expectation $\mathbf{E}[\gamma(\mathcal{X}_n, N)]$, depends on n, F, and N; $0 \leq \mathbf{E}[\gamma(\mathcal{X}_n, N)] \leq n$. In general, the variance of γ_n satisfies, $1 \leq \mathbf{Var}[\gamma(\mathcal{X}_n, N)] \leq n^2/4$.

4.2 The Asymptotic Distribution of Domination Number of the CCCD for Uniform Data on Compact Intervals in $\mathbb{R}$

Recall that for $\mathcal{Y} = \{0, 1\} \subset \mathbb{R}$, $\kappa(N_S) = 2$ (see Section 3.5) where $N_S(x) = B(x, r(x))$ with $r(x) = \min(x, 1 - x)$, so it suffices to find $P(\gamma(\mathcal{X}_n, N_S) > 1) = P(\gamma(\mathcal{X}_n, N_S) = 2)$ to get the distribution of $\gamma(\mathcal{X}_n, N_S)$. Note that $\gamma(\mathcal{X}_n, N_S) > 1$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_S) = \emptyset$ where $\Gamma_1(\mathcal{X}_n, N_S) = \left[\frac{X_{n:n}}{2}, \frac{1+X_{1:n}}{2}\right]$.

For the proximity maps we have defined in $\mathbb{R}^2$, i.e., for $N \in \{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$ (see Chapter 3), $\kappa(N) > 2$, so finding only $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N) > 1)$ or $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N) = 2)$ does not suffice to find the asymptotic distribution of $\gamma(\mathcal{X}_n, N)$. In particular, if $\kappa(N) = 3$, finding — for example — both $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N) > 1)$ and $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N) = 2)$ will yield the asymptotic distribution of $\gamma(\mathcal{X}_n, N)$. The latter probability will be found by a method similar to the below alternative calculation — alternative to the one given in [35]— for $\gamma(\mathcal{X}_n, N_S)$ in $\mathbb{R}$ where $N_S(x) = B(x, r(x))$.

Let $\mathcal{X}_n$ be a random sample from $F = \mathcal{U}(0, 1)$, we will calculate $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N_\mathcal{Y}) = 1)$ as follows. Let

$$X_{1/2}^-(n) \in \text{argmin}_{X \in \mathcal{X}_n \cap [0,1/2]} (1/2 - X) \text{ and } X_{1/2}^+(n) \in \text{argmin}_{X \in \mathcal{X}_n \cap [1/2,1]} (X - 1/2)$$

whenever they exist. Note that either one exists with probability one.

As $n \to \infty$, $\left\{X_{1/2}^-(n), X_{1/2}^+(n)\right\} \subsetneq \mathcal{X}_n$ both uniquely exist with probability 1. Then

$\gamma(\mathcal{X}_n, N_S) = 1$ iff $\mathcal{X}_n \subset N_S\left(X^-_{1/2}(n)\right)$ or $\mathcal{X}_n \subset N_S\left(X^+_{1/2}(n)\right)$. Let the events $E^\pm$ de defined as $\mathcal{X}_n \subset N_S\left(X^\pm_{1/2}\right)$. Then

$$P(\gamma(\mathcal{X}_n, N_S) = 1) = P(E^-_n) + P(E^+_n) - P(E^-_n \cap E^+_n).$$

By symmetry $P(E^-_n) = P(E^+_n)$.

The asymptotically accurate pdf of $X^-_{1/2}(n)$ (see Remark A.1.1) is

$$f^-(x) = n\left(1 - (1/2 - x)\right)^{n-1}\mathbf{I}(0 \le x \le 1/2) = n\,(x + 1/2)^{n-1}\,\mathbf{I}(0 \le x \le 1/2),$$

then $P(E^-_n) = P\left(\mathcal{X}_n \subset N_S\big(X^-_{1/2}(n)\big)\right) = P\left(\mathcal{X}_n \subset \left[0, 2\,X^-_{1/2}(n)\right] \setminus \left[X^-_{1/2}(n), 1/2\right]\right)$. For sufficiently large n,

$$\begin{aligned} P(E^-_n) &\approx \int_{1/4}^{1/2} \left(\frac{2\,x - (1/2 - x)}{x + 1/2}\right)^{n-1} f^-(x)\,dx = \int_{1/4}^{1/2} \left(\frac{3\,x - 1/2}{x + 1/2}\right)^{n-1} f^-(x)\,dx \\ &= \int_{1/4}^{1/2} n\,(3\,x - 1/2)^{n-1} dx = \frac{1}{3} - \frac{1}{3}\,4^{-n}. \end{aligned}$$

Hence $\lim_{n\to\infty} P(E^-_n) = \lim_{n\to\infty} P(E^+_n) = 1/3$.

Next, we find $P(E^-_n \cap E^+_n)$. The asymptotically accurate joint pdf of $X^-_{1/2}(n), X^+_{1/2}(n)$ is

$$f^\pm(x_1, x_2) = n\,(n-1)\left(1 - (x_2 - x_1)\right)^{n-2}\mathbf{I}\big[(x_1, x_2) \in (0, 1/2) \times (1/2, 1)\big].$$

Let $\varepsilon > 0$ be sufficiently small (say, $\varepsilon \le 1/6$) that $x_1 \ge 1/2 - \varepsilon$ and $x_2 \le 1/2 + \varepsilon$ implies that $[x_1, x_2] \subseteq [2\,x_2 - 1, 2\,x_1]$. Then for sufficiently large n

$$\begin{aligned} P(E^-_n \cap E^+_n) &\approx P\left(\mathcal{X}_n \subset \left[N_S\big(X^-_{1/2}(n)\big) \cap N_S\big(X^+_{1/2}(n)\big)\right]\right) = P\big(\mathcal{X}_n \subset [2\,x_2 - 1, 2\,x_1] \setminus [x_1, x_2]\big) \\ &\approx \int_{1/2-\varepsilon}^{1/2}\int_{1/2}^{1/2+\varepsilon} \left(\frac{2\,x_1 - 2\,x_2 + 1 - (x_2 - x_1)}{1 - (x_2 - x_1)}\right)^{n-2} f^\pm(x_1, x_2)\,dx_2\,dx_1 \\ &= \int_{1/2-\varepsilon}^{1/2}\int_{1/2}^{1/2+\varepsilon} n\,(n-1)\,(3\,x_1 - 3\,x_2 + 1)^{n-2}\,dx_2\,dx_1, \end{aligned}$$

since $P\left(X^-_{1/2}(n) \ge 1/2 - \varepsilon,\ X^+_{1/2}(n) \le 1/2 + \varepsilon\right) \to 1$ as $n \to \infty$. The integrand is critical at $x_1 = x_2 = 1/2$, since $(3\,x_1 - 3\,x_2 + 1) = 1$ when $x_1 = x_2 = 1/2$. So we make a change of variables

$x_1 = 1/2 - z_1$ and $x_2 = 1/2 + z_2$, then for sufficiently large n

$$P(E_n^- \cap E_n^+) \approx \int_0^{\varepsilon} \int_0^{\varepsilon} n\,(n-1)\,(1 - 3\,z_1 - 3\,z_2)^{n-2}\,dz_2\,dz_1.$$

The new integral is critical when $z_1 = z_2 = 0$, so we make the change of variables $z_j = w_j/n$, then for sufficiently large n

$$\begin{aligned} P(E_n^- \cap E_n^+) &\approx \int_0^{n\,\varepsilon} \int_0^{n\,\varepsilon} \frac{n\,(n-1)}{n^2} \left[1 - \frac{1}{n}\,(3\,w_1 + 3\,w_2)\right]^{n-2} dw_2\,dw_1, \\ \text{letting } n \to \infty &\approx \int_0^{\infty} \int_0^{\infty} \exp(-3\,w_1 - 3\,w_2)\,dw_2\,dw_1 = 1/9. \end{aligned}$$

Hence $\lim_{n\to\infty} P(E_n^- \cap E_n^+) = 1/9$. Then $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N_S) = 1) = 2/3 - 1/9 = 5/9$, which agrees with the asymptotic result given in [35].

Below is a general result for the limiting distribution of $\gamma(\mathcal{X}_n, N)$ for $\mathcal{X}_n$ from a very broad family of distributions and for general $N(\cdot)$.

Theorem 4.2.1. *Let $\mathscr{R}_S(N)$ be the superset region for the proximity map $N(\cdot)$. Let $\mathcal{X}_n$ be a random sample from F with $P_F(X \in \mathscr{R}_S(N)) > 0$, then* $\lim_{n\to\infty} P_F(\gamma(\mathcal{X}_n, N) = 1) = 1$.

Proof: Suppose $P_F(X \in \mathscr{R}_S(N)) > 0$. Recall that for any $x \in \mathscr{R}_S(N)$, $N(x) = \Omega$, so $\mathcal{X}_n \subseteq N(x)$, hence if $\mathcal{X}_n \cap \mathscr{R}_S(N) \neq \emptyset$ then $\gamma(\mathcal{X}_n, N) = 1$. Then $P(\mathcal{X}_n \cap \mathscr{R}_S(N) \neq \emptyset) \leq P(\gamma(\mathcal{X}_n, N) = 1)$. But

$$P\big(\mathcal{X}_n \cap \mathscr{R}_S(N) \neq \emptyset\big) = 1 - P\big(\mathcal{X}_n \cap \mathscr{R}_S(N) = \emptyset\big) = 1 - \big[1 - P_F\big(X \in \mathscr{R}_S(N)\big)\big]^n \to 1 \text{ as } n \to \infty,$$

since $P_F\big(X \in \mathscr{R}_S(N)\big) > 0$. Hence $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N) = 1) = 1$. ■

Remark 4.2.2. In particular, for $F = \mathcal{U}(T(\mathcal{Y}))$, the inequality $P_F(X \in \mathscr{R}_S(N)) > 0$ holds iff $A(\mathscr{R}_S(N)) > 0$ and $P(\mathcal{X}_n \cap \mathscr{R}_S(N) \neq \emptyset) \to 1$ at rate $O\Big(\Big(1 - \frac{A(\mathscr{R}_S(N))}{A(T(\mathcal{Y}))}\Big)^n\Big)$ as $n \to \infty$. □

For $\mathcal{Y} = \{0, 1\} \subset \mathbb{R}$, $\mathscr{R}_S(N_S) = \{1/2\}$, so Theorem 4.2.1 does not apply for N_S in $\mathbb{R}$.

4.3 The Asymptotic Distribution of $\gamma(\mathcal{X}_n, N_{AS}, M)$

Recall that any given triangle $T(\mathcal{Y})$ can be mapped to the basic triangle T_b (see Section 2.2). For $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}))$, the transformed data set is also a random sample from

$\mathcal{U}(T_b)$. So without loss of generality, we can assume $T(\mathcal{Y}) = T_b$. The domination number of the PCD based on $N_{AS}(\cdot, M)$ is denoted as $\gamma(\mathcal{X}_n, N_{AS}, M)$ to make the dependence on M explicit.

Recall also that $\kappa(N_{AS}) \leq 5$, then $1 \leq \mathbf{E}[\gamma(\mathcal{X}_n, N_{AS}, M_{CC})] \leq 5$ and $0 \leq \mathbf{Var}[\gamma(\mathcal{X}_n, N_{AS}, M_{CC})] \leq 25/4$.

For $N_{AS}(x) = \overline{B}(x, r(x)) \cap T(\mathcal{Y})$, where $r(x) = \min_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$, we are implicitly using the CC-vertex regions. See Section 3.2.1. We can also use other M-vertex regions. The superset region for $N_{AS}(\cdot, M_{CC})$ is $\{M_{CC}\}$, so Theorem 4.2.1 does not apply to $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$.

If $T(\mathcal{Y})$ is an acute triangle, then $M_{CC} \in T(\mathcal{Y})^o$. For a given $\mathcal{X}_n$, $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) \subseteq H(\mathcal{X}_n, \mathcal{Y}) = \cap_{\mathsf{y} \in \mathcal{Y}} \Gamma_1(X_{\mathsf{y}}(n), N_{AS})$, where $X_{\mathsf{y}}(n) \in \operatorname{argmin}_{X \in \mathcal{X}_n} d(X, \mathsf{y})$; i.e., the vertex extremum that is closest to vertex y which is unique a.s. if $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$.

Proposition 4.3.1. *Suppose $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$. If $T(\mathcal{Y})$ is an obtuse triangle, then $P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) > 1) \to 1$ as $n \to \infty$.*

Proof: In an obtuse triangle, $M_{CC} \notin T(\mathcal{Y})$. Hence $H(\mathcal{X}_n, \mathcal{Y}) = \emptyset$ with probability 1 as $n \to \infty$ which implies that $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) = \emptyset$ with probability 1 as $n \to \infty$. Therefore $P(\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) = \emptyset) = P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) > 1) \to 1$ as $n \to \infty$. ■

The above proposition indicates that the distribution of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ depends on the inner angles of $T(\mathcal{Y})$ or (c_1, c_2). Observe that, the distribution of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ is not geometry invariant, in particular, under $\phi_e(\cdot)$, which transforms T_b to T_e.

Now, suppose $T(\mathcal{Y})$ is an acute triangle and $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$, then $M_{CC} \in T(\mathcal{Y})^o$ and $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC})$ is nonempty and has positive area with probability 1 for all finite n. For sufficiently large n, $H(\mathcal{X}_n, \mathcal{Y})$ is a (convex) hexagon bounded by the perpendicular lines at the midpoints of the line segments joining $X_{\mathsf{y}_j}(n)$ and y_k, for all $k \neq j$ with probability 1. See also Figure 4.3.1.

Given $X_{\mathsf{y}_j}(n) = x_{\mathsf{y}_j} = (x_j, y_j)$, for $j \in \{1, 2, 3\}$, let $\ell_j^{\perp}(u, x)$ be the line perpendicular to the line segment joining u and y_j at the midpoint of the line segment for $j \in \{1, 2, 3\}$. See Figure 4.3.1. Below are the explicit forms of these lines.

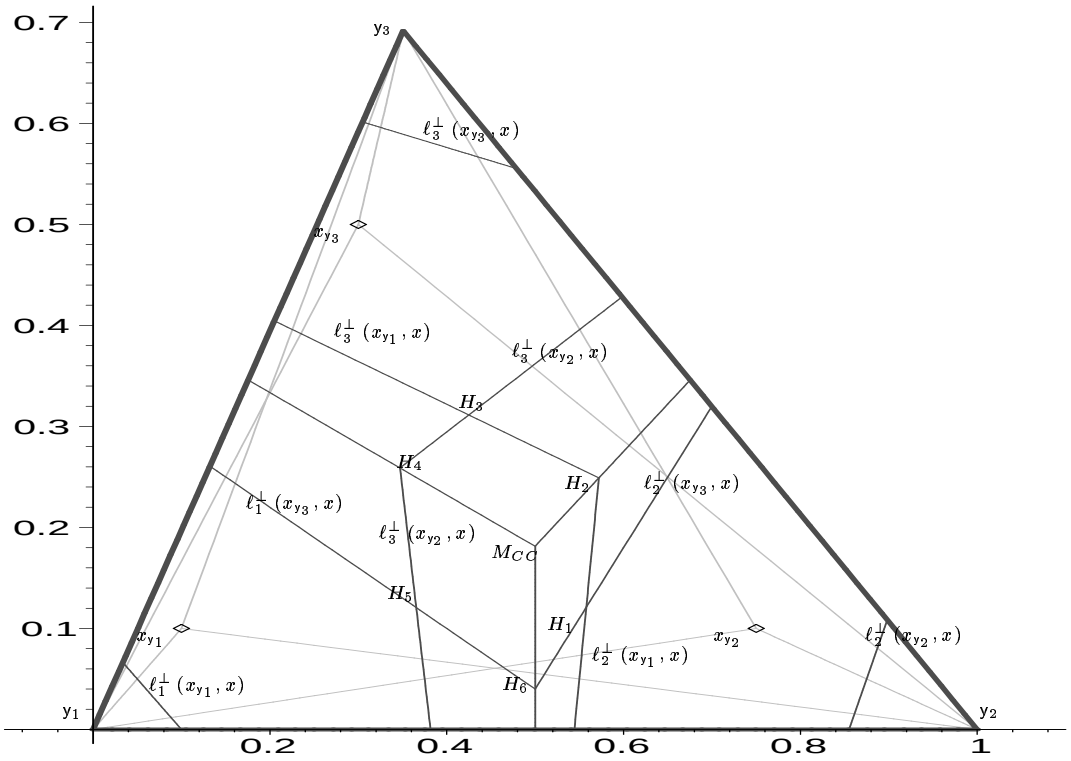


Figure 4.3.1: A realization of $H(\mathcal{X}_n, \mathcal{Y})$ based on a given set of vertex extrema $X_{\mathrm{y}_j}(n) = x_{\mathrm{y}_j}$ from $\mathcal{X}_n$.

$$\ell_2^{\perp}(x_{\mathrm{y}_1}, x) = \frac{y_1^2 + x_1(2x - x_1) + 2x - 1}{2y_1}, \qquad \ell_1^{\perp}(x_{\mathrm{y}_3}, x) = \frac{y_3^2 - x_3(2x - x_3)}{2y_3},$$

$$\ell_3^{\perp}(x_{\mathrm{y}_2}, x) = \frac{c_2^2 - y_2^2 + x_2(2x - x_2) + c_1(c_1 - 2x)}{2(c_2 - y_2)}, \quad \ell_1^{\perp}(x_{\mathrm{y}_2}, x) = \frac{y_2^2 - x_2(2x - x_2)}{2y_2},$$

$$\ell_3^{\perp}(x_{\mathrm{y}_1}, x) = \frac{c_2^2 - y_2^2 + x_2(2x - x_2) + c_1(c_1 - 2x)}{2(c_2 - y_2)}, \quad \ell_2^{\perp}(x_{\mathrm{y}_3}, x) = \frac{y_3^2 - x_3(2x - x_3) - (2x - 1)}{2y_3},$$

$$\ell_1^{\perp}(x_{\mathrm{y}_1}, x) = \frac{y_1 - x_1(2x - x_1)}{2y_1}, \qquad \ell_2^{\perp}(x_{\mathrm{y}_2}, x) = \frac{y_2^2 - x_2(2x - x_2) + 2x - 1}{2y_2},$$

$$\ell_3^{\perp}(x_{\mathrm{y}_3}, x) = \frac{c_2^2 - y_3^2 + x_3(2x - x_3) + c_1(c_1 - 2x)}{2(c_2 - y_3)}.$$

The area of a planar (convex or non-convex) polygon with vertices $(a_1, b_1),\ (a_2, b_2), \ldots,\ (a_n, b_n)$

arranged in counterclockwise order is

$$\frac{1}{2}\cdot\left(\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}+\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}+\ldots+\begin{vmatrix} a_n & a_1 \\ b_n & b_1 \end{vmatrix}\right). \tag{4.3.1}$$

If the vertices are arranged in a clockwise order, the above expression is negative. For example, the area of a triangle with coordinates (a_1, b_1), (a_2, b_2), (a_3, b_3) in counterclockwise order is

$$\frac{1}{2}\cdot\left(\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}+\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}+\begin{vmatrix} a_3 & a_1 \\ b_3 & b_1 \end{vmatrix}\right)=\frac{1}{2}\cdot(a_1\,b_2-a_2\,b_1+a_2\,b_3-a_3\,b_2+a_3\,b_1-a_1\,b_3).$$

Observe that $\ell_j^{\perp}(x_{e_k}, x)$ and $\ell_j^{\perp}(x_{e_l}, x)$ for $k, l \neq j$ are the only (active) lines —there are six such lines— that determine the boundary of $H(\mathcal{X}_n, \mathcal{Y})$. Then $H(\mathcal{X}_n, \mathcal{Y})$ is the hexagon with vertices H_j for $j \in \{1, \ldots, 6\}$ (see Figure 4.3.1), where

$$H_1=\ell_2^{\perp}(x_{\mathsf{y}_1}, x)\cap\ell_2^{\perp}(x_{\mathsf{y}_3}, x),\quad H_2=\ell_2^{\perp}(x_{\mathsf{y}_1}, x)\cap\ell_3^{\perp}(x_{\mathsf{y}_1}, x),\quad H_3=\ell_3^{\perp}(x_{\mathsf{y}_1}, x)\cap\ell_3^{\perp}(x_{\mathsf{y}_2}, x),$$
$$H_4=\ell_3^{\perp}(x_{\mathsf{y}_2}, x)\cap\ell_1^{\perp}(x_{\mathsf{y}_2}, x),\quad H_5=\ell_1^{\perp}(x_{\mathsf{y}_2}, x)\cap\ell_1^{\perp}(x_{\mathsf{y}_3}, x),\quad H_6=\ell_1^{\perp}(x_{\mathsf{y}_3}, x)\cap\ell_2^{\perp}(x_{\mathsf{y}_3}, x).$$

Below are the explicit forms of the coordinates of H_j for $j = 1, \ldots, 6$.

$$H_1=\left(\frac{y_3\,(y_1^2+x_1^2)-y_1\,(y_3^2+x_3^2)+y_1-y_3}{2\,(y_3\,(x_1-)-y_1\,(x_3-1))},\ \frac{(x_1^2+y_1^2)\,(x_3-1)+(x_3^2+y_3^2)\,(1-x_1)+x_1-x_3}{2\,(y_1\,(x_3-1)-y_3\,(x_1-1))}\right),$$
$$H_2=\left(\frac{c_2\,(x_1^2+y_1^2)+y_1-c_2-y_1\,(c_1^2+c_2^2)}{2\,(c_2\,(x_1-1)+y_1\,(1-c_1))},\ \frac{(1-c_1)\,(x_1^2+y_1^2+c_1)-(1-c_1^2-c_2^2)\,x_1+c_2^2}{2\,(c_2\,(x_1-1)+y_1\,(1-c_1))}\right),$$
$$H_3=\left(\frac{(c_2-y_2)\,(x_1^2+y_1^2)+(y_1-c_1)\,x_2^2+(y_2^2-c_1^2-c_2^2)\,y_1+(c_1^2+c_2^2)\,y_2-y_2^2\,c_2}{2\,((c_2-y_2)\,x_1+(y_1-c_2)\,x_2+c_1\,(y_2-y_1))},\right.$$
$$\left.\frac{(x_2-c_1)\,(x_1^2+y_1^2)+(c_1-x_1)\,(x_2^2+y_2^2)+(c_1^2+c_2^2)\,(x_1-x_2)}{2\,((c_2-y_2)\,x_1+(y_1-c_2)\,x_2+c_1\,(y_2-y_1))}\right),$$
$$H_4=\left(\frac{c_2\,(x_2^2+y_2^2)-(c_1^2+c_2^2)\,y_2}{2\,(x_2\,c_2-y_2\,c_1)},\ \frac{c_1\,(x_2^2+y_2^2)-(c_1^2+c_2^2)\,x_2}{2\,(y_2\,c_1-c_2\,x_2)}\right),$$
$$H_5=\left(\frac{y_3\,(x_2^2+y_2^2)-y_2\,(x_3^2+y_3^2)}{2\,(x_2\,y_3-x_3\,y_2)},\ \frac{x_3\,(x_2^2+y_2^2)-x_2\,(x_3^2+y_3^2)}{2\,(x_3\,y_2-x_2\,y_3))}\right),$$
$$H_6=\left(\frac{1}{2},\ \frac{y_3^2+x_3\,(x_3-1)}{2\,y_3}\right).$$

The area of the hexagon $H(\mathcal{X}_n, \mathcal{Y})$, which is a quite lengthy expression, can be obtained by

using Equation 4.3.1.

Since $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) \subseteq H(\mathcal{X}_n, \mathcal{Y})$, the intersection $\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset$ implies that $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) = \emptyset$. So

$$P(\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset) \leq P(\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) = \emptyset) = P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) > 1).$$

Hence $P(\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset)$ serves as a lower bound for $P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) > 1)$.

For sufficiently large n and $\mathcal{X}_n$ a random sample from a continuous distribution F on $T(\mathcal{Y})$, $H(\mathcal{X}_n, \mathcal{Y}) \subsetneq T(\mathcal{Y})^o$ with probability 1 in an acute triangle. Let $\varepsilon > 0$ sufficiently small so that $d(x_{\mathsf{y}}, \mathsf{y}) < \varepsilon$ for each $\mathsf{y} \in \mathcal{Y}$ will imply the vertex extrema are distinct and $H(\mathcal{X}_n, \mathcal{Y}) \subsetneq T(\mathcal{Y})^o$. Then the vertex extrema are distinct and $H(\mathcal{X}_n, \mathcal{Y}) \subsetneq T(\mathcal{Y})^o$ with probability 1, since $d(X_{\mathsf{y}}, \mathsf{y}) < \varepsilon$ for each $\mathsf{y} \in \mathcal{Y}$ occurs with probability 1 as $n \to \infty$. Denote the associated event $d(X_{\mathsf{y}}(n), \mathsf{y}) < \varepsilon$ as $E_n^{\varepsilon}(\mathcal{Y})$. See Figure 4.3.2 (left).

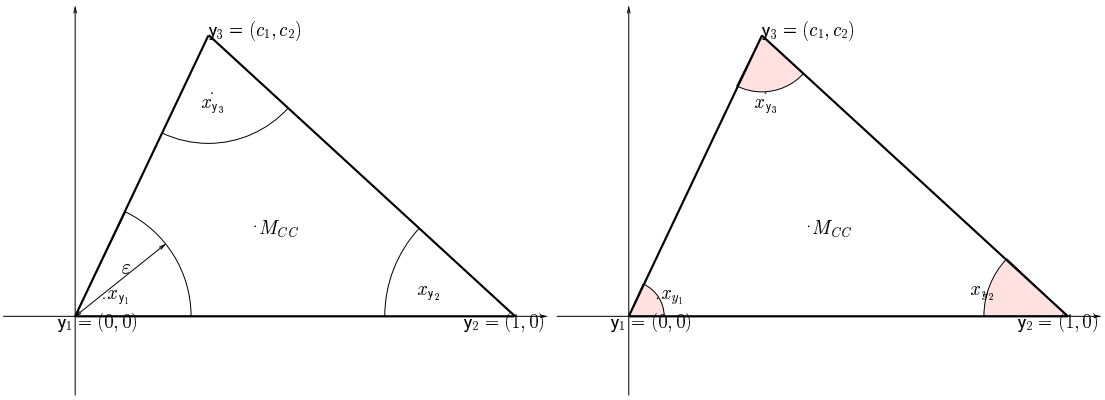


Figure 4.3.2: A figure for the description of the event $E_n^{\varepsilon}(\mathcal{Y})$ (left) and the pdf of $X_{\mathsf{y}}(n)$ (right) given three distinct vertex extrema $X_{\mathsf{y}_j}(n) = x_{\mathsf{y}_j}$ for $j \in \{1, 2, 3\}$.

Since $H(\mathcal{X}_n, \mathcal{Y})$ is determined by $X_{\mathsf{y}}(n)$, to find the expected area of $H(\mathcal{X}_n, \mathcal{Y})$ we need the exact joint pdf of the (closest) vertex extrema. But to find the limit of the expected area of $H(\mathcal{X}_n, \mathcal{Y})$ as $n \to \infty$ we only need the asymptotically accurate joint pdf of the vertex extrema. (see Remark A.1.1).

Now suppose $F = \mathcal{U}(T(\mathcal{Y}))$. Let the inner angles of the triangle $T(\mathcal{Y})$ be

$$\hat{\mathsf{y}}_1 = \arctan(c_2/c_1), \;\; \hat{\mathsf{y}}_2 = \arctan(c_2/(1-c_1)),$$

$$\hat{\mathsf{y}}_3 = \pi - (\hat{\mathsf{y}}_1 + \hat{\mathsf{y}}_2) = \pi - (\arctan(c_2/c_1 + \arctan(c_2/(1-c_1))) = \arctan(c_1/c_2 + \arctan((1-c_1)/c_2)).$$

Given $X_{\mathsf{y}_j}(n) = x_{\mathsf{y}_j} = (x_j, y_j)$ for $j \in \{1,2,3\}$, let $S_U(\vec{\zeta})$ be the region $T(\mathcal{Y}) \setminus \left[\cup_{\mathsf{y}\in\mathcal{Y}} B(\mathsf{y}, d(x_{\mathsf{y}}, \mathsf{y}))\right]$, an example of which is the unshaded region in Figure 4.3.2 (right) where $\vec{\zeta} = (x_1, y_1, x_2, y_2, x_3, y_3)$. So the area of $S_U(\vec{\zeta})$ is

$$A(S_U(\vec{\zeta})) = A(T(\mathcal{Y})) - 1/2\Big(\big(x_1^2 + y_1^2\big)\widehat{\mathsf{y}}_1 + \big((x_2-1)^2 + y_2\big)\,\widehat{\mathsf{y}}_2 + \big((x_3 - c_1)^2 + (y_3 - c_2)^2\big)\,\widehat{\mathsf{y}}_3\Big),$$

where $A(T(\mathcal{Y})) = c_2/2$. So the asymptotically accurate joint pdf of the vertex extrema of an $\mathcal{X}_n$ which is a random sample from $\mathcal{U}(T(\mathcal{Y}))$ is

$$f_3^{\mathcal{Y}}(\vec{\zeta}) = \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3}\left[\frac{A(S_U)}{A(T(\mathcal{Y}))}\right]^{n-3},$$

where the domain $D_S = \{x_{\mathsf{y}_j} = (x_j, y_j) : d(x_{\mathsf{y}_j}, \mathsf{y}_j) < \varepsilon, \text{ for } j \in \{1,2,3\}\}$, where ε is sufficiently small that x_{y_j} are distinct and $H(\mathcal{X}_n, \mathcal{Y}) \subsetneq T(\mathcal{Y})^o$.

We switch to the polar coordinates as

$$\begin{array}{lll}
x_1 = r_1\cos(\theta_1), & y_1 = r_1\sin(\theta_1), & \text{for } 0 \le \theta_1 \le \widehat{\mathsf{y}}_1, \\
x_2 = 1 + r_2\cos(\theta_2), & y_2 = r_2\sin(\theta_2), & \text{for } \widehat{\mathsf{y}}_1 + \widehat{\mathsf{y}}_3 \le \theta_2 \le \pi, \\
x_3 = c_1 - r_3\sin(\theta_3), & y_3 = c_2 - r_3\cos(\theta_3), & \text{for } \widehat{\mathsf{y}}_1 \le \theta_3 \le \widehat{\mathsf{y}}_1 + \widehat{\mathsf{y}}_3.
\end{array} \tag{4.3.2}$$

Then

$$A(S_U) = \frac{c_2}{2} - \frac{1}{2}\Big[\arctan(c_2/c_1){r_1}^2 + \arctan(c_2/(1-c_1))r_2^2 + r_3^2\,(\arctan(c_1/c_2) + \arctan((1-c_1)/c_2)\Big]$$

and $D_S = \{(\vec{r}, \vec{\theta}) \in \mathbb{R}^6 : r_j < \varepsilon \text{ for } j \in \{1,2,3\}\}$, where $\vec{r} = (r_1, r_2, r_3)$ and $\vec{\theta} = (\theta_1, \theta_2, \theta_3)$ and the ranges of θ_j are given in Equation 4.3.2.

In the standard equilateral triangle T_e and $A(S_U) = \frac{\sqrt{3}}{4} - \frac{\pi}{6}\,(r_1^2 + r_2^2 + r_3^2)$. $A(H(\mathcal{X}_n, \mathcal{Y}))$ is a rather lengthy expression in polar coordinates, so the general version (for T_b) is not provided.

But in T_e,

$$A(H(\mathcal{X}_n,\mathcal{Y})) = \frac{r_1\,r_3}{6}\left(\cos(\theta_3-\theta_1)+\cos(\theta_3+\theta_1)-\frac{1}{\sqrt{3}}\sin(\theta_3-\theta_1)+\frac{1}{\sqrt{3}}\sin(\theta_3+\theta_1)\right)$$
$$-\frac{r_1\,r_2}{6}\left(\frac{1}{\sqrt{3}}\cos(\theta_1-\theta_2)+\frac{2}{\sqrt{3}}\cos(\theta_1+\theta_2)-\sin(\theta_1-\theta_2)\right)-\frac{r_2\,r_3}{6}\left(\cos(\theta_3-\theta_2)+\cos(\theta_3+\theta_2)\right.$$
$$\left.-\frac{1}{\sqrt{3}}\sin(\theta_3-\theta_2)+\frac{1}{\sqrt{3}}\sin(\theta_3+\theta_2)\right)+O\left(r_1^2\right)+O\left(r_2^2\right)+O\left(r_3^2\right).$$

Furthermore, the Jacobian $J\left(\vec{r},\vec{\theta}\right)$ is

$$J\left(\vec{r},\vec{\theta}\right) = \begin{vmatrix} \cos(\theta_1) & -r_1\sin(\theta_1) & 0 & 0 & 0 & 0 \\ \sin(\theta_1) & r_1\cos(\theta_1) & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\theta_2) & -r_2\sin(\theta_2) & 0 & 0 \\ 0 & 0 & \sin(\theta_2) & r_2\cos(\theta_2) & 0 & 0 \\ 0 & 0 & 0 & 0 & -\cos(\theta_3) & r_3\sin(\theta_3) \\ 0 & 0 & 0 & 0 & -\sin(\theta_3) & -r_3\cos(\theta_3) \end{vmatrix} = r_1\,r_2\,r_3.$$

Hence $\left|J\left(\vec{r},\vec{\theta}\right)\right| = r_1\,r_2\,r_3$.

Remark 4.3.2. We can give a proof —more explicit than Theorem 3.4.26— for $\lim_{n\to\infty}\mathbf{E}\left[A(\Gamma_1(\mathcal{X}_n,N_{AS},M_{CC}))\right]=0$, using the explicit form of the area of the hexagon $A(H(\mathcal{X}_n,\mathcal{Y}))$.

Remark 4.3.3. Note that transforming the basic triangle T_b to the standard equilateral triangle T_e with $u(x,y)=x+\frac{1-2\,c_1}{\sqrt{3}}\,y$ and $v(x,y)=\frac{\sqrt{3}}{2\,c_2}\,y$ (see Section 2.2) does preserve the uniformity of the data but not the circular structure of the balls unless $T_b=T_e$, but rather a ball is mapped to an ellipse. For example, the ball

$$B(\mathsf{y}_1,r)=\left\{(x,y)\in\mathbb{R}^2\,:\,x^2+y^2\le r^2\right\}$$

is mapped to the ellipse

$$\left\{(u,v)\in\mathbb{R}^2\,:\,u^2+\left(\frac{2\,(2\,c_1-2)}{\sqrt{3}}\right)u\,v+\left(\frac{4\,(c_1\,(c_1-1)+c_2^2)+1}{3}\right)v^2\le r^2\right\}.$$

Moreover, the circumcenter of T_b $M_{CC}=\left(\frac{1}{2},\frac{c_2^2-c_1\,(1-c_1)}{2\,c_2}\right)$ is mapped to $\left(\frac{(c_1^2+c_2^2)\,(3-2\,c_1)-c_1}{4\,c_2^2},\right.$ $\left.\frac{\sqrt{3}\,(c_2^2-c_1(1-c_1))}{4\,c_2^2}\right)$. Hence the distribution of $\gamma(\mathcal{X}_n,N_{AS},M_{CC})$ is not geometry invariant. □

Given $X_{\mathsf{y}_j}(n) = x_{\mathsf{y}_j} = (x_j, y_j)$ for $j \in \{1,2,3\}$,

$$P(\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset) = \left(\frac{A(S_U) - A(H(\mathcal{X}_n, \mathcal{Y}))}{A(S_U)}\right)^{n-3}$$

so

$$\begin{aligned} P(\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset) &= \int \left(\frac{A(S_U) - A(H(\mathcal{X}_n, \mathcal{Y}))}{A(S_U)}\right)^{n-3} f_3^{\mathcal{Y}}(\vec{\zeta})\, d\vec{\zeta} \\ &= \int \frac{n(n-1)(n-2)}{A(T(\mathcal{Y}))^3} \left(\frac{A(S_U) - A(H(\mathcal{X}_n, \mathcal{Y}))}{A(T(\mathcal{Y}))}\right)^{n-3} d\vec{\zeta}, \end{aligned}$$

where $d\vec{\zeta} = dx_1\, dy_1\, dx_2\, dy_2\, dx_3\, dy_3$.

Let

$$G(\vec{x}, \vec{y}) = \frac{A(S_U) - A(H(\mathcal{X}_n, \mathcal{Y}))}{A(T(\mathcal{Y}))},$$

where $\vec{x} = (x_1, x_2, x_3)$ and $\vec{y} = (y_1, y_2, y_3)$.

If the vertex extrema are in fact at the vertices, then $G(\vec{x}, \vec{y}) = 1$, hence the integrand is critical when $x_{\mathsf{y}} = \mathsf{y}$, for each $\mathsf{y} \in \mathcal{Y}$; i.e., when $(x_1, y_1) = (0,0)$, $(x_2, y_2) = (1,0)$, and $(x_3, y_3) = (c_1, c_2)$. Switching to polar coordinates as in Equation 4.3.2, $G(\vec{x}, \vec{y})$ becomes $G\left(\vec{r}, \vec{\theta}\right)$ and

$$P(\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset) = \int \frac{n(n-1)(n-2)}{A(T(\mathcal{Y}))^3} G\left(\vec{r}, \vec{\theta}\right)^{n-3} r_1 r_2 r_3 d\vec{r} d\vec{\theta}.$$

The new integrand is critical at $(r_1, r_2, r_3) = (0,0,0)$, since $G\left(\vec{r} = (0,0,0), \vec{\theta}\right) = 1$. Letting $r_j = w_j/\sqrt{n}$, for $j \in \{1,2,3\}$, and $\vec{w} = (w_1, w_2, w_3)$ for sufficiently large n we get

$$\begin{aligned} P(\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset) &\approx \int_0^{\sqrt{n}\,\varepsilon} \int_0^{\sqrt{n}\,\varepsilon} \int_0^{\sqrt{n}\,\varepsilon} \frac{n(n-1)(n-2)}{n^3} w_1\, w_2\, w_3\, G\left(\vec{w}, \vec{\theta}\right)^{n-3} d\vec{w}, \\ &\approx \int_0^{\sqrt{n}\,\varepsilon} \int_0^{\sqrt{n}\,\varepsilon} \int_0^{\sqrt{n}\,\varepsilon} \left(1 - \frac{1}{n} g\left(\vec{w}, \vec{\theta}\right)\right)^{n-3} w_1\, w_2\, w_3\, dw_1\, dw_2\, dw_3 \end{aligned}$$

where $d\vec{w} = dw_1 dw_2 dw_3$, $g\left(\vec{w}, \vec{\theta}\right) = K_1\left(\vec{\theta}\right) w_1^2 + K_2\left(\vec{\theta}\right) w_2^2 + K_{12}\left(\vec{\theta}\right) w_1\, w_2 + K_{13}\left(\vec{\theta}\right) w_1\, w_3 + K_{23}\left(\vec{\theta}\right) w_2\, w_3)$, with $K_1\left(\vec{\theta}\right)$, $K_2\left(\vec{\theta}\right)$, $K_{12}\left(\vec{\theta}\right)$, $K_{13}\left(\vec{\theta}\right)$, and $K_{23}\left(\vec{\theta}\right)$ are functions of only $\vec{\theta}$. For example, in T_e,

$$K_1\left(\vec{\theta}\right) = K_2\left(\vec{\theta}\right) = 2\,\pi\,\sqrt{3}/9, \quad K_{12}\left(\vec{\theta}\right) = \sqrt{3}\,\sin(\theta_2 - \theta_1) + \cos(\theta_1 + \theta_2) + \cos(\theta_2)\,\cos(\theta_1),$$
$$K_{13}\left(\vec{\theta}\right) = -4\,\cos(\theta_3)\left(\sin(\theta_1) + \sqrt{3}\,\cos(\theta_1)\right)/9, \quad K_{23}\left(\vec{\theta}\right) = 4\,\cos(\theta_3)\left(\sqrt{3}\,\cos(\theta_2) - \sin(\theta_2)\right)/9$$

So, letting $n \to \infty$

$$P(\mathcal{X}_n \cap H(\mathcal{X}_n, \mathcal{Y}) = \emptyset) \approx \int_0^\infty \int_0^\infty \int_0^\infty w_1\, w_2\, w_3\ \exp\Bigg(-\Big[K_1\big(\vec{\theta}\big)\, w_1^2 + K_2\big(\vec{\theta}\big)\, w_2^2 + K_{12}\big(\vec{\theta}\big)\, w_1\, w_2 + K_{13}\big(\vec{\theta}\big)\, w_1\, w_3 + K_{23}\big(\vec{\theta}\big)\, w_2\, w_3\Big]\Bigg)\, d\vec{w},$$

which is not analytically integrable but yields a bounded constant $K(c_1, c_2)$.

For $T_b \neq T_e$, $P\big(\mathcal{X}_n \cap \big[\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) \cap R_{CC}\,(\mathsf{y}_j)\big] \neq \emptyset\big)$ is maximized when $j = 3$, since in T_b

$$A\big(\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) \cap R_{CC}\,(\mathsf{y}_k)\big) \leq^{ST} A\big(\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) \cap R_{CC}\,(\mathsf{y}_3)\big) \text{ for } k = 1, 2.$$

Furthermore,

$$A(N_{AS}\,(X, M_{CC}), X \in R_{CC}\,(\mathsf{y}_k)) \leq^{ST} A(N_{AS}\,(X, M_{CC}), X \in R_{CC}\,(\mathsf{y}_3)) \text{ for } k = 1, 2.$$

If $T_b = T_e$, $A(\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC}) \cap R_{CC}\,(\mathsf{y}_j))$ and $A(N_{AS}\,(X, M_{CC}), X \in R_{CC}\,(\mathsf{y}_j))$ are all identically distributed for $j \in \{1, 2, 3\}$.

Following are some a.s. upper bounds for the asymptotic values of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ for $N_{AS}\,(\cdot\,, M)$.

Proposition 4.3.4. *Suppose $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$. Then $P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \leq 3) \to 1$ as $n \to \infty$.*

Proof: See Appendix Section A.2.1 for the proof. ■

Below is an improved upper bound for acute triangles with $F = \mathcal{U}(T(\mathcal{Y}))$.

Theorem 4.3.5. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and $T(\mathcal{Y})$ is an acute triangle, then $P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \leq 2) \to 1$ as $n \to \infty$.*

Proof: See Appendix Section A.2.2 for the proof. ■

We conjecture the following on the asymptotic distribution of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ for uniform data on triangles.

Conjecture 4.3.6. Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and $T(\mathcal{Y})$ is an acute triangle.

Then for $N_{AS}(\cdot\,, M_{CC})$,

$$\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \to \begin{cases} 1 & \text{w.p. } \pi(c_1, c_2), \\ 2 & \text{w.p. } 1 - \pi(c_1, c_2), \end{cases} \quad \text{as } n \to \infty.$$

where $\pi(c_1, c_2) \in (0, 1)$.

Conjecture 4.3.7. Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and $T(\mathcal{Y})$ is a right triangle. Then for $N_{AS}(\cdot\,, M_{CC})$, $P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \leq 2) \to 1$ as $n \to \infty$.

Conjecture 4.3.8. Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and $T(\mathcal{Y})$ is an obtuse triangle. Then for $N_{AS}(\cdot\,, M_{CC})$, $P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) = 2) \to 1$ as $n \to \infty$.

We estimate the distribution of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ for various n empirically and present empirical estimates of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ with $n = 10, 20, 30, 50, 100$ based on 1000 Monte Carlo replicates. The estimates for T_e are presented in Table 4.3.1 (left) (see Figure 4.3.3 (left) for the corresponding graph); for $T_b = \big((0,0),(1,0),(1/2,1/2)\big)$ which is an isosceles right triangle, are presented in Table 4.3.1 (right) (see Figure 4.3.3 (right) for the corresponding graph); for $T_b = \big((0,0),(1,0),(1/2,1/4)\big)$, are presented in Table 4.3.2 (see Figure 4.3.4 for the corresponding graph).

$k \diagdown n$	10	20	30	50	100
1	697	704	697	705	712
2	284	284	300	293	288
3	19	12	3	2	0

$k \diagdown n$	10	20	30	50	100
1	593	583	590	597	618
2	384	413	403	400	382
3	23	4	7	3	0

Table 4.3.1: The number of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) = k$ out of $N = 1000$ Monte Carlo replicates for the equilateral triangle T_e (left) for and the isosceles right triangle $T_b = \big((0,0),(1,0),(1/2,1/2)\big)$ (right).

Proposition 4.3.9. *Suppose $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$. Then for $M \in T(\mathcal{Y}) \setminus \{M_{CC}\}$, $P(\gamma(\mathcal{X}_n, N_{AS}, M) = 1) \to 1$ as $n \to \infty$.*

Proof: For $M \in T(\mathcal{Y}) \setminus \{M_{CC}\}$, $\mathscr{R}_S(N_{AS}, M)$ has positive area, hence the result follows by Theorem 4.2.1. ■

Remark 4.3.10. For $N_{AS}(\cdot\,, M)$ with $M \in \{M_I, M_C\}$, $P(\gamma(\mathcal{X}_n, N_{AS}, M) = 1) \to 1$ as $n \to \infty$ in non-equilateral triangles.

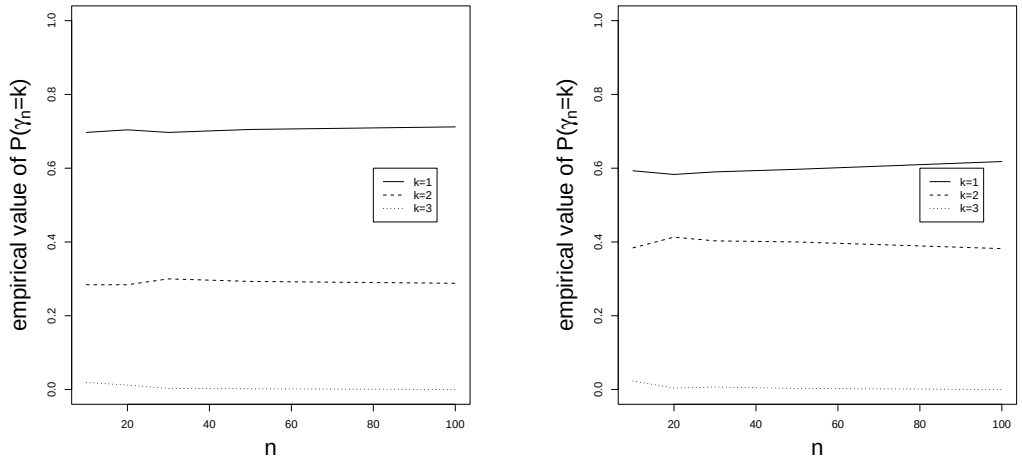


Figure 4.3.3: The empirical estimates of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ versus various n values for T_e (left) and for $T_b = \big((0,0),(1,0),(1/2,1/2)\big)$ (right).

$k \diagdown n$	10	20	30	50	100	200
1	30	0	0	0	0	0
2	783	867	926	958	991	997
3	185	133	74	42	9	3
4	2	0	0	0	0	0

Table 4.3.2: The number of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) = k$ out of $N = 1000$ Monte Carlo replicates for the obtuse triangle $T_b = \big((0,0),(1,0),(1/2,1/4)\big)$.

4.4 The Asymptotic Distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$

Recall that $\kappa\left(N_{PE}^r\right) \leq 3$, then

$$1 \leq \mathbf{E}\left[\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)\right] \leq 3 \text{ and } 0 \leq \mathbf{Var}\left[\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)\right] \leq 9/4.$$

Furthermore, there is a stochastic ordering for $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$.

Theorem 4.4.1. *Suppose $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$. Then for $r_1 < r_2$, we have $\gamma\left(\mathcal{X}_n, N_{PE}^{r_2}, M\right) \leq^{ST} \gamma\left(\mathcal{X}_n, N_{PE}^{r_1}, M\right)$.*

Proof: Suppose $r_1 < r_2$. Then $P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{r_2}, M\right) = 1\right) > P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{r_1}, M\right) = 1\right)$ since $\Gamma_1(\mathcal{X}_n, N_{PE}^{r_1}) \subsetneq \Gamma_1(\mathcal{X}_n, N_{PE}^{r_2})$ for any realization of $\mathcal{X}_n$ and by a similar argument $P\big(\gamma\left(\mathcal{X}_n, N_{PE}^{r_2}, M\right) = 2\big) > P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{r_1}, M\right) = 2\right)$ so $P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{r_2}, M\right) = 3\right) < P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{r_1}, M\right) = 3\right)$. Hence the desired result follows. ■

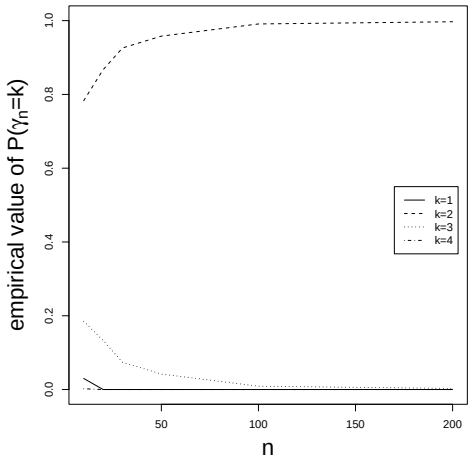


Figure 4.3.4: The empirical estimates of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ versus various n values for $T_b = \big((0,0),(1,0),(1/2,1/4)\big)$.

We present a "geometry invariance" result for $N_{PE}^r(\cdot, M)$ which will simplify our subsequent analysis by allowing us to consider the special case of the equilateral triangle.

Theorem 4.4.2. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for any $r \in [1,\infty]$ the distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ is independent of $\mathcal{Y}$ and hence the geometry of $T(\mathcal{Y})$.*

Proof: A composition of translation, rotation, reflections, and scaling will take any given triangle $T_o = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ to the basic triangle $T_b = T((0,0),(1,0),(c_1,c_2))$ with $0 < c_1 \leq 1/2$, $c_2 > 0$, and $(1-c_1)^2 + c_2^2 \leq 1$, preserving uniformity. The transformation $\phi_e : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\phi_e(u,v) = \left(u + \frac{1-2\,c_1}{\sqrt{3}}\,v, \frac{\sqrt{3}}{2\,c_2}\,v\right)$ takes T_b to the equilateral triangle $T_e = \big((0,0),(1,0),(1/2,\sqrt{3}/2)\big)$. Investigation of the Jacobian shows that ϕ_e also preserves uniformity. Furthermore, the composition of ϕ_e, with the scaling and rigid body transformations, maps the boundary of the original triangle, T_o, to the boundary of the equilateral triangle, T_e, the lines joining M to y_j in T_b to the lines joining $\phi_e(M)$ to $\phi_e(\mathsf{y}_j)$ in T_e, and lines parallel to the edges of T_o to lines parallel to the edges of T_e. Since the distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ involves only probability content of unions and intersections of regions bounded by precisely such lines and the probability content of such regions is preserved since uniformity is preserved; the desired result follows. ■

Note that geometry invariance of $\gamma\left(\mathcal{X}_n, N_{PE}^{r=\infty}, M\right)$ follows trivially since for $r = \infty$, $\gamma\left(\mathcal{X}_n, N_{PE}^{r=\infty}, M\right) = 1$ a.s. for all $T(\mathcal{Y})$.

Theorem 4.4.2 generalizes to multiple dimensions so that any simplex $\mathfrak{S}$ in $\mathbb{R}^d$ can be transformed into a regular polytope (with egdes being equal in length and faces being equal in volume) preserving uniformity. Delaunay triangulation becomes Delaunay tessellation in $\mathbb{R}^d$, provided no more than $d+1$ points being cospherical (lying on the boundary of the same sphere). In particular, with $d=3$, the general simplex is a tetrahedron (4 vertices, 4 triangular faces and 6 edges), which can be mapped into a regular tetrahedron (4 faces are equilateral triangles) with vertices $(0,0,0)\,(1,0,0)\,\big(1/2,\sqrt{3}/2,0\big),\ \big(1/2,\sqrt{3}/6,\sqrt{6}/3\big)$.

Remark 4.4.3. The orthogonal projections from $M \in T(\mathcal{Y})^o$ to the edges e_j are not mapped to the orthogonal projections from $\phi_e(M)$ to $\phi_e(e_j)$, since orthogonal projections depend on the inner angles (or equivalently on (c_1, c_2)). Hence the distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ with vertex regions obtained by orthogonal projections of M to the edges, e.g., the distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, M_{CC})$, is not geometry invariant. □

Based on Theorem 4.4.2 we may assume that $T(\mathcal{Y})$ is a standard equilateral triangle with $\mathcal{Y} = \big\{(0,0),(1,0),\big(1/2,\sqrt{3}/2\big)\big\}$ for $N_{PE}^r(\cdot, M)$ with M-vertex regions.

Theorem 4.4.4. *Suppose $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$. If $M \notin \mathscr{T}^r$ (see Figure 3.3.4 and Equation 3.3.1 for $\mathscr{T}^r$), then* $\lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = 1\right) = 1$ *for all $M \in \mathbb{R}^2 \setminus \mathcal{Y}$.*

Proof: $M \notin \mathscr{T}^r$ implies that $\mathscr{R}_S\left(N_{PE}^r, M\right)$ has positive area. Hence the result follows by Theorem 4.2.1. ■

Corollary 4.4.5. *Suppose $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$. Then for $r > 3/2$,* $\lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = 1\right) = 1$ *for all $M \in R^2 \setminus \mathcal{Y}$.*

Proof: For $r > 3/2$, $\mathscr{T}^r = \emptyset$, so $M \notin \mathscr{T}^r$. Hence the result follows by Theorem 4.4.4. ■

We estimate the distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ with $r = 2$ and $M = M_C$ for various n empirically. In Table 4.4.1, we present the empirical estimates of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ with $n = 10, 20, 30, 50, 100$ based on 1000 Monte Carlo replicates in T_e. See Figure 4.4.1 for the corresponding graph.

In particular, for $N_{PE}^2(\cdot, M_C)$, with CM-vertex regions defined by median lines, the superset region is $\mathscr{R}_S\left(N_{PE}^2, M_C\right) = T(M_1, M_2, M_3)$ and $P(\mathcal{X}_n \cap \mathscr{R}_S\left(N_{PE}^2, M_C\right) \neq \emptyset) = P(\mathcal{X}_n \cap T(M_1, M_2, M_3) \neq \emptyset) \to 1$ at rate $O\big((3/4)^n\big)$. The same holds for $N_{PE}^2(\cdot, M_{CC})$ in acute triangles.

$k \diagdown n$	10	20	30	50	100
1	961	1000	1000	1000	1000
2	34	0	0	0	0
3	5	0	0	0	0

Table 4.4.1: The number of $\gamma(\mathcal{X}_n, N_{PE}^r, M) = k$ out of $N = 1000$ Monte Carlo replicates with $r = 2$ and $M = M_C$.

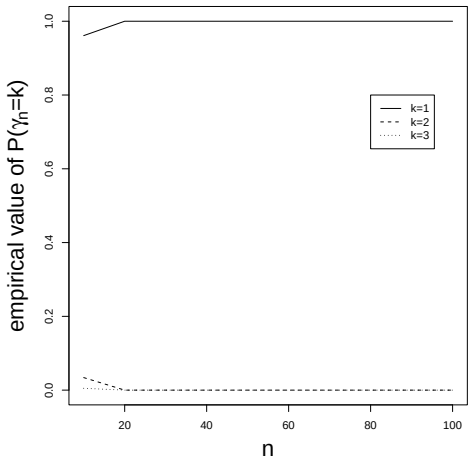


Figure 4.4.1: The empirical estimates of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{PE}^r, M)$ versus various n values with $r = 2$ and $M = M_C$.

The asymptotic distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, M)$ for $r < 3/2$ depends on the relative position of M with respect to $\mathscr{T}^r$. See Figure 3.3.4 and Equation 3.3.1 for $\mathscr{T}^r$.

Remark 4.4.6. Let $h_j := d(\mathsf{y}_j, e_j)$ (i.e., the altitude of $T(\mathcal{Y})$ associated with vertex y_j), for $j \in \{1, 2, 3\}$. Then for any triangle $T(\mathcal{Y})$ with M-vertex regions, if $d(M, e_j) < (\frac{r-1}{r}) \cdot h_j$ for some $j \in \{1, 2, 3\}$ then $M \notin \mathscr{T}^r$, hence $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N_{PE}^r, M) = 1) = 1$. □

For an example of $M_{CC} \notin \mathscr{T}^r$, see Figure 3.3.4 and for $M_{CC} \in \mathscr{T}^r$, consider the equilateral triangle. In non-acute triangles, the above corollary holds for $M = M_{CC}$, since in right and obtuse triangles, $M_{CC} \notin T(\mathcal{Y})^o$ which implies $M_{CC} \notin \mathscr{T}^r$.

Theorem 4.4.7. *Suppose $\mathcal{X}_n$ is a random sample from a continuous distribution F on $T(\mathcal{Y})$. If $M \in (\mathscr{T}^r)^o$, then $P(\gamma(\mathcal{X}_n, N_{PE}^r, M) = 3) \to 1$ as $n \to \infty$.*

Proof: Note that $(\mathscr{T}^r)^o \neq \emptyset$ iff $r < 3/2$. Suppose $M \in (\mathscr{T}^r)^o$. Then for any point u in $R_M(\mathsf{y}_j)$, $N_{PE}^r(u, M) \subsetneq T(\mathcal{Y})$, because there is a tiny strip adjacent to edge e_j not covered

by $N_{PE}^r(u, M)$, for each $j \in \{1, 2, 3\}$. Then, $N_{PE}^r(u, M) \cup N_{PE}^r(v, M) \subsetneq T(\mathcal{Y})$ for all $(u, v) \in R_M(\mathsf{y}_1) \times R_M(\mathsf{y}_2)$. Pick

$$\sup_{(u,v)\in R_M(\mathsf{y}_1)\times R_M(\mathsf{y}_2)} N_{PE}^r(u, M) \cup N_{PE}^r(v, M) \subsetneq T(\mathcal{Y}).$$

Then

$$T(\mathcal{Y}) \setminus \left[\sup_{(u,v)\in R_M(\mathsf{y}_1)\times R_M(\mathsf{y}_2)} N_{PE}^r(u, M) \cup N_{PE}^r(v, M)\right]$$

has positive area. So

$$\mathcal{X}_n \cap \left[T(\mathcal{Y}) \setminus \left[\sup_{(u,v)\in R_M(\mathsf{y}_1)\times R_M(\mathsf{y}_2)} N_{PE}^r(u, M) \cup N_{PE}^r(v, M)\right]\right] \neq \emptyset$$

with probability 1 for sufficiently large n. (The supremum of a set functional $A(x)$ over a range B is defined as the set $S := \sup_{x\in B} A(x)$ such that S is the smallest set satisfying $A(x) \subseteq S$ for all $x \in B$.) Then at least three points—one for each vertex region— are required to dominate $\mathcal{X}_n$. Hence for sufficiently large n, $\gamma(\mathcal{X}_n, N_{PE}^r, M) \geq 3$ with probability 1, but $\kappa(N_{PE}^r) = 3$ by Theorem 3.5.3. Then $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N_{PE}^r, M) = 3) = 1$ for $r < 3/2$. ■

As a corollary to this, in an acute triangle, if $d(M_{CC}, e_j) > (\frac{r-1}{r}) h_j$ for all $j \in \{1, 2, 3\}$, then $M_{CC} \in (\mathscr{T}^r)^o$, which implies $P(\gamma(\mathcal{X}_n, N_{PE}^r, M_{CC}) = 3) \to 1$ as $n \to \infty$.

We estimate the distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, M)$ with $r = 5/4$ and $M = M_C$ for various n empirically. In Table 4.4.2, we present the empirical estimates of $\gamma(\mathcal{X}_n, N_{PE}^r, M)$ with $n = 10, 20, 30, 50, 100$ based on 1000 Monte Carlo replicates in T_e. See Figure 4.4.2 for the corresponding graph.

$k \backslash n$	10	20	30	50	100
1	9	0	0	0	0
2	293	110	30	8	0
3	698	890	970	992	1000

Table 4.4.2: The number of $\gamma(\mathcal{X}_n, N_{PE}^r, M) = k$ out of $N = 1000$ Monte Carlo replicates with $r = 5/4$ and $M = M_C$.

Theorem 4.4.8. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. If $M \in \partial(\mathscr{T}^r)$, then $P(\gamma(\mathcal{X}_n, N_{PE}^r, M) > 1) \to 1$ as $n \to \infty$.*

Proof: See Appendix Section A.2.3 for the proof. ■

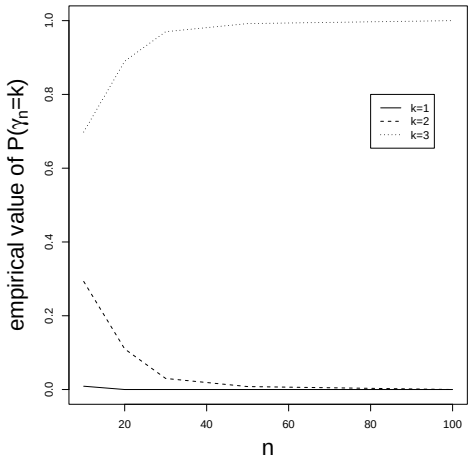


Figure 4.4.2: The empirical estimates of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{PE}^r, M)$ versus various n values with $r = 5/4$ and $M = M_C$.

For $M \in \partial(\mathscr{T}^r)$, there are two separate cases:

(i) $M \in \partial(\mathscr{T}^r) \setminus \{t_1, t_2, t_3\}$ where t_j with $j \in \{1, 2, 3\}$ are the vertices of $\mathscr{T}^r$ whose explicit forms are $t_1 = \left(\frac{(r-1)(1+c_1)}{r}, \frac{c_2(r-1)}{r}\right)$, $t_2 = \left(\frac{2-r+c_1(r-1)}{r}, \frac{c_2(r-1)}{r}\right)$, and $t_3 = \left(\frac{c_1(2-r)+r-1}{r}, \frac{c_2(r-2)}{r}\right)$.

(ii) $M \in \{t_1, t_2, t_3\}$.

Theorem 4.4.9. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. If $M \in \partial(\mathscr{T}^r) \setminus \{t_1, t_2, t_3\}$, then $P(\gamma(\mathcal{X}_n, N_{PE}^r, M) = 3) \to 1$ as $n \to \infty$.*

Proof: See Appendix Section A.2.4 for the proof. ■

We estimate the distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, M)$ with $r = 5/4$ and $M = (3/5, \sqrt{3}/10)$ for various n empirically. In Table 4.4.3 we present empirical estimates of $\gamma(\mathcal{X}_n, N_{PE}^r, M)$ with $n = 10, 20, 30, 50, 100, 500, 1000$ based on 1000 Monte Carlo replicates in T_e. See Figure 4.4.3 for the corresponding graph.

Theorem 4.4.10. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. If $M \in \{t_1, t_2, t_3\}$, then $P(\gamma(\mathcal{X}_n, N_{PE}^r, M) = 2) \to p_r$ as $n \to \infty$ where $p_r \in (0, 1)$.*

Proof: See Appendix Section A.2.5 for the proof. ■

For example,$p_{r=\sqrt{2}} \approx .4826$ and $p_{r=5/4} \approx .6514$.

$k \backslash n$	10	20	30	50	100	500	1000
1	118	60	51	39	15	1	2
2	462	409	361	299	258	100	57
3	420	531	588	662	727	899	941

Table 4.4.3: The number of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = k$ out of $N = 1000$ Monte Carlo replicates with $r = 5/4$ and $M = \left(3/5, \sqrt{3}/10\right)$.

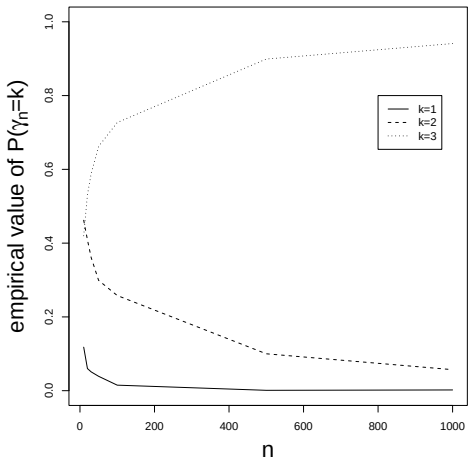


Figure 4.4.3: The empirical estimates of $P\left(\gamma_n = k\right)$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ versus various n values with $r = 5/4$ and $M = \left(3/5, \sqrt{3}/10\right)$.

So the asymptotic distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ with $r \in [1, 3/2)$ and $M \in \{t_1, t_2, t_3\}$ is

$$\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = \begin{cases} 2 & \text{wp} \quad p_r, \\ 3 & \text{wp} \quad 1 - p_r, \end{cases} \qquad \text{as } n \to \infty. \tag{4.4.1}$$

See Figure 4.4.4 for p_r as a function of $r \in (1, 1.5)$.

We estimate the distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ with $r = 5/4$ and $M = \left(7/10, \sqrt{3}/10\right)$ for various n empirically. In Table 4.4.4, we present the empirical estimates of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)$ with $n = 10, 20, 30, 50, 100, 500, 1000$ based on 1000 Monte Carlo replicates in T_e. See Figure 4.4.5 for the corresponding graph.

Theorem 4.4.11. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for $r = 3/2$, as $n \to \infty$ $P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M_C\right) > 1\right) \to 1$ at rate $O\left(n^{-1}\right)$.*

Proof: See Appendix Section A.2.6 for the proof. ■

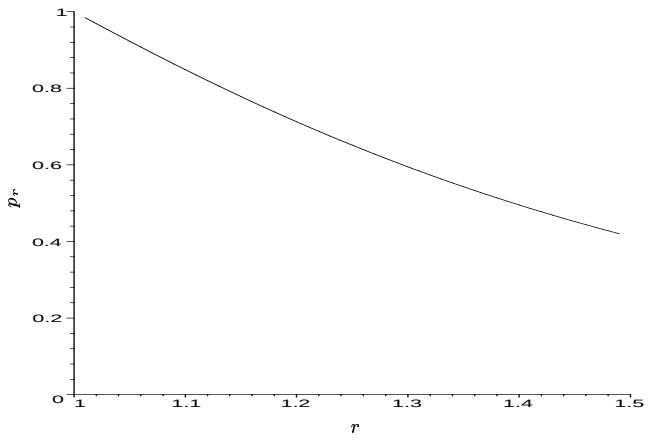


Figure 4.4.4: Plotted is the probability $p_r = \lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = 2\right)$ as a function of r for $r \in [1, 3/2)$ and $M \in \{t_1, t_2, t_3\}$.

$k \backslash n$	10	20	30	50	100	500	1000
1	174	118	82	61	22	5	1
2	532	526	548	561	611	617	633
3	294	356	370	378	367	378	366

Table 4.4.4: The number of $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = k$ out of $N = 1000$ Monte Carlo replicates with $r = 5/4$ and $M = \left(7/10, \sqrt{3}/10\right)$.

Theorem 4.4.12. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for $r = 3/2$, $\gamma\left(\mathcal{X}_n, N_{PE}^r, M_C\right)$ has the following asymptotic distribution.*

$$\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) = \begin{cases} 2 & wp \approx .7413, \\ 3 & wp \approx .2487, \end{cases} \quad as\ n \to \infty. \tag{4.4.2}$$

Proof: See Appendix Section A.2.7 for the proof. ■

Then

$$\lim_{n\to\infty} \mathbf{E}\left[\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right] \to \mu \approx 2.2587 \tag{4.4.3}$$

and

$$\mathbf{Var}\left[\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right] \to \sigma^2 \approx .1917. \tag{4.4.4}$$

Indeed, the finite sample distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)$ hence the finite sample mean and variance can be obtained by numerical methods.

We can also estimate the distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)$ for various n empirically. The

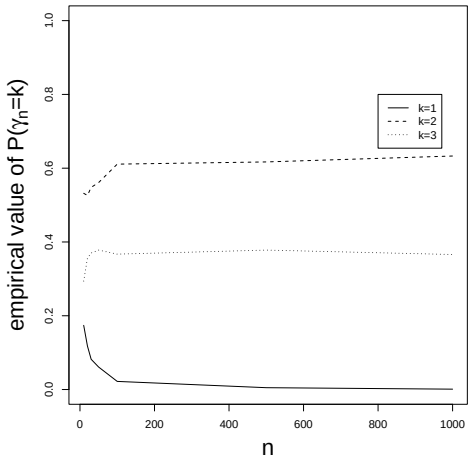


Figure 4.4.5: The empirical estimates of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{PE}^r, M)$ versus various n values with $r = 5/4$ and $M = \left(7/10, \sqrt{3}/10\right)$.

empirical estimates for $n = 10, 20, \ldots, 100, 200, 300, 500$ based on 1000 Monte Carlo replicates are given in Table 4.4.5. See Figure 4.4.6 for the corresponding graph.

$k \diagdown n$	10	20	30	40	50	60	70	80	90	100	200	300	500
1	151	82	61	67	50	24	29	21	15	27	10	7	2
2	602	636	688	670	693	714	739	708	723	718	715	730	753
3	247	282	251	263	257	262	232	271	262	255	275	263	245

Table 4.4.5: The number of $\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) = k$ out of $N = 1000$ Monte Carlo replicates.

4.5 Summary of the Asymptotic Distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, M)$

We can summarize the asymptotic distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, M)$ in a succinct form as follows:

Theorem 4.5.1. *The domination number of the PCD based on N_{PE}^r with $\mathcal{X}_n$, a random sample from $\mathcal{U}(T(\mathcal{Y}))$, with M-vertex regions has the following asymptotic distribution. As $n \to \infty$,*

$$\gamma(\mathcal{X}_n, N_{PE}^r, M) \to \begin{cases} \textit{nondegenerate}, & \textit{for } M \in \{t_1, t_2, t_3\} \textit{ and } r \in [1, 3/2], \\ 1, & \textit{for } r > 3/2, \\ 3, & \textit{for } M \in \mathscr{T}^r \setminus \{t_1, t_2, t_3\} \textit{ and } r \in [1, 3/2). \end{cases} \tag{4.5.1}$$

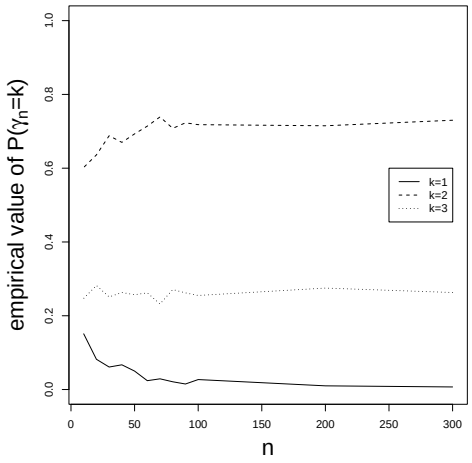


Figure 4.4.6: The empirical values of $P(\gamma_n = k)$ where $\gamma_n = \gamma(\mathcal{X}_n, N_{PE}^r, M_C)$ for various n.

The proof is given in the previous section. The nondegenerate asymptotic distributions are provided in Equation 4.4.1 for $r \in [1, 3/2)$ with $M \in \{t_1, t_2, t_3\}$ and Equation 4.4.2 for $r = 3/2$ with $M = M_C$.

Let $\gamma(\mathcal{X}_n, N_{PE}^r, d)$ be the domination number of the PCD based on the extension of $N_{PE}^r(\cdot, M_C)$ to $\mathbb{R}^d$. Then it is easy to see that $\gamma(\mathcal{X}_n, N_{PE}^r, 3)$ is nondegenerate as $n \to \infty$ for $r = 4/3$. In $\mathbb{R}^d$, it can be seen that $\gamma(\mathcal{X}_n, N_{PE}^r, d)$ is nondegenerate only for $r = (d+1)/d$ in the limit. Furthermore, for large d, asymptotic distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, d)$ is nondegenerate at values of r closer to 1. Moreover, it can be shown that $\lim_{n\to\infty} P\big(2 \le \gamma(\mathcal{X}_n, N_{PE}^{r=(d+1)/d}, d) \le d+1\big) = 1$, and we conjecture the following.

Conjecture 4.5.2. Suppose $\mathcal{X}_n$ is random sample from the uniform distribution on a simplex in $\mathbb{R}^d$. Then the domination number $\gamma(\mathcal{X}_n, N_{PE}^r, M_C)$ in the simplex satisfies

$$\lim_{n\to\infty} P\left(d \le \gamma(\mathcal{X}_n, N_{PE}^{(d+1)/d}, d) \le d+1\right) = 1.$$

For instance, with $d = 3$ we estimate the empirical distribution of $\gamma(\mathcal{X}_n, 4/3)$ for various n. The empirical estimates for $n = 10$, 20, 30, 40, 50, 100, 200, 500, 1000, 2000 based on 1000 Monte Carlo replicates for each n are given in Table 4.5.1. See Figure 4.5.1 for the corresponding graph.

$k \diagdown n$	10	20	30	40	50	100	200	500	1000	2000
1	52	18	5	5	4	0	0	0	0	0
2	385	308	263	221	219	155	88	41	31	19
3	348	455	557	609	621	725	773	831	845	862
4	215	219	175	165	156	120	139	128	124	119

Table 4.5.1: The number of $\gamma\left(\mathcal{X}_n, N_{PE}^{4/3}, M_C\right) = k$ out of $N = 1000$ Monte Carlo replicates.

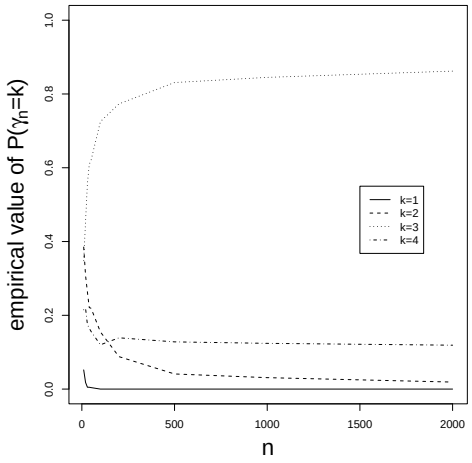


Figure 4.5.1: The empirical values of $\gamma_n = k$ versus various n values where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{PE}^{4/3}, M_C\right)$.

4.6 The Use of the Domination Number $\gamma(\mathcal{X}_n, N_{PE}^r, M_C)$ for Testing Spatial Patterns of Segregation and Association

4.6.1 Null Distribution of Domination Number

A detailed description of spatial patterns of segregation and association is given in Section 2.4. Consider the null hypothesis

$$H_0 : X_i \overset{iid}{\sim} \mathcal{U}\left(T(\mathcal{Y})\right),$$

which is a form of *complete spatial randomness*. If it is desired to have the sample size be a random variable, we may consider a spatial Poisson point process on $T(\mathcal{Y})$ as our null hypothesis.

Based on Theorem 4.4.2 and our uniform null hypothesis, we may assume that $T(\mathcal{Y})$ is a standard equilateral triangle with $\mathcal{Y} = \left\{(0,0), (1,0), \left(1/2, \sqrt{3}/2\right)\right\}$ henceforth.

In our r-factor proportional-edge proximity map and uniform null hypothesis, the asymptotic

null distribution of $\gamma(\mathcal{X}_n, N_{PE}^r, M_C)$ as a function of r can be derived, see main result in Section 4.5.

4.6.2 The Null Distribution of Mean Domination Number in the Multiple Triangle Case

Suppose $\mathcal{Y}$ is a finite collection of points in $\mathbb{R}^2$ with $|\mathcal{Y}| \geq 3$. Consider the Delaunay triangulation (assumed to exist) of $\mathcal{Y}$, where T_j denotes the j^{th} Delaunay triangle, J denotes the number of triangles, and $C_H(\mathcal{Y})$ denotes the convex hull of $\mathcal{Y}$ (see [30]). We wish to investigate $H_0 : X_i \overset{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}))$ against segregation and association alternatives.

Figure 4.6.1 (middle) presents a realization of 1000 observations independent and identically distributed according to $\mathcal{U}(C_H(\mathcal{Y}))$ for $|\mathcal{Y}| = 10$ and $J = 13$.

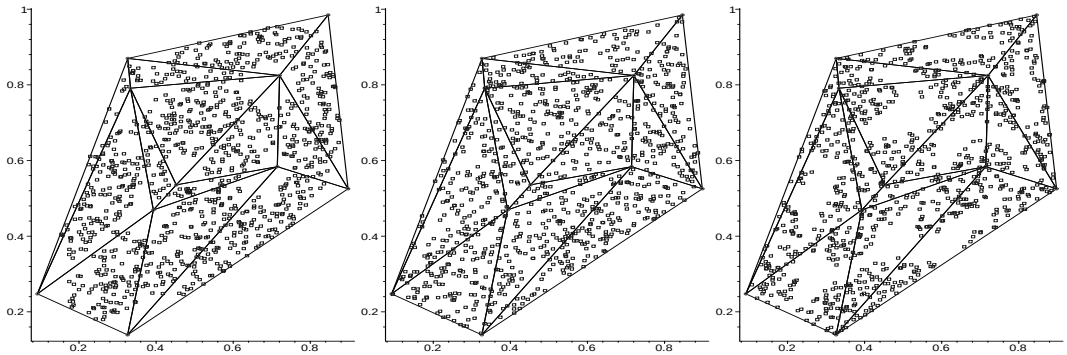


Figure 4.6.1: Realizations of segregation (left), null case (middle), and association (right) for $|\mathcal{Y}| = 10$, $J = 13$, and $n = 1000$.

The PCD is constructed using $N_{PE}^r(\cdot, M_C, \mathcal{Y}_j)$ as described above where for $X_i \in T_j$ the three points in $\mathcal{Y}$ defining the Delaunay triangle T_j are used as $\mathcal{Y}_j$ (to make the dependence explicit, $\mathcal{Y}_j$ is included in notation). Let $\gamma_{n_j}(r)$ be the domination number of the component of the digraph in T_j, where $n_j = |\mathcal{X}_n \cap T_j|$.

Theorem 4.6.1. *(Asymptotic Normality) Suppose $n_j \gg 1$ and J is sufficiently large. Then the null distribution of the mean domination number of the components of the PCD associated with $N_{PE}^{3/2}(\cdot, M_C)$, $\overline{G}_J := \frac{1}{J}\sum_{j=1}^{J} \gamma_{n_j}(3/2)$, is given by*

$$\overline{G}_J \overset{approx}{\sim} \mathcal{N}(\mu, \sigma^2/J)$$

where μ and σ^2 are given in Equations 4.4.3 and 4.4.4, respectively.

Sketch of the proof: For fixed J sufficiently large and each n_j sufficiently large, $\gamma_{n_j}(3/2)$ are approximately independent identically distributed as in Equation 4.4.2. ■

Figure 4.6.2 indicates that, for $J = 13$ with the realization of $\mathcal{Y}$ given in Figure 4.6.1 and $n = 100$ the normal approximation is not appropriate (even though the distribution looks symmetric) since not all n_j are large enough. But for $n = 1000$ the histogram and the corresponding normal curve are similar indicating that this sample size is large enough to allow the use of the asymptotic normal approximation, since all n_j are sufficiently large. However, larger J values require larger sample sizes in order to obtain approximate normality. Figure 4.6.3 indicates that, for $J = 30$ and $n = 3000$, the normal approximation is accurate, although skewness may be indicated, and for $n = 5000$, the approximation is closer.

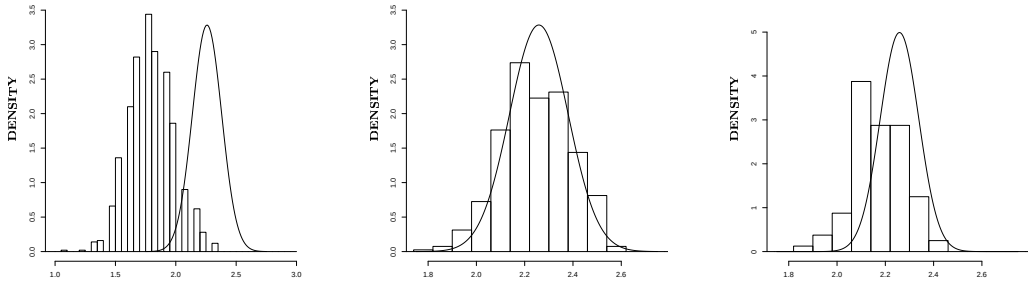


Figure 4.6.2: Depicted are $\overline{G}_J \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu \approx 2.2587, \sigma^2/J \approx .1917/J\right)$ for $J = 13$, $n = 100$ (left); $J = 13$, $n = 1000$ (middle); and $J = 30$, $n = 1000$ (right). Histograms are based on 1000 Monte Carlo replicates and the curves are the associated asymptotic normal curves.

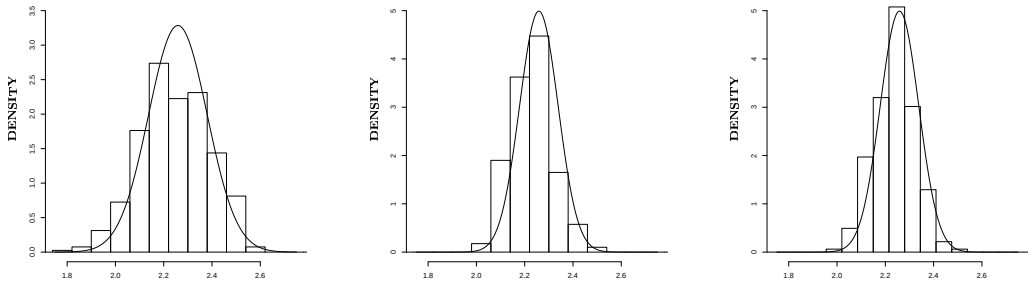


Figure 4.6.3: Depicted are $\overline{G}_J \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu \approx 2.2587, \sigma^2/J \approx .1917/J\right)$ for $J = 13$, $n = 1000$ (left); $J = 30$, $n = 3000$ (middle); and $J = 30$, $n = 5000$ (right). Histograms are based on 1000 Monte Carlo replicates and the curves are the associated asymptotic normal curves.

For finite n, let $\overline{G}_J\left(N_{PE}^r\right)$ be the mean domination number of the components of the PCD

based on N_{PE}^r. Then as a corollary to Theorem 4.4.1, it follows that for $r_1 < r_2$, we have $\overline{G}_J(N_{PE}^{r_2}) \leq^{ST} \overline{G}_J(N_{PE}^{r_1})$.

4.6.3 Segregation and Association Alternatives

We define two simple classes of alternatives, H_ε^S and H_ε^A, with $\varepsilon \in \left(0, \sqrt{3}/3\right)$, for segregation and association, respectively. Let $\mathcal{Y} = \{(0,0), (0,1), \left(1/2, \sqrt{3}/2\right)\}$ and $T(\mathcal{Y}) = T_e$. For $\mathsf{y} \in \mathcal{Y}$, let $e(\mathsf{y})$ denote the edge of $T(\mathcal{Y})$ opposite vertex y, and for $x \in T(\mathcal{Y})$ let $\ell_{\mathsf{y}}(x)$ denote the line parallel to $e(\mathsf{y})$ through x. Then define $T(\mathsf{y}, \varepsilon) = \{x \in T(\mathcal{Y}) : d(\mathsf{y}, \ell_{\mathsf{y}}(x)) \leq \varepsilon\}$. Let H_ε^S be the model under which $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}) \setminus \cup_{\mathsf{y}\in\mathcal{Y}} T(\mathsf{y}, \varepsilon))$ and H_ε^A be the model under which $X_i \overset{iid}{\sim} \mathcal{U}(\cup_{\mathsf{y}\in\mathcal{Y}} T(\mathsf{y}, \sqrt{3}/3 - \varepsilon))$. Thus the segregation model excludes the possibility of any X_i occurring near a y_j, and the association model requires that all X_i occur near y_j. The $\sqrt{3}/3 - \varepsilon$ in the definition of the association alternative is so that $\varepsilon = 0$ yields H_0 under both classes of alternatives.

Note that $T(\mathcal{Y}) \setminus \cup_{\mathsf{y}\in\mathcal{Y}} T(\mathsf{y}, \varepsilon)$ is a hexagon for $\varepsilon \in \left(0, \sqrt{3}/4\right)$ and is a triangle for $\varepsilon \in \left[\sqrt{3}/4, \sqrt{3}/3\right)$, is the center of mass for $\varepsilon = \sqrt{3}/3$ and empty set for $\varepsilon \in \left(\sqrt{3}/3, \infty\right)$.

Remark 4.6.2. These definitions of the alternatives are given for the standard equilateral triangle. The geometry invariance result of Theorem 4.4.2 still holds under the alternatives, in the following sense. If, in an arbitrary triangle, a small percentage $\delta \cdot 100\%$ where $\delta \in (0, 4/9)$ of the area is carved away as forbidden from each vertex using line segments parallel to the opposite edge, then under the transformation to the standard equilateral triangle this will result in the alternative $H^S_{\sqrt{3\delta/4}}$. This argument is for segregation; a similar construction is available for association. □

We parametrize the distributions under the segregation (association) alternative as follows. Let $\mathscr{F}^S$ ($\mathscr{F}^A$) be the family of continuous distributions, with support $T(\mathcal{Y})$ and let F_ε, be the distribution F restricted to $T(\mathcal{Y}) \setminus \cup_{j=1}^3 T(\mathsf{y}, \varepsilon)$ ($\cup_{j=1}^3 T(\mathsf{y}, \varepsilon)$). The distribution for complete spatial randomness in $\mathscr{F}^S$ ($\mathscr{F}^A$) is with $\varepsilon = 0$ and $F_{\varepsilon=0} = \mathcal{U}(T(\mathcal{Y}))$ and $\varepsilon > 0$ for $F_\varepsilon \in \mathscr{F}^S$ ($F_\varepsilon \in \mathscr{F}^A$) implies segregation (association).

Theorem 4.6.3. *(Stochastic Ordering) Let $\gamma_\varepsilon^S(\mathcal{X}_n, N_{PE}^r)$ be the domination number under the segregation alternative with $\varepsilon > 0$. Then with $\varepsilon_j \in \left(0, \sqrt{3}/3\right)$, $j \in \{1, 2\}$, $\varepsilon_1 > \varepsilon_2$ implies that $\gamma_{\varepsilon_1}^S(\mathcal{X}_n, N_{PE}^r) \leq^{ST} \gamma_{\varepsilon_2}^S(\mathcal{X}_n, N_{PE}^r)$.*

Proof: $P\left(\gamma_{\varepsilon_1}^S\left(\mathcal{X}_n, N_{PE}^{3/2}\right)=1\right) > P\left(\gamma_{\varepsilon_2}^S\left(\mathcal{X}_n, N_{PE}^{3/2}\right)=1\right)$ and $P\left(\gamma_{\varepsilon_1}^S\left(\mathcal{X}_n, N_{PE}^{3/2}\right)=2\right) > P\left(\gamma_{\varepsilon_2}^S\left(\mathcal{X}_n, N_{PE}^{3/2}\right)=2\right)$. The desired result follows. ■

Note that, the stochastic ordering in the limiting case requires $\varepsilon_2 \in \left(0, \sqrt{3}/4\right]$ and $\varepsilon_1 \in \left(\sqrt{3}/4, \sqrt{3}/3\right)$. For $\varepsilon \in \left(0, \sqrt{3}/4\right]$, $\gamma_\varepsilon^S\left(\mathcal{X}_n, N_{PE}^{3/2}\right) \to 2$ in probability as $n \to \infty$, and for $\varepsilon \in \left(\sqrt{3}/4, \sqrt{3}/3\right)$, $\gamma_\varepsilon^S\left(\mathcal{X}_n, N_{PE}^{3/2}\right) \to 1$ in probability as $n \to \infty$.

Similarly, the stochastic ordering result of Theorem 4.6.3 holds for association for all ε and $n < \infty$, with the inequalities being reversed.

Remark 4.6.4. The stochastic ordering in Theorem 4.6.3 also holds for all $r < \infty$ and all M. □

Notice that under segregation with $\varepsilon \in \left(0, \sqrt{3}/4\right)$, $\gamma_\varepsilon^S\left(\mathcal{X}_n, N_{PE}^r\right)$ is degenerate in the limit except for $r = \left(3 - \sqrt{3}\,\varepsilon\right)/2$. With $\varepsilon \in \left(\sqrt{3}/4, \sqrt{3}/3\right)$, $\gamma_\varepsilon^S\left(\mathcal{X}_n, N_{PE}^r\right)$ is degenerate in the limit except for $r = \sqrt{3}/\varepsilon - 2$. Furthermore, under association with $\varepsilon \in \left(0, \sqrt{3}/4\right)$, $\gamma_\varepsilon^A\left(\mathcal{X}_n, N_{PE}^r\right)$ is degenerate in the limit except for $r = \frac{3}{2\left(1-\sqrt{3}\,\varepsilon\right)}$.

The mean domination number of the components of the proximity catch digraph, $\overline{G}_J := \frac{1}{J}\sum_{j=1}^J \gamma_{n_j}\left(N_{PE}^{3/2}\right)$, is a test statistic for the segregation/association alternative; rejecting for extreme values of $\overline{G}_J$ is appropriate since under segregation, we expect $\overline{G}_J$ to be small, while under association, we expect $\overline{G}_J$ to be large. Using the equivalent test statistic

$$S = \sqrt{J}\left(\overline{G}_J - \mu\right)/\sigma, \tag{4.6.1}$$

the asymptotic critical value for the one-sided level α test against segregation is given by

$$z_{1-\alpha} = \Phi^{-1}(\alpha), \tag{4.6.2}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Against segregation, the test rejects for $S < z_{1-\alpha}$. Against association, the test rejects for $S > z_\alpha$.

Depicted in Figure 4.6.1 are the segregation with $\delta = 1/16$, null, and association with $\delta = 1/4$ realizations for $|\mathcal{Y}| = 10$ and $J = 13$, and $n = 1000$. The associated mean domination numbers are 1.923, 2.308, and 3.000, respectively, yielding p-values .003, .660 and .000. We also present a Monte Carlo power investigation in Section 4.6.4 for this case.

Theorem 4.6.5. *(Consistent Test) Let* $J^*(\alpha, \varepsilon) := \left\lceil \left(\frac{\sigma \cdot z_\alpha}{\mu - \overline{G}_J}\right)^2 \right\rceil$ *where* $\lceil \cdot \rceil$ *is the ceiling function and* ε*-dependence is through* $\overline{G}_J$ *under a given alternative. Then the test against* H_ε^S *which*

rejects for $S < z_{1-\alpha}$ *is consistent for all* $\varepsilon \in \left(0, \sqrt{3}/3\right)$ *and* $J \geq J^*(1-\alpha, \varepsilon)$ *and the test against* H_ε^A *which rejects for* $S > z_\alpha$ *is consistent for all* $\varepsilon \in \left(0, \sqrt{3}/3\right)$ *and* $J \geq J^*(\alpha, \varepsilon)$.

Proof: Let $\varepsilon > 0$. Under H_ε^S, $\gamma_\varepsilon^S\left(\mathcal{X}_n, N_{PE}^{3/2}\right)$ is degenerate in the limit as $n \to \infty$, which implies $\overline{G}_J$ is a constant a.s.. In particular, for $\varepsilon \in \left(0, \sqrt{3}/4\right]$, $\overline{G}_J = 2$ a.s. and for $\varepsilon \in \left(\sqrt{3}/4, \sqrt{3}/3\right)$, $\overline{G}_J = 1$ a.s. as $n \to \infty$. Then the test statistic $S = \sqrt{J}(\overline{G}_J - \mu)/\sigma$ is a constant a.s. and $J \geq J^*(1-\alpha, \varepsilon)$ implies that $S < z_{1-\alpha}$ a.s.. Hence consistency follows for segregation.

Under H_ε^A, as $n \to \infty$, $\overline{G}_J = 3$ for all $\varepsilon \in \left(0, \sqrt{3}/3\right)$, a.s.. Then $J \geq J^*(\alpha, \varepsilon)$ implies that $S > z_\alpha$ a.s., hence consistency follows for association. ■

4.6.4 Monte Carlo Power Analysis

In Figure 4.6.4, we observe empirically that even under mild segregation we obtain considerable separation for moderate J and n values suggesting high power at $\alpha = .05$. A similar result is observed for association (see Figure 4.6.5). With $J = 13$ and $n = 1000$, under H_0, the estimated significance level is $\widehat{\alpha}_S = .09$ relative to segregation, and $\widehat{\alpha}_A = .07$ relative to association. Under $H^S_{\sqrt{3}/8}$, the empirical power (using the asymptotic critical value) is $\widehat{\beta}_S = .97$, and under $H^A_{\sqrt{3}/21}$, $\widehat{\beta}_A = 1.00$. With $J = 30$ and $n = 5000$, under H_0, the estimated significance level is $\widehat{\alpha}_S = .06$ relative to segregation and $\widehat{\alpha}_A = .04$ relative to association. The empirical power is $\widehat{\beta} = 1.00$ for both alternatives.

We also estimate the empirical power by using the empirical critical values. With $J = 13$ and $n = 1000$, under $H^S_{\sqrt{3}/8}$, the empirical power is $\widehat{\beta}^S_{mc} = .72$ at empirical level $\widehat{\alpha}^S_{mc} = .033$ and under $H^A_{\sqrt{3}/21}$ the empirical power is $\widehat{\beta}^A_{mc} = 1.00$ at empirical level $\widehat{\alpha}^A_{mc} = .03$. With $J = 30$ and $n = 5000$, under $H^S_{\sqrt{3}/8}$, the empirical power is $\widehat{\beta}^S_{mc} = 1.00$ at empirical level $\widehat{\alpha}^S_{mc} = .034$ and under $H^A_{\sqrt{3}/21}$ the empirical power is $\widehat{\beta}^A_{mc} = 1$ at empirical level $\widehat{\alpha}^A_{mc} = .04$.

4.6.5 Asymptotic Efficacy Analysis

Pitman asymptotic efficacy (PAE) provides for an investigation of "local asymptotic power"– local around H_0. This involves the limit as $n \to \infty$ as well as the limit as $\varepsilon \to 0$. A detailed discussion of PAE is available in Section 2.7.2.1. For segregation or association alternatives the PAE is not applicable for $\overline{G}_J$, because none of the Pitman conditions in Section 2.7.2.1 is satisfied by the test statistic, $\overline{G}_J$.

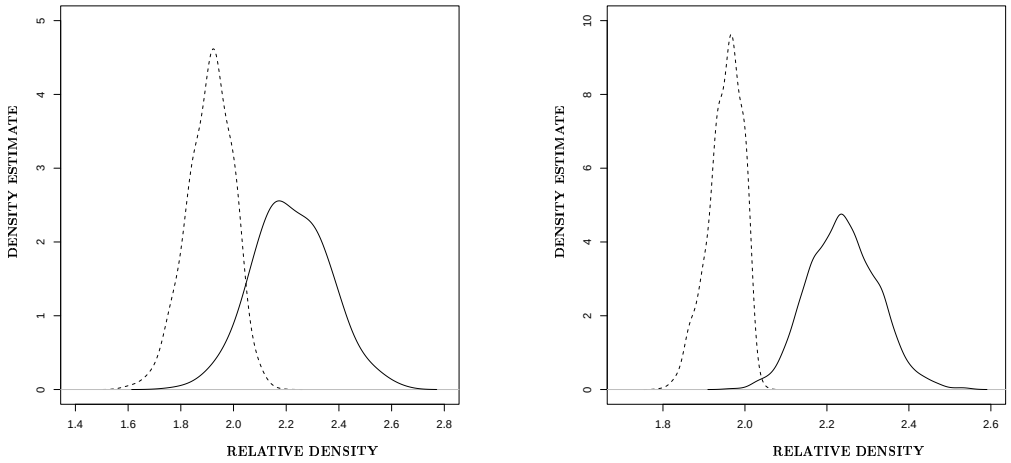


Figure 4.6.4: Two Monte Carlo experiments against the segregation alternatives $H^S_{\sqrt{3}/8}$ i.e., $\delta = 1/16$. Depicted are kernel density estimates of $\overline{G}_J$ for $J = 13$ and $n = 1000$ with 1000 Monte Carlo replicates (left) and $J = 30$ and $n = 5000$ with 1000 Monte Carlo replicates (right) under the null (solid) and the segregation alternative (dashed).

Hodges-Lehmann asymptotic efficacy (see [16]) is unbounded under each alternative and asymptotic power function analysis (see [22]) is not applicable for $\overline{G}_J$. However, for ε small and n large, this test is very sensitive for both alternatives because $\gamma^S_\varepsilon\left(\mathcal{X}_n, N^{3/2}_{PE}\right) \to 2$ in probability as $n \to \infty$ for segregation and $\gamma^A_\varepsilon\left(\mathcal{X}_n, N^{3/2}_{PE}\right) \to 3$ in probability as $n \to \infty$ for association. That is, the test statistic become degenerate in the limit for all $\varepsilon > 0$, but in the right direction for both alternatives.

4.7 The Asymptotic Distribution of $\gamma\left(\mathcal{X}_n, N^\tau_{CS}, M\right)$

For a general family of continuous distributions $\mathcal{F}_C$ on $T(\mathcal{Y})$ we have the following result.

Proposition 4.7.1. *Let $\mathcal{X}_n$ be a data set of size n from an $F \in \mathcal{F}_C$, then $P\left(\gamma\left(\mathcal{X}_n, N^\tau_{CS}, M\right) = n\right) > 0$ for all $n < \infty$.*

Proof: Similar to the proof of Theorem 3.5.4. ■

Then $1 \leq \mathbf{E}\left[\gamma\left(\mathcal{X}_n, N^\tau_{CS}, M\right)\right] \leq n$ and $0 \leq \mathbf{Var}\left[\gamma\left(\mathcal{X}_n, N^\tau_{CS}, M\right)\right] \leq n^2/4$.

There is also a stochastic ordering for $\gamma\left(\mathcal{X}_n, N^\tau_{CS}, M\right)$.

Theorem 4.7.2. *Suppose $\mathcal{X}_n$ is a random sample from an $F \in \mathcal{F}_C$. Then for $\tau_1 < \tau_2$, we have $\gamma\left(\mathcal{X}_n, N^{\tau_2}_{CS}, M\right) \leq^{ST} \gamma\left(\mathcal{X}_n, N^{\tau_1}_{CS}, M\right)$.*

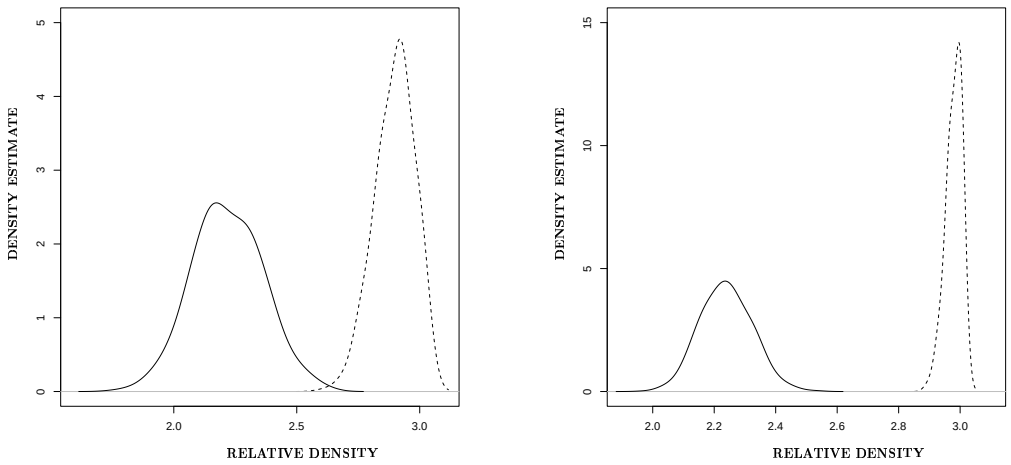


Figure 4.6.5: Two Monte Carlo experiments against the association alternatives $H^{A}_{\sqrt{3}/21}$, i.e., $\delta = 16/49$. Depicted are kernel density estimates of $\overline{G}_J$ for $J = 13$ and $n = 1000$ with 1000 Monte Carlo replicates (left) and $J = 30$ and $n = 5000$ with 100 Monte Carlo replicates (right) under the null (solid) and the association alternative (dashed).

Proof: For $\tau_1 < \tau_2$, $P_F\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau_2}, M\right) \leq k\right) \geq P_F\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau_1}, M\right) \leq k\right)$ for all $k \in \{0, 1, 2, \ldots, n\}$. In particular, for $k = 1$ strict inequality holds. Hence the desired result follows. ■

We present a "geometry invariance" result for X_n a random sample from $\mathcal{U}(T(\mathcal{Y}))$, which will simplify our subsequent analysis by allowing us to consider the special case of the standard equilateral triangle.

Theorem 4.7.3. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for any $\tau \in [0, 1]$ the distribution of $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right)$ is independent of $\mathcal{Y}$ and hence the geometry of $T(\mathcal{Y})$.*

Proof: Recall T_o, T_b, ϕ_e, and T_e in the proof of Theorem 4.4.2. A composition of scaling and rigid motion transformations will take any given triangle T_o to T_b, preserving uniformity. The transformation ϕ_e takes T_b to T_e and preserves uniformity. Furthermore, the composition of ϕ_e with the rigid motion transformations maps the boundary of the original triangle T_o to the boundary of the equilateral triangle T_e, the median lines of T_o to the median lines of T_e, lines parallel to the edges of T_o to lines parallel to the edges of T_e, and straight lines that cross T_o to the lines that cross T_e. Since the distribution of $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right)$ involves only probability content of unions and intersections of regions bounded by precisely such lines and the probability

content of such regions is preserved since uniformity is preserved; the desired result follows. ■

In fact, for $\tau = 0$, geometry invariance follows trivially since $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=0}, M\right) = n$ a.s. in any triangle $T(\mathcal{Y})$ for any $\mathcal{X}_n$ from $F \in \mathcal{F}_C$.

Theorem 4.7.3 generalizes to multiple dimensions, so that any simplex $\mathfrak{S}$ in $\mathbb{R}^d$ can be transformed into a regular polytope, e.g., a regular tetrahedron with vertices $(0,0,0)$ $(1,0,0)$ $\left(1/2, \sqrt{3}/2, 0\right)$, $\left(1/2, \sqrt{3}/6, \sqrt{6}/3\right)$ in $\mathbb{R}^3$.

Theorem 4.7.4. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $P(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) > 1) \to 1$ as $n \to \infty$.*

Proof: See Appendix Section A.2.8 for the proof. ■

In particular, $P(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M_C\right) > 1) \to 1$ as $n \to \infty$. See Figure A.2.12 (right) for a realization of $\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau}, M_C\right)$ with $M = M_C$.

As a corollary, we have the following result for $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right)$ for all $\tau \in [0,1]$.

Corollary 4.7.5. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right) > 1\right) \to 1$ as $n \to \infty$ for all $\tau \in [0,1]$.*

Proof: Above in Theorem 4.7.4, we have proved the result for $\tau = 1$. By Theorem 4.7.2, for all $\tau \in [0,1)$, $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right) \leq 1\right) \leq P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) \leq 1\right) \to 0$ as $n \to \infty$. Hence the desired result follows. ■

Next, we conjecture that for $\tau \in (0,1]$, $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M_C\right) = 1\right) \to 0$ as $n \to \infty$ at rate $O\left(\frac{\tau^{2(n-1)}}{n}\right)$ based on the following theorem.

Theorem 4.7.6. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Whenever $\mathcal{X}_n \cap R_{CM}(\mathsf{y}_j) \neq \emptyset$, let*

$$X_{e_j}^f(n) \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_{CM}(\mathsf{y}_j)} d\left(X, e_j\right),$$

i.e., the data point furthest from edge e_j in the edge-region $R_{CM}\left(\mathsf{y}_j\right)$ for $j \in \{1,2,3\}$. Note that $X_{e_j}^f(n)$ are unique a.s. as $n \to \infty$. Then for $\tau \in (0,1]$, $P\left(\mathcal{X}_n \subset N_{CS}^{\tau}\left(X_{e_j}^f(n), M_C\right)\right) \to 0$ as $n \to \infty$ at rate $O\left(\frac{\tau^{2(n-1)}}{n}\right)$.

Proof: See Appendix Section A.2.9 for the proof. ■

The result in Theorem 4.7.6 also holds for $N_{CS}^{\tau}(\cdot, M)$ for general $M \in T(\mathcal{Y})^o$, which can be shown similarly.

Next, we conjecture that $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right) > 2\right) \to 1$ as $n \to \infty$ based on the following theorem.

Theorem 4.7.7. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $P\Big(\mathcal{X}_n \subset N_{CS}^{\tau=1}\left(X_{e_i}^f, M_C\right) \cup N_{CS}^{\tau=1}\left(X_{e_j}^f(n), M_C\right)\Big) \to 0$ as $n \to \infty$ for all $(i,j) \in \{(1,2),(1,3),(2,3)\}$.*

Proof: See Appendix Section A.2.10 for the proof. ■

If the conjecture holds, then it follows that for all $\tau_1 \in (0,1)$, by Theorem 4.7.2,

$$P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau_1}, M_C\right) \leq 2\right) \leq P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right) \leq 2\right) \to 0 \text{ as } n \to \infty.$$

Hence $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M_C\right) > 2\right) \to 1$ as $n \to \infty$ for all $\tau \in [0,1]$. The result also holds for $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right)$ for general M which can be proven similarly.

We estimate $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right) = k\right)$ for $k = 1, 2$, and $k \geq 3$ with $\tau = 1$ and $M = M_C$ for various n empirically and present empirical estimates of $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M\right)$ with $n =$ 10, 20, 30, 40, 50, 100, 500 based on 1000 Monte Carlo replicates in T_e in Table 4.7.1 (see also Figure 4.7.1).

$k \backslash n$	10	20	30	40	50	100	500
1	41	23	19	11	16	4	0
2	160	110	112	113	120	79	42
≥ 3	799	867	869	876	864	917	958

Table 4.7.1: The number of $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) = k$ for $k = 1$, 2, and $k \geq 3$ out of $N = 1000$ Monte Carlo replicates with $\tau = 1$ and $M = M_C$.

Now, we prove an upper bound for $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau}, M_C\right)$.

Theorem 4.7.8. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right)\right) \leq 6) \to 1$ as $n \to \infty$.*

Proof: See Appendix Section A.2.11 for the proof. ■

Note that Theorem 4.7.8 does imply that $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau_1}, M_C\right) \leq 6\right) \leq P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right) \leq 6\right)$ for all $\tau_1 \in [0,1)$, but $P\left(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right) \leq 6\right) \to 1$ as $n \to \infty$ does not imply $P\big(\gamma\left(\mathcal{X}_n, N_{CS}^{\tau_1}, M_C\right) \leq 6\big) \to 1$ as $n \to \infty$.

In the proof of Theorem 4.7.8, we show that $\lim_{n\to\infty} P(\gamma_{ind}(\mathcal{X}_n', N_{CS}^{\tau=1}, M_C) = 1) = 1$ where $\mathcal{X}_n' = \mathcal{X}_m \cap T_s$ (conditional on the event that $|\mathcal{X}_m \cap T_s| = n$) .

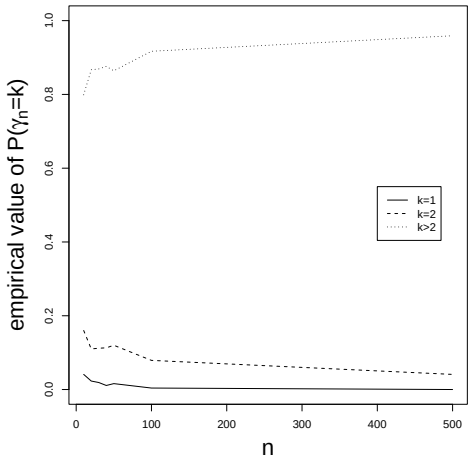


Figure 4.7.1: The empirical estimates of $P(\gamma_n = k)$ for $k = 1, 2$ and $k > 2$ versus various n values with $\tau = 1$ and $M = M_C$ where $\gamma_n = \gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)$.

We estimate the distribution of $\gamma_{ind}\left(\mathcal{X}_n', N_{CS}^{\tau}, M\right)$ with $\tau = 1$ and $M = M_C$ for various n values empirically. In Table 4.7.2, we present empirical estimates of $\gamma_{ind}\left(\mathcal{X}_n', N_{CS}^{\tau=1}, M_C\right)$ with $n \in \{10, 20, 30, 40, 50, 100, 500, 1000, 2000\}$ based on 1000 Monte Carlo replicates in T_s. See Figure 4.7.2 for the corresponding graph.

$k \diagdown n$	10	20	30	40	50	100	500	1000	2000
1	765	849	881	885	900	942	989	990	995
2	207	147	113	114	99	58	11	10	5
≥ 3	28	4	6	1	1	0	0	0	0

Table 4.7.2: The number of $\gamma_{ind}\left(\mathcal{X}_n', N_{CS}^{\tau=1}, M_C\right) = k$ out of $N = 1000$ Monte Carlo replicates.

The fact that $\lim_{n\to\infty} P(\gamma_{ind}(\mathcal{X}_n', N_{CS}^{\tau=1}, M_C) = 1) = 1$, from proof of Theorem 4.7.8, does also imply $\lim_{n\to\infty} P(\gamma_{ind}(\mathcal{X}_n'', N_{CS}^{\tau=1}, M_C) \leq 2) = 1$ where $\mathcal{X}_n'' = \mathcal{X}_m \cap T(\mathsf{y}_1, \mathsf{y}_2, M_C)$ (conditional on the event that $|\mathcal{X}_m \cap T(\mathsf{y}_1, \mathsf{y}_2, M_C)| = n$). We also estimate the distribution of $\gamma_{ind}\left(\mathcal{X}_n'', N_{CS}^{\tau}, M\right)$ with $\tau = 1$ and $M = M_C$ for various n empirically. In Table 4.7.3, we present empirical estimates of $\gamma_{ind}\left(\mathcal{X}_n'', N_{CS}^{\tau}, M\right)$ with $n \in \{10, 20, 30, 40, 50, 100, 500, 1000\}$ based on 1000 Monte Carlo replicates in $T(\mathsf{y}_1, \mathsf{y}_2, M_C)$. See Figure 4.7.3 for the corresponding graph.

As a summary of this section, we have shown that $P\left(2 \leq \gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right) \leq 6\right) \to 1$, and we conjecture that $P\left(3 \leq \gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right) \leq 6\right) \to 1$ as $n \to \infty$.

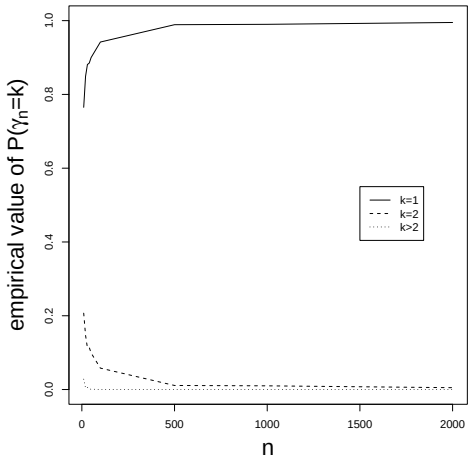


Figure 4.7.2: The empirical estimates of $P(\gamma_n = k)$ versus various n values with $\tau = 1$ and $M = M_C$ where $\gamma_n = \gamma_{ind}(\mathcal{X}_n \cap T_s, N_{CS}^{\tau}, M)$.

$k \backslash n$	10	20	30	40	50	100	500	1000
1	365	378	347	352	342	342	372	365
2	434	478	550	572	579	609	619	631
≥ 3	201	144	103	76	79	49	9	4

Table 4.7.3: The number of $\gamma_{ind}\left(\mathcal{X}_n'', N_{CS}^{\tau=1}, M_C\right) = k$ out of $N = 1000$ Monte Carlo replicates.

In the multi-dimensional case with $d > 2$, let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(\mathfrak{S})$ with $\mathfrak{S} \subsetneq \mathbb{R}^d$ and denote the domination number associated with N_{CS} based on $\mathcal{X}_n$ as $\gamma_d\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)$. Then we conjecture that

$$P\left(d+1 \leq \gamma_d\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) \leq d\,(d+1)\right) \to 1 \text{ as } n \to \infty.$$

4.7.1 Summary of Asymptotic Distribution of $\gamma(\mathcal{X}_n, N)$

Here is a summary of asymptotic distribution of $\gamma(\mathcal{X}_n, N)$ for the proximity maps we have defined, i.e., for $N \in \{N_{AS}, N_{PE}^{r}, N_{CS}^{\tau}\}$.

Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}))$ and $\gamma(\mathcal{X}_n, N, M)$ be the domination number of the PCD associated with $N \in \{N_{AS}, N_{PE}^{r}, N_{CS}^{\tau}\}$. Then as $n \to \infty$,

- for N_{AS},

  $\gamma(\mathcal{X}_n, N_{AS}, M) = 1$, w.p. 1, for all $M \in T(\mathcal{Y}) \setminus \{M_{CC}\}$,

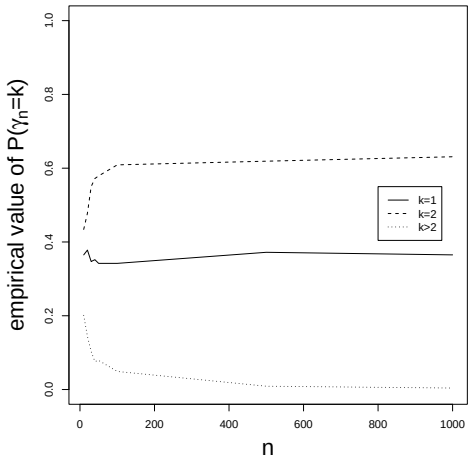


Figure 4.7.3: The empirical estimates of $P(\gamma_n = k)$ versus various n values with $\tau = 1$ and $M = M_C$ where $\gamma_n = \gamma_{ind}\left(\mathcal{X}_n^{''}, N_{CS}^{\tau}, M\right)$.

$\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \in \{1, 2\}$ w.p. 1 for $M = M_{CC}$ and acute $T(\mathcal{Y})$,

$\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \in \{1, 2, 3\}$ w.p. 1 for $M = M_{CC}$ and right $T(\mathcal{Y})$,

$\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \in \{2, 3\}$ w.p. 1 for $M = M_{CC}$ and obtuse $T(\mathcal{Y})$;

- for N_{PE}^{r},

  $\gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right) = 1$ w.p. 1 for all $M \notin \mathscr{T}^r$,

  $\gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right) = 3$ w.p. 1 for all $M \in \mathscr{T}^r \setminus \{t_1, t_2, t_3\}$,

  $\gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right) \in \{2, 3\}$, w.p. 1, for all $M \in \{t_1, t_2, t_3\}$;

- for N_{CS},

  $\gamma(\mathcal{X}_n, N_{CS}^{\tau=1}) \in \{2, 3, 4, 5, 6\}$ w.p. 1 for all $M \in T(\mathcal{Y})^o$,

  $\gamma(\mathcal{X}_n, N_{CS}^{\tau_1}) \geq 2$ w.p. 1 for all $M \in T(\mathcal{Y})^o$ and $\tau_1 \in [0, 1)$.

The exact and asymptotic distribution of $\gamma(\mathcal{X}_n, N_{AS}, M_{CC})$ and $\gamma(\mathcal{X}_n, N_{CS}^{\tau_1}, M)$ for $\tau \in (0, 1]$ are open problems. The exact distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right)$ can be found by numerical methods. The asymptotic distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^{r}, M\right)$ for $M \in \{t_1, t_2, t_3\}$ and $r \in [1, 3/2)$ is given in Equation 4.4.1 and asymptotic distribution of $\gamma\left(\mathcal{X}_n, N_{PE}^{r=3/2}, M_C\right)$ is given in Equation 4.4.2.

CHAPTER 5

Relative Density of Proximity Catch Digraphs

5.1 Preliminaries and Foundation

Let $D = (\mathcal{V}, \mathcal{A})$ be the digraph with vertex set $\mathcal{V}$ of size $|\mathcal{V}| = n$ and arc set $\mathcal{A}$. The *relative density* of D, denoted $\rho(D)$, is defined as

$$\rho(D) := \frac{|\mathcal{A}|}{n\,(n-1)}.$$

See e.g., [18]. That is, $\rho(D)$ is the ratio of number of arcs in D to the number of arcs in a *complete symmetric digraph* of order n, which is $n\,(n-1)$.

For $X_i \overset{iid}{\sim} F$, relative density $\rho(D)$ of the PCD, $D = D(\mathcal{X}_n, \mathcal{A})$, associated with the proximity map $N(\cdot)$ is denoted as $\rho(\mathcal{X}_n; h, N)$ and is a U-statistic;

$$\rho(\mathcal{X}_n; h, N) = \frac{1}{n\,(n-1)} \sum \sum_{i<j} h\big(X_i, X_j; N\big)$$

where

$$h\big(X_i, X_j; N\big) = \mathbf{I}\big(X_i X_j \in \mathcal{A}\big) + \mathbf{I}\big(X_j X_i \in \mathcal{A}\big) = \mathbf{I}\big(X_i \in N(X_j)\big) + \mathbf{I}\big(X_j \in N(X_i)\big)$$

is the number of arcs between points X_i and X_j. We will denote $h\big(X_i, X_j; N\big)$ as $h_{ij}(N)$ for brevity of notation. Since the digraph is not symmetric, $h_{ij}(N)$ is defined as the number of arcs in D between vertices X_i and X_j in order to produce a symmetric kernel with finite variance [27]. Notice that the variance is finite, since $0 \leq h\big(X_i, X_j; N\big) \leq 2$.

By definition, $X_i X_j \in \mathcal{A}$ iff $X_j \in N(X_i)$ for distinct $X_i, X_j \in \mathcal{X}_n$ in digraph $D = D(\mathcal{V} = \mathcal{X}_n, \mathcal{A})$ based on the proximity region $N(\cdot)$. Below we define a probability related to the random variable $\mathbf{I}\big(X_i X_j \in \mathcal{A}\big)$.

Definition 5.1.1. The *arc probability* between any two distinct vertices $X_i, X_j \in \mathcal{X}_n$ is defined to be $\mu\big(X_i, X_j; N\big) := \mathbf{E}\left[\mathbf{I}\big(X_i X_j \in \mathcal{A}\big)\right] = P(X_i X_j \in \mathcal{A})$.

Note that this probability could also be called *domination probability*, since $X_i X_j \in \mathcal{A}$ iff X_i dominates X_j. But by definition any vertex dominates itself and vv is generally not allowed in a digraph, hence the preference of the name *arc probability*. Furthermore, for the PCD based on the data set $\mathcal{X}_n$ and the proximity map $N(\cdot)$,

$$\mu\big(X_i, X_j; N\big) = P\big(X_i X_j \in \mathcal{A}\big) = P\big(X_j \in N(X_i)\big).$$

For $X_i \overset{iid}{\sim} F$ in Ω, $P\big(X_j \in N(X_i)\big) = P\big(X_k \in N(X_l)\big)$ for all X_i, X_j, X_k, X_l with $i \neq j$ and $k \neq l$ (the equality holds trivially if $i = j$ and $k = l$), so for $i \neq j$, $\mu\big(X_i, X_j; N\big) = \mu\big(X_1, X_2; N\big) = P\big(X_2 \in N(X_1)\big)$ does depend on the type of the proximity region $N(\cdot)$ and the distribution of X_j. In short, we will denote the arc probability as $\mu(N)$ for proximity region $N(\cdot)$ and $X_i \overset{iid}{\sim} F$.

If $x \in \mathscr{R}_S(N)$, the superset region associated with $N(\cdot)$, then $N(x) = \Omega$. So

$$\begin{aligned} \mu(N) = P\big(X_2 \in N(X_1)\big) &= P\big(X_2 \in N(X_1), X_1 \in \mathscr{R}_S(N)\big) + P\big(X_2 \in N(X_1), X_1 \notin \mathscr{R}_S(N)\big) \\ &= P\big(X_1 \in \mathscr{R}_S(N)\big) + P\big(X_2 \in N(X_1), X_1 \notin \mathscr{R}_S(N)\big). \end{aligned} \tag{5.1.1}$$

Notice that the random variable $\rho_n(N) := \rho\big(\mathcal{X}_n; h, N\big)$, the relative density of PCD based on $N(\cdot)$, depends on $\mathcal{X}_n$, h, and $N(\cdot)$ explicitly and on F implicitly. The expectation $\mathbf{E}\left[\rho_n\right]$, however, is independent of n, but depends on only F, h and $N(\cdot)$:

$$0 \leq \mathbf{E}\left[\rho_n(N)\right] = \frac{1}{n\,(n-1)} \sum_{i<j} \sum \mathbf{E}\left[h_{ij}(N)\right] = \frac{1}{2}\mathbf{E}\left[h_{12}(N)\right] = \mu(N). \tag{5.1.2}$$

The variance $\mathbf{Var}\left[\rho_n(N)\right]$ simplifies to

$$0 \leq \mathbf{Var}\left[\rho_n(N)\right] = \frac{1}{2n(n-1)}\mathbf{Var}\left[h_{12}(N)\right] + \frac{n-2}{n(n-1)}\mathbf{Cov}\left[h_{12}(N), h_{13}(N)\right] \leq 1/4. \tag{5.1.3}$$

A central limit theorem for U-statistics (see [27]) yields

$$\sqrt{n}\big(\rho_n(N) - \mathbf{E}\left[\rho_n(N)\right]\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \mathbf{Cov}\left[h_{12}(N), h_{13}(N)\right]\big) \tag{5.1.4}$$

provided that $\mathbf{Cov}\,[h_{12}(N), h_{13}(N)] > 0$. The asymptotic variance of $\rho_n(N)$, $\mathbf{Cov}\,[h_{12}(N), h_{13}(N)]$, depends on only F, h, and $N(\cdot)$. Thus, we need determine only $\mathbf{E}\,[h_{12}(N)]$ and $\mathbf{Cov}\,[h_{12}(N), h_{13}(N)]$ in order to obtain the normal approximation

$$\rho_n(N) \overset{\text{approx}}{\sim} \mathcal{N}\big(\mathbf{E}\,[\rho_n(N)], \mathbf{Var}\,[\rho_n(N)]\big) = \mathcal{N}\left(\frac{\mathbf{E}\,[h_{12}(N)]}{2}, \frac{\mathbf{Cov}\,[h_{12}(N), h_{13}(N)]}{n}\right) \text{ for large } n. \tag{5.1.5}$$

Now, let A_{ij} be the event that $\{X_i X_j \in \mathcal{A}\} = \{X_i \in N(X_j)\}$, then $h_{ij}(N) = \mathbf{I}\,(A_{ij}) + \mathbf{I}\,(A_{ji})$. In particular $h_{12}(N) = \mathbf{I}\,(A_{12}) + \mathbf{I}\,(A_{21})$. Then

$$\mathbf{Var}\,[h_{12}(N)] = \mathbf{E}\left[(h_{12}(N))^2\right] - \big(\mathbf{E}\,[h_{12}(N)]\big)^2$$

where $\mathbf{E}\,[h_{12}(N)] = 2\,\mu(N)$. Furthermore,

$$\begin{aligned}\mathbf{E}\left[(h_{12}(N))^2\right] &= \mathbf{E}\left[\big(\mathbf{I}\,(A_{12}) + \mathbf{I}\,(A_{21})\big)^2\right] = \mathbf{E}\left[\mathbf{I}\,(A_{12}) + 2\,\mathbf{I}\,(A_{12})\,\mathbf{I}\,(A_{21}) + \mathbf{I}\,(A_{21})\right] \\ &= P\,(A_{12}) + 2\,P\big(A_{12} \cap A_{21}\big) + P\,(A_{21}) = 2\,\mu(N) + 2\,\pi_{sa}(N).\end{aligned}$$

where

$$\pi_{sa}(N) = P\big(A_{12} \cap A_{21}\big) = P\big(X_1 \in N(X_2) \cap X_2 \in N(X_1)\big) = P\big(X_2 \in N(X_1) \cap \Gamma_1(X_1, N)\big),$$

which is the probability of having a *symmetric arc* between X_1 and X_2 (or X_i and X_j, for general $i \neq j$), hence the notation $\pi_{sa}(N)$. Then

$$\mathbf{Var}\,[h_{12}(N)] = 2\,\mu(N) + 2\,\pi_{sa}(N) - \big[2\,\mu(N)\big]^2,$$

and

$$\mathbf{Cov}\,[h_{12}(N), h_{13}(N)] = \mathbf{E}\,[h_{12}(N)\,h_{13}(N)] - \mathbf{E}\,[h_{12}(N)]\,\mathbf{E}\,[h_{13}(N)]$$

where $\mathbf{E}[h_{12}(N)] = \mathbf{E}[h_{13}(N)] = 2\,\mu(N)$. Furthermore,

$$\begin{aligned}\mathbf{E}[h_{12}(N)\,h_{13}(N)] &= \mathbf{E}\left[\big(\mathbf{I}(A_{12}) + \mathbf{I}(A_{21})\big)\big(\mathbf{I}(A_{13}) + \mathbf{I}(A_{31})\big)\right] \\ &= \mathbf{E}\left[\mathbf{I}(A_{12}\cap A_{13}) + \mathbf{I}(A_{12}\cap A_{31}) + \mathbf{I}(A_{21}\cap A_{13}) + \mathbf{I}(A_{21}\cap A_{31})\right] \\ &= P(A_{12}\cap A_{13}) + P(A_{12}\cap A_{31}) + P(A_{21}\cap A_{13}) + P(A_{21}\cap A_{31})\end{aligned}$$

where

$$P(A_{12}\cap A_{13}) = P\big(X_2 \in N(X_1) \wedge X_3 \in N(X_1)\big) = P\big(\{X_2, X_3\} \subset N(X_1)\big)$$

and

$$P(A_{21}\cap A_{31}) = P\big(X_2 \in \Gamma_1(X_1, N) \wedge X_3 \in \Gamma_1(X_1, N)\big) = P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N)\big).$$

Furthermore, by symmetry $P(A_{12}\cap A_{31}) = P(A_{21}\cap A_{13})$ and

$$P(A_{12}\cap A_{31}) = P\big(X_2 \in N(X_1), X_3 \in \Gamma_1(X_1, N)\big).$$

Let $P_{2N} := P\big(\{X_2, X_3\} \subset N(X_1)\big)$, $P_{2G} := P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N)\big)$, and $P_M := P\big(X_2 \in N(X_1), X_3 \in \Gamma_1(X_1, N)\big)$. Then,

$$\mathbf{Cov}[h_{12}(N), h_{13}(N)] = P_{2N} + 2\,P_M + P_{2G} - \big[2\,\mu(N)\big]^2.$$

Remark 5.1.2. The joint distribution of $(h_{12}(N), h_{13}(N))$ can be computed by finding

$$P\big((h_{12}(N), h_{13}(N)) = (i, j)\big)$$

for each $(i, j) \in \big\{(0,0), (1,0), (0,1), (1,1), (0,2), (2,0), (1,2), (2,1), (2,2)\big\}$. □

5.2 Relative Density of the CCCD Based on N_S and Uniform Data on Compact Intervals in $\mathbb{R}$

In $\mathbb{R}$, let $\mathcal{Y} = \{\mathtt{y}_1, \mathtt{y}_2\}$ with $\mathtt{y}_1 < \mathtt{y}_2$ and $F = \mathcal{U}(\mathtt{y}_1, \mathtt{y}_2)$, by geometry invariance, without loss of generality, we can assume $\mathtt{y}_1 = 0$ and $\mathtt{y}_2 = 1$, so $F = \mathcal{U}(0, 1)$. Recall that $N_S(x) = B(x, r(x))$

where $r(x) = \min(x, 1-x)$. Then the arc probability of the CCCD (based on $N_S(\cdot)$) is given by

$$\mu(N_S) = P\big(X_2 \in N_S(X_1)\big) = P\big(X_2 \in N_S(X_1), X_1 \le 1/2\big) + P\big(X_2 \in N_S(X_1), X_1 > 1/2\big).$$

By symmetry $P\big(X_2 \in N_S(X_1), X_1 \le 1/2\big) = P\big(X_2 \in N_S(X_1), X_1 > 1/2\big)$. Then

$$\begin{aligned} \mu(N_S) = 2\,P\big(X_2 \in N_S(X_1), X_1 \le 1/2\big) &= 2\,P\big(X_2 \le 2\,X_1,\ X_1 \le 1/2\big) \\ &= 2\int_0^{1/2} 2\,x\,dx = 2\,(1/4) = 1/2. \end{aligned}$$

The symmetric arc probability of the CCCD is

$$\begin{aligned} \pi_{sa}(N_S) &= P\big(X_2 \in N_S(X_1) \cap \Gamma_1(X_1, N_S)\big) \\ &= P\big(X_2 \in N_S(X_1) \cap \Gamma_1(X_1, N_S), X_1 \le 1/2\big) + P\big(X_2 \in N_S(X_1) \cap \Gamma_1(X_1, N_S), X_1 \ge 1/2\big) \\ &= 2\,\big(P\big(X_2 \in N_S(X_1) \cap \Gamma_1(X_1, N_S), X_1 \le 1/2\big)\big). \end{aligned}$$

For $x_1 \le 1/2$, $N_S(x_1) = [0, 2\,x_1]$ and $\Gamma_1(x_1, N_S) = [x_1/2, (1+x_1)/2]$, which are both connected intervals. Then for $x_1 \le 1/2$,

$$N_S(x_1) \cap \Gamma_1(x_1, N_S) = \big[x_1/2, \min(2\,x_1, (1+x_1)/2)\big]$$

where $\min(2\,x_1, x_1/2) = 2\,x_1\,\mathbf{I}(0 \le x_1 \le 1/3) + (1+x_1)/2\,\mathbf{I}(1/3 \le x_1 \le 1/2)$. Hence

$$\begin{aligned} \pi_{sa}(N_S) &= 2\,P\big(X_2 \in N_S(X_1) \cap \Gamma_1(X_1, N_S), X_1 \le 1/2\big) \\ &= 2\left(\int_0^{1/3} (2\,x_1 - x_1/2)\,dx_1 + \int_{1/3}^{1/2} \big((1+x_1)/2 - x_1/2\big)\,dx_1\right) \\ &= 2\,(1/12 + 1/12) = 1/3. \end{aligned}$$

Hence

$$\mathbf{Var}\,[h_{12}(N_S)] = 2\,\mu(N_S) + 2\,\pi_{sa}(N_S) - \big[2\,\mu(N_S)\big]^2 = 2\,(1/2) + 2\,(1/3) - [2\,(1/2)]^2 = 2/3.$$

Next, we find $\mathbf{E}\,[h_{12}(N_S)\,h_{13}(N_S)]$. By symmetry,

$$P\big(\{X_2, X_3\} \subset N_S(X_1), X_1 \le 1/2\big) = P\big(\{X_2, X_3\} \subset N_S(X_1), X_1 \ge 1/2\big)$$

so $P_{2N} = 2\,P\big(\{X_2, X_3\} \subset N_S(X_1), X_1 \le 1/2\big) = 2\left(\int_0^{1/2} (2\,x_1)^2\,dx_1\right) = 2\,(1/6) = 1/3.$

Similarly, $P_{2G} = \int_0^1 \big((1+x_1)/2 - x_1/2\big)^2\,dx_1 = (1/2)^2 = 1/4.$

Next,

$$P_M = P\big(X_2 \in N_S(X_1),\ X_3 \in \Gamma_1(X_1, N_S), X_1 \le 1/2\big) + P\big(X_2 \in N_S(X_1),\ X_3 \in \Gamma_1(X_1, N_S), X_1 > 1/2\big)$$
$$= 2\,\big(P\big(X_2 \in N_S(X_1),\ X_3 \in \Gamma_1(X_1, N_S), X_1 \le 1/2\big)\big) = 2\left(\int_0^{1/2} (2\,x_1)\,(1/2)\,dx_1\right) = 2\,(1/8) = 1/4,$$

since, given X_1, the events $X_2 \in N_S(X_1)$ and $X_3 \in \Gamma_1(X_1, N_S)$ are independent.

Hence

$$\mathbf{E}\,[h_{12}(N_S)\,h_{13}(N_S)] = 1/3 + 2\,(1/4) + 1/4 = 13/12 = 1.08\bar{3},$$

which yields

$$\mathbf{Cov}\,[h_{12}(N_S),\ h_{13}(N_S)] = 13/12 - [2\,(1/2)]^2 = 1/12 = 0.08\bar{3}.$$

Thus

$$\sqrt{n}\big(\rho_n(N_S) - 1/2\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, 1/12\big) \text{ as } n \to \infty.$$

5.3 Relative Density of Proximity Catch Digraphs Based on Arc-Slice Proximity Maps

The asymptotic distribution of $\rho_n(N_{AS})$ is not analytically tractable, so we only calculate the arc probability $\mu(N_{AS}, M_{CC})$ by numerical methods. Throughout this section, N_{AS} is defined with CC-vertex regions, hence dependence on M_{CC} is omitted.

Recall that $\mathscr{R}_S(N_{AS}) = \{M_{CC}\}$, so $A(\mathscr{R}_S(N_{AS})) = 0$, hence

$$\mu(N_{AS}) = \sum_{\mathsf{y} \in \mathcal{Y}} P\big(X_2 \in N_{AS}(X_1), X_1 \in R_{CC}(\mathsf{y})\big).$$

As in Remark 5.2.3, the distribution of $\rho_n(N_{AS})$ is not geometry invariant; so we can only use the basic triangle T_b for $T(\mathcal{Y})$.

Observe that, by symmetry, for any proximity map $N(\cdot)$

$$\mu(N) = P\big(X_2 \in N(X_1)\big) = P\big(X_1 \in \Gamma_1(X_2, N)\big) = P\big(X_2 \in \Gamma_1(X_1, N)\big).$$

In general the calculation of $\mu(N)$ with the direct method which uses $P\big(X_2 \in N(X_1)\big)$ is simpler

than the indirect method which uses $P\big(X_2 \in \Gamma_1(X_1, N)\big)$. However, $\mu(N_{AS})$ is easier to compute numerically with the latter (i.e., the indirect) method, since in the direct method the integrands contain inverse trigonometric functions and are lengthier.

For simplicity, we use the equilateral triangle T_e. By symmetry in T_e, we only consider the triangle $T_s := T(\mathsf{y}_1, M_3, M_{CC})$. There are two cases for $\Gamma_1(x, N_{AS})$ for $x \in T_s$. If $x \in R_1$ in Figure 5.3.1, $\Gamma_1(x, N_{AS})$ is a pentagon and if $x \in R_2$ in Figure 5.3.1, then $\Gamma_1(x, N_{AS})$ is a hexagon. See Figure 5.3.2 for the pentagonal and hexagonal Γ_1-regions, $\Gamma_1(x, N_{AS})$. Then

$$P\big(X_2 \in \Gamma_1(X_1, N_{AS}), X_1 \in T_s\big) = \sum_{j=1}^{2} P\big(X_2 \in \Gamma_1(X_1, N_{AS}), X_1 \in R_j\big).$$

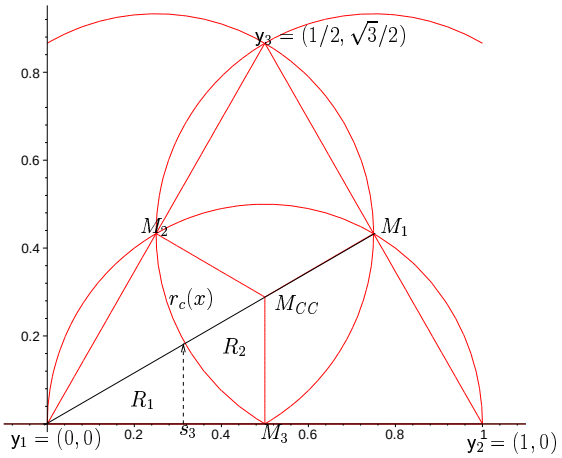


Figure 5.3.1: The two regions in T_s for which $\Gamma_1(x, N_{AS})$ is a pentagon or a hexagon.

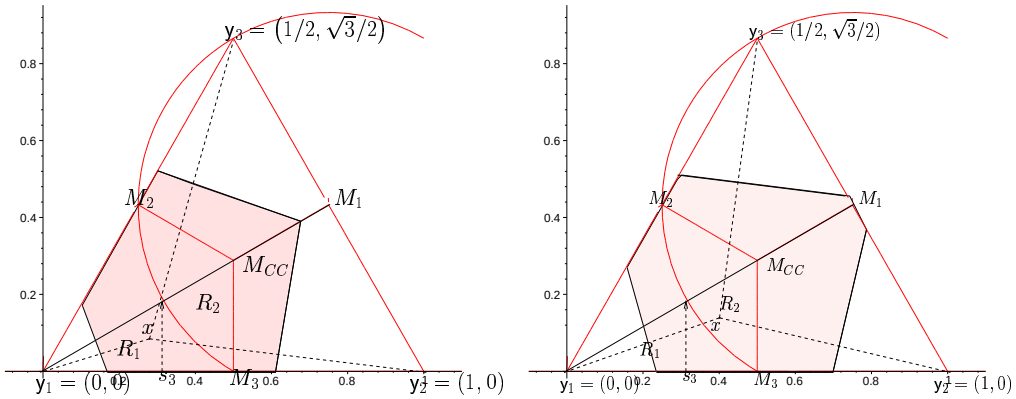


Figure 5.3.2: A sample figure for pentagonal $\Gamma_1(x, N_{AS})$ (left) and a hexagonal $\Gamma_1(x, N_{AS})$ (right).

First,

$$P(X_2 \in \Gamma_1(X_1, N_{AS}), X_1 \in R_1) = \int_0^{s_3} \int_0^{x/\sqrt{3}} \frac{A\big(\Gamma_1(x_1, N_{AS})\big)}{A(T(\mathcal{Y}))^2} dydx + \int_{s_3}^{1/2} \int_0^{r_c(x)} \frac{A\big(\Gamma_1(x_1, N_{AS})\big)}{A(T(\mathcal{Y}))^2} dydx \approx .04865,$$

which is obtained by numerical integration, where $s_3 = \left(3 - \sqrt{3}\right)/4$, $r_c(x) = \frac{\sqrt{3} - \sqrt{(4\,x-1)\,(5-4\,x)}}{4}$ (see Figure 5.3.1), and

$$A\big(\Gamma_1(x_1, N_{AS})\big) = \frac{\sqrt{3}\,(x^2+y^2-1)^2}{8\,(x-1)\,\left(3\,x+\sqrt{3}\,y-3\right)} + \frac{(x^2+y^2-1)^2\sqrt{3}}{4\,\left(3\,x+\sqrt{3}\,y-3\right)\left(x-2+\sqrt{3}\,y\right)} + \frac{3\,(x^2+y^2-1)(x^2+y^2)\,y}{8\,\left(x-2+\sqrt{3}\,y\right)x\,\left(x+\sqrt{3}\,y\right)} - \frac{\sqrt{3}\,(x^2+y^2)^2}{8\,x^2}.$$

(See Figure 5.3.2 (left)).

Next,

$$P\big(X_2 \in \Gamma_1(X_1, N_{AS}), X_1 \in R_2\big) = \int_{s_3}^{1/2} \int_{r_c(x)}^{x/\sqrt{3}} \frac{A\big(\Gamma_1(x_1, N_{AS})\big)}{A(T(\mathcal{Y}))^2} dydx \approx .0417$$

which is obtained by numerical integration, where

$$A\big(\Gamma_1(x_1, N_{AS})\big) = \frac{\sqrt{3}\,(x^2+y^2-1)\,(-2\,x+1+x^2+y^2)}{8\,(x-1)\,\left(-x+1+\sqrt{3}\,y\right)} - \frac{\sqrt{3}\,\left(-2\,y^2-2\,x^2-1+3\,x+\sqrt{3}\,y\right)}{4\,\left(-x+1+\sqrt{3}\,y\right)\left(-x-1+\sqrt{3}\,y\right)} + \frac{\sqrt{3}\,(x^2+y^2-1)\,\left(-y^2-1-x^2+\sqrt{3}\,y+x\right)}{4\,\left(-x-1+\sqrt{3}\,y\right)\left(x-2+\sqrt{3}\,y\right)} + \frac{3\,(x^2+y^2-1)(x^2+y^2)\,y}{8\,\left(x-2+\sqrt{3}\,y\right)x\,\left(x+\sqrt{3}\,y\right)} - \frac{\sqrt{3}\,(x^2+y^2)^2}{8\,x^2}.$$

Hence,

$$\mu(N_{AS}) \approx 6\,(.04865 + .0417) \approx .5421$$

in T_e. The empirical estimate of $\mu(N_{AS})$ with 10000 Monte Carlo replicates is $\widehat{\mu}(N_{AS}) = .5393$. Given a triangle $T(\mathcal{Y})$, i.e., given (c_1, c_2), $\mu(N_{AS})$ can be obtained similarly. However the computation is longer, since symmetry does not necessarily hold.

The asymptotic variance can also be calculated similarly, but it requires more effort and computation time.

5.4 Relative Density of Random Proximity Catch Digraphs for Testing Spatial Patterns of Segregation and Association

In Section 4.6, we use the domination number for testing segregation and association. In this section, we employ a different statistic, namely the *relative density*, $\rho(\mathcal{X}_n; h, N_{\mathcal{Y}})$ where $N_{\mathcal{Y}}(\cdot)$ is a proximity map based on $\mathcal{Y}$. Two simple classes of alternative hypotheses, for segregation and association, are defined in Section 4.6.3. This test is related to the available tests of spatial patterns in the literature, such as Pielou's test and Ripley's test. See the discussion in Chapter 6 for further detail. Our approach is valid for data in any dimension, but for simplicity of expression and visualization, will be described for the two-dimensional data.

The phenomena known as *segregation* and *association* are defined in Section 2.4. For statistical testing of spatial patterns of segregation and association, we consider a form of *complete spatial randomness* for the null hypothesis;

$$H_0 : X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y})).$$

5.4.1 Relative Density of Proximity Catch Digraphs Based in r-Factor Proportional-Edge Proximity Maps

We have defined the r-factor proportional-edge proximity maps and the associated PCDs in Section 3.3.2.

The test statistic used here is the relative density $\rho_n(N_{PE}^r) = \rho(\mathcal{X}_n; h, N_{PE}^r)$ where N_{PE}^r is defined with CM-vertex regions throughout this section, hence dependence on M_C is suppressed for brevity of notation. The asymptotic distributions under both the null and the alternative hypotheses are determined in Sections 5.4.2 and 5.4.3, respectively, by using the standard U-statistic central limit theory. Pitman and Hodges-Lehman asymptotic efficacies are analyzed in Sections 5.4.8 and 5.4.9, respectively.

5.4.2 Asymptotic Normality Under the Null Hypothesis

First, we present a "geometry invariance" result which allows us to assume $T(\mathcal{Y})$ is the standard equilateral triangle, $T\left((0,0),(1,0),\left(1/2,\sqrt{3}/2\right)\right)$, thereby simplifying our subsequent analyses.

Theorem 5.4.1. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for any $r \in [1,\infty]$ the distribution of $\rho_n(N_{PE}^r)$ is independent of $\mathcal{Y}$ and hence the geometry of $T(\mathcal{Y})$.*

Proof: Similar to the proof of Theorem 4.4.2. ■

Based on Theorem 5.4.1 and our uniform null hypothesis, we may assume that $T(\mathcal{Y})$ is the standard equilateral triangle with $\mathcal{Y} = \{(0,0),(1,0),(1/2,\sqrt{3}/2)\}$ for $\rho_n(N_{PE}^r)$.

For our r-factor proportional-edge proximity map and uniform null hypothesis, the asymptotic null distribution of $\rho_n(N_{PE}^r)$ can be derived as a function of r. Let $\mu(N_{PE}^r) = \mathbf{E}[\rho_n(N_{PE}^r)]$ and $\nu(N_{PE}^r) := \mathbf{Cov}[h_{12}(N_{PE}^r), h_{13}(N_{PE}^r)]$.

By detailed geometric probability calculations, the asymptotic variance of the relative density of the r-factor proximity catch digraph can be explicitly computed. The central limit theorem for U-statistics then establishes the asymptotic normality under the uniform null hypothesis. These results are summarized in the following theorem.

Theorem 5.4.2. *For $r \in [1,\infty)$,*

$$\frac{\sqrt{n}\left(\rho_n(N_{PE}^r) - \mu(N_{PE}^r)\right)}{\sqrt{\nu(N_{PE}^r)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \tag{5.4.1}$$

where

$$\mu(N_{PE}^r) = \begin{cases} \frac{37}{216}\, r^2, & \text{for} \quad r \in [1,3/2), \\ -\frac{1}{8}\, r^2 + 4 - 8r^{-1} + \frac{9}{2}\, r^{-2}, & \text{for} \quad r \in [3/2,2), \\ 1 - \frac{3}{2}\, r^{-2}, & \text{for} \quad r \in [2,\infty), \end{cases} \tag{5.4.2}$$

and

$$\begin{aligned} \nu(N_{PE}^r) = \nu_1(r)\,\mathbf{I}(r \in [1,4/3)) + \nu_2(r)\,\mathbf{I}(r \in [4/3,3/2)) \\ + \nu_3(r)\,\mathbf{I}(r \in [3/2,2)) + \nu_4(r)\,\mathbf{I}(r \in [2,\infty]) \end{aligned} \tag{5.4.3}$$

with

$$\begin{aligned} \nu_1(r) &= \left[3007\, r^{10} - 13824\ r^9 + 898\, r^8 + 77760\, r^7 - 117953\, r^6 + 48888\, r^5 - 24246\, r^4 + 60480\, r^3 - 38880\, r^2 \right. \\ &\quad \left. +3888\right] \Big/ \left[58320\, r^4\right], \\ \nu_2(r) &= \frac{5467\, r^{10} - 37800\, r^9 + 61912\, r^8 + 46588\, r^6 - 191520\, r^5 + 13608\, r^4 + 241920\, r^3 - 155520\, r^2 + 15552}{233280\, r^4}, \\ \nu_3(r) &= -\left[7\, r^{12} - 72\, r^{11} + 312\, r^{10} - 5332\, r^8 + 15072\, r^7 + 13704\ r^6 - 139264\ r^5 + 273600\, r^4 - 242176\, r^3 \right. \\ &\quad \left. +103232\, r^2 - 27648\, r + 8640\right] \Big/ \left[960\, r^6\right], \end{aligned}$$

$$\nu_4(r) \quad = \quad \frac{15\,r^4 - 11\,r^2 - 48\,r + 25}{15\,r^6}.$$

For $r = \infty$, $\rho_n(N_{PE}^r)$ is degenerate.

See Appendix Sections B.1 and B.2 for the derivation of mean and variance, respectively.

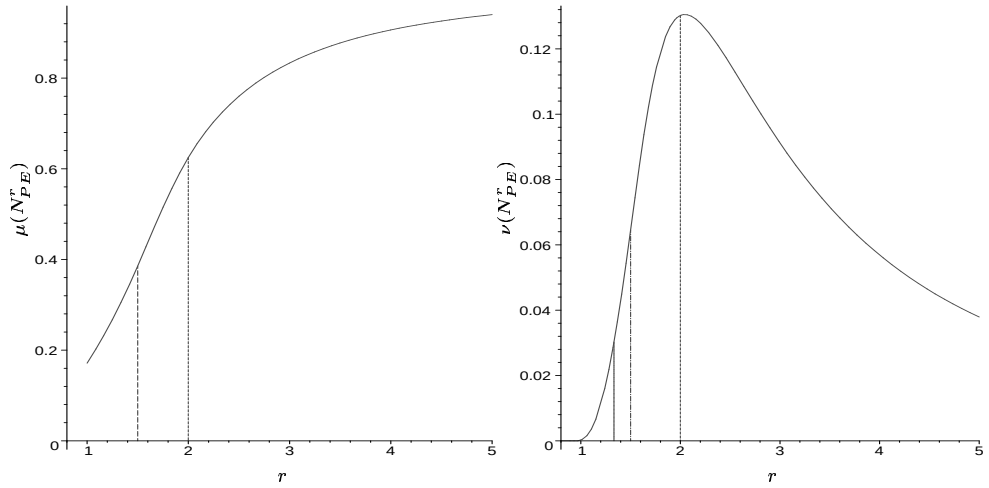


Figure 5.4.1: Asymptotic null mean $\mu(N_{PE}^r)$ (left) and variance $\nu(N_{PE}^r)$ (right) from Theorem 5.4.2 for $r \in [1, 5]$. The vertical lines indicate the endpoints of the intervals in the piecewise definition of the functions. Notice that the vertical axes are differently scaled.

Consider the forms of the mean and asymptotic variance functions, which are depicted in Figure 5.4.1. Note that $\mu(N_{PE}^r)$ is monotonically increasing in r, since $N_{PE}^r(x)$ gets larger with r for all $x \in R_{CM}(\mathsf{y}_j) \setminus \mathscr{R}_S(N_{PE}^r)$. In addition, $\mu(N_{PE}^r) \to 1$ as $r \to \infty$ (at rate $O\left(r^{-2}\right)$), since the digraph becomes complete asymptotically, which explains why $\rho_n(N_{PE}^r)$ becomes degenerate as $r \to \infty$, i.e., $\nu(N_{PE}^{r=\infty}) = 0$. Note also that $\mu(N_{PE}^r)$ is continuous, with $\mu(N_{PE}^{r=1}) = 37/216 \approx .1713$. By similar arguments, similar results hold for $N_{PE}^r(\cdot, M)$ with general M.

Regarding the asymptotic variance, note that $\nu(N_{PE}^r)$ is also continuous in r with $\lim_{r\to\infty} \nu(N_{PE}^r) = 0$ and $\nu(N_{PE}^{r=1}) = 34/58320 \approx .000583$ and observe that $\operatorname{argsup}_{r\geq 1} \nu(N_{PE}^r) \approx 2.045$ with $\sup_{r\geq 1} \nu(N_{PE}^r) \approx .1305$.

To illustrate the limiting distribution, $r = 2$ yields

$$\frac{\sqrt{n}\left(\rho_n(N_{PE}^{r=2}) - \mu(N_{PE}^{r=2})\right)}{\sqrt{\nu(N_{PE}^{r=2})}} = \sqrt{\frac{192\,n}{25}}\left(\rho_n(N_{PE}^{r=2}) - \frac{5}{8}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

or equivalently,

$$\rho_n(N_{PE}^{r=2}) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{5}{8}, \frac{25}{192\,n}\right).$$

The finite sample variance and skewness may be derived analytically in much the same way as was $\mathbf{Cov}\left[h_{12}(N_{PE}^r), h_{13}(N_{PE}^r)\right]$ for the asymptotic variance. In particular, the variance of $h_{12}(N_{PE}^r)$ is

$$\omega(N_{PE}^r) = \mathbf{Var}\left[h_{12}(N_{PE}^r)\right] = \omega_{1,1}(r)\,\mathbf{I}(r \in [1, 4/3)) + \omega_{1,2}(r)\,\mathbf{I}(r \in [4/3, 3/2)) + \omega_{1,3}(r)\,\mathbf{I}(r \in [3/2, 2)) + \omega_{1,4}(r)\,\mathbf{I}(r \in [2, \infty))$$

where

$$\omega_{1,1}(r) = \frac{-(1369\,r^8 + 4107\,r^7 + 902\,r^6 - 78084\ r^5 + 161784\ r^4 - 182736\,r^3 - 23328\,r^2 + 155520\,r - 55296)}{11664\ (r+2)(r+1)r^2},$$

$$\omega_{1,2}(r) = -\frac{1369\,r^7 + 4107\,r^6 + 9650\,r^5 - 98496\,r^4 + 132624\ r^3 - 79056\,r^2 - 57888\,r + 72576}{11664\ (r+2)(r+1)r},$$

$$\omega_{1,3}(r) = -\frac{r^{10} + 3\,r^9 - 62\,r^8 + 968\,r^6 - 1704\ r^5 - 1824\ r^4 + 5424\ r^3 - 1168\,r^2 - 3856\,r + 2208}{16\,(r+2)(r+1)r^4},$$

$$\omega_{1,4}(r) = \frac{3\,r^3 + 3\,r^2 + 3\,r - 13}{r^4(r+1)}.$$

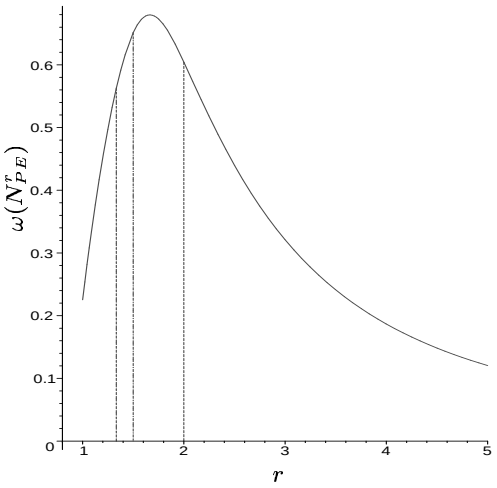


Figure 5.4.2: The plot of $\omega(N_{PE}^r) = \mathbf{Var}\left[h_{12}(N_{PE}^r)\right]$ as a function of r for $r \in [1, 5]$.

In Figure 5.4.2 is the graph of $\omega(N_{PE}^r)$ for $r \in [1, 5]$. Note that $\omega(r = 1) = 2627/11664 \approx .2252$, and $\lim_{r\to\infty} \omega(N_{PE}^r) = 0$ (at rate $O\left(r^{-2}\right)$), $\text{argsup}_{r\in[1,\infty)}\, \omega(N_{PE}^r) \approx 1.66$ with $\sup_{r\in[1,\infty)}$

$\omega(N_{PE}^r) \approx .6796$.

In fact, the exact distribution of $\rho_n(N_{PE}^r)$ is, in principle, available by successively conditioning on the values of X_i. Alas, while the joint distribution of $h_{12}(N_{PE}^r), h_{13}(N_{PE}^r)$ is available, the joint distribution of $\{h_{ij}(N_{PE}^r)\}_{1 \le i < j \le n}$, and hence the calculation for the exact distribution of $\rho_n(N_{PE}^r)$, is extraordinarily tedious and lengthy for even small values of n.

Figure 5.4.3 indicates that, for $r = 2$, the normal approximation is accurate even for small n (although kurtosis may be indicated for $n = 10$). Figure 5.4.4 demonstrates, however, that severe skewness obtains for small values of n and extreme values of r.

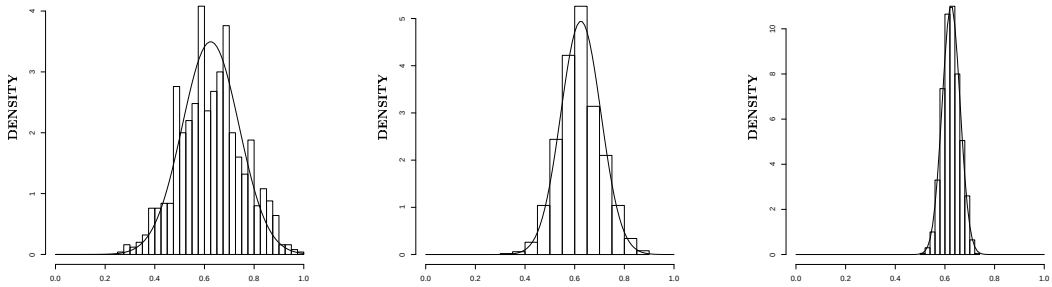


Figure 5.4.3: Depicted are the distributions of $\rho_n(N_{PE}^2) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{5}{8}, \frac{25}{192\,n}\right)$ for $10, 20, 100$ (left to right). Histograms are based on 1000 Monte Carlo replicates. Solid curves represent the approximating normal densities given in Theorem 5.4.2. Note that the vertical axes are differently scaled.

Letting $H_n(N_{PE}^r) = \sum_{i=1}^n h_{i,n+1}\big(N_{PE}^r\big)$, the exact distribution of $\rho_n(N_{PE}^r)$ can be written as the recurrence

$$(n+1)\,n\,\rho_{n+1}(N_{PE}^r) \overset{d}{=} n\,(n-1)\,\rho_n(N_{PE}^r) + H_n(N_{PE}^r)$$

by noting that the conditional random variable $H_n(N_{PE}^r)|X_{n+1}$ is the sum of n independent and identically distributed random variables. Alas, this calculation is also tedious for large n.

5.4.3 Asymptotic Normality Under the Alternatives

Asymptotic normality of relative density of the proximity catch digraphs under the alternative hypotheses of segregation and association can be established by the same method as under the

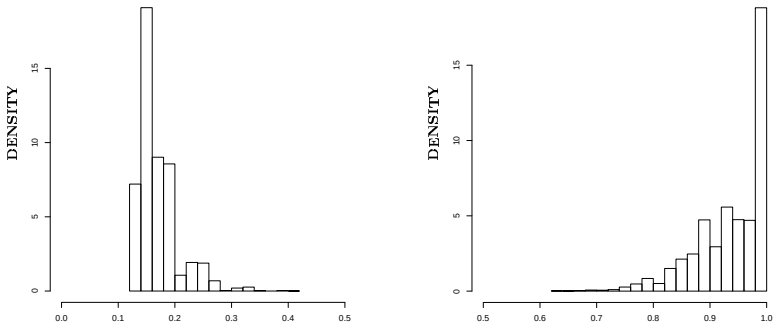


Figure 5.4.4: Depicted are the histograms for 10000 Monte Carlo replicates of $\rho_{10}(N^1_{PE})$ (left) and $\rho_{10}(N^5_{PE})$ (right) indicating severe small sample skewness for extreme values of r.

null hypothesis. Let $\mathbf{E}^S_\varepsilon[\cdot]$ ($\mathbf{E}^A_\varepsilon[\cdot]$) be the expectation with respect to the uniform distribution under the segregation (association) alternatives with $\varepsilon \in \left(0, \sqrt{3}/3\right)$.

Theorem 5.4.3. *Let $\mu_S(N^r_{PE}, \varepsilon)$ be the mean $\mathbf{E}^S_\varepsilon[h_{12}(N^r_{PE})]$, $\nu_S(N^r_{PE}, \varepsilon)$ be the covariance, $\mathbf{Cov}^S_\varepsilon[h_{12}(N^r_{PE}), h_{13}(N^r_{PE})]$ for $r \in [1, \infty]$, and $\varepsilon \in \left[0, \sqrt{3}/3\right)$ under H^S_ε. Then $\sqrt{n}\big(\rho_n(N^r_{PE}) - \mu_S(N^r_{PE}, \varepsilon)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_S(N^r_{PE}, \varepsilon))$ at the values of (r, ε) for which $\nu_S(N^r_{PE}, \varepsilon) > 0$. Likewise for H^A_ε.*

Sketch of Proof: Under the alternatives, i.e., $\varepsilon > 0$, $\rho_n(N^r_{PE})$ is a U-statistic with the same symmetric kernel $h_{ij}(N^r_{PE})$ as in the null case. Under H^S_ε, the mean $\mu_S(N^r_{PE}, \varepsilon) = \mathbf{E}^S_\varepsilon[\rho_n(N^r_{PE})] = \mathbf{E}^S_\varepsilon[h_{12}(N^r_{PE})]/2$, now a function of both r and ε, is again in $[0, 1]$. The asymptotic variance $\nu_S(N^r_{PE}, \varepsilon) = \mathbf{Cov}^S_\varepsilon[h_{12}(N^r_{PE}), h_{13}(N^r_{PE})]$, also a function of both r and ε, is bounded above by $1/4$, as before. Thus asymptotic normality obtains provided $\nu_S(N^r_{PE}, \varepsilon) > 0$; otherwise $\rho_n(N^r_{PE})$ is degenerate. Likewise for H^A_ε.

The explicit forms of $\mu_S(N^r_{PE}, \varepsilon)$ and $\mu_A(N^r_{PE}, \varepsilon)$ are given, defined piecewise, in Appendix Section B.4. Sample values of $\mu_S(N^r_{PE}, \varepsilon)$, $\nu_S(N^r_{PE}, \varepsilon)$, and $\mu_A(N^r_{PE}, \varepsilon)$, $\nu_A(N^r_{PE}, \varepsilon)$ are given in Section 5.4.9.1 under segregation with $\varepsilon \in \left\{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\right\}$ and in Section 5.4.9.2 under association with $\varepsilon \in \left\{5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21\right\}$. Note that under H^S_ε,

$$\nu_S(N^r_{PE}, \varepsilon) > 0 \text{ iff } (r, \varepsilon) \in \left[1, \sqrt{3}/(2\varepsilon)\right) \times \left(0, \sqrt{3}/4\right] \cup \left[1, \sqrt{3}/\varepsilon - 2\right) \times \left(\sqrt{3}/4, \sqrt{3}/3\right),$$

and under H_ε^A,

$$\nu_A(N_{PE}^r, \varepsilon) > 0 \text{ iff } (r, \varepsilon) \in (1, \infty) \times \left(0, \sqrt{3}/3\right) \cup \{1\} \times \left(0, \sqrt{3}/12\right). \blacksquare$$

Notice that under the association alternatives any $r \in (1, \infty)$ yields asymptotic normality for all $\varepsilon \in \left(0, \sqrt{3}/3\right)$, while under the segregation alternatives only $r = 1$ yields this universal asymptotic normality.

5.4.4 The Test and the Analysis

The relative density of the proximity catch digraph is a test statistic for the segregation/association alternative; rejecting for extreme values of $\rho_n(N_{PE}^r)$ is appropriate since under segregation we expect $\rho_n(N_{PE}^r)$ to be large, while under association we expect $\rho_n(N_{PE}^r)$ to be small. Using the test statistic

$$R(N_{PE}^r) = \frac{\sqrt{n}\left(\rho_n(N_{PE}^r) - \mu(N_{PE}^r)\right)}{\sqrt{\nu(N_{PE}^r)}}, \tag{5.4.4}$$

the asymptotic critical value for the one-sided level α test against segregation is given by $z_\alpha = \Phi^{-1}(1-\alpha)$. Against segregation, the test rejects for $R(N_{PE}^r) > z_\alpha$ and against association, the test rejects for $R(N_{PE}^r) < z_{1-\alpha}$.

5.4.5 Consistency

Theorem 5.4.4. *The test against H_ε^S which rejects for $R(N_{PE}^r) > z_{1-\alpha}$ and the test against H_ε^A which rejects for $R(N_{PE}^r) < z_\alpha$ are consistent for $r \in [1, \infty)$ and $\varepsilon \in \left(0, \sqrt{3}/3\right)$.*

Proof: Since the variance of the asymptotically normal test statistic, under both the null and the alternatives, converges to 0 as $n \to \infty$ (or is degenerate), it remains to show that the mean under the null, $\mu(N_{PE}^r) = \mathbf{E}\left[\rho_n(N_{PE}^r)\right]$, is less than (greater than) the mean under the alternative, $\mu_S(N_{PE}^r, \varepsilon) = \mathbf{E}_\varepsilon^S[\rho_n(N_{PE}^r)]$ against segregation ($\mu_A(N_{PE}^r, \varepsilon) = \mathbf{E}_\varepsilon^A[\rho_n(N_{PE}^r)]$ against association) for $\varepsilon > 0$. Whence it will follow that power converges to 1 as $n \to \infty$.

Detailed analysis of $\mu_S(N_{PE}^r, \varepsilon)$ in Appendix Section B.4.1 indicates that under segregation $\mu_S(N_{PE}^r, \varepsilon) > \mu(N_{PE}^r)$ for all $\varepsilon > 0$ and $r \in [1, \infty)$. Likewise, detailed analysis of $\mu_A(N_{PE}^r, \varepsilon)$ in Appendix Section B.4.2 indicates that under association $\mu_A(N_{PE}^r, \varepsilon) < \mu(N_{PE}^r)$ for all $\varepsilon > 0$ and $r \in [1, \infty)$. Hence the desired result follows for both alternatives. $\blacksquare$

Remark 5.4.5. In fact, the analyses of $\mu_S(N_{PE}^r, \varepsilon)$ and $\mu_A(N_{PE}^r, \varepsilon)$ under the alternatives reveal more than what is required for consistency. Under segregation, the analysis indicates that $\mu_S(N_{PE}^r, \varepsilon_1) < \mu_S(N_{PE}^r, \varepsilon_2)$ for $\varepsilon_1 < \varepsilon_2$. Likewise, under association, the analysis indicates that $\mu_A(N_{PE}^r, \varepsilon_1) > \mu_A(N_{PE}^r, \varepsilon_2)$ for $\varepsilon_1 < \varepsilon_2$. □

Remark 5.4.6. Note that the consistency holds for segregation (association) for all $F_\varepsilon \in \mathscr{F}^S$ ($F_\varepsilon \in \mathscr{F}^A$) (see Section 4.6.3 for $\mathscr{F}^S$ and $\mathscr{F}^A$). Moreover, let $\mathscr{G}^S$ ($\mathscr{G}^A$) be the family of continuous distributions, G_ε, whose support is $S_\varepsilon \subsetneq T(\mathcal{Y})$ with $d(S_\varepsilon, \mathcal{Y}) = \min_{(x,y)\in S_\varepsilon \times \mathcal{Y}} d(x,y) = \varepsilon$ $\left(d(S_\varepsilon, \mathcal{Y}) = \min_{(x,y)\in S_\varepsilon \times \mathcal{Y}} d(x,y) = \sqrt{3}/3 - \varepsilon\right)$ and whose densities are nondecreasing along straight lines towards the center of mass (the vertices). Then $\varepsilon > 0$ for $G_\varepsilon \in \mathscr{G}^S$ ($G_\varepsilon \in \mathscr{G}^A$) implies segregation (association). The consistency holds for the alternatives with $X_i \overset{iid}{\sim} G_\varepsilon$ with $\varepsilon > 0$. □

5.4.6 Monte Carlo Power Analysis Under Segregation

In segregation alternatives with $\varepsilon > 0$, we implement the above described Monte Carlo experiment for various values of r (for which the associated relative density $\rho_n(N_{PE}^r)$ is non-degenerate). Recall that $\rho_n(N_{PE}^r)$ is degenerate for large r at each $\varepsilon > 0$. In particular, with $\varepsilon = \sqrt{3}/8$, $\rho_n(N_{PE}^r)$ is degenarate for $r \geq 4$, with $\varepsilon = \sqrt{3}/4$, $\rho_n(N_{PE}^r)$ is degenarate for $r \geq 2$, and with $\varepsilon = 2\sqrt{3}/7$, $\rho_n(N_{PE}^r)$ is degenarate for $r \geq 3/2$.

Let $\rho_k(n, N_{PE}^r)$ be the empirical relative density for experiment k and $\rho_{j:N}(n, N_{PE}^r)$ be the j^{th} (ordered) empirical relative density for $j = 1, \ldots, N$. Then for each r value, we estimate the empirical critical value $\widehat{c}_n^S(r) := \rho_{\lceil (1-\alpha)\, N \rceil : N}(n, N_{PE}^r)$ and the empirical significance level $\widehat{\alpha}_{mc}^S(n,r) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}\left(\rho_j(n, N_{PE}^r) > \widehat{c}_n^S(r)\right)$ under H_0 and the empirical power $\widehat{\beta}_{mc}^S(n,r,\varepsilon) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}\left(\rho_j(n, N_{PE}^r) > \widehat{c}_n^S(r)\right)$ under H_ε^S with $\varepsilon \in \left\{\sqrt{3}/8,\, \sqrt{3}/4,\, 2\sqrt{3}/7\right\}$.

For segregation with $\varepsilon = \sqrt{3}/8 \approx .2165$, we run the Monte Carlo experiments for eight r values: 1, 11/10, 6/5, 4/3, $\sqrt{2}$, 3/2, 2, and 3. In Figure 5.4.5 are the kernel density estimates for the null case and the segregation alternative for the eight r values with $\varepsilon = \sqrt{3}/8$, $n = 10$, and $N = 10000$. Observe that under both H_0 and $H_{\sqrt{3}/8}^S$, kernel density estimates are skewed right for $r \in \{1,\, 11/10\}$, (with skewness increasing as r gets smaller) and kernel density estimates are almost symmetric for $r \in \left\{6/5,\, 4/3,\, \sqrt{2}, 3/2,\, 2\right\}$, with most symmetry occurring at $r = 3/2$, kernel density estimate is skewed left for $r = 3$ (with skewness increasing as r gets larger).

The empirical critical values, empirical significance levels, and empirical power estimates

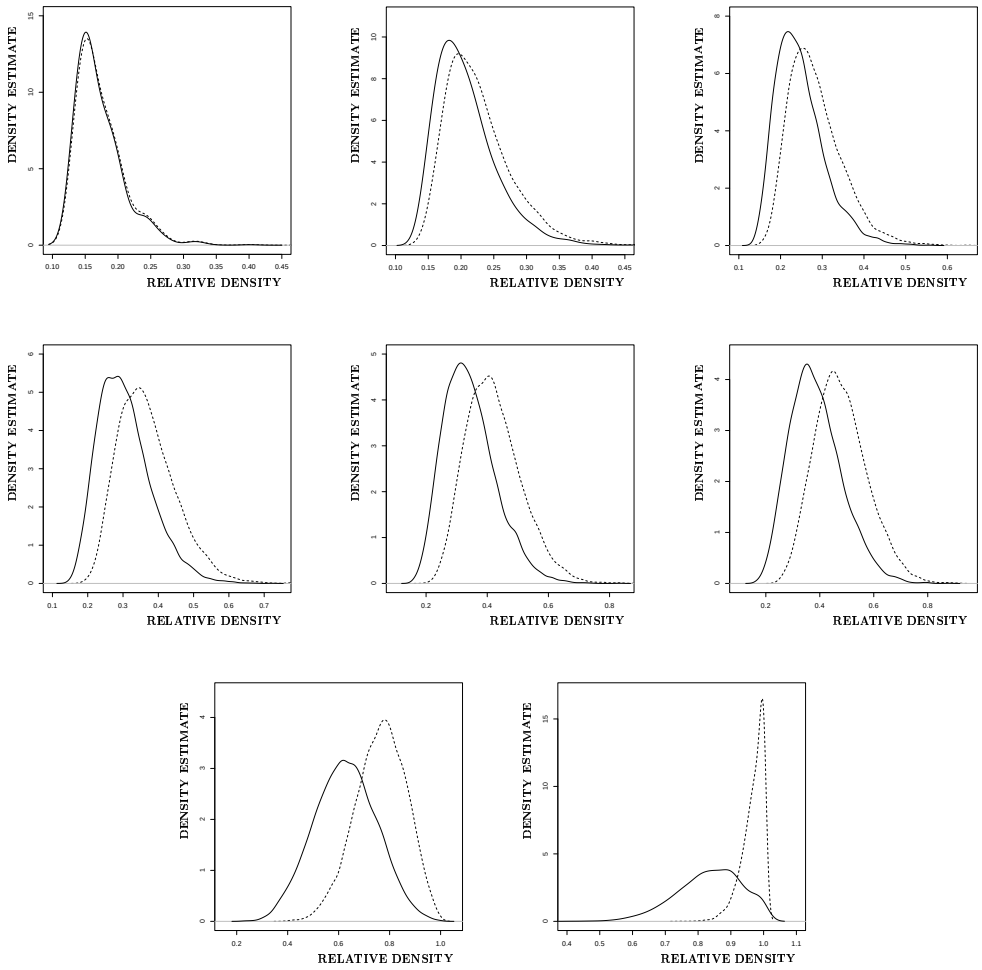


Figure 5.4.5: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/8}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2, 3\}$ (left-to-right).

under $H^S_{\sqrt{3}/8}$ are presented in Table 5.4.1.

In Figure 5.4.6, we present a Monte Carlo investigation against the segregation alternative $H^S_{\sqrt{3}/8}$ for $r = 11/10$, and $n = 10$, $N = 10000$ (left), $n = 100$, $N = 1000$ (right). With $n = 10$, the null and alternative probability density functions for $\rho_{10}(11/10)$ are very similar, implying small power (10000 Monte Carlo replicates yield $\widehat{\beta}^S_{mc}\left(11/10,\ 10,\ \sqrt{3}/8\right) = 0.0787$, $\widehat{c}^S_n(11/10) = 0.1\bar{5}$, and $\widehat{\alpha}^S_{mc}(11/10,\ 10) = 0.0484$). With $n = 100$, there is more separation between null and alternative probability density functions; for this case, 1000 Monte Carlo replicates yield $\widehat{\beta}^S_{mc}\left(11/10,\ 100,\ \sqrt{3}/8\right) = 0.77$, $\widehat{c}^S_n(11/10) = 0.2203$, and $\widehat{\alpha}^S_{mc}(11/10, 100) = 0.05$. Notice also that the probability density functions are more skewed for $n = 10$, while approximate normality

r	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3
$\widehat{c}_n^S(r)$	$0.2\overline{4}$	.3	$.3\overline{5}$	$.\overline{4}$	.5	$.\overline{5}$	$.8\overline{2}$	$0.9\overline{8}$
$\widehat{\alpha}_{mc}^S(r, 10)$	.0324	.0403	.0484	.0442	.0446	.0492	.049	.0389
$\widehat{\beta}_{mc}^S\left(r, 10, \sqrt{3}/8\right)$	.0381	.0787	.122	.1571	.1719	.1955	.2791	.2901

Table 5.4.1: The empirical critical values, empirical significance levels, and empirical power estimates under $H^S_{\sqrt{3}/8}$, $N = 10000$, and $n = 10$ at $\alpha = .05$.

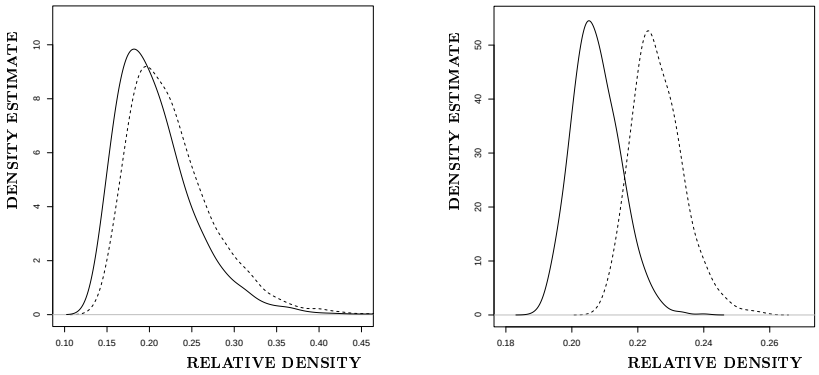


Figure 5.4.6: Two Monte Carlo experiments against the segregation alternative $H^S_{\sqrt{3}/8}$. Depicted are kernel density estimates for $\rho_n(N_{PE}^{11/10})$ for $n = 10$ (left) and $n = 100$ (right) under the null (solid) and alternative (dashed).

holds for $n = 100$.

For segregation with $\varepsilon = \sqrt{3}/4 \approx .433$, we run the Monte Carlo experiments for six r values: 1, 11/10, 6/5, 4/3, $\sqrt{2}$, 3/2. In Figure 5.4.7, are the kernel density estimates for the null case and the segregation alternative for the six r values with $\varepsilon = \sqrt{3}/4$, $n = 10$, and $N = 10000$. Observe that under $H^S_{\sqrt{3}/4}$, kernel density estimate is skewed right for $r = 1$ and kernel density estimates are almost symmetric for $r \in \{11/10, 6/5, 4/3, \sqrt{2}\}$, with most symmetry occurring at $r = 4/3$, kernel density estimate is skewed left for $r = 3/2$.

The empirical critical values, empirical significance levels, and empirical power estimates under $H^S_{\sqrt{3}/4}$ are presented in Table 5.4.2.

For segregation with $\varepsilon = 2\sqrt{3}/7 \approx .495$, we run the Monte Carlo experiments for six r values, 1, 21/20, 11/10, 6/5, 4/3, $\sqrt{2}$. In Figure 5.4.8, are the kernel density estimates for the null case and the segregation alternative with $\varepsilon = 2\sqrt{3}/7$ for the six r values with $n = 10$ and $N = 10000$. Observe that under $H^S_{2\sqrt{3}/7}$, kernel density estimate is skewed right for $r = 1$ and kernel density estimates are almost symmetric for $r \in \{21/20, 11/10, 6/5\}$ with most symmetry

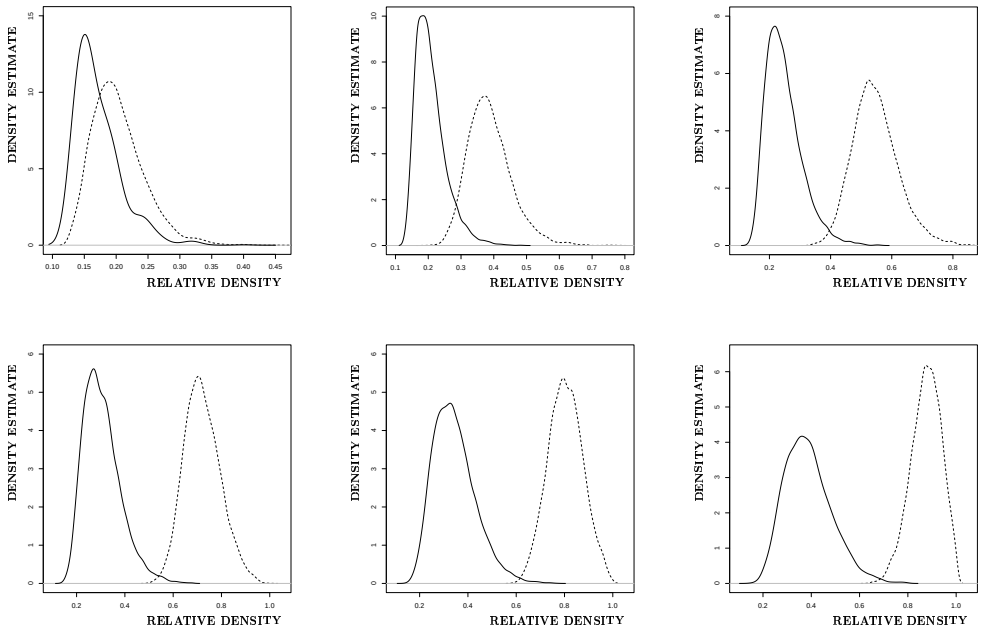


Figure 5.4.7: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/4}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2\}$ (left-to-right).

r	1	11/10	6/5	4/3	$\sqrt{2}$	3/2
$\widehat{c}^S_n(r)$	$.2\overline{4}$	.3	$.3\overline{5}$	$.\overline{4}$	.5	$.\overline{5}$
$\widehat{\alpha}^S_{mc}(r, 10)$	.0318	.0411	.0479	.0484	.0481	.043
$\widehat{\beta}^S_{mc}\left(r, 10, \sqrt{3}/4\right)$	.1247	.9138	.998	1.0	1.0	1.0

Table 5.4.2: The empirical critical values, empirical significance levels, and empirical power estimates under $H^S_{\sqrt{3}/4}$, $N = 10000$, and $n = 10$ at $\alpha = .05$.

occurring at $r = 6/5$, kernel density estimate is skewed left for $r = \sqrt{2}$.

The empirical critical values, empirical significance levels, and empirical power estimates under $H^S_{2\sqrt{3}/7}$ are presented in Table 5.4.3.

We also plot the empirical power as a function of r in Figure 5.4.9. Let $r^*_S(\varepsilon)$ be the value of r at which maximum Monte Carlo power estimate occurs, then $r^*_S(\sqrt{3}/8) = 3$. Furthermore, Monte Carlo power estimate increases as r gets larger and then decreases, due to the magnitude of r and n. Because for small n and large r the critical value is approximately 1 under H_0, as we get a complete digraph with high probability.

Furthermore, $r^*_S(\sqrt{3}/4) \in \{4/3,\ \sqrt{2},\ 3/2\}$ and $r^*_S(2\sqrt{3}/7) \in \{11/10,\ 6/5,\ 4/3,\ \sqrt{2}\}$. Monte Carlo power estimates increase as r gets larger. The phenomenon happened above for $\varepsilon = \sqrt{3}/8$

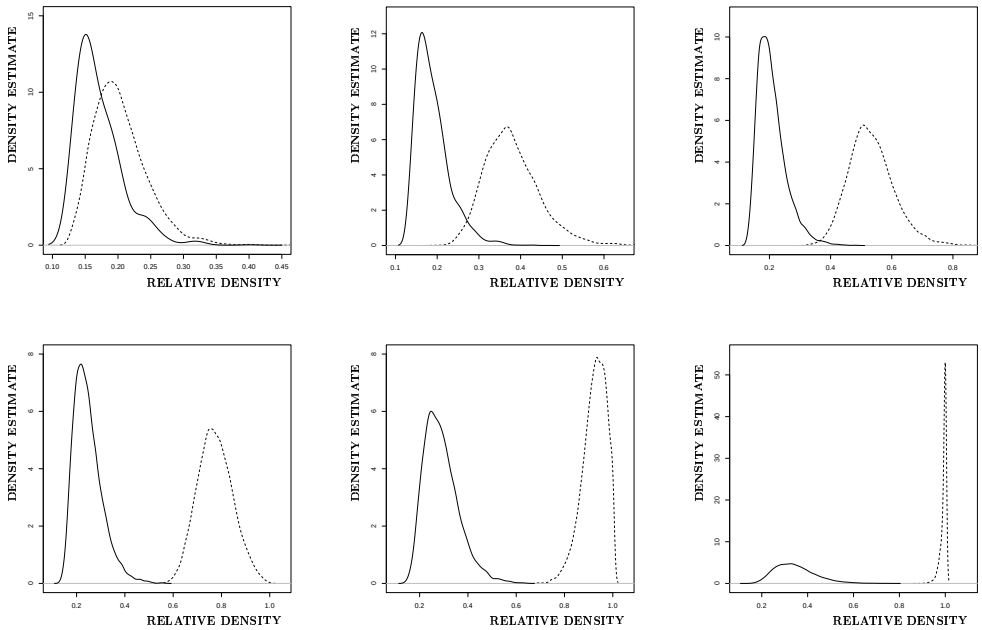


Figure 5.4.8: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{2\sqrt{3}/7}$ (dashed) for $r \in \{1,\ 21/20,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2}\}$ (left-to-right).

r	1	21/20	11/10	6/5	4/3	$\sqrt{2}$
$\widehat{c}^S_n(r)$	$.2\overline{4}$	$.2\overline{8}$	.3	$.3\overline{5}$	$.4\overline{2}$	.5
$\widehat{\alpha}^S_{mc}(r, 10)$	.0318	.0447	.0411	.0479	.0477	.0481
$\widehat{\beta}^S_{mc}\left(r, 10, 2\sqrt{3}/7\right)$	.1247	.9728	1.0	1.0	1.0	1.0

Table 5.4.3: The empirical critical values, empirical significance levels, and empirical power estimates under $H^S_{2\sqrt{3}/7}$, $N = 10000$, and $n = 10$ at $\alpha = .05$.

does not occur, because r values are not large enough to yield complete digraphs under H_0 with high probability.

For a given alternative and sample size, we may consider analyzing the power of the test — using the asymptotic critical value— as a function of the proximity factor r. Let $R_j(N^r_{PE}) := \frac{\sqrt{n}\left(\rho_j(N^r_{PE}, n) - \mu(N^r_{PE})\right)}{\sqrt{\nu(N^r_{PE})}}$ be the standardized relative density for experiment j with sample size n for $j = 1, 2, \ldots, N$. For any $r \in [1, \infty)$, the level α asymptotic critical value is $\mu(N^r_{PE}) + z_{(1-\alpha)} \cdot \sqrt{\nu(N^r_{PE})/n}$. We estimate the empirical power as $\widehat{\beta}^S_n(r, \varepsilon) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}\left(R_j(N^r_{PE}) > z_{1-\alpha}\right)$. In Figure 5.4.10, we present a Monte Carlo investigation of power $\widehat{\beta}^S_n(r, \varepsilon)$ against $H^S_{\sqrt{3}/8}$, $H^S_{\sqrt{3}/4}$, and $H^S_{2\sqrt{3}/7}$ as a function of r for $n = 10$. The empirical significance level is $\widehat{\alpha}_S(r, n) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}\left(R_j(N^r_{PE}) > z_{1-\alpha} | H_0\right)$. Then $\widehat{\alpha}_S(r, 10)$ is about .05 for $r \in \{2, 3\}$ and the empirical

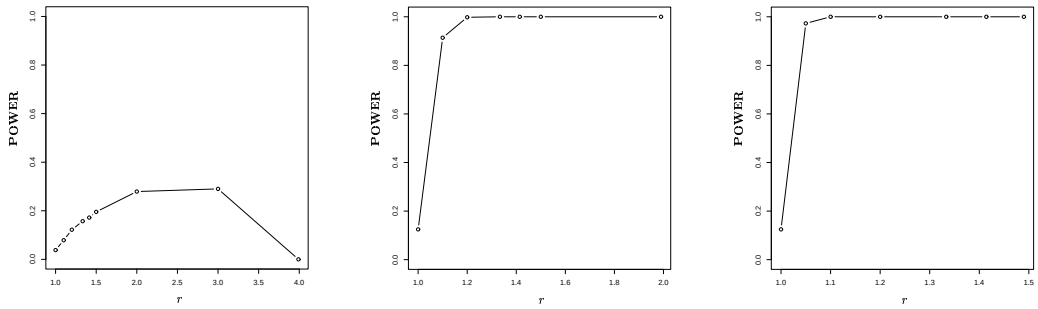


Figure 5.4.9: Monte Carlo power using the empirical critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of r for $n = 10$.

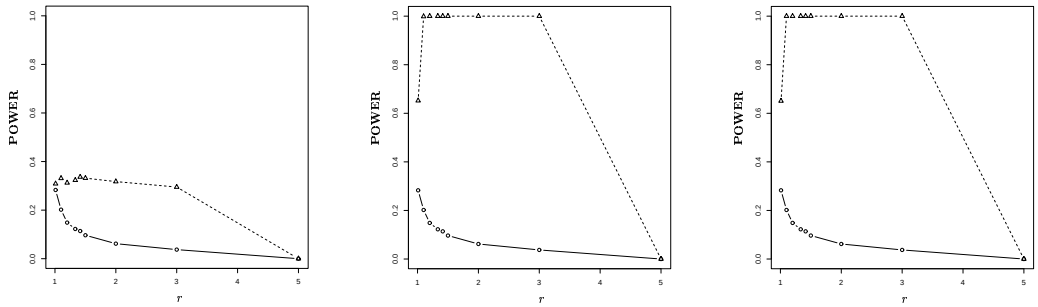


Figure 5.4.10: Monte Carlo power using the asymptotic critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of r for $n = 10$. The circles represent the empirical significance levels while triangles represent the empirical power values.

powers are $\widehat{\beta}^S_{10}\left(r, \sqrt{3}/8\right) \approx .35$ and $\widehat{\beta}^S_{10}(r, \varepsilon) = 1$ for $\varepsilon \in \left\{\sqrt{3}/4, 2\sqrt{3}/7\right\}$. So, for small sample sizes, moderate values of r are more appropriate for normal approximation, as they yield the desired significance level and have the property that more severe segregation yields higher power at each r.

The empirical significance levels and empirical power values $\widehat{\beta}^S_n(r, \varepsilon)$ under H^S_ε for $\varepsilon \in \left\{\sqrt{3}/8,\ \sqrt{3}/4,\ 2\sqrt{3}/7\right\}$ are presented in Table 5.4.4. Note that even for $n = 10$, the plots of the empirical power $\widehat{\beta}^S_n(r, \varepsilon)$ resemble the curves of the asymptotic power function $\Pi_S(r)$ in Section 5.4.10.

5.4.7 Monte Carlo Power Analysis Under Association

In association alternatives with $\varepsilon > 0$, we implement the Monte Carlo experiment for $r \in \left\{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2,\ 3,\ 5,\ 10\right\}$. Then for each r value, we estimate the empirical

r	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5
$\widehat{\alpha}_S(r,n)$	.2829	.2019	.1486	.1224	.1139	.0966	.0619	.0374	.000
$\widehat{\beta}_n^S\left(r,\sqrt{3}/8\right)$	.3086	.3309	.3123	.3233	.3365	.3317	.3175	.2950	.0000
$\widehat{\beta}_n^S\left(r,\sqrt{3}/4\right)$	.6519	.9985	1.000	1.000	1.000	1.000	1.000	1.000	.0000
$\widehat{\beta}_n^S\left(r,2\sqrt{3}/7\right)$	.6508	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.0000

Table 5.4.4: The empirical significance level and empirical power values under H_ε^S for $\varepsilon \in \left\{\sqrt{3}/8,\ \sqrt{3}/4,\ 2\sqrt{3}/7\right\}$, $N = 10000$, and $n = 10$ at $\alpha = .05$.

critical value $\widehat{c}_n^A(r) := \rho_{\lfloor \alpha N \rfloor:N}(N_{PE}^r, n)$ where $\lfloor\cdot\rfloor$ is the floor functional, and the empirical significance level $\widehat{\alpha}_{mc}^A(r,n) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}\left(\rho_j(N_{PE}^r,n) < \widehat{c}_n^A(r)\right)$ under H_0 and the empirical power $\widehat{\beta}_{mc}^A(n,r,\varepsilon) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}\left(\rho_j(r,n) < \widehat{c}_n^A(r)\right)$. We implement the Monte Carlo simulation for $\varepsilon \in \left\{5\sqrt{3}/24,\ \sqrt{3}/12,\ \sqrt{3}/21\right\}$.

The empirical critical values, empirical significance levels, and empirical power estimates under H_ε^A are presented in Table 5.4.5.

For association with $\varepsilon = 5\sqrt{3}/24 \approx .36$, in Figure 5.4.11, are the kernel density estimates for the null case and the segregation alternative for the ten r values with $n = 10$ and $N = 10000$. Observe that, under H_0, kernel density estimates are skewed right for $r \in \{1,\ 11/10\}$, (with skewness increasing as r gets smaller) and kernel density estimates are almost symmetric for $r \in \left\{6/5,\ 4/3,\ \sqrt{2}, 3/2,\ 2\right\}$, with most symmetry occurring at $r = 3/2$, kernel density estimates are skewed left for $r \in \{3,\ 5,\ 10\}$, (with skewness increasing as r gets larger). Under $H_{5\sqrt{3}/24}^A$, kernel density estimates are skewed right for $r \in \left\{1,\ 11/10,\ 6/5,\ 4/3,\ 3/2,\ 2,\ 3\right\}$, (with skewness increasing as r gets smaller) and kernel density estimate is almost symmetric for $r = 5$, kernel density estimate is skewed left for $r = 10$.

For association with $\varepsilon = \sqrt{3}/12 \approx .144$, in Figure 5.4.12, are the kernel density estimates for the null case and the segregation alternative for the ten r values with $n = 10$ and $N = 10000$. Observe that under $H_{\sqrt{3}/12}^A$, kernel density estimates are skewed right for $r \in \{1,\ 11/10,\ 6/5,\ 4/3\}$, (with skewness increasing as r gets smaller) and kernel density estimates are almost symmetric for $r \in \{\sqrt{2}, 3/2,\ 2\}$, with most symmetry occurring at $r = 2$, kernel density estimates are skewed left for $r \in \{3,\ 5,\ 10\}$, (with skewness increasing as r gets larger).

Note also that for $r = 11/10$ with $n = 10$ and $N = 1000$, the kernel density estimates are very similar, implying small power. With $N = 10000$, $\widehat{\beta}_{mc}^A\left(11/10, 10, \sqrt{3}/12\right) = 0.0921$, $\widehat{c}_n^A(11/10) = 0.1\bar{5}$, and $\widehat{\alpha}_{mc}^A(11/10, 10) = 0.0484$. See Figure 5.4.13. Note that for large n, there is more separation between null and alternative kernel densities, which implies higher

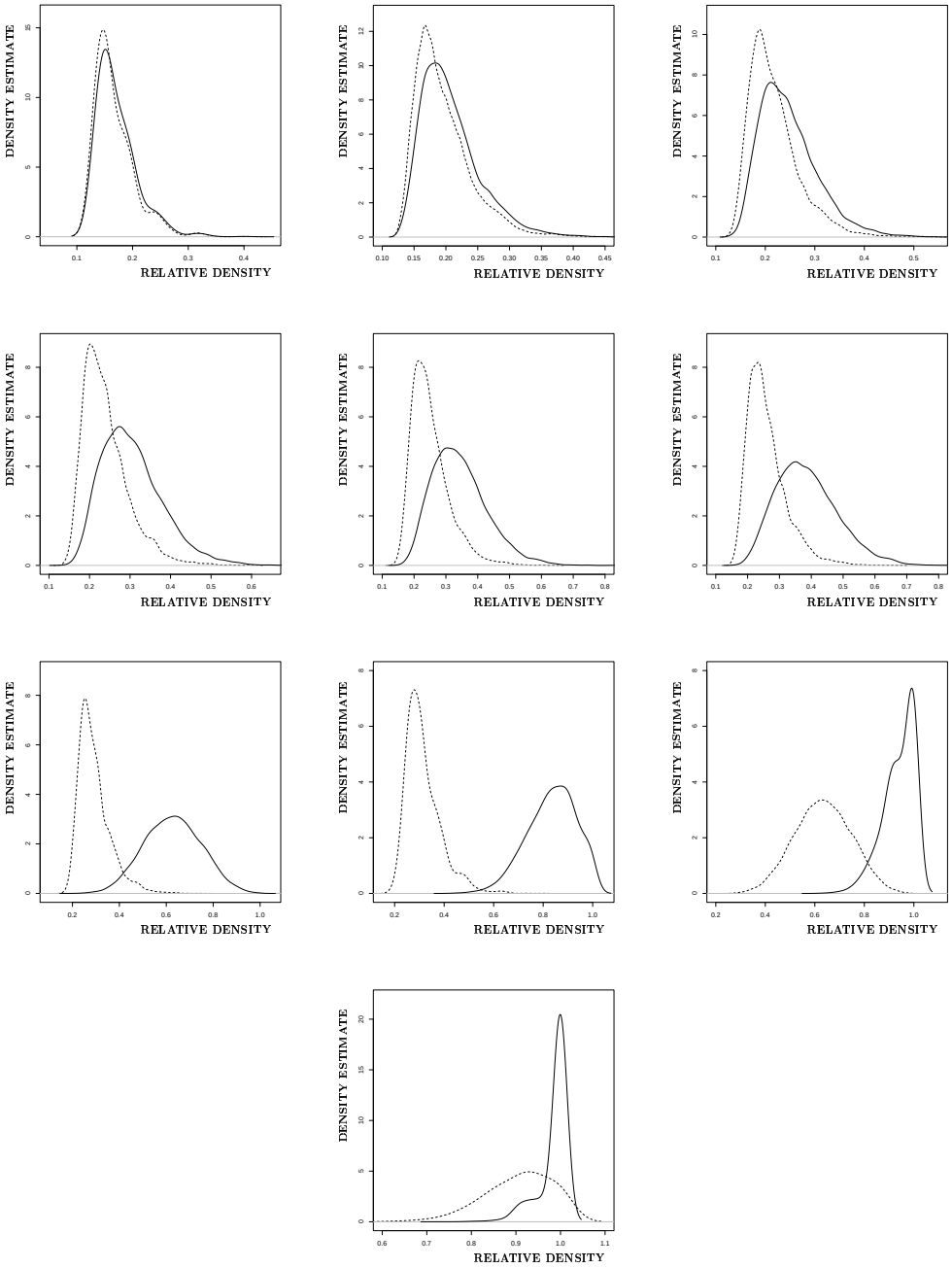


Figure 5.4.11: Kernel density estimates for the null (solid) and the association alternative $H^A_{5\sqrt{3}/24}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2,\ 3,\ 5,\ 10\}$ (left-to-right).

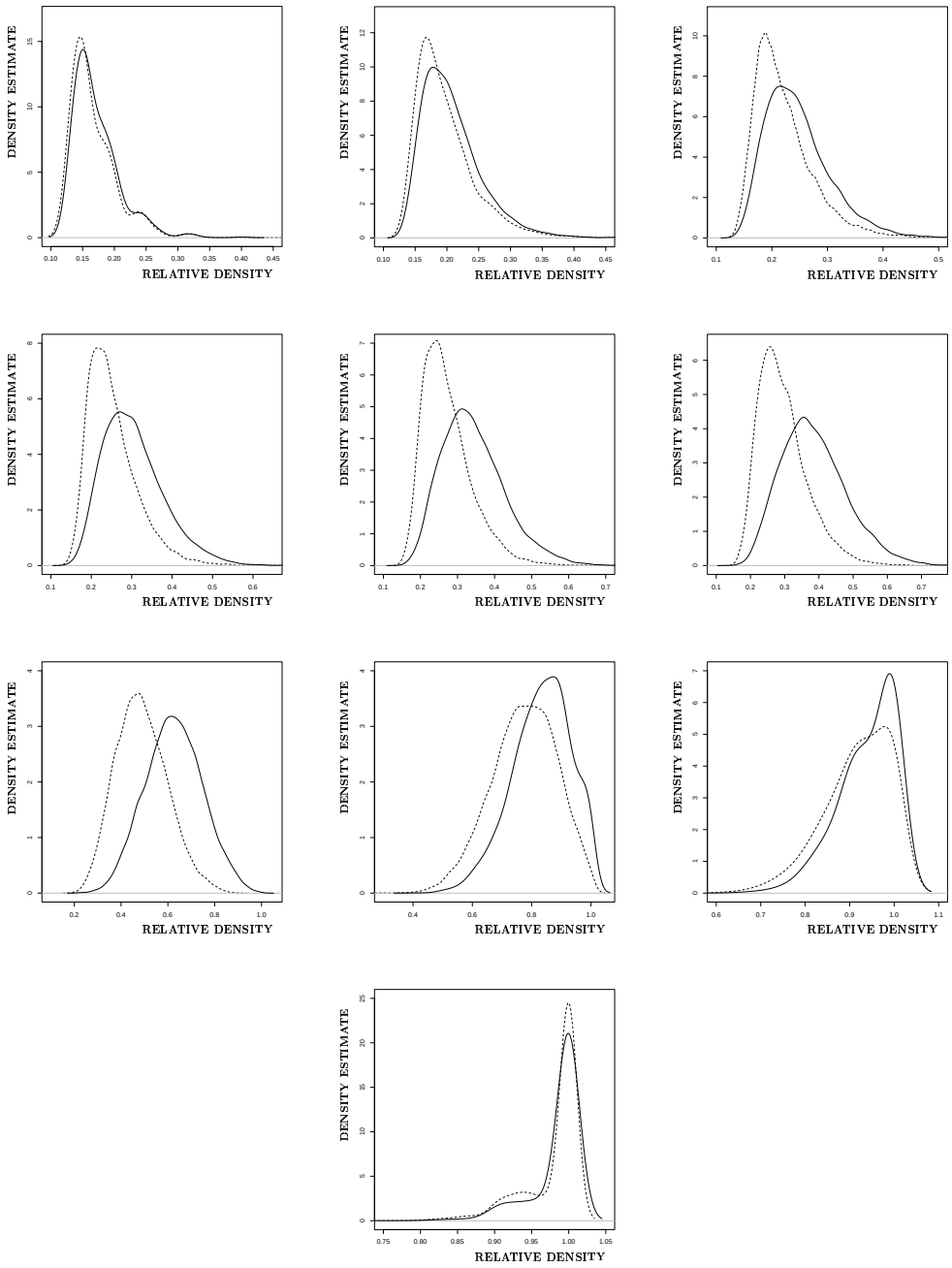


Figure 5.4.12: Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/12}$ (dashed) for $r \in \{1, 11/10, 6/5, 4/3, \sqrt{2}, 3/2, 2, 3, 5, 10\}$ (left-to-right).

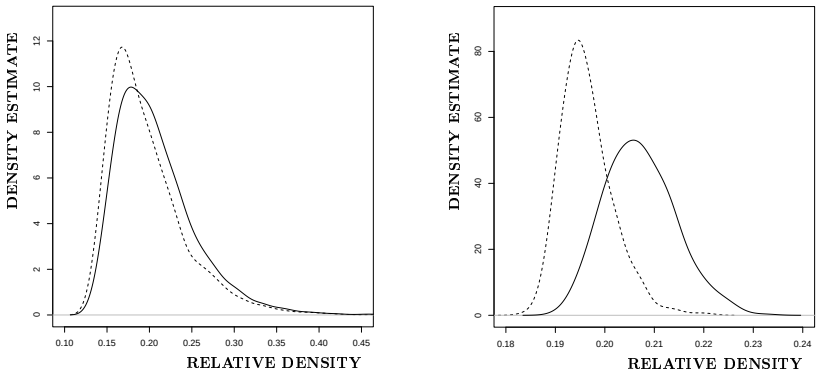


Figure 5.4.13: Two Monte Carlo experiments against the association alternative $H^A_{\sqrt{3}/12}$. Depicted are kernel density estimates for $\rho_n(N^{11/10}_{PE})$ for $n = 10$ (left) and $n = 100$ (right) under the null (solid) and alternative (dashed).

power. With $n = 100$, $N = 1000$ and get $\widehat{c}^A_n(11/10) = 0.1963$, $\widehat{\alpha}^A_{mc}(11/10, 100) = 0.049$, and $\widehat{\beta}^A_{mc}\left(11/10, 100, \sqrt{3}/12\right) = 0.56$.

For association with $\varepsilon = \sqrt{3}/21 \approx .0825$, in Figure 5.4.14, are the kernel density estimates for the null case and the segregation alternative for the ten r values with $n = 10$ and $N = 10000$. Observe that under $H^A_{\sqrt{3}/21}$, kernel density estimates are skewed right for $r \in \{1, 11/10, 6/5, 4/3\}$, (with skewness increasing as r gets smaller) and kernel density estimates are almost symmetric for $r \in \{\sqrt{2}, 3/2, 2\}$, with most symmetry occurring at $r = 3/2$, kernel density estimates are skewed left for $r \in \{3, 5, 10\}$, (with skewness increasing as r gets larger).

We also plot the empirical power as a function of r in Figure 5.4.15. Let $r^*_A(\varepsilon)$ be the value at which maximum Monte Carlo power estimate occurs. Then $r^*_A\left(5\sqrt{3}/24\right) = 3$, $r^*_A\left(\sqrt{3}/12\right) = 2$, and for $r^*_A\left(\sqrt{3}/21\right) = 3/2$. Notice that the more severe the association the larger the value of $r^*_A(\varepsilon)$. Based on the analysis of the Monte Carlo power estimates, we suggest moderate r values for moderate association.

We also estimate the power using the asymptotic critical value in association alternatives for various values of r. For each r value, the level α asymptotic critical value is $\mu(N^r_{PE}) + z_\alpha \cdot \sqrt{\nu(N^r_{PE})/n}$. We estimate the empirical power as $\widehat{\beta}^A_n(r, \varepsilon) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}(R_j(N^r_{PE}) < z_\alpha)$.

In Figure 5.4.16, we present a Monte Carlo investigation of power against $H^A_{\sqrt{3}/21}$, $H^A_{\sqrt{3}/12}$, and $H^A_{5\sqrt{3}/24}$ as a function of r for $n = 10$. The empirical significance level is $\widehat{\alpha}_A(r, n) := \frac{1}{N}\sum_{j=1}^N \mathbf{I}(R_j(N^r_{PE}) < z_\alpha | H_0)$. Then $\widehat{\alpha}_A(r, 10)$, is about .05 for $r \in \{\sqrt{2}, 3/2, 2, 3, 5\}$ which

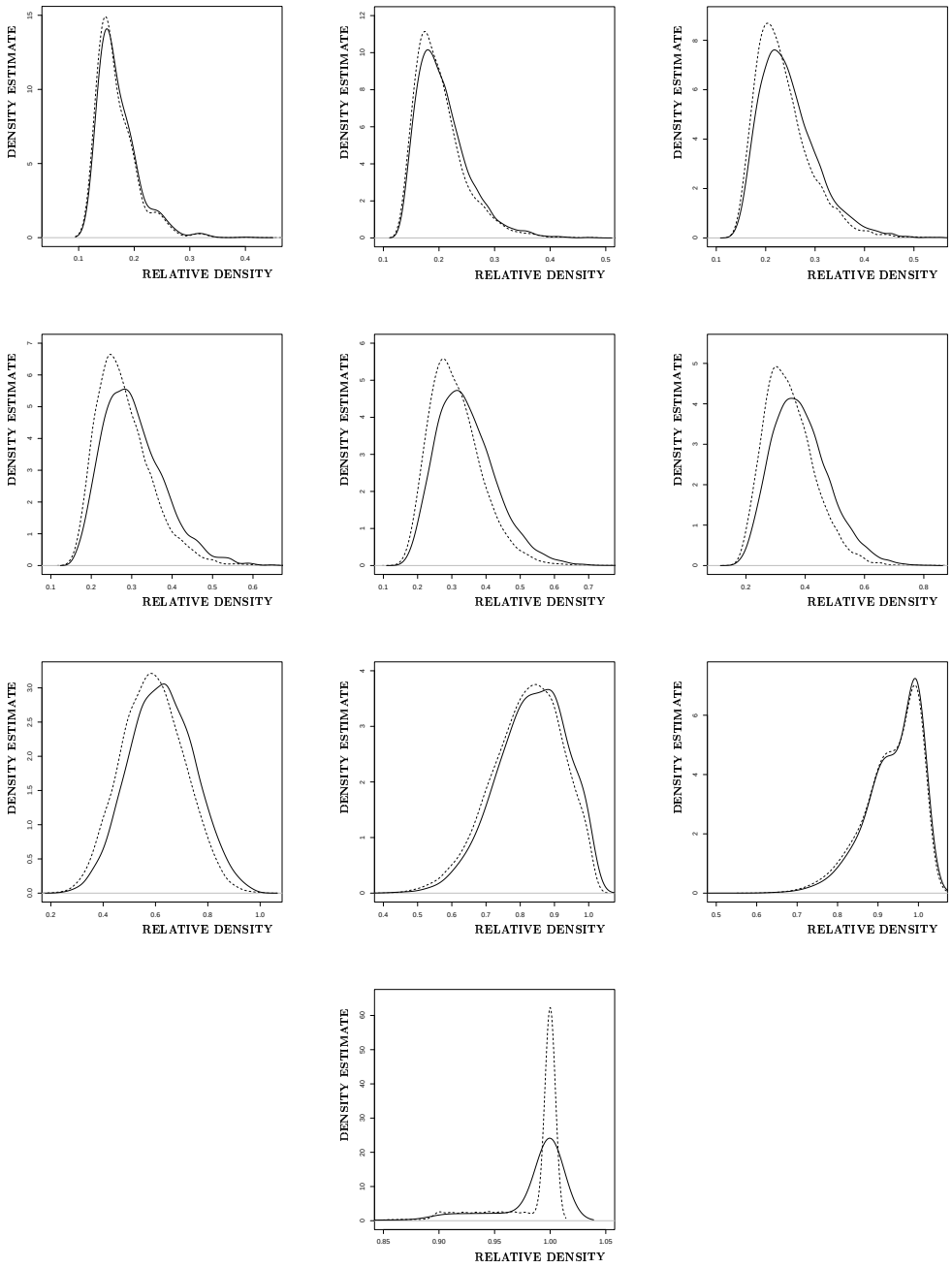


Figure 5.4.14: Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/21}$ (dashed) for $r \in \{1,\ 11/10,\ 6/5,\ 4/3,\ \sqrt{2},\ 3/2,\ 2,\ 3,\ 5,\ 10\}$ (left-to-right).

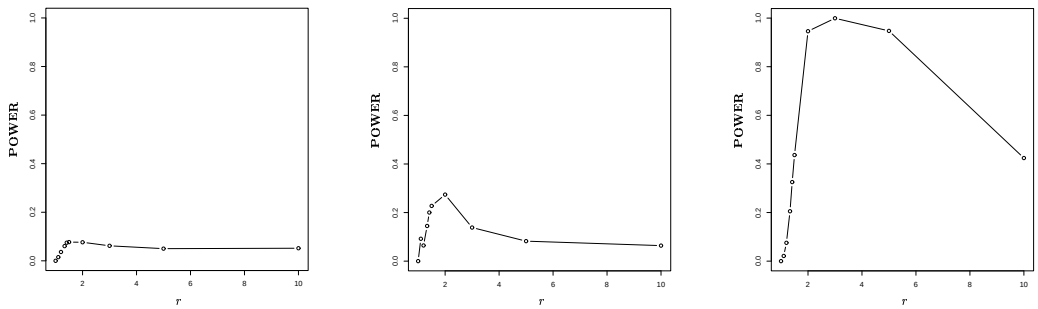


Figure 5.4.15: Monte Carlo power using the empirical critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of r for $n = 10$.

r	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5	10
$\widehat{c}^A_n(r)$	$.1\overline{3}$	$.1\overline{4}$	$.1\overline{6}$	.2	$.\overline{2}$	$.2\overline{4}$	$.4\,\overline{2}$	.65	$.8\overline{2}$	$.9\overline{1}$
$\widehat{\alpha}^A_{mc}(r, 10)$	0	.0112	.0208	.0308	.0363	.0359	.0392	.0413	.0478	.0398
$\widehat{\beta}^A_{mc}(r,\ 10,\ 5\sqrt{3}/24)$	0	.0213	.0754	.2052	.3253	.4365	.946	.9993	.9473	.4242
$\widehat{\beta}^A_{mc}(r,\ 10,\ \sqrt{3}/12)$	0	.0921	.0645	.1448	.2002	.2274	.2739	.1383	.0823	.0639
$\widehat{\beta}^A_{mc}(r,\ 10,\ \sqrt{3}/21)$	0	.0151	.0364	.0605	.0746	.0771	.0764	.0618	.0501	.0518

Table 5.4.5: The empirical critical values, empirical significance levels, and empirical power estimates under H^A_ε for $\varepsilon \in \{5\sqrt{3}/24,\ \sqrt{3}/12,\ \sqrt{3}/21\}$ and $n = 10$ at $\alpha = .05$.

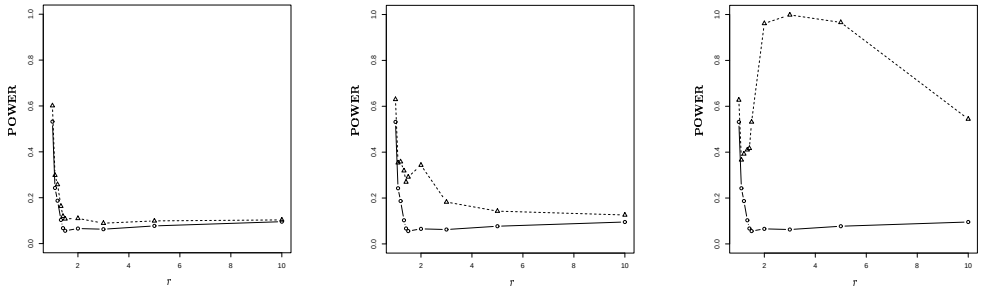


Figure 5.4.16: Monte Carlo power using the asymptotic critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of r for $n = 10$. The circles represent the empirical significance levels while triangles represent the empirical power values.

r	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5	10
$\widehat{\alpha}_A(r,10)$	.5318	.2426	.1869	.1031	.0673	.0559	.0656	.0627	.0771	.0955
$\widehat{\beta}^A_{10}\left(r, 5\sqrt{3}/24\right)$	.6273	.3663	.3923	.4103	.4167	.5316	.9610	.9983	.9656	.5443
$\widehat{\beta}^A_{10}\left(r, \sqrt{3}/12\right)$	.6300	.3537	.3583	.3190	.2698	.2919	.3433	.1825	.1429	.1261
$\widehat{\beta}^A_{10}\left(r, \sqrt{3}/21\right)$	.6012	.2979	.2574	.1629	.1190	.1077	.1098	.0889	.0989	.1033

Table 5.4.6: The empirical significance level and empirical power values under H^A_ε for $\varepsilon \in \{5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21\}$ with $N = 10000$ and $n = 10$ at $\alpha = .05$.

have the empirical power $\widehat{\beta}^A_{10}\left(r, \sqrt{3}/12\right) \leq .36$ with maximum power is .3433 and is attained at $r = 2$, and $\widehat{\beta}^A_{10}\left(r = 3, 5\sqrt{3}/24\right) = 1$. So, for small sample sizes, moderate values of r are more appropriate for normal approximation, as they yield the desired significance level and have the property that more severe association yields higher power.

The empirical significance levels and empirical power $\widehat{\beta}^S_n(r,\varepsilon)$ values under H^A_ε for $\varepsilon \in \left\{5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21\right\}$ are presented in Table 5.4.6.

Note that even for $n = 10$, the plots of the empirical power $\widehat{\beta}^A_{10}(r)$ resembles the curves of the asymptotic power function $\Pi_A(N^r_{PE})$ in Section 5.4.10.

5.4.8 Pitman Asymptotic Efficacy

For a detailed discussion of PAE, see [22] and [13]. Under segregation or association alternatives the PAE of $\rho_n(N^r_{PE})$ is given by $\text{PAE}(N^r_{PE}) = \frac{\left(\mu^{''}(N^r_{PE}, \varepsilon=0)\right)^2}{\nu(N^r_{PE})}$ since $\mu^{''}(N^r_{PE}, \varepsilon = 0) \neq 0$ but $\mu^{'}(N^r_{PE}, \varepsilon = 0) = 0$.

5.4.8.1 Pitman Asymptotic Efficacy Under Segregation Alternatives

Consider the test sequences $\rho(r) = \left\{\rho_n(N^r_{PE})\right\}$ for sufficiently small $\varepsilon > 0$ and $r \in \left[1, \sqrt{3}/(2\,\varepsilon)\right)$.

In the PAE framework of Section 2.7.2.1, $\theta = \varepsilon$ and $\theta_0 = 0$. Suppose, $\mu_n(\varepsilon) = E^S_\varepsilon[\rho_n(N^r_{PE})] = \mu_S(N^r_{PE}, \varepsilon)$. For $\varepsilon \in \left[0, \sqrt{3}/8\right)$,

$$\mu_S(N^r_{PE}, \varepsilon) = \sum_{j=1}^{5} \varpi_{1,j}(r,\varepsilon)\, \mathbf{I}(r \in \mathcal{I}_j)$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 3/2 - \sqrt{3}\,\varepsilon\right)$, $\mathcal{I}_2 = \left[3/2 - \sqrt{3}\,\varepsilon, 3/2\right)$, $\mathcal{I}_3 = \left[3/2, 2 - 4\,\varepsilon/\sqrt{3}\right)$, $\mathcal{I}_4 = \left[2 - 4\,\varepsilon/\sqrt{3}, 2\right)$, $\mathcal{I}_5 = \left[2, \sqrt{3}/(2\,\varepsilon)\right)$. See Appendix Section B.4.1 for the explicit form of $\mu(N^r_{PE}, \varepsilon)$ and Appendix Section B.3 for derivation. Notice that as $\varepsilon \to 0$, only $\mathcal{I}_1 =$

$\left[1, 3/2 - \sqrt{3}\,\varepsilon\right)$, $\mathcal{I}_3 = \left[3/2, 2 - 4\,\varepsilon/\sqrt{3}\right)$, $\mathcal{I}_5 = \left[2, \sqrt{3}/(2\,\varepsilon)\right)$ do not vanish, so we only keep the components of $\mu_S(N_{PE}^r, \varepsilon)$ on these intervals.

Furthermore,

$$\sigma_S^2(n,\varepsilon) = \mathbf{Var}_{\varepsilon}^{S}(\rho_n(N_{PE}^r)) = \frac{1}{2\,n\,(n-1)}\mathbf{Var}_{\varepsilon}^{S}[h_{12}(N_{PE}^r)] + \frac{(n-2)}{n\,(n-1)}\mathbf{Cov}_{\varepsilon}^{S}[h_{12}(N_{PE}^r), h_{13}(N_{PE}^r)].$$

The explicit forms of $\mathbf{Var}_{\varepsilon}^{S}[h_{12}(N_{PE}^r)]$ and $\mathbf{Cov}_{\varepsilon}^{S}[h_{12}(N_{PE}^r), h_{13}(N_{PE}^r)]$ are not calculated, since we only need $\lim_{n\to\infty}\sigma_n^2(\varepsilon = 0) = \nu(N_{PE}^r)$ which is given in Equation 5.4.3.

Notice that $\mathbf{E}_{\varepsilon}^{S}|h_{12}(N_{PE}^r)|^3 \leq 8 < \infty$ and $\mathbf{E}_{\varepsilon}^{S}[h_{12}\,h_{13}] - \mathbf{E}_{\varepsilon}^{S}[h_{12}]^2 = \mathbf{Cov}_{\varepsilon}^{S}[h_{12}, h_{13}] > 0$ then by Callaert and Janssen [2],

$$\sup_{t\in\mathbb{R}}\left|P_{\varepsilon}\left(\sqrt{n}\frac{(\rho_n(N_{PE}^r) - \mu_S(N_{PE}^r,\varepsilon))}{\sqrt{\nu_S(N_{PE}^r,\varepsilon)}} \leq t\right) - \Phi(t)\right| \leq C\,\mathbf{E}_{\varepsilon}^{S}\,|h_{12}(N_{PE}^r)|^3\,[\nu_S(N_{PE}^r,\varepsilon)]^{-3/2}\,n^{-1/2}$$

where C is a constant and $\Phi(\cdot)$ is the standard normal distribution function. Then (PC1) follows for each $r \in \left[1, \sqrt{3}/(2\,\varepsilon)\right)$ and $\varepsilon \in \left[0, \sqrt{3}/4\right)$ (see Section 2.7.2.1 for Pitman's conditions).

Differentiating $\mu_S(N_{PE}^r, \varepsilon)$ with respect to ε yields

$$\mu_S'(N_{PE}^r,\varepsilon) = \varpi_{1,1}'(r,\varepsilon)\,\mathbf{I}\left(r \in \left[1, 3/2 - \sqrt{3}\,\varepsilon\right)\right) + \varpi_{1,3}'(r,\varepsilon)\,\mathbf{I}\left(r \in [3/2, 2 - 4\,\varepsilon/\sqrt{3})\right) + \varpi_{1,5}'(r,\varepsilon)\,\mathbf{I}\left(r \in \left[2, \sqrt{3}/(2\,\varepsilon)\right)\right)$$

where

$$\varpi_{1,1}'(r,\varepsilon) = \frac{2\,\varepsilon\,(144\,\varepsilon^2\,(r^2-1) + 36 - 37\,r^2)}{27\,(2\,\varepsilon-1)^3(2\,\varepsilon+1)^3},$$

$$\varpi_{1,3}'(r,\varepsilon) = \Big[2\sqrt{3}\Big((2\,r-3)\,64\,\varepsilon^3 + (7\,r^2 + r^4 - 24\,r + 20)\,16\sqrt{3}\varepsilon^2 + (r-3)\,48\,\varepsilon + 3\sqrt{3}\,r^4 + 96\sqrt{3}\,r - 36\sqrt{3} - 60\sqrt{3}\,r^2\Big)\varepsilon\Big] \Big/ \Big[9\,(2\,\varepsilon+1)^3(2\,\varepsilon-1)^3 r^2\Big],$$

$$\varpi_{1,5}'(r,\varepsilon) = \frac{8\sqrt{3}\,\varepsilon\,\left(48\,\varepsilon^3 + (3\,r^4 + 3\,r^2 - 20)\,4\,\sqrt{3}\,\varepsilon^2 + 36\,\varepsilon + 9\sqrt{3} - 9\sqrt{3}\,r^2\right)}{27\,r^2(2\,\varepsilon+1)^3(2\,\varepsilon-1)^3}.$$

Hence, $\mu_S'(N_{PE}^r, \varepsilon = 0) = 0$, so we need higher order derivatives for (PC2). A detailed discussion is available in Kendall and Stuart [22].

Differentiating $\mu_S'(N_{PE}^r, \varepsilon)$ with respect to ε yields

$$\mu_S''(N_{PE}^r, \varepsilon) = \varpi_{1,1}''(r,\varepsilon)\,\mathbf{I}\left(r \in \left[1, 3/2 - \sqrt{3}\,\varepsilon\right)\right) + \varpi_{1,3}''(r,\varepsilon)\,\mathbf{I}\left(r \in [3/2, 2 - 4\,\varepsilon/\sqrt{3})\right) + \varpi_{1,5}''(r,\varepsilon)\,\mathbf{I}\left(r \in \left[2, \sqrt{3}/(2\,\varepsilon)\right)\right)$$

where

$$\varpi_{1,1}''(r,\varepsilon) = -\frac{2\,(r^2-1)\,1728\,\varepsilon^4 + (72 - 77\,r^2)\,4\,\varepsilon^2 + 36 - 37\,r^2}{27\,(4\,\varepsilon^2 - 1)^4},$$

$$\varpi_{1,3}''(r,\varepsilon) = -2\Big[(2\,r-3)\,512\,\sqrt{3}\,\varepsilon^5 + (20 + r^4 + 7\,r^2 - 24\,r)\,576\,\varepsilon^4 + (2\,r-3)\,1024\,\sqrt{3}\,\varepsilon^3 + (20 - 108\,r^2 + 96\,r + 9\,r^4)\,36\,\varepsilon^2 + (-3 + 2\,r)\,96\,\sqrt{3}\,\varepsilon - 108 + 9\,r^4 - 180\,r^2 + 288\,r\Big] \Big/ \Big[9\,r^2(2\,\varepsilon+1)^4(2\,\varepsilon-1)^4\Big],$$

$$\varpi_{1,5}''(r,\varepsilon) = -8\Big[128\,\sqrt{3}\,\varepsilon^5 + (-20 + 3\,r^4 + 3\,r^2)\,48\,\varepsilon^4 + 256\,\sqrt{3}\,\varepsilon^3 + (-5 - 12\,r^2 + 3\,r^4)\,12\,\varepsilon^2 + 24\,\varepsilon\,\sqrt{3} + 9 - 9\,r^2\Big] \Big/ \Big[9\,r^2(2\,\varepsilon+1)^4(2\,\varepsilon-1)^4\Big].$$

Thus,

$$\mu_S''(N_{PE}^r, \varepsilon = 0) = \begin{cases} -\frac{8}{3} + \frac{74}{27}\,r^2, & \text{for} \quad r \in [1, 3/2), \\ -2\,\frac{(r^2 - 4\,r + 2)(r^2 + 4\,r - 6)}{r^2}, & \text{for} \quad r \in [3/2, 2), \\ -\frac{8\,(1-r^2)}{r^2}, & \text{for} \quad r \in [2, \sqrt{3}/(2\,\varepsilon)). \end{cases} \tag{5.4.5}$$

Observe that $\mu_S''(N_{PE}^r, \varepsilon = 0) > 0$ for all $r \in \left[1, \sqrt{3}/(2\,\varepsilon)\right)$, so (PC2) holds with the second derivative. (PC3) in the second derivative form follows from continuity of $\mu_S''(N_{PE}^r, \varepsilon)$ in ε and (PC4) follows from continuity of $\sigma_n^2(r,\varepsilon)$ in ε.

Next, we find

$$C_S(\rho(r)) = \lim_{n\to\infty} \frac{\mu_S''(N_{PE}^r, \varepsilon = 0)}{\sqrt{n}\,\sigma_n(r, \varepsilon = 0)} = \frac{\mu_S''(N_{PE}^r, \varepsilon = 0)}{\sqrt{\nu(N_{PE}^r)}},$$

where numerator is given in Equation 5.4.5 and denominator is given in Equation 5.4.3. We can easily see that $C_S(\rho(r)) > 0$, since $C_S(\rho(r))$ is increasing in r and $C_S(\rho(r=1)) > 0$. Then (PC5) follows. So under segregation alternatives H_ε^S, the PAE of $\rho_n(N_{PE}^r)$ is given by

$$\mathrm{PAE}^S(N_{PE}^r) = C_S^2(\rho(r)) = \frac{\left(\mu_S''(N_{PE}^r, \varepsilon = 0)\right)^2}{\nu(N_{PE}^r)}.$$

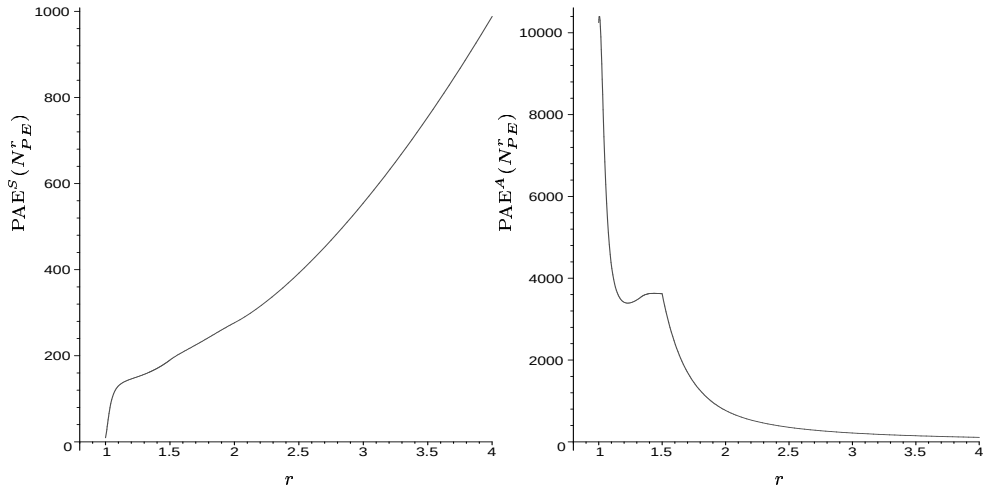


Figure 5.4.17: Pitman asymptotic efficacy against segregation (left) and against association (right) as a function of r.

In Figure 5.4.17 (left), we present the PAE as a function of r for segregation. Notice that $\mathrm{PAE}^S(N_{PE}^{r=1}) = 160/7 \approx 22.8571$, $\lim_{r\to\infty} \mathrm{PAE}^S(N_{PE}^r) = \infty$. Based on the PAE analysis, we suggest, for large n and small ε, choosing r large for testing against segregation. However, for small and moderate values of n, normal approximation is not appropriate due to the skewness in the density of $\rho_n(N_{PE}^r)$, Therefore, for small n, we suggest moderate r values.

PAE analysis is local (around $\varepsilon = 0$) and for arbitrarily large n. The comparison would hold in general provided that $\mu(N_{PE}^r, \varepsilon)$ is convex in ε for all $\varepsilon \in \left[0, \sqrt{3}/3\right)$. As an alternative, we fix an ε and then compare the asymptotic behaviour of $\rho_n(N_{PE}^r)$ with Hodges-Lehmann asymptotic efficacy (HLAE) in Section 5.4.9.1.

5.4.8.2 Pitman Asymptotic Efficacy Under Association Alternatives

Consider the test sequences $\rho(r) = \left\{\rho_n(N_{PE}^r)\right\}$ for sufficiently small $\varepsilon > 0$ and $r \in [1, \infty)$.

In the PAE framework of Section 2.7.2.1, $\theta = \varepsilon$ and $\theta_0 = 0$. Suppose, $\mu_n(\varepsilon) = \mathbf{E}_{\,\varepsilon}[\rho_n(N_{PE}^r)] = \mu_A(N_{PE}^r, \varepsilon)$. For $\varepsilon \in \left[0, \left(7\sqrt{3} - 3\sqrt{15}\right)/12 \approx .042\right)$,

$$\mu_A(N_{PE}^r, \varepsilon) = \sum_{j=1}^{6} \varpi_{1,j}(r, \varepsilon)\, \mathbf{I}(r \in \mathcal{I}_j)$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, \left(1 + 2\sqrt{3}\,\varepsilon\right) / \left(1 - \sqrt{3}\,\varepsilon\right)\right)$, $\mathcal{I}_2 = \left[\left(1 + 2\sqrt{3}\,\varepsilon\right) / \left(1 - \sqrt{3}\,\varepsilon\right),\right.$ $\left. 4\ \left(1 - \sqrt{3}\,\varepsilon\right)/3\right)$, $\mathcal{I}_3 = \left[4\ \left(1 - \sqrt{3}\,\varepsilon\right)/3, 4\ \left(1 + 2\sqrt{3}\,\varepsilon\right)/3\right)$, $\mathcal{I}_4 = \left[4\ \left(1 + 2\sqrt{3}\,\varepsilon\right)/3, 3/(2\ \left(1 - \sqrt{3}\,\varepsilon\right))\right)$,

$\mathcal{I}_5 = \left[3/(2\ (1-\sqrt{3}\,\varepsilon)), 2\right)$, and $\mathcal{I}_6 = [2,\infty)$. Notice that as $\varepsilon \to 0$, only $\mathcal{I}_j$ for $j \in \{2,4,5,6\}$ do not vanish, so we only keep the components of $\mu_A(N_{PE}^r, \varepsilon)$ on these intervals. See Appendix Section B.4.2 for the explicit form of $\mu_A(N_{PE}^r, \varepsilon)$ and Appendix Section B.3 for derivation.

Furthermore,

$$\sigma_n^2(\varepsilon) = \mathbf{Var}_{\varepsilon}^{A}(\rho_n(N_{PE}^r)) = \frac{1}{2\,n\,(n-1)}\mathbf{Var}_{\varepsilon}^{A}[h_{12}(N_{PE}^r)] + \frac{(n-2)}{n\,(n-1)}\,\mathbf{Cov}_{\varepsilon}^{A}[h_{12}(N_{PE}^r), h_{13}(N_{PE}^r)]$$

whose explicit form is not calculated, since we only need $\lim_{n\to\infty}\sqrt{n}\,\sigma_n(\varepsilon=0) = \nu(N_{PE}^r)$ which is given Equation 5.4.3.

(PC1) follows for each $r \in [1,\infty)$ and $\varepsilon \in \left[0, \sqrt{3}/3\right)$ as in the segregation case.

Differentiating $\mu_A(N_{PE}^r, \varepsilon)$ with respect to ε, then we get

$$\mu_A'(N_{PE}^r, \varepsilon) = \varpi_{1,2}'(r,\varepsilon)\,\mathbf{I}(r \in [1,4/3)) + \varpi_{1,4}'(r,\varepsilon)\,\mathbf{I}(r \in [4/3,3/2)) + \varpi_{1,5}'(r,\varepsilon)\,\mathbf{I}(r \in [3/2,2)) + \varpi_{1,6}'(r,\varepsilon)\,\mathbf{I}(r \in [2,\infty))$$

where

$$\begin{aligned}
\varpi_{1,2}'(r,\varepsilon) &= -2\left[\sqrt{3}\Big(-1152\,r^4\varepsilon^3 + 720\,\sqrt{3}\,r^4\varepsilon^2 - 288\,r^4\,\varepsilon + 11\,\sqrt{3}\,r^4 + 2592\,\sqrt{3}\,r^2\varepsilon^2 - 10368\,\sqrt{3}\,r\varepsilon^2\right.\\
&\quad \left. + 432\,\sqrt{3}\,r^2 + 6480\,\sqrt{3}\varepsilon^2 - 864\,\sqrt{3}\,r + 432\,\sqrt{3}\Big)\varepsilon\right] \Big/ \left[\left(-6\,\varepsilon + \sqrt{3}\right)^3\left(6\,\varepsilon + \sqrt{3}\right)^3 r^2\right],\\
\varpi_{1,4}'(r,\varepsilon) &= -2\left[\sqrt{3}\Big(-1152\,r^4\varepsilon^3 + 720\,\sqrt{3}\,r^4\varepsilon^2 - 288\,r^4\,\varepsilon + 11\,\sqrt{3}\,r^4 - 1296\,\sqrt{3}\,r^2\varepsilon^2 + 108\,\sqrt{3}\,r^2\right.\\
&\quad \left. - 2160\,\sqrt{3}\varepsilon^2 - 144\,\sqrt{3}\Big)\varepsilon\right] \Big/ \left[\left(-6\,\varepsilon + \sqrt{3}\right)^3\left(6\,\varepsilon + \sqrt{3}\right)^3 r^2\right],\\
\varpi_{1,5}'(r,\varepsilon) &= \frac{2\,\varepsilon\,(3\,r^4 - 72\,r^2 - 240\,\varepsilon^2 + 192\,r - 124)}{r^2(12\,\varepsilon^2 - 1)^3},\\
\varpi_{1,6}'(r,\varepsilon) &= -\frac{40\,\varepsilon}{r^2(12\,\varepsilon^2 - 1)^2}.
\end{aligned}$$

Hence $\mu_A'(N_{PE}^r, \varepsilon = 0) = 0$, so we differentiate $\mu_A'(N_{PE}^r, \varepsilon)$ with respect to ε and get

$$\mu_A''(N_{PE}^r, \varepsilon) = \varpi_{1,2}''(r,\varepsilon)\,\mathbf{I}(r \in [1,4/3)) + \varpi_{1,4}''(r,\varepsilon)\,\mathbf{I}(r \in [4/3,3/2)) + \varpi_{1,5}''(r,\varepsilon)\,\mathbf{I}(r \in [3/2,2)) + \varpi_{1,6}''(r,\varepsilon)\,\mathbf{I}(r \in [2,\infty))$$

where

$$\varpi''_{1,2}(r,\varepsilon) = -6\left[\sqrt{3}\Big(-27648\,r^4\varepsilon^5 + 25920\sqrt{3}\,r^4\varepsilon^4 - 18432\,r^4\varepsilon^3 + 2820\sqrt{3}\,r^4\varepsilon^2 + 93312\sqrt{3}\,r^2\varepsilon^4 - 576\,r^4\,\varepsilon - 373248\sqrt{3}\,r\varepsilon^4 + 11\sqrt{3}\,r^4 + 33696\sqrt{3}\,r^2\varepsilon^2 + 233280\sqrt{3}\varepsilon^4 - 82944\sqrt{3}\,r\varepsilon^2 + 432\sqrt{3}\,r^2 + 45360\sqrt{3}\varepsilon^2 - 864\sqrt{3}\,r + 432\sqrt{3}\Big)\right] \Big/ \left[\Big(6\,\varepsilon + \sqrt{3}\Big)^4\,\Big(-6\,\varepsilon + \sqrt{3}\Big)^4\,r^2\right],$$

$$\varpi''_{1,4}(r,\varepsilon) = -6\left[\sqrt{3}\Big(-27648\,r^4\varepsilon^5 + 25920\sqrt{3}\,r^4\varepsilon^4 - 18432\,r^4\varepsilon^3 + 2820\sqrt{3}\,r^4\varepsilon^2 - 46656\sqrt{3}\,r^2\varepsilon^4 - 576\,r^4\,\varepsilon + 11\sqrt{3}\,r^4 + 2592\sqrt{3}\,r^2\varepsilon^2 - 77760\sqrt{3}\varepsilon^4 + 108\sqrt{3}\,r^2 - 15120\sqrt{3}\varepsilon^2 - 144\sqrt{3}\Big)\right] \Big/ \left[\Big(6\,\varepsilon + \sqrt{3}\Big)^4\,\Big(-6\,\varepsilon + \sqrt{3}\Big)^4\,r^2\right],$$

$$\varpi''_{1,5}(r,\varepsilon) = -\frac{2\,(180\,r^4\varepsilon^2 + 3\,r^4 - 4320\,r^2\varepsilon^2 - 8640\,\varepsilon^4 + 11520\,r\varepsilon^2 - 72\,r^2 - 8160\,\varepsilon^2 + 192\,r - 124)}{r^2(12\,\varepsilon^2 - 1)^4},$$

$$\varpi''_{1,6}(r,\varepsilon) = \frac{40\,(36\,\varepsilon^2 + 1)}{r^2(12\,\varepsilon^2 - 1)^3}.$$

Thus,

$$\mu''_A(N^r_{PE}, \varepsilon = 0) = \begin{cases} -\frac{22}{9}\,r^2 + 192\,r^{-1} - 96\,r^{-2} - 96, & \text{for} \quad r \in [1, 4/3), \\ -\frac{22}{9}\,r^2 + 32\,r^{-2} - 24, & \text{for} \quad r \in [4/3, 3/2), \\ -6\,r^2 - 384\,r^{-1} + 248\,r^{-2} + 144, & \text{for} \quad r \in [3/2, 2), \\ -40\,r^{-2}, & \text{for} \quad r \in [2, \infty). \end{cases} \tag{5.4.6}$$

Note that $\mu''_A(N^r_{PE}, \varepsilon = 0) < 0$ for all $r \in [1, \infty)$, so (PC2) follows with the second derivative. (PC3) and (PC4) follow from continuity of $\mu''_A(N^r_{PE}, \varepsilon)$ and $\sigma^2_n(r, \varepsilon)$ in ε.

Next, we find $C_A(\rho(r)) = \lim_{n\to\infty} \frac{\mu''_A(N^r_{PE}, \varepsilon=0)}{\sqrt{n}\,\sigma_n(r, \varepsilon=0)} = \frac{\mu''_A(N^r_{PE}, 0)}{\sqrt{\nu(N^r_{PE})}}$, by substituting the numerator from Equation 5.4.6 and denominator from Equation 5.4.3. We can easily see that $C_A(\rho(r)) < 0$, for all $r \geq 1$. Then (PC5) holds, so under association alternatives H^A_ε, the PAE of $\rho_n(N^r_{PE})$ is

$$\mathrm{PAE}^A(N^r_{PE}) = C^2_A(\rho(r)) = \frac{\big(\mu''_A(N^r_{PE}, \varepsilon = 0)\big)^2}{\nu(N^r_{PE})}.$$

In Figure 5.4.17 (right), we present the PAE as a function of r for association. Notice that $\mathrm{PAE}^A(N^{r=1}_{PE}) = 174240/17 \approx 10249.4118$, $\lim_{r\to\infty} \mathrm{PAE}^A(N^r_{PE}) = 0$, $\mathrm{argsup}_{r\in[1,\infty)}\,\mathrm{PAE}^A(N^r_{PE}) \approx$

1.006 with supremum ≈ 10399.7726. $\mathrm{PAE}^A(N_{PE}^r)$ has also a local supremum at $r_l \approx 1.4356$ with local supremum ≈ 3630.8932. Based on the PAE analysis, we suggest, for large n and small ε, choosing r small for testing against association. However, for small and moderate values of n normal approximation is not appropriate due to the skewness in the density of $\rho_n(N_{PE}^r)$. Therefore, for small n, we suggest moderate r values.

We also calculate Hodges-Lehmann asymptotic efficacy for fixed alternatives in Section 5.4.9.2.

5.4.9 Hodges-Lehmann Asymptotic Efficacy

Unlike PAE, HLAE does not involve the limit as $\varepsilon \to 0$. Since this requires the mean and, especially, the asymptotic variance of $\rho_n(N_{PE}^r)$ *under an alternative*, we investigate HLAE for specific values of ε.

5.4.9.1 Hodges-Lehmann Asymptotic Efficacy Under Segregation Alternatives

In the HLAE framework of Section 2.7.2.2, $\theta = \mathbf{E}_\varepsilon^S[\rho_n(N_{PE}^r)] = \mu_S(N_{PE}^r, \varepsilon)$ and $\theta_0 = \mu(N_{PE}^r)$. Then testing $H_0 : \varepsilon = 0$ versus $H_\varepsilon^S : \varepsilon > 0$ is equivalent to $H_0 : \mathbf{E}[\rho_n(N_{PE}^r)] = \mu(N_{PE}^r)$ versus $H_\varepsilon^S : \mathbf{E}_\varepsilon^S[\rho_n(N_{PE}^r)] = \mu_S(N_{PE}^r, \varepsilon) > \mu(N_{PE}^r)$. Let $\delta(r,\varepsilon) = \frac{\mu_S(N_{PE}^r,\varepsilon) - \mu(N_{PE}^r)}{\sqrt{\nu_S(N_{PE}^r,\varepsilon)}}$ and $\widetilde{R}(r) = \frac{\rho_n(N_{PE}^r) - \mu(N_{PE}^r)}{\sqrt{\nu_S(N_{PE}^r,\varepsilon)}}$, then $\widetilde{R}(r) \xrightarrow{\mathcal{L}} \mathcal{N}(\delta(r,\varepsilon), 1)$.

Then HLAE of $\rho_n(N_{PE}^r)$ is given by

$$\mathrm{HLAE}^S(N_{PE}^r, \varepsilon) := \delta(r,\varepsilon)^2 = \frac{\left(\mu_S(N_{PE}^r, \varepsilon) - \mu(N_{PE}^r)\right)^2}{\nu_S(N_{PE}^r, \varepsilon)}.$$

We calculate HLAE of $\rho_n(N_{PE}^r)$ under H_ε^S for $\varepsilon = \sqrt{3}/8$, $\varepsilon = \sqrt{3}/4$, and $\varepsilon = 2\sqrt{3}/7$.

With $\varepsilon = \sqrt{3}/8$, $\rho_n(N_{PE}^r)$ is non-degenerate for $r \in [1,4)$, and the mean is given by

$$\mu_S\left(N_{PE}^r, \sqrt{3}/8\right) = \begin{cases} \frac{2287}{9126}\, r^2 - \frac{1}{13}, & \text{for} \quad r \in [1, 9/8), \\ -\frac{5905\, r^4 - 36864\ r^3 + 62910\, r^2 - 46656\, r + 13122}{9126\, r^2}, & \text{for} \quad r \in [9/8, 3/2), \\ \frac{61\, r^4 - 768\, r^3 + 3494\ r^2 - 5120\, r + 2466}{338\, r^2}, & \text{for} \quad r \in [3/2, 2), \\ -\frac{3\, r^4 - 422\, r^2 + 606}{338\, r^2}, & \text{for} \quad r \in [2, 3), \\ \frac{3\, r^4 - 48\, r^3 + 530\, r^2 - 768}{338\, r^2}, & \text{for} \quad r \in [2, 4], \end{cases}$$

and the variance is given by $\nu_S\left(N_{PE}^r r, \sqrt{3}/8\right) = \sum_{j=1}^{12} \nu_j\left(r, \sqrt{3}/8\right)\, \mathbf{I}(\mathcal{I}_j)$ where

$$\begin{aligned}
\nu_1\left(r, \sqrt{3}/8\right) &= \Big[9959911\, r^{10} - 46006272\, r^9 - 430526\, r^8 + 258785280\, r^7 - 385799609\, r^6 + 162699264\ r^5 \\
&\quad - 83976048\, r^4 + 201277440\, r^3 - 129392640\, r^2 + 12939264\Big] \Big/ \Big[104104845\, r^4\Big], \\
\nu_2\left(r, \sqrt{3}/8\right) &= \Big[9959911\, r^{10} - 46006272\, r^9 - 430526\, r^8 + 258785280\, r^7 - 415110891\, r^6 + 272331072\, r^5 \\
&\quad - 158725008\, r^4 - 16174080\, r^3 + 315394560\, r^2 - 310542336\, r + 90574848\Big] \Big/ \Big[104104845\, r^4\Big], \\
\nu_3\left(r, \sqrt{3}/8\right) &= \Big[3144167\, r^{12} + 15335424\ r^{11} - 378655166\, r^{10} + 2750459904\ r^9 - 11800111467\, r^8 \\
&\quad + 31878202752\, r^7 - 54792387144\, r^6 + 60339341664\ r^5 - 42745183272\, r^4 + 19903426272\, r^3 \\
&\quad - 6790168926\, r^2 + 1989715104\ r - 373071582\Big] \Big/ \Big[104104845\, r^6\Big], \\
\nu_4\left(r, \sqrt{3}/8\right) &= -\Big[8177689\, r^{12} - 54153216\, r^{11} + 320428478\, r^{10} - 2459326464\ r^9 + 11854698987\, r^8 \\
&\quad - 32751603072\, r^7 + 55010737224\ r^6 - 59029241184\ r^5 + 42131073672\, r^4 - 20886001632\, r^3 \\
&\quad + 7379714142\, r^2 - 1694942496\, r + 170415414\Big] \Big/ \Big[104104845\, r^6\Big], \\
\nu_5\left(r, \sqrt{3}/8\right) &= -\Big[8177689\, r^{12} - 54153216\, r^{11} + 320428478\, r^{10} - 2459326464\ r^9 + 12509010411\, r^8 \\
&\quad - 37904305536\, r^7 + 71918042184\ r^6 - 88617024864\ r^5 + 71256548232\, r^4 - 36176875776\, r^3 \\
&\quad + 10724592861\, r^2 - 1694942496\, r + 170415414\Big] \Big/ \Big[104104845\, r^6\Big]. \\
\nu_6\left(r, \sqrt{3}/8\right) &= -\Big[2718937\, r^{12} - 39596544\, r^{11} + 434455742\, r^{10} - 3154811904\ r^9 + 14086429683\, r^8 \\
&\quad - 39680803584\ r^7 + 72881433288\, r^6 - 88893062496\, r^5 + 71547681672\, r^4 \\
&\quad - 36487418112\, r^3 + 10828106973\, r^2 - 1694942496\, r + 170415414\Big] \Big/ \Big[104104845\, r^6\Big], \\
\nu_7\left(r, \sqrt{3}/8\right) &= -\Big[1027\, r^{12} - 19968\, r^{11} + 295626\, r^{10} - 3265792\, r^9 + 23210081\, r^8 \\
&\quad - 103077696\, r^7 + 289042360\, r^6 - 511170304\ r^5 + 553668600\, r^4 - 343186304\ r^3 \\
&\quad + 109133095\, r^2 - 20431008\, r + 5845554\Big] \Big/ \Big[428415\, r^6\Big], \\
\nu_8\left(r, \sqrt{3}/8\right) &= -\Big[637\, r^{12} - 19968\, r^{11} + 299370\, r^{10} - 3265792\, r^9 + 23199551\, r^8 \\
&\quad - 103077696\, r^7 + 289042360\, r^6 - 511170304\ r^5 + 553700190\, r^4 - 343186304\ r^3 \\
&\quad + 109133095\, r^2 - 20431008\, r + 5788692\Big] \Big/ \Big[428415\, r^6\Big], \\
\nu_9\left(r, \sqrt{3}/8\right) &= -\Big[637\, r^{12} - 19968\, r^{11} + 299370\, r^{10} - 3265792\, r^9 + 24051519\, r^8 \\
&\quad - 112023360\, r^7 + 328179640\, r^6 - 602490624\ r^5 + 673558110\, r^4 - 427086848\, r^3 \\
&\quad + 133604087\, r^2 - 20431008\, r + 5788692\Big] \Big/ \Big[428415\, r^6\Big], \\
\nu_{10}\left(r, \sqrt{3}/8\right) &= \Big[130\, r^{12} - 2496\, r^{11} + 22134\ r^{10} - 122720\, r^9 + 452225\, r^8 - 1010880\, r^7 + 1075400\, r^6 \\
&\quad + 26624\ r^5 - 1993566\, r^4 + 5324800\, r^3 - 5083895\, r^2 + 303264\ r - 37908\Big] \Big/ \Big[428415\, r^6\Big],
\end{aligned}$$

$$\nu_{11}\left(r,\sqrt{3}/8\right) = -\Big[330\,r^8 - 8896\,r^7 + 85445\,r^6 - 342624\,r^5 + 332000\,r^4 + 1148560\,r^3 - 1180986\,r^2 - 5324800\,r + 6678947\Big]\Big/\Big[428415\,r^4\Big],$$

$$\nu_{12}\left(r,\sqrt{3}/8\right) = -\frac{(330\,r^5 - 4936\,r^4 + 12453\,r^3 + 47388\,r^2 - 12992\,r - 128256)(r-4)^3}{428415\,r^4},$$

and the corresponding intervals are $\mathcal{I}_1 = [1, 12/11)$, $\mathcal{I}_2 = [12/11, 9/8)$, $\mathcal{I}_3 = \left[9/8, \sqrt{6}/2\right)$, $\mathcal{I}_4 = \left[\sqrt{6}/2, 21/16\right)$, $\mathcal{I}_5 = [21/16, 4/3)$, $\mathcal{I}_6 = [4/3, 3/2)$, $\mathcal{I}_7 = \left[3/2, \sqrt{3}\right)$, $\mathcal{I}_8 = \left[\sqrt{3}, 7/4\right)$, $\mathcal{I}_9 = [7/4, 2)$, $\mathcal{I}_{10} = [2, 3)$, $\mathcal{I}_{11} = [3, 7/2)$, and $\mathcal{I}_{12} = [7/2, 4)$. See Section B.3.1 for derivation and Figure 5.4.19 for the graph of $\mu_S\left(N_{PE}^r, \sqrt{3}/8\right)$ and $\nu_S\left(N_{PE}^r, \sqrt{3}/8\right)$.

Then we get $\mathrm{HLAE}^S\left(N_{PE}^r, \sqrt{3}/8\right) = \frac{\left(\mu_S\left(N_{PE}^r, \sqrt{3}/8\right) - \mu(N_{PE}^r)\right)^2}{\nu_S\left(N_{PE}^r r, \sqrt{3}/8\right)}$ by substituting the relevant terms. See Figure 5.4.18.

With $\varepsilon = \sqrt{3}/4$, $\rho_n(N_{PE}^r)$ is non-degenerate for $r \in [1, 2)$, and the mean is given by

$$\mu_S\left(N_{PE}^r, \sqrt{3}/4\right) = \begin{cases} -\frac{67}{54}\,r^2 + \frac{40}{9}\,r - 3, & \text{for} \quad r \in [1, 3/2), \\ \frac{7\,r^4 - 48\,r^3 + 122\,r^2 - 128\,r + 48}{2\,r^2}, & \text{for} \quad r \in [3/2, 2), \end{cases}$$

and the variance is given by

$$\nu_S\left(N_{PE}^r r, \sqrt{3}/4\right) = \sum_{j=1}^{5} \nu_j\left(r, \sqrt{3}/4\right)\,\mathbf{I}(\mathcal{I}_j)$$

where

$$\nu_1\left(r,\sqrt{3}/4\right) = -\Big[14285\,r^7 - 28224\,r^6 - 233266\,r^5 + 1106688\,r^4 - 2021199\,r^3 + 1876608\,r^2 - 880794\,r + 165888\Big]\Big/\Big[3645\,r\Big],$$

$$\nu_2\left(r,\sqrt{3}/4\right) = -\Big[14285\,r^{10} - 28224\,r^9 - 233266\,r^8 + 1106688\,r^7 - 1234767\,r^6 - 3431808\,r^5 + 14049126\,r^4 - 22228992\,r^3 + 18895680\,r^2 - 8503056\,r + 1594323\Big]\Big/\Big[3645\,r^4\Big],$$

$$\nu_3\left(r,\sqrt{3}/4\right) = -\Big[14285\,r^{10} - 28224\,r^9 - 233266\,r^8 + 1106688\,r^7 - 2545713\,r^6 + 5903280\,r^5 - 13456044\,r^4 + 20636208\,r^3 - 18305190\,r^2 + 8503056\,r - 1594323\Big]\Big/\Big[3645\,r^4\Big],$$

$$\nu_4\left(r,\sqrt{3}/4\right) = \Big[104920\,r^8 - 111072\,r^7 + 1992132\,r^6 - 15844032\,r^5 + 50174640\,r^4 + 6377292 - 34012224\,r + 73220760\,r^2 - 81881280\,r^3 + 1909\,r^{10} - 27072\,r^9\Big]\Big/\Big[14580\,r^4\Big],$$

$$\nu_5\left(r,\sqrt{3}/4\right) = -\Big[-1187904\ r^5 + 1331492\, r^6 + 433304\ r^2 + 611163\, r^{10} - 850240\, r^9 - 198144\, r$$
$$+ 955392\, r^4 - 705536\, r^3 - 387680\, r^{11} + 1118472\, r^8 - 1308960\, r^7 + 175984\ r^{12}$$
$$- 46176\, r^{13} + 5120\, r^{14} + 56016\Big] \Big/ \Big[20\, r^4\Big],$$

and the corresponding intervals are $\mathcal{I}_1 = [1, 9/8)$, $\mathcal{I}_2 = [9/8, 9/7)$, $\mathcal{I}_3 = [9/7, 4/3)$, $\mathcal{I}_4 = [4/3, 3/2)$, and $\mathcal{I}_5 = [3/2, 2)$. See Figure 5.4.19 for the graph of $\mu_S\left(N_{PE}^r, \sqrt{3}/4\right)$ and $\nu_S\left(N_{PE}^r r, \sqrt{3}/4\right)$.

Then we get $\mathrm{HLAE}^S\left(N_{PE}^r, \sqrt{3}/4\right) = \frac{\left(\mu_S\left(N_{PE}^r, \sqrt{3}/4\right) - \mu(N_{PE}^r)\right)^2}{\nu_S\left(N_{PE}^r r, \sqrt{3}/4\right)}$ by substituting the relevant terms. See Figure 5.4.18.

With $\varepsilon = 2\sqrt{3}/7$, $\rho_n(N_{PE}^r)$ is non-degenerate for $r \in [1, 3/2)$, and the mean is given by

$$\mu_S\left(N_{PE}^r, 2\sqrt{3}/7\right) = \begin{cases} -\frac{241}{54}\, r^2 + \frac{38}{3}\, r - 8, & \text{for} \quad r \in [1, 9/7), \\ \frac{80\, r^4 - 432\, r^3 + 866\, r^2 - 756\, r + 243}{2\, r^2}, & \text{for} \quad r \in [9/7, 3/2), \end{cases}$$

and the variance is given by

$$\nu_S\left(N_{PE}^r r, 2\sqrt{3}/7\right) = \sum_{j=1}^{6} \nu_j\left(r, 2\sqrt{3}/7\right) \mathbf{I}(\mathcal{I}_j)$$

where

$$\nu_1\left(r, 2\sqrt{3}/7\right) = -\Big[2495087\, r^7 - 5067342\, r^6 - 29145379\, r^5 + 134149248\, r^4 - 230713503\, r^3$$
$$+ 202262778\, r^2 - 90317349\, r + 16336404\Big] \Big/ \Big[14580\, r\Big],$$
$$\nu_2\left(r, 2\sqrt{3}/7\right) = -\Big[2495087\, r^{10} - 5067342\, r^9 - 29145379\, r^8 + 134149248\, r^7 - 140359071\, r^6$$
$$- 378587142\, r^5 + 1465530651\, r^4 - 2206303596\, r^3 + 1786050000\, r^2 - 765450000\, r$$
$$+ 136687500\Big] \Big/ \Big[14580\, r^4\Big],$$
$$\nu_3\left(r, 2\sqrt{3}/7\right) = -\Big[2495087\, r^{10} - 5067342\, r^9 - 29145379\, r^8 + 134149248\, r^7 - 309668679\, r^6$$
$$+ 731864538\, r^5 - 1559738349\, r^4 + 2174176404\, r^3 - 1767825000\, r^2 + 765450000\, r$$
$$- 136687500\Big] \Big/ \Big[14580\, r^4\Big],$$
$$\nu_4\left(r, 2\sqrt{3}/7\right) = \Big[1000147\, r^8 - 654768\, r^7 + 77561559\, r^6 - 527363136\, r^5 + 1468526760\, r^4$$
$$+ 1767825000\, r^2 - 765450000\, r - 2157840000\, r^3 + 136687500 + 24337\, r^{10}$$
$$- 321426\, r^9\Big] \Big/ \Big[14580\, r^4\Big],$$

$$\nu_5\left(r, 2\sqrt{3}/7\right) = \frac{24337}{14580}r^6 - \frac{17857}{810}r^5 + \frac{1000147}{14580}r^4 - \frac{18188}{405}r^3 - \frac{174113}{1620}r^2 + \frac{8176}{45}r - 78,$$
$$\nu_6\left(r, 2\sqrt{3}/7\right) = -\frac{\left(8r^6 - 106r^5 + 8709r^4 - 39684r^3 + 68000r^2 - 51192r + 14256\right)(2r-3)^4}{20r^4}.$$

The corresponding intervals are $\mathcal{I}_1 = [1, 15/14)$, $\mathcal{I}_2 = [15/14, 15/13)$, $\mathcal{I}_3 = [15/13, 7/6)$, $\mathcal{I}_4 = [7/6, 5/4)$, $\mathcal{I}_5 = [5/4, 9/7)$, $\mathcal{I}_6 = [9/7, 3/2)$. See Figure 5.4.19 for the graph of $\mu_S\left(N_{PE}^r, 2\sqrt{3}/7\right)$ and $\nu_S\left(N_{PE}^r r, 2\sqrt{3}/7\right)$.

Then we get $\text{HLAE}^S\left(N_{PE}^r, 2\sqrt{3}/7\right) = \frac{\left(\mu_S\left(N_{PE}^r, 2\sqrt{3}/7\right) - \mu_S\left(N_{PE}^r, 2\sqrt{3}/7\right)\right)^2}{\nu_S\left(N_{PE}^r r, 2\sqrt{3}/7\right)}$ by substituting the relevant terms. In Figure 5.4.18 are the graphs of $\text{HLAE}^S(N_{PE}^r, \varepsilon)$ for $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$.

From Figure 5.4.18, we see that, under H_ε^S, $\text{HLAE}^S(N_{PE}^r, \varepsilon)$ appears to be an increasing function, dependent on ε, of r. Let $r_\delta(\varepsilon)$ be the minimum r such that $\rho_n(N_{PE}^r)$ becomes degenerate under the alternative H_ε^S. Then $r_\delta\left(\sqrt{3}/8\right) = 4$, $r_\delta\left(\sqrt{3}/4\right) = 2$, and $r_\delta\left(2\sqrt{3}/7\right) = 3/2$. In fact, for $\varepsilon \in \left(0, \sqrt{3}/4\right]$, $r_\delta(\varepsilon) = \sqrt{3}/(2\varepsilon)$ and for $\varepsilon \in \left(\sqrt{3}/4, \sqrt{3}/3\right)$, $r_\delta(\varepsilon) = \sqrt{3}/\varepsilon - 2$. Notice that $\lim_{r \to r_\delta(\varepsilon)} \text{HLAE}^S(\rho_n(N_{PE}^r), \varepsilon) = \infty$, which is in agreement with PAE analysis because as $\varepsilon \to 0$ HLAE becomes PAE, and as $\varepsilon \to 0$, $r_\delta(\varepsilon) \to \infty$ and under H_0, $\rho_n(N_{PE}^r)$ is degenerate for $r = \infty$. The above result for HLAE can also be generalized for arbitrary ε as follows.

Proposition 5.4.7. *Let* $\widetilde{r} := \text{argsup}_{r \in [1, r_\delta(\varepsilon)]} \text{HLAE}^S(N_{PE}^r, \varepsilon)$ *where* $r_\delta(\varepsilon)$ *is the minimum value of* r *at which* $\rho_n(N_{PE}^r)$ *is degenerate under* H_ε^S. *Then* $\widetilde{r} = r_\delta(\varepsilon)$. *In particular, for* $\varepsilon \in \left[0, \sqrt{3}/4\right]$, $r_\delta = \sqrt{3}/(2\varepsilon)$ *and for* $\varepsilon \in \left(\sqrt{3}/4, \sqrt{3}/3\right]$, $r_\delta = \sqrt{3}/\varepsilon - 2$.

Proof: Recall that $\text{HLAE}^S(N_{PE}^r, \varepsilon) = \frac{\left(\mu_S(N_{PE}^r, \varepsilon) - \mu(N_{PE}^r)\right)^2}{\nu_S(N_{PE}^r, \varepsilon)}$. For $\varepsilon \in \left[0, \sqrt{3}/4\right]$, $\mu_S(N_{PE}^r, \varepsilon) \to 1$ and $\nu(N_{PE}^r, \varepsilon) \to 0$ as $r \to r_\delta(\varepsilon) = \sqrt{3}/(2\varepsilon)$. Hence $\text{HLAE}^S(N_{PE}^r, \varepsilon) \to \infty$ as $r \to r_\delta(\varepsilon) = \sqrt{3}/(2\varepsilon)$. So for $\varepsilon \in \left[0, \sqrt{3}/4\right]$, $\widetilde{r} = \sqrt{3}/(2\varepsilon)$. For $\varepsilon \in \left(\sqrt{3}/4, \sqrt{3}/3\right]$, the result follows similarly. ■

So HLAE suggests choosing r larger as the segregation gets more severe, but choosing r too large will reduce power since $r \geq r_\delta(\varepsilon)$ guarantees the complete digraph under the alternative and, as r increases therefrom, provides an ever greater probability of seeing the complete digraph under the null.

In Figure 5.4.19, we plot the graphs of mean and asymptotic variance for $r \in [1, 4]$ under segregation with $\varepsilon \in \{0, \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$. Notice that $\mu_S(N_{PE}^r, \varepsilon)$ gets larger as ε gets larger at each r which is in agreement with the $\mu_S(N_{PE}^r, \varepsilon)$ expressions in Appendix Section B.4.1.

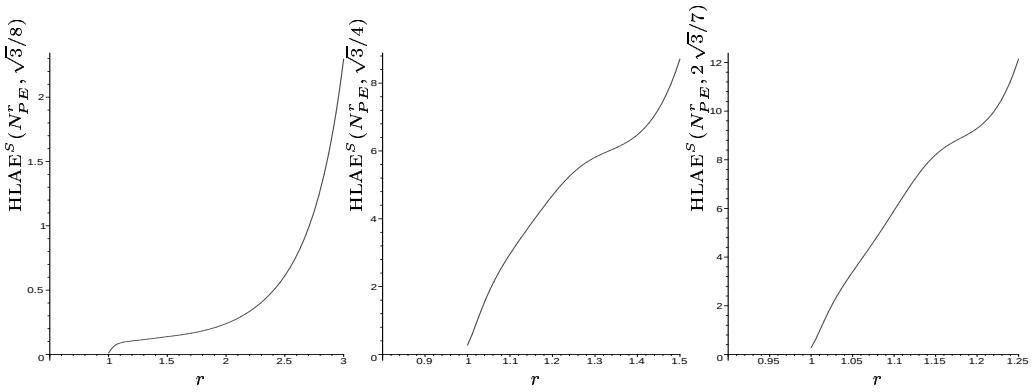


Figure 5.4.18: Hodges-Lehmann asymptotic efficacy against segregation alternative H_ε^S as a function of r for $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$ (left to right).

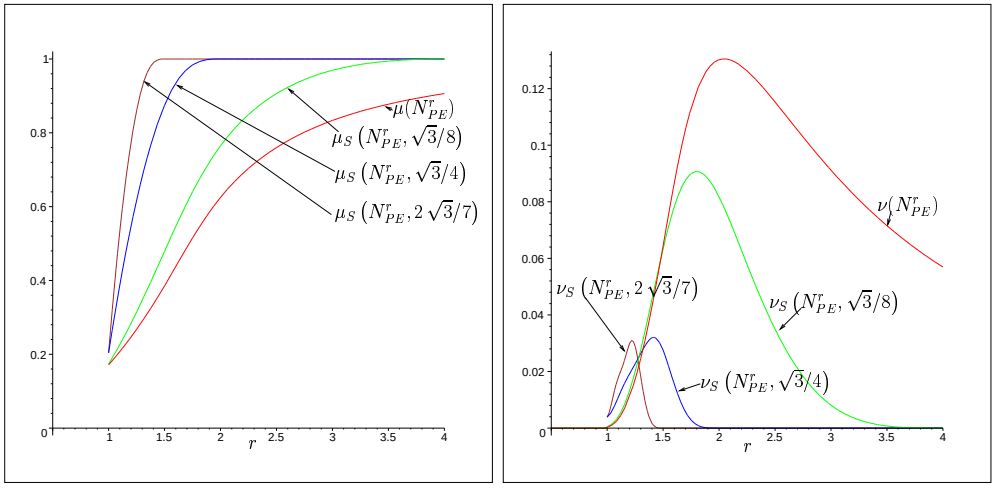


Figure 5.4.19: The mean $\mu_S(N_{PE}^r, \varepsilon)$ (left) and asymptotic variance $\nu_S(N_{PE}^r, \varepsilon)$ (right) as a function of r under segregation with $\varepsilon \in \{0, \sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$.

However, the same ordering holds for $\nu_S(N_{PE}^r, \varepsilon)$ at each r only for large r, but for small r the ordering is reversed. Furthermore, both the $\sup_{r\in[1,\infty]} \nu_S(N_{PE}^r, \varepsilon)$ and $\text{argsup}_{r\in[1,\infty]} \nu_S(N_{PE}^r, \varepsilon)$ seem to decrease as ε increases.

5.4.9.2 Hodges-Lehmann Asymptotic Efficacy Under Association Alternatives

In the HLAE framework of Section 2.7.2.2, $\theta = \mathbf{E}_\varepsilon^A[\rho_n(N_{PE}^r)] = \mu_A(N_{PE}^r, \varepsilon)$ and $\theta_0 = \mu_A(N_{PE}^r, \varepsilon = 0)$. Then testing $H_0 : \varepsilon = 0$ versus $H_a : \varepsilon > 0$ is equivalent to $H_0 : \mathbf{E}[\rho_n] = \mu(N_{PE}^r)$ versus $H_a : \mathbf{E}_\varepsilon^A[\rho_n] = \mu_A(N_{PE}^r, \varepsilon) < \mu(N_{PE}^r)$. Let $\delta(r, \varepsilon) = \frac{\mu_A(N_{PE}^r, \varepsilon) - \mu(N_{PE}^r)}{\sigma_n(r,\varepsilon)}$ and $\widetilde{R}(r) = \frac{\rho_n(N_{PE}^r) - \mu(N_{PE}^r)}{\sigma_n(r,\varepsilon)}$, then $\widetilde{R}(r) \xrightarrow{\mathcal{L}} \mathcal{N}(\delta(r, \varepsilon), 1)$.

Hodges-Lehmann asymptotic efficacy (HLAE) [16] is given by

$$\mathrm{HLAE}^A(N_{PE}^r,\varepsilon) := \delta(r,\varepsilon)^2 = \frac{\left(\mu_A(N_{PE}^r,\varepsilon)-\mu(N_{PE}^r)\right)^2}{\nu_A(N_{PE}^r,\varepsilon)}.$$

Rather than an arbitrary ε we pick specific values: $5\sqrt{3}/24$, $\sqrt{3}/12$, and $\sqrt{3}/21$. Recall that $\rho_n(N_{PE}^r)$ is degenerate as $r\to\infty$. Furthermore, $\rho_n(N_{PE}^r)$ is degenerate when $r=1$.

With $\varepsilon = 5\sqrt{3}/24$, the mean is given by

$$\mu_A\left(N_{PE}^r, 5\sqrt{3}/24\right) = \begin{cases} -\frac{1}{6}r^{-2}+\frac{1}{3}, & \text{for} \quad r\in[1,3), \\ \frac{1}{3}r^2-\frac{8}{3}r-\frac{55}{6}r^{-2}+\frac{19}{3}, & \text{for} \quad r\in[3,4), \\ -\frac{55}{6}r^{-2}+1 & \text{for,} \quad r\in[4,\infty), \end{cases}$$

and the variance is given by

$$\nu_A\left(N_{PE}^r, 5\sqrt{3}/24\right) = \sum_{j=1}^{5}\nu_j\left(r, 5\sqrt{3}/24\right)\mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned}
\nu_1\left(r,5\sqrt{3}/24\right) &= \frac{r^4-2r^2+1}{27r^6}, \\
\nu_2\left(r,5\sqrt{3}/24\right) &= -\Big[120r^{10}-2176r^9+15340r^8-50304r^7+58754r^6+74880r^5-248577r^4 \\
&\quad +138240r^3+47172r^2+23328r-7305\Big]\Big/\Big[405r^6\Big], \\
\nu_3\left(r,5\sqrt{3}/24\right) &= -\Big[120r^{10}-2176r^9+15180r^8-48960r^7+58754r^6+47440r^5-176547r^4 \\
&\quad +138240r^3-70477r^2+23328r-7305\Big]\Big/\Big[405r^6\Big], \\
\nu_4\left(r,5\sqrt{3}/24\right) &= \Big[10r^{12}-192r^{11}+1320r^{10}-2944r^9-7590r^8+49920r^7-69986r^6-46480r^5 \\
&\quad +184137r^4-143360r^3+71917r^2-23520r+7315\Big]\Big/\Big[405r^6\Big], \\
\nu_5\left(r,5\sqrt{3}/24\right) &= \frac{787r^4-7601r^2-16032r+9265}{135r^6}.
\end{aligned}$$

The corresponding intervals are $\mathcal{I}_1=(1,3)$, $\mathcal{I}_2=[3,7/2)$, $\mathcal{I}_3=\left[7/2,2+\sqrt{3}\right)$, $\mathcal{I}_4=\left[2+\sqrt{3},4\right)$, and $\mathcal{I}_5=[4,\infty)$. See Figure 5.4.21 for the graph of $\mu_A\left(N_{PE}^r,5\sqrt{3}/24\right)$ and $\nu_A\left(N_{PE}^r,5\sqrt{3}/24\right)$.

Then we get $\mathrm{HLAE}^A\left(N_{PE}^r,5\sqrt{3}/24\right) = \frac{\left(\mu_A\left(N_{PE}^r,5\sqrt{3}/24\right)-\mu(N_{PE}^r)\right)^2}{\nu_A\left(N_{PE}^r,5\sqrt{3}/24\right)}$ by substituting the rele-

vant terms. See Figure 5.4.20.

With $\varepsilon = \sqrt{3}/12$, the mean is given by

$$\mu_A\left(N_{PE}^r, \sqrt{3}/12\right) = \begin{cases} \frac{6\,r^4 - 16\,r^3 + 18\,r^2 - 5}{18\,r^2}, & \text{for} \quad r \in [1,2), \\ -\frac{37}{18}\,r^{-2} + 1, & \text{for} \quad r \in [2,\infty), \end{cases}$$

and the variance is given by

$$\nu_A\left(N_{PE}^r, \sqrt{3}/12\right) = \sum_{j=1}^{3} \nu_j\left(r, \sqrt{3}/12\right)\,\mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned} \nu_1\left(r, \sqrt{3}/12\right) &= \left[10\,r^{12} - 96\,r^{11} + 240\,r^{10} + 192\,r^9 - 1830\,r^8 + 3360\,r^7 - 2650\,r^6 + 240\,r^5\right. \\ &\quad \left. + 1383\,r^4 - 1280\,r^3 + 540\,r^2 - 144\,r + 35\right] \Big/ \left[405\,r^6\right], \\ \nu_2\left(r, \sqrt{3}/12\right) &= \left[10\,r^{12} - 96\,r^{11} + 240\,r^{10} + 192\,r^9 - 1670\,r^8 + 2784\,r^7 - 2650\,r^6 + 2400\,r^5\right. \\ &\quad \left. - 1047\,r^4 - 1280\,r^3 + 1269\,r^2 - 144\,r + 35\right] \Big/ \left[405\,r^6\right], \\ \nu_3\left(r, \sqrt{3}/12\right) &= \frac{537\,r^4 - 683\,r^2 - 2448\,r + 1315}{405\,r^6}. \end{aligned}$$

The corresponding intervals are $\mathcal{I}_1 = [1, 3/2)$, $\mathcal{I}_2 = [3/2, 2)$, and $\mathcal{I}_3 = [2, \infty)$. See Figure 5.4.21 for the graph of $\mu_A\left(N_{PE}^r, \sqrt{3}/12\right)$ and $\nu_A\left(N_{PE}^r, \sqrt{3}/12\right)$.

Then we get $\mathrm{HLAE}^A\left(N_{PE}^r, \sqrt{3}/12\right) = \frac{\left(\mu_A\left(N_{PE}^r, \sqrt{3}/12\right) - \mu(N_{PE}^r)\right)^2}{\nu_A\left(N_{PE}^r, \sqrt{3}/12\right)}$ by substituting the relevant terms. See Figure 5.4.20.

With $\varepsilon = \sqrt{3}/21$, the mean is given by

$$\mu_A\left(N_{PE}^r, \sqrt{3}/21\right) = \begin{cases} \frac{7839\,r^4 - 27648\,r^3 + 49152\,r^2 - 35840\,r + 9216}{16200\,r^2}, & \text{for} \quad r \in [1, 8/7), \\ \frac{2719\,r^4 - 5592\,r^3 + 5760\,r^2 - 1536}{8100\,r^2}, & \text{for} \quad r \in (8/7, 3/2), \\ \frac{53\,r^4 + 2744\,r^3 - 7296\,r^2 + 8064\,r - 3104}{2700\,r^2}, & \text{for} \quad r \in (3/2, 12/7), \\ \frac{2719\,r^4 - 1440\,r^2 + 2112}{16200\,r^2}, & \text{for} \quad r \in (12/7, 7/4), \\ -\frac{2401\,r^4 - 73824\,r^2 + 153664\,r - 88548}{16200\,r^2}, & \text{for} \quad r \in (7/4, 2), \\ 1 - \frac{89}{54}\,r^{-2}, & \text{for} \quad r \in [2, \infty), \end{cases}$$

and the variance is given by

$$\nu_A\left(N_{PE}^r, \sqrt{3}/21\right) = \sum_{j=1}^{10} \nu_j\left(r, \sqrt{3}/21\right)\, \mathbf{I}(\mathcal{I}_j)$$

where

$$\begin{aligned}
\nu_1\left(r, \sqrt{3}/21\right) &= \Big[4124031\, r^{12} - 22708224\ r^{11} - 389826\, r^{10} + 369129408\, r^9 - 1592672721\, r^8 \\
&+ 3532359672\, r^7 - 4721848374\ r^6 + 4050858048\, r^5 - 2387433568\, r^4 + 995033088\, r^3 \\
&- 209048784\ r^2 - 43352064\ r + 25952256\Big] \Big/ \Big[65610000\, r^6\Big], \\
\nu_2\left(r, \sqrt{3}/21\right) &= \Big[6594660\, r^{12} - 31178952\, r^{11} - 14911074\ r^{10} + 441735648\, r^9 - 1578842961\, r^8 \\
&+ 3311083512\, r^7 - 4669163574\ r^6 + 4366966848\, r^5 - 2522908768\, r^4 + 778272768\, r^3 \\
&- 93443280\, r^2 + 14450688\, r - 8650752\Big] \Big/ \Big[65610000\, r^6\Big], \\
\nu_3\left(r, \sqrt{3}/21\right) &= \Big[826701\, r^{12} - 7118748\, r^{11} + 14155864\ r^{10} + 18467640\, r^9 - 104968680\, r^8 \\
&+ 165877272\, r^7 - 128355690\, r^6 + 27338184\ r^5 + 47304144\, r^4 - 52684800\, r^3 \\
&+ 24413592\, r^2 - 7225344\, r + 1966080\Big] \Big/ \Big[32805000\, r^6\Big], \\
\nu_4\left(r, \sqrt{3}/21\right) &= \Big[826701\, r^{12} - 7118748\, r^{11} + 14155864\ r^{10} + 18467640\, r^9 + 20074008\, r^8 \\
&- 671672808\, r^7 + 2194076310\, r^6 - 3382581816\, r^5 + 2840904144\, r^4 - 1262284800\, r^3 \\
&+ 240413592\, r^2 - 7225344\, r + 1966080\Big] \Big/ \Big[32805000\, r^6\Big], \\
\nu_5\left(r, \sqrt{3}/21\right) &= \Big[826701\, r^{12} - 7118748\, r^{11} + 14155864\ r^{10} + 18467640\, r^9 - 137116617\, r^8 \\
&+ 512952192\, r^7 - 1511673690\, r^6 + 2773418184\ r^5 - 2883095856\, r^4 + 1560115200\, r^3 \\
&- 335586408\, r^2 - 7225344\, r + 1966080\Big] \Big/ \Big[32805000\, r^6\Big], \\
\nu_6\left(r, \sqrt{3}/21\right) &= \Big[826701\, r^{12} - 7118748\, r^{11} + 14155864\ r^{10} + 18467640\, r^9 - 91939401\, r^8 \\
&+ 125718912\, r^7 - 128697690\, r^6 + 139178184\ r^5 - 60695856\, r^4 - 52684800\, r^3 \\
&+ 48413592\, r^2 - 7225344\, r + 1966080\Big] \Big/ \Big[32805000\, r^6\Big], \\
\nu_7\left(r, \sqrt{3}/21\right) &= \Big[226415\, r^{12} - 1426740\, r^{11} + 334536\, r^{10} + 17196648\, r^9 - 87678147\, r^8 \\
&+ 311364480\, r^7 - 711864862\, r^6 + 944809880\, r^5 - 684036240\, r^4 + 238099456\, r^3 \\
&- 24048504\ r^2 - 7633920\, r + 4761088\Big] \Big/ \Big[10935000\, r^6\Big],
\end{aligned}$$

$$\nu_8\left(r,\sqrt{3}/21\right) = \Big[5786907\,r^{12} - 42712488\,r^{11} + 76274888\,r^{10} + 51865788\,r^8 - 300043296\,r^7$$
$$+ 132202536\,r^6 + 171413760\,r^5 - 93614976\,r^4 + 147517440\,r^3 - 194460480\,r^2$$
$$+ 67608576\,r - 29061120\Big] \Big/ \Big[262440000\,r^6\Big],$$
$$\nu_9\left(r,\sqrt{3}/21\right) = -\Big[2470629\,r^{12} - 25412184\,r^{11} + 112001848\,r^{10} - 1958438076\,r^8$$
$$+ 5449924256\,r^7 + 6150612888\,r^6 - 55820599296\,r^5 + 109663683136\,r^4$$
$$- 97335694848\,r^3 + 40552466112\,r^2 - 9825887232\,r + 3078523200\Big] \Big/ \Big[262440000\,r^6\Big],$$
$$\nu_{10}\left(r,\sqrt{3}/21\right) = \frac{493829\,r^4 - 433645\,r^2 - 1765008\,r + 929955}{455625\,r^6}.$$

The corresponding intervals are $\mathcal{I}_1 = \left(1, 2\sqrt{14}/7\right)$, $\mathcal{I}_2 = \left[2\sqrt{14}/7, 8/7\right)$, $\mathcal{I}_3 = [8/7, 5/4)$, $\mathcal{I}_4 = [5/4, 4/3)$, $\mathcal{I}_5 = [4/3, 10/7)$, $\mathcal{I}_6 = [10/7, 3/2)$, $\mathcal{I}_7 = [3/2, 12/7)$, $\mathcal{I}_8 = [12/7, 7/4)$, $\mathcal{I}_9 = [7/4, 2)$, and $\mathcal{I}_{10} = [2, \infty)$. See Figure 5.4.21 for the graph of $\mu_A\left(N_{PE}^r, \sqrt{3}/21\right)$ and $\nu_A\left(N_{PE}^r, \sqrt{3}/21\right)$.

Then we get $\mathrm{HLAE}^A\left(N_{PE}^r, \sqrt{3}/21\right) = \frac{\left(\mu_A\left(N_{PE}^r, \sqrt{3}/21\right) - \mu(N_{PE}^r)\right)^2}{\nu_A\left(N_{PE}^r, \sqrt{3}/21\right)}$ by substituting the relevant terms. See Figure 5.4.20.

Notice that for $\varepsilon = 5\sqrt{3}/24$ and $\varepsilon = \sqrt{3}/12$, $\mathrm{argsup}_{r\geq 1}\,\mathrm{HLAE}^A(N_{PE}^r, \varepsilon) = 1$. This result for HLAE can be generalized for arbitrary ε as follows.

Proposition 5.4.8. *Let* $r^* := \mathrm{argsup}_{r\geq 1}\,\mathrm{HLAE}^A(N_{PE}^r, \varepsilon)$ *and* $\varepsilon \geq \sqrt{3}/12$. *Then* $r^* = 1$.

Proof: Recall that $\mathrm{HLAE}^A(N_{PE}^r, \varepsilon) = \frac{(\mu_A(N_{PE}^r, \varepsilon) - \mu(N_{PE}^r))^2}{\nu_A(N_{PE}^r, \varepsilon)}$. For $\varepsilon \in \left[\sqrt{3}/12, \sqrt{3}/3\right)$, $\mu_A(N_{PE}^r = 1, \varepsilon) \to 1$ and $\nu_A(N_{PE}^r = 1, \varepsilon) \to 0$. Hence $\mathrm{HLAE}^A(N_{PE}^r, \varepsilon) \to \infty$ as $r \to 1$. So the desired result follows. ■

For $\varepsilon \in \left[0, \sqrt{3}/12\right]$, it seems that for a while $r^* = 1$ with respect to HLAE, e.g.,for $\varepsilon = \sqrt{3}/21$. But for sufficiently small ε, $r^* > 1$ holds. This can also be seen as $\varepsilon \to 0$ in which case HLAE becomes PAE and the optimal value is about 1.006 with respect to PAE. Furthermore, observe that the argsup for HLAE gets closer to 1 as $\varepsilon \to 0$ and $\nu_A(N_{PE}^r, \varepsilon) > 0$ for $\varepsilon \in \left(0, \sqrt{3}/12\right)$ and $\nu(N_{PE}^r, \varepsilon)$ gets larger as $\varepsilon \to 0$.

Figure 5.4.20 contains a graph of HLAE against association as a function of r for $\varepsilon \in \{5\sqrt{3}/24,\ \sqrt{3}/12,\ \sqrt{3}/21\}$. Notice that since $\nu_A(N_{PE}^{r=1}, \varepsilon) = 0$ for $\varepsilon \geq \sqrt{3}/12$, $\mathrm{HLAE}^A(N_{PE}^{r=1}, \varepsilon) = \infty$ for $\varepsilon \geq \sqrt{3}/12$ and $\lim_{r\to\infty} \mathrm{HLAE}^A(N_{PE}^r, \varepsilon) = 0$. In Figure 5.4.20 we see that, against H_ε^A, $\mathrm{HLAE}^A(N_{PE}^r, \varepsilon)$ has a local supremum for some $r > 1$. Let $\widetilde{r}_l$ be the value at which this local supremum is attained. Then $\widetilde{r}_l\left(5\sqrt{3}/24\right) \approx 3.2323$, $\widetilde{r}_l\left(\sqrt{3}/12\right) \approx 1.5676$, and $\widetilde{r}_l\left(\sqrt{3}/21\right) \approx$

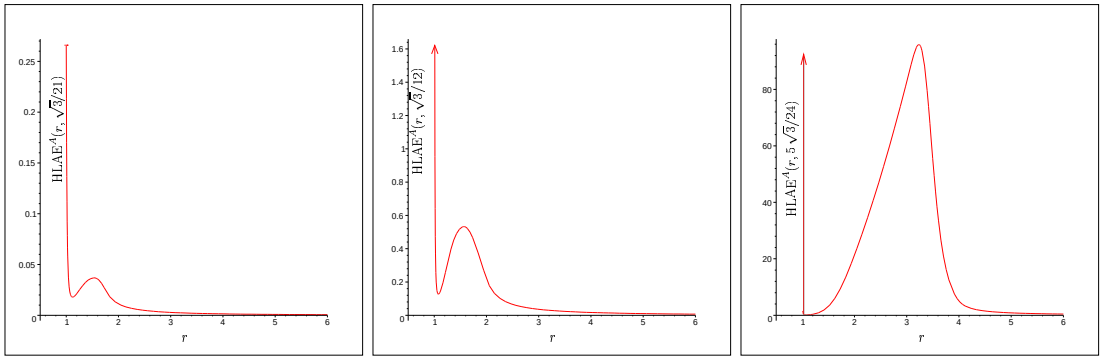


Figure 5.4.20: Hodges-Lehmann asymptotic efficacy against association alternative H_ε^A as a function of r for $\varepsilon \in \left\{\sqrt{3}/21,\ \sqrt{3}/12,\ 5\sqrt{3}/24\right\}$ (left to right).

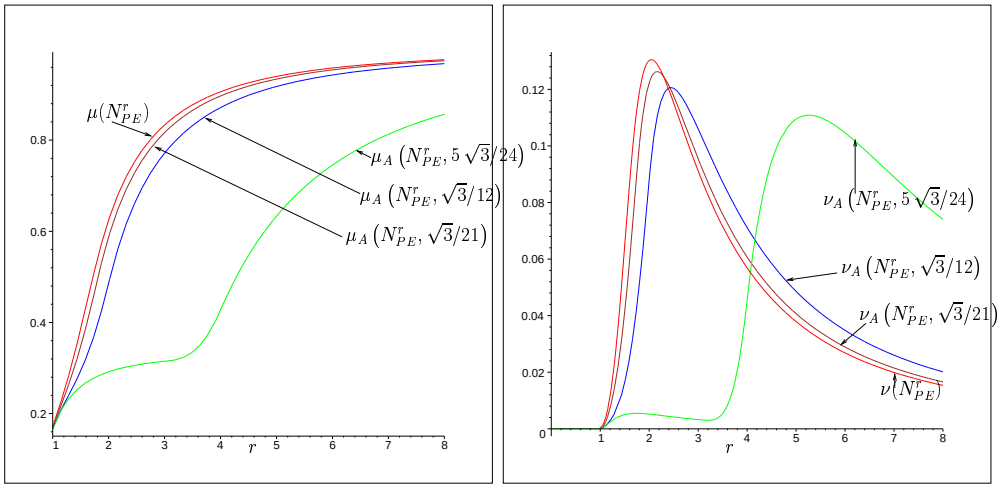


Figure 5.4.21: The mean $\mu_A(N_{PE}^r, \varepsilon)$ (left) and asymptotic variance $\nu_A(N_{PE}^r, \varepsilon)$ (right) as a function of r under association with $\varepsilon \in \left\{0, \sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24\right\}$.

1.533. Note that, as ε gets smaller, $\widetilde{r}_l$ gets smaller. Furthermore, $\text{HLAE}^A\left(r = 1, \sqrt{3}/21\right) < \infty$ and as $\varepsilon \to 0$, so $\widetilde{r}_l$ becomes the global supremum and $\text{PAE}^A(N_{PE}^{r=1}) = 0$ and $\text{argsup}_{r \geq 1} \text{PAE}^A($ $N_{PE}^{r=1}) \approx 1.006$. So HLAE suggests choosing moderate r when testing against association, whereas PAE suggests choosing small r.

Derivation of $\mu_A(N_{PE}^r, \varepsilon)$ and $\nu_A(N_{PE}^r, \varepsilon)$ or association with $\varepsilon \in \left\{5\sqrt{3}/24,\ \sqrt{3}/12,\ \sqrt{3}/21\right\}$ are similar to — with the supports being the complements of— the corresponding segregation cases.

In Figure 5.4.21, we plot the graphs of mean and asymptotic variance for $r \in [1, 8]$ under association with $\varepsilon \in \{0, \sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24\}$. Notice that $\mu_A(N_{PE}^r, \varepsilon)$ gets smaller as ε gets larger at each r which is in agreement with the $\mu_A(N_{PE}^r, \varepsilon)$ expressions in Appendix Section

B.4.2. However, the same ordering does not hold for $\nu_A(N_{PE}^r, \varepsilon)$ at each r. For small r the ordering is same as in $\mu_A(N_{PE}^r, \varepsilon)$, but for large r the ordering is reversed. Furthermore, $\sup_{r\in[1,\infty]} \nu_A(N_{PE}^r, \varepsilon)$ seems to decrease as ε increases while $\text{argsup}_{r\in[1,\infty]} \nu_A(N_{PE}^r, \varepsilon)$ seems to increase as ε increases.

5.4.10 Asymptotic Power Function Analysis

The asymptotic power function (see e.g.,[22]) can also be investigated as a function of r, n, and ε using the asymptotic critical value and an appeal to normality.

5.4.10.1 Asymptotic Power Function Analysis Under Segregation

Under segregation, for sufficiently large n, we reject H_0 when $\sqrt{n}\left(\frac{\rho_n(N_{PE}^r)-\mu_S(N_{PE}^r,\varepsilon)}{\sqrt{\nu_S(N_{PE}^r,\varepsilon)}}\right) > z_{(1-\alpha)}$. Then size α critical region for large samples is

$$\rho_n(N_{PE}^r) > \mu(N_{PE}^r) + z_{(1-\alpha)} \cdot \sqrt{\nu(N_{PE}^r)/n}.$$

Under a specific segregation alternative H_ε^S, the asymptotic power function is given by

$$\begin{aligned}\Pi_S(N_{PE}^r, n, \varepsilon) &:= P\left(\rho_n(N_{PE}^r) > \mu(N_{PE}^r) + z_{(1-\alpha)} \cdot \sqrt{\nu(N_{PE}^r)/n}\right) \\ &= 1 - \Phi\left(\frac{z_{(1-\alpha)}\sqrt{\nu(N_{PE}^r)}}{\sqrt{\nu_S(N_{PE}^r,\varepsilon)}} + \frac{\sqrt{n}\,(\mu(N_{PE}^r) - \mu_S(N_{PE}^r,\varepsilon))}{\sqrt{\nu_S(N_{PE}^r,\varepsilon)}}\right).\end{aligned}$$

With $\varepsilon = \sqrt{3}/8$, $\Pi_S(N_{PE}^r, n, \varepsilon)$ at level $\alpha = .05$ is plotted in Figure 5.4.22. Observe that $\Pi_S(N_{PE}^r, n, \sqrt{3}/8) \to 0$ as $r \to 4$ for $n \in \{5, 10, 15\}$. Let $r_g^*(n, \varepsilon)$ be the the value at which $\Pi_S(N_{PE}^r, n, \varepsilon)$ attains its global supremum and $r_l^*(n, \varepsilon)$ be the the value at which $\Pi_S(N_{PE}^r, n, \varepsilon)$ attains its local supremum. Then $r_g^*(5, \sqrt{3}/8) \approx 1.260$, $r_g^*(10, \sqrt{3}/8) \approx 1.3741$ and $r_l^*(10, \sqrt{3}/8) \approx 2.3818$, $r_g^*(15, \sqrt{3}/8) \approx 3.3724$ and $r_l^*(15, \sqrt{3}/8) \approx 1.45$, $r_g^*(20, \sqrt{3}/8) = 4$ and $r_l^*(20, \sqrt{3}/8) \approx 1.5$. Finally, $r_g^*(n, \sqrt{3}/8) = 4$ for $n \in \{20, 50, 100\}$ and $\Pi_S(N_{PE}^r, n, \sqrt{3}/8)$ has a hump for $n = 10$ and $n = 15$.

With $\varepsilon = \sqrt{3}/4$, $\Pi_S(N_{PE}^r, n, \varepsilon)$ at level $\alpha = .05$ is plotted in Figure 5.4.23. Observe that $\Pi_S(N_{PE}^r, n, \sqrt{3}/4) \to 1$ as $r \to 2$ for $n \in \{3, 5\}$. Moreover, $r_g^*(n, \sqrt{3}/4) = 2$, for $n = 3, 5$.

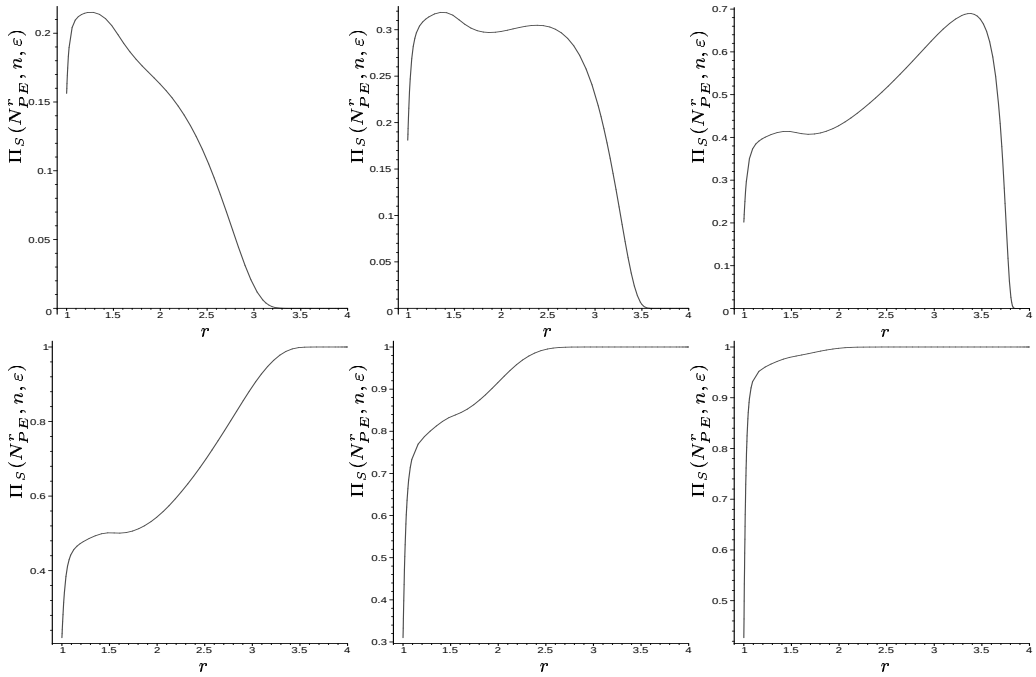


Figure 5.4.22: Asymptotic power function against segregation alternative $H^S_{\sqrt{3}/8}$ as a function of r for $n \in \{5, 10, 15, 20, 50, 100\}$.

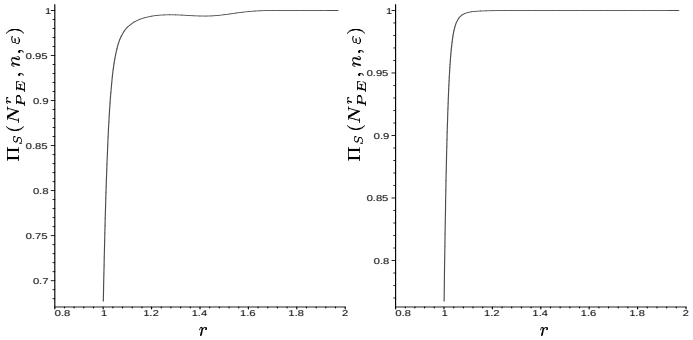


Figure 5.4.23: Asymptotic power function against segregation alternative $H^S_{\sqrt{3}/4}$ as a function of r for $n = 3$ (left) and $n = 5$ (right).

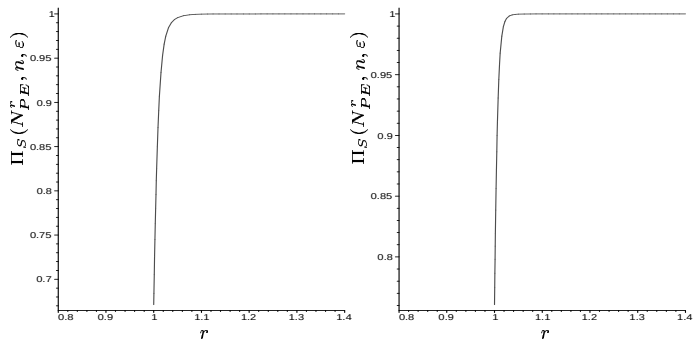


Figure 5.4.24: Asymptotic power function against segregation alternative $H^S_{\sqrt{3}/7}$ as a function of r for $n = 3$ (left) and $n = 5$ (right).

With $\varepsilon = 2\sqrt{3}/7$, $\Pi_S(N_{PE}^r, n, \varepsilon)$ at level $\alpha = .05$ is plotted in Figure 5.4.24. Observe that $\Pi_S(N_{PE}^r, n, 2\sqrt{3}/7) \to 1$ as $r \to 3/2$ for $n \in \{3, 5\}$ and $r_g^*(n, 2\sqrt{3}/7) = 2$, for $n \in \{3, 5\}$.

5.4.10.2 Asymptotic Power Function Analysis Under Association

Under association, for sufficiently large n, we reject H_0 when $\sqrt{n}\left(\frac{\rho_n(N_{PE}^r) - \mu_A(N_{PE}^r, \varepsilon)}{\sqrt{\nu_A(N_{PE}^r, \varepsilon)}}\right) < z_\alpha$. Then size α critical region for large samples is

$$\rho_n(N_{PE}^r) < \mu(N_{PE}^r) + z_\alpha \cdot \sqrt{\nu(N_{PE}^r)/n}.$$

Under H_ε^A, we have

$$\begin{aligned} \Pi_A(N_{PE}^r, n, \varepsilon) &:= P\left(\rho_n(N_{PE}^r) < \mu(N_{PE}^r) + z_\alpha \cdot \sqrt{\nu(N_{PE}^r)/n}\right) \\ &= \Phi\left(\frac{z_\alpha \sqrt{\nu(N_{PE}^r)}}{\sqrt{\nu_A(N_{PE}^r, \varepsilon)}} + \frac{\sqrt{n}\left(\mu(N_{PE}^r) - \mu_A(N_{PE}^r, \varepsilon)\right)}{\sqrt{\nu_A(N_{PE}^r, \varepsilon)}}\right). \end{aligned}$$

With $\varepsilon = \sqrt{3}/21$, $\Pi_A(N_{PE}^r, n, \varepsilon)$ at level $\alpha = .05$ is plotted in Figure 5.4.25. Observe that $\Pi_A\left(n, r, \sqrt{3}/21\right) \to .057$ as $r \to \infty$ for $n = 5, 10, 100$. Let $\widehat{r}\,(n, \varepsilon)$ be the value at which $\Pi_A(N_{PE}^r, n, \varepsilon)$ attains its supremum. Then, $\widehat{r}\,\left(5, \sqrt{3}/21\right) \approx 2.01$, and $\widehat{r}\,\left(10, \sqrt{3}/21\right) \approx 1.875$, and $\widehat{r}\,\left(100, \sqrt{3}/21\right) \approx 1.645$. Moreover, $\Pi_A\left(r, 100, \sqrt{3}/21\right)$ attains a local infimum at ≈ 1.065.

With $\varepsilon = \sqrt{3}/12$, $\Pi_A(N_{PE}^r, n, \varepsilon)$ at level $\alpha = .05$ is plotted in Figure 5.4.26. Observe that $\Pi_A\left(n, r, \sqrt{3}/12\right) \to .0766$ as $r \to \infty$ for $n = 5, 10, 100$. Moreover, $\widehat{r}\,\left(5, \sqrt{3}/12\right) \approx 1.99$, $\widehat{r}\,\left(10, \sqrt{3}/12\right) \approx 1.75$, and $\widehat{r}\,\left(100, \sqrt{3}/12\right) \approx 1.60$. Moreover, $\Pi_A\left(r, 100, \sqrt{3}/21\right)$ attains a local infimum at ≈ 1.105.

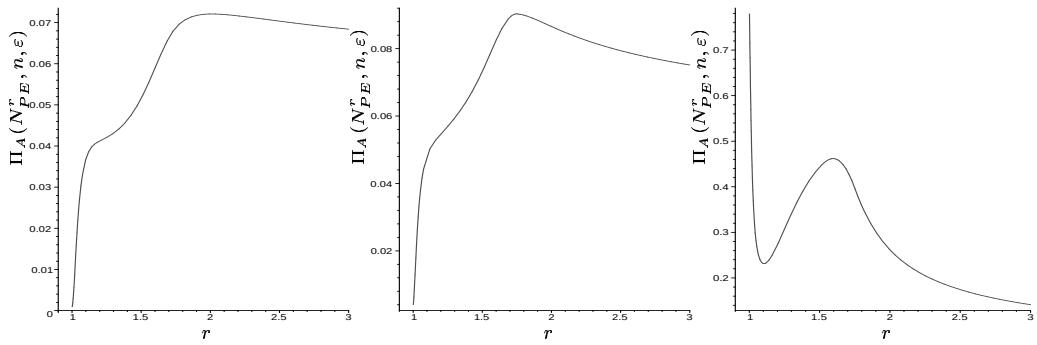


Figure 5.4.25: Asymptotic power function against association alternative $H^A_{\sqrt{3}/21}$ as a function of r for $n = 5, 10, 100$

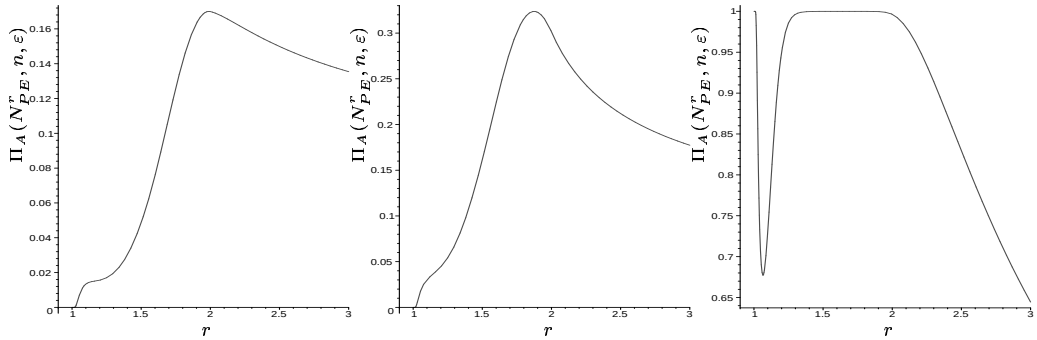


Figure 5.4.26: Asymptotic power function against association alternative $H^A_{\sqrt{3}/12}$ as a function of r for $n \in \{5, 10, 100\}$

With $\varepsilon = 5\sqrt{3}/24$, $\Pi_A(N^r_{PE}, n, \varepsilon)$ at level $\alpha = .05$ is plotted in Figure 5.4.27. Observe that $\Pi\left(n, r, 5\sqrt{3}/24\right) \to 1$ as $r \to 2$ for $n \in \{5, 10\}$.

5.4.11 Multiple Triangle Case

Suppose $\mathcal{Y}$ is a finite collection of points in $\mathbb{R}^2$ with $|\mathcal{Y}| \geq 3$. Consider the Delaunay triangulation (assumed to exist) of $\mathcal{Y}$, where T_j denotes the j^{th} Delaunay triangle, J denotes the number of triangles, and $C_H(\mathcal{Y})$ denotes the convex hull of $\mathcal{Y}$. We wish to test

$$H_0 : X_i \overset{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}))$$

against segregation and association alternatives.

Figure 5.4.28 and Figure 4.6.1 are graphs of realizations of $n = 100$ and $n = 1000$ observations which are independent and identically distributed according to $\mathcal{U}(C_H(\mathcal{Y}))$ for $|\mathcal{Y}| = 10$ and

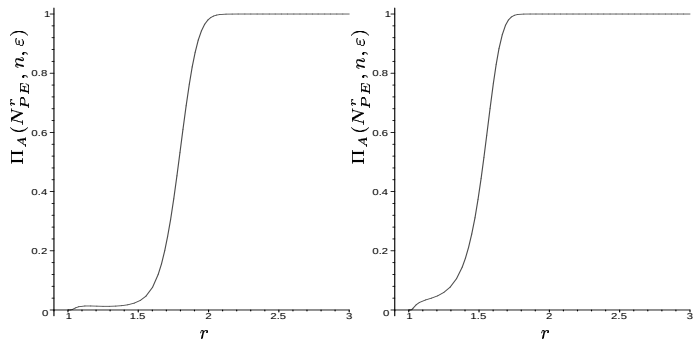


Figure 5.4.27: Asymptotic power function against association alternative $H^A_{5\sqrt{3}/24}$ as a function of r for $n \in \{5, 10\}$.

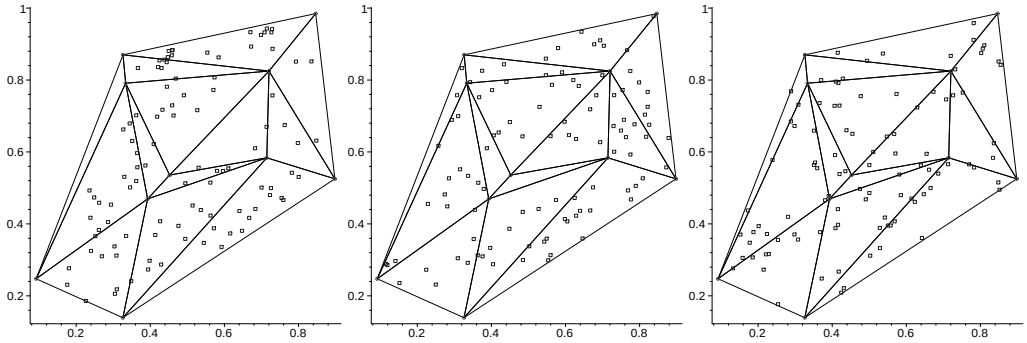


Figure 5.4.28: Realization of segregation (left), H_0 (middle), and association (right) for $|\mathcal{Y}| = 10$, $J = 13$, and $n = 100$.

$J = 13$ and realizations under segregation and association for the same $\mathcal{Y}$, respectively.

The digraph D is constructed using $N_{PE}^r(\cdot, \mathcal{Y}_j)$ as described in Section 1.1, here for $X_i \in T_j$ the three points in $\mathcal{Y}$ defining the Delaunay triangle T_j are used as $\mathcal{Y}_j$. Let $\rho_n(N_{PE}^r, J)$ be the relative density of the digraph based on $\mathcal{X}_n$ and $\mathcal{Y}$ which yields J Delaunay triangles, and let $w_j = A(T_j)/A(C_H(\mathcal{Y}))$ for $j \in \{1, \ldots, J\}$ where $A(C_H(\mathcal{Y})) = \sum_{j=1}^J A(T_j)$. Then we obtain the following as a corollary to Theorem 5.4.2.

Corollary 5.4.9. *The asymptotic null distribution for $\rho_n(N_{PE}^r, J)$ conditional on $\mathcal{W} = \{w_1, \ldots, w_J\}$ for $r \in [1, \infty)$ is given by $\mathcal{N}(\mu(N_{PE}^r, J), \nu(N_{PE}^r, J)/n)$ provided that $\nu(N_{PE}^r, J) > 0$ with*

$$\begin{aligned} \mu(N_{PE}^r, J) &:= \mu(N_{PE}^r) \sum_{j=1}^J w_j^2 \quad \textit{and} \\ \nu(N_{PE}^r, J) &:= \nu(N_{PE}^r) \sum_{j=1}^J w_j^3 + 4\mu(N_{PE}^r)^2 \left[\sum_{j=1}^J w_j^3 - \left(\sum_{j=1}^J w_j^2 \right)^2 \right], \end{aligned} \tag{5.4.7}$$

where $\mu(N_{PE}^r)$ and $\nu(N_{PE}^r)$ are given in Equations (5.4.2) and (5.4.3), respectively.

Proof: In the multiple triangle case,

$$\mu(N_{PE}^r, J) = \mathbf{E}\left[\rho_n(N_{PE}^r, J)\right] = \frac{1}{n\,(n-1)} \sum_{i<j} \sum \mathbf{E}\left[h_{ij}(N_{PE}^r)\right] = \frac{1}{2}\mathbf{E}\left[h_{12}(N_{PE}^r)\right]$$

$$= \mathbf{E}\left[\mathbf{I}\left(A_{12}\right)\right] = P\big(X_2 \in N_{PE}^r(X_1)\big).$$

By definition of $N_{PE}^r(\cdot)$, $P\big(X_2 \in N_{PE}^r(X_1)\big) = 0$ if X_1 and X_2 are in different triangles. So by the law of total probability

$$\mu(N_{PE}^r, J) := P\big(X_2 \in N_{PE}^r(X_1)\big) = \sum_{j=1}^{J} P\big(X_2 \in N_{PE}^r(X_1) \,|\, \{X_1, X_2\} \subset T_j\big)\, P\big(\{X_1, X_2\} \subset T_j\big)$$

$$= \sum_{j=1}^{J} \mu(N_{PE}^r)\, P\big(\{X_1, X_2\} \subset T_j\big) \quad \left(\text{since } P\big(X_2 \in N_{PE}^r(X_1) \,|\, \{X_1, X_2\} \subset T_j\big) = \mu(N_{PE}^r)\right)$$

$$= \mu(N_{PE}^r) \sum_{j=1}^{J} \Big[A(T_j)/A(C_H(\mathcal{Y}))\Big]^2 \quad \left(\text{since } P\big(\{X_1, X_2\} \subset T_j\big) = \big(A(T_j)/A(C_H(\mathcal{Y}))\big)^2\right)$$

Then $\mu(N_{PE}^r, J) = \mu(N_{PE}^r) \cdot \left(\sum_{j=1}^{J} w_j^2\right)$ where $\mu(N_{PE}^r)$ is given in Equation (5.4.2).

Furthermore, the asymptotic variance is

$$\begin{aligned}
\nu(N_{PE}^r, J) &= \mathbf{E}\left[h_{12}(N_{PE}^r)\, h_{13}(N_{PE}^r)\right] - \mathbf{E}\left[h_{12}(N_{PE}^r)\right]\mathbf{E}\left[h_{13}(N_{PE}^r)\right] \\
&= P\big(\{X_2, X_3\} \subset N_{PE}^r(X_1)\big) + 2\,P\big(X_2 \in N_{PE}^r(X_1), X_3 \in \Gamma_1(X_1, N_{PE}^r)\big) \\
&\quad + P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r)\big) - 4\,\left(\mu(N_{PE}^r, J)\right)^2.
\end{aligned}$$

Let $P_{2N}(N_{PE}^r) := P\big(\{X_2, X_3\} \subset N_{PE}^r(X_1)\big)$, $P_{2G}(N_{PE}^r) := P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r)\big)$, and $P_M(N_{PE}^r) := P\big(X_2 \in N_{PE}^r(X_1), X_3 \in \Gamma_1(X_1, N_{PE}^r)\big)$. Then for $J > 1$, we have

$$\begin{aligned}
P\big(\{X_2, X_3\} \subset N_{PE}^r(X_1)\big) &= \sum_{j=1}^{J} P\big(\{X_2, X_3\} \subset N_{PE}^r(X_1) \,|\, \{X_1, X_2, X_3\} \subset T_j\big)\, P\big(\{X_1, X_2, X_3\} \subset T_j\big) \\
&= \sum_{j=1}^{J} P_{2N}(N_{PE}^r)\left(A(T_j)/A(C_H(\mathcal{Y}))\right)^3 = P_{2N}(N_{PE}^r)\left(\sum_{j=1}^{J} w_j^3\right).
\end{aligned}$$

Similarly, $P\big(X_2 \in N_{PE}^r(X_1), X_3 \in \Gamma_1(X_1, N_{PE}^r)\big) = P_M(N_{PE}^r)\left(\sum_{j=1}^{J} w_j^3\right)$ and $P\big(\{X_2, X_3\} \subset$

$\Gamma_1(X_1, N_{PE}^r)) = P_{2G}(N_{PE}^r) \left(\sum_{j=1}^J w_j^3\right)$, hence,

$$\nu(N_{PE}^r, J) = \left(P_{2N}(N_{PE}^r) + 2\,P_M^r + P_{2G}(N_{PE}^r)\right) \left(\sum_{j=1}^J w_j^3\right) - 4\;\mu(N_{PE}^r, J)^2 =$$

$$\nu(N_{PE}^r) \left(\sum_{j=1}^J w_j^3\right) + 4\;\mu(N_{PE}^r)^2 \left(\sum_{j=1}^J w_j^3 - \left(\sum_{j=1}^J w_j^2\right)^2\right),$$

so conditional on $\mathcal{W}$, if $\nu(N_{PE}^r, J) > 0$ then $\sqrt{n}\left(\rho_n(N_{PE}^r, J) - \mu(N_{PE}^r, J)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu(N_{PE}^r, J))$.

■

By an appropriate application of Jensen's inequality, we see that $\sum_{j=1}^J w_j^3 \geq \left(\sum_{j=1}^J w_j^2\right)^2$. Therefore, the covariance $\nu(N_{PE}^r, J) = 0$ iff both $\nu(N_{PE}^r) = 0$ and $\sum_{j=1}^J w_j^3 = \left(\sum_{j=1}^J w_j^2\right)^2$ hold, so asymptotic normality may hold even when $\nu(N_{PE}^r) = 0$.

Similarly, for the segregation (association) alternatives with $4\,\varepsilon^2/3 \times 100\%$ of the triangles around the vertices of each triangle is forbidden (allowed), we obtain the above asymptotic distribution of $\rho_n(N_{PE}^r, J)$ with $\mu(N_{PE}^r, J)$ being replaced by $\mu(N_{PE}^r, J, \varepsilon)$, $\nu(N_{PE}^r, J)$ by $\nu(N_{PE}^r, J, \varepsilon)$, $\mu(N_{PE}^r)$ by $\mu(N_{PE}^r, \varepsilon)$, and $\nu(N_{PE}^r)$ by $\nu(N_{PE}^r, \varepsilon)$.

Thus in the case of $J > 1$, we have a (conditional) test of $H_0 : X_i \overset{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}))$ which once again rejects against segregation for large values of $\rho_n(N_{PE}^r, J)$ and rejects against association for small values of $\rho_n(N_{PE}^r, J)$.

The segregation (with $\delta = 1/16$, i.e., $\varepsilon = \sqrt{3}/8$), null, and association (with $\delta = 1/4$, i.e., $\varepsilon = \sqrt{3}/12$) realizations (from left to right) are depicted in Figure 5.4.28 with $n = 100$ and in Figure 4.6.1 with $n = 1000$. For both $n = 100$ and $n = 1000$, for the null realization, the p-value is greater than 0.1 for all r values and both alternatives. For the segregation realization, with $n = 100$ we obtain $p < 0.001$ for $r \leq 3$, $p = 0.025$ for $r = 5$, and $p > 0.1$ for $r \geq 10$ and with $n = 1000$ we obtain $p < 0.0031$ for $1 < r \leq 5$ and $p > 0.24$ for $r = 1$ and $r \geq 10$. For the association realization, with $n = 100$, we obtain $p < 0.05$ for $r = 1.5$, 2, and $p > 0.06$ for for other values of r and with $n = 1000$, we obtain $p < 0.0135$ for $1 < r \leq 3$, $p = .14$ for $r = 1$, and $p > 0.25$ for for $r \geq 5$. Note that this is only for one realization of $\mathcal{X}_n$.

We implement the above described Monte Carlo experiment 1000 times with $n = 100$, $n = 200$, and $n = 500$ and find the empirical significance levels $\widehat{\alpha}_S(n, J)$ and $\widehat{\alpha}_A(n, J)$ and the empirical powers $\widehat{\beta}_n^S(r, \sqrt{3}/8, J)$ and $\widehat{\beta}_n^A(r, \sqrt{3}/12, J)$. These empirical estimates are presented

r	1	11/10	6/5	4/3	$\sqrt{2}$	3/2	2	3	5	10
$n = 100$, $N = 1000$										
$\widehat{\alpha}_S(n, r, J)$	.144	.141	.124	.101	.095	.087	.070	.075	.071	.072
$\widehat{\beta}^S_n(r, \sqrt{3}/8, J)$	.191	.383	.543	.668	.714	.742	.742	.625	.271	.124
$\widehat{\alpha}_A(n, r, J)$	.118	.111	.089	.081	.065	.062	.067	.064	.068	.071
$\widehat{\beta}^A_n(r, \sqrt{3}/12, J)$	.231	.295	.356	.338	.269	.209	.148	.095	.113	.167
$n = 200$, $N = 1000$										
$\widehat{\alpha}_S(n, r, J)$	.095	.092	.087	.077	.073	.076	.072	.071	.074	.073
$\widehat{\beta}^S_n(r, \sqrt{3}/8, J)$	.135	.479	.743	.886	.927	.944	.959	.884	.335	.105
$\widehat{\alpha}_A(n, J)$	.071	.071	.062	.057	.055	.047	.038	.035	.036	.040
$\widehat{\beta}^A_n(r, \sqrt{3}/12, J)$	.182	.317	.610	.886	.952	.985	.972	.386	.143	.068
$n = 500$, $N = 1000$										
$\widehat{\alpha}_S(n, r, J)$	.089	.092	.087	.086	.080	.078	.079	.079	.076	.081
$\widehat{\beta}^S_n(r, \sqrt{3}/8, J)$	.145	.810	.981	.997	.999	1.000	1.000	1.000	.604	.130
$\widehat{\alpha}_A(n, r, J)$	.087	.085	.076	.075	.073	.075	.072	.067	.066	.061
$\widehat{\beta}^A_n(r, \sqrt{3}/12, J)$	.241	.522	.937	1.000	1.000	1.000	1.000	.712	.187	.063

Table 5.4.7: The empirical significance level and empirical power values under $H^S_{\sqrt{3}/8}$ and $H^A_{\sqrt{3}/12}$, $N = 1000$, $n = 100$, and $J = 13$, at $\alpha = .05$ for the realization of $\mathcal{Y}$ in Figure 4.6.1.

in Table 5.4.7 and plotted in Figures 5.4.29 and 5.4.30. Notice that the empirical significance levels are all larger than .05 for both alternatives, so for the given realization of $\mathcal{Y}$ and n values this test is liberal in rejecting H_0 against both alternatives. The smallest empirical significance levels and highest empirical power estimates occur at moderate r values ($r = 3/2$, 2, 3) against segregation and at smaller r values ($r = \sqrt{2}$, $3/2$) against association. Based on this analysis, for the given realization of $\mathcal{Y}$, we suggest the use of moderate r values for segregation and slightly smaller for association. Notice also that as n increases, the empirical power estimates gets larger for both alternatives.

Remark 5.4.10. The conditional test presented here is appropriate when $w_j \in \mathcal{W}$ are fixed, not random. An unconditional version requires the joint distribution of the number and relative size of Delaunay triangles when $\mathcal{Y}$ is, for instance, a Poisson point pattern. Alas, this joint distribution is not available (see [30]). □

5.4.11.1 Related Test Statistics in Multiple Triangle Case

For $J > 1$, we have derived the asymptotic distribution of $\rho_n(N^r_{PE}, J) = \frac{|\mathcal{A}|}{(n(n-1))}$. Let $\mathcal{A}_j$ be the number of arcs and $\rho_{n_j}(N^r_{PE})$ be the relative density for the component of the PCD on

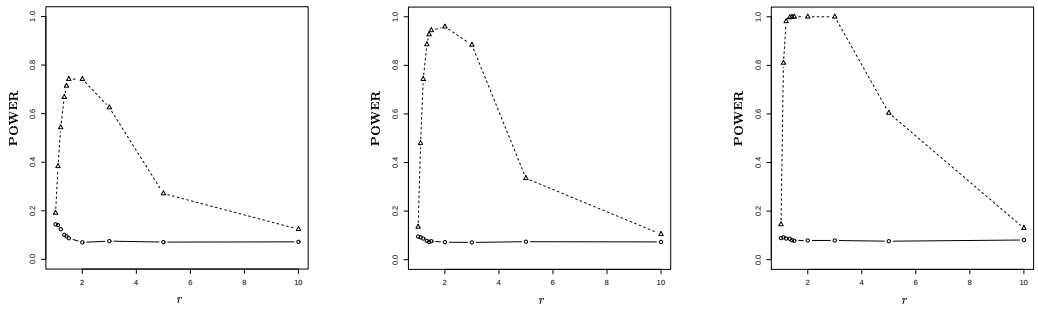


Figure 5.4.29: Monte Carlo power using the asymptotic critical value against $H^S_{\sqrt{3}/8}$, as a function of r, for $n = 100$ (left), $n = 200$ (middle), and $n = 500$ (right) conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. The circles represent the empirical significance levels while triangles represent the empirical power values.

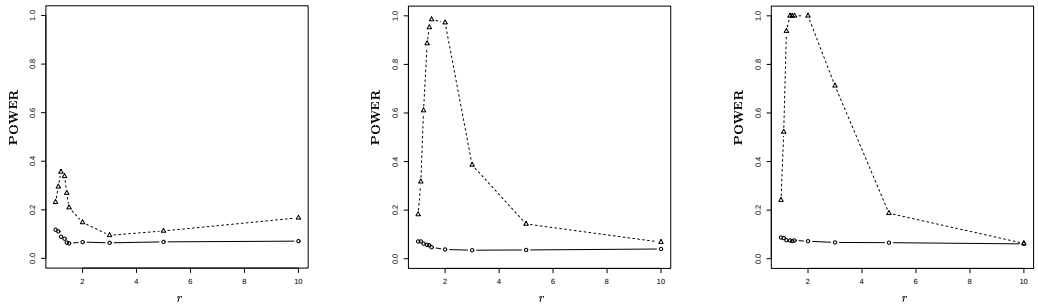


Figure 5.4.30: Monte Carlo power using the asymptotic critical value against $H^A_{\sqrt{3}/12}$ as a function of r, for $n = 100$ (left), $n = 200$ (middle), and $n = 500$ (right) conditional on the realization of $\mathcal{Y}$ in Figure 4.6.1. The circles represent the empirical significance levels while triangles represent the empirical power values.

triangle T_j and $n_j := |\mathcal{X}_n \cap T_j|$, for $j \in \{1, \ldots, J\}$. So

$$\sum_{j=1}^{J} \frac{n_j\,(n_j-1)}{n\,(n-1)} \rho_{n_j}(N^r_{PE}) = \rho_n(N^r_{PE}, J),$$

since

$$\sum_{j=1}^{J} \frac{n_j\,(n_j-1)}{n\,(n-1)} \rho_{n_j}(N^r_{PE}) = \frac{\sum_{j=1}^{J} |\mathcal{A}_j|}{n\,(n-1)} = \frac{|\mathcal{A}|}{n\,(n-1)} = \rho_n(N^r_{PE}, J).$$

Let $\widehat{U}_n := \sum_{j=1}^{J} w_j^2 \cdot \rho_{n_j}(N^r_{PE})$ where $w_j = A(T_j)/A(C_H(\mathcal{Y}))$. Since $\rho_{n_j}(N^r_{PE})$ are asymptotically independent, $\sqrt{n}\left(\widehat{U}_n - \mu(N^r_{PE}, J)\right)$ and $\sqrt{n}\,(\rho_n(N^r_{PE}, J) - \mu(N^r_{PE}, J))$ both converge in distribution to $\mathcal{N}(0, \nu(N^r_{PE}, J))$.

In the denominator of $\rho_n(N_{PE}^r, J)$, we use $n(n-1)$ as the maximum number of arcs possible. However, by definition, we can at most have a digraph with J complete symmetric components of order n_j, for $j \in \{1, \ldots, J\}$. Then the maximum number possible is $n_t := \sum_{j=1}^{J} n_j (n_j - 1)$. So the (adjusted) relative density is $\rho_{n,J}^{adj}(N_{PE}^r) := \frac{|\mathcal{A}|}{n_t}$ and

$$\rho_{n,J}^{adj}(N_{PE}^r) = \frac{\sum_{j=1}^{J} |\mathcal{A}_j|}{n_t} = \sum_{j=1}^{J} \frac{n_j (n_j - 1)}{n_t} \rho_{n_j}(N_{PE}^r).$$

Since $\frac{n_j (n_j-1)}{n_t} \geq 0$ for each j, and $\sum_{j=1}^{J} \frac{n_j (n_j-1)}{n_t} = 1$, $\rho_{n,J}^{adj}(N_{PE}^r)$ is a mixture of $\rho_{n_j}(N_{PE}^r)$. Then $\mathbf{E}\left[\rho_{n,J}^{adj}(N_{PE}^r)\right] = \mu(N_{PE}^r, J)$ and the asymptotic variance of $\rho_{n,J}^{adj}(N_{PE}^r)$ is

$$\frac{1}{n}\left[\nu(N_{PE}^r)\left(\sum_{j=1}^{J} w_j^3 \Big/ \left(\sum_{j=1}^{J} w_j^2\right)^2\right) + 4\,\mu(N_{PE}^r)^2\left(\sum_{j=1}^{J} w_j^3 \Big/ \left(\sum_{j=1}^{J} w_j^2\right)^2 - 1\right)\right].\square$$

5.4.11.2 Asymptotic Efficacy Analysis for $J > 1$

The PAE, HLAE, and asymptotic power function analysis are given for $J = 1$. For $J > 1$, the analyses will depend on both the number of triangles as well as the relative sizes of the triangles. So the optimal r values with respect to these efficacy criteria for $J = 1$ do not necessarily hold for $J > 1$, so the analyses need to be updated, given the values of J and $\mathcal{W}$.

Under segregation alternative H_ε^S, the PAE is given by

$$\begin{aligned} \mathrm{PAE}_J^S(N_{PE}^r) &= \frac{\left(\mu_S''(N_{PE}^r, J, \varepsilon = 0)\right)^2}{\nu(N_{PE}^r, J)} \\ &= \frac{\left(\mu_S''(N_{PE}^r, \varepsilon = 0)\sum_{j=1}^{J} w_j^2\right)^2}{\nu(N_{PE}^r)\sum_{j=1}^{J} w_j^3 + 4\,\mu(N_{PE}^r)^2\left(\sum_{j=1}^{J} w_j^3 - \left(\sum_{j=1}^{J} w_j^2\right)^2\right)}. \end{aligned} \quad (5.4.8)$$

Under association alternative H_ε^A the PAE is similar.

In Figure 5.4.31, we present the PAE as a function of r for both segregation and association conditional on the realization of $\mathcal{Y}$ in Figure 4.6.1. Notice that, unlike $J = 1$ case, $\mathrm{PAE}_J^S(N_{PE}^r)$ is bounded. Some values of interest are $\mathrm{PAE}_J^S(N_{PE}^{r=1}) = .3884$, $\lim_{r\to\infty} \mathrm{PAE}_J^S(N_{PE}^r) = (8\sum_{j=1}^{J} w_j^2) \Big/ \left(256\left(\sum_{j=1}^{J} w_j^3 - \left(\sum_{j=1}^{J} w_j^2\right)^2\right)\right) \approx 139.34$, $\mathrm{argsup}_{r\in[1,2]} \mathrm{PAE}_J^S(N_{PE}^r) \approx 1.974$. As for association, $\mathrm{PAE}_J^A(N_{PE}^{r=1}) = 422.9551$, $\lim_{r\to\infty} \mathrm{PAE}_J^A(N_{PE}^r) = 0$, $\mathrm{argsup}_{r\geq 1} \mathrm{PAE}_J^A(N_{PE}^r) \approx 1.5$ with $\mathrm{PAE}_J^A(N_{PE}^{r=1.5}) \approx 1855.9672$. Based on the asymptotic efficacy analysis, we suggest, for large n

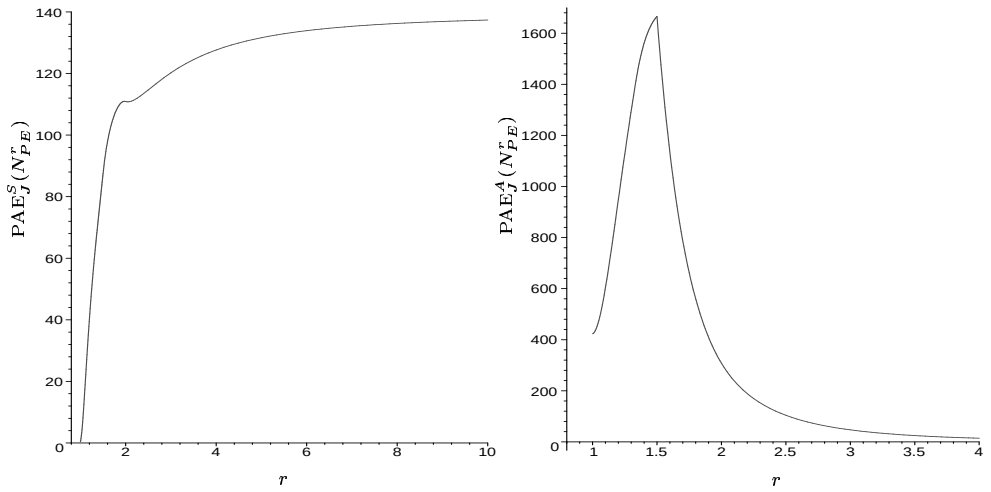


Figure 5.4.31: Pitman asymptotic efficacy against segregation (left) and association (right) as a function of r with the realization of $\mathcal{Y}$ in Figure 5.4.28. Notice that vertical axes are differently scaled.

and small ε, choosing moderate r for testing against segregation and association.

Under segregation, the HLAE is given by

$$\begin{aligned} \mathrm{HLAE}_J^S(N_{PE}^r),\varepsilon) &:= \frac{(\mu_S(N_{PE}^r,J,\varepsilon)-\mu(N_{PE}^r,J))^2}{\nu_S(N_{PE}^r,J,\varepsilon)} \qquad (5.4.9)\\ &= \frac{\left(\mu_S(N_{PE}^r,\varepsilon)\ \left(\sum_{j=1}^J w_j^2\right)-\mu(N_{PE}^r)\ \left(\sum_{j=1}^J w_j^2\right)\right)^2}{\nu_S(N_{PE}^r,\varepsilon)\ \sum_{j=1}^J w_j^3+4\,\mu_S(N_{PE}^r,\varepsilon)^2\left(\sum_{j=1}^J w_j^3-\left(\sum_{j=1}^J w_j^2\right)^2\right)}. \end{aligned}$$

Notice that $\mathrm{HLAE}_J^S(N_{PE}^r),\varepsilon=0)=0$ and $\lim_{\to\infty}\mathrm{HLAE}_J^S(N_{PE}^r),\varepsilon)=0$.

We calculate HLAE of $\rho_n(N_{PE}^r,J)$ under H_ε^S for $\varepsilon=\sqrt{3}/8$, $\varepsilon=\sqrt{3}/4$, and $\varepsilon=2\sqrt{3}/7$. In Figure 5.4.32 we present $\mathrm{HLAE}_J^S(N_{PE}^r),\varepsilon)$ for these ε values conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28.

Note that with $\varepsilon=\sqrt{3}/8$, $\mathrm{HLAE}_J^S\left(r=1,\sqrt{3}/8\right)\approx .0004$ and $\mathrm{argsup}_{r\in[1,\infty]}\mathrm{HLAE}_J^S\big(N_{PE}^r$, $\sqrt{3}/8\big)\approx 1.8928$ with the supremum $\approx .0544$. With $\varepsilon=\sqrt{3}/4$, $\mathrm{HLAE}_J^S\left(N_{PE}^r)=1,\sqrt{3}/4\right)\approx$ $.0450$ and $\mathrm{argsup}_{r\in[1,\infty]}\mathrm{HLAE}_J^S\left(N_{PE}^r,\sqrt{3}/4\right)\approx 1.3746$ with the supremum $\approx .6416$. With $\varepsilon=$ $2\sqrt{3}/7$, $\mathrm{HLAE}_J^S\left(N_{PE}^r\right)=1,2\sqrt{3}/7\big)\approx .045$ and $\mathrm{argsup}_{r\in[1,\infty]}\mathrm{HLAE}_J^S\left(N_{PE}^r\right),2\sqrt{3}/7\big)\approx 1.3288$ with the supremum $\approx .9844$. Furthermore, we observe that $\mathrm{HLAE}_J^S\left(N_{PE}^r\right),2\sqrt{3}/7\big)>\mathrm{HLAE}_J^S$ $\left(r,\sqrt{3}/4\right)>\mathrm{HLAE}_J^S\left(N_{PE}^r,\sqrt{3}/8\right)$ at each r. Based on the HLAE analysis for the given $\mathcal{Y}$ we suggest moderate r values for moderate segregation, and small r values for severe segregation.

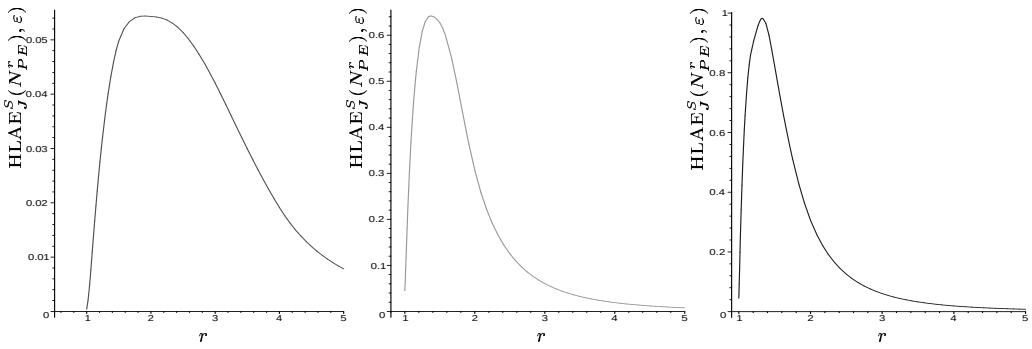


Figure 5.4.32: Hodges-Lehmann asymptotic efficacy against segregation alternative H_ε^S as a function of r for $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$ (left to right) conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28.

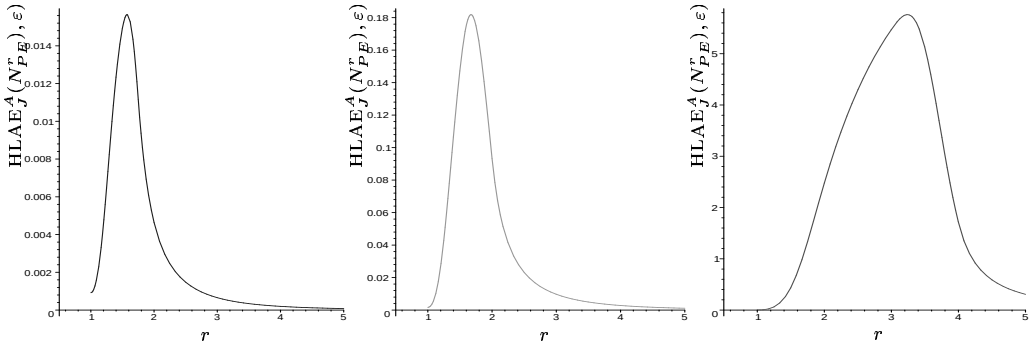


Figure 5.4.33: Hodges-Lehmann asymptotic efficacy against association alternative H_ε^A as a function of r for $\varepsilon \in \{\sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24\}$ (left to right) conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28.

The explicit form of $\text{HLAE}_J^A(N_{PE}^r), \varepsilon)$ is similar which implies $\text{HLAE}_J^A(N_{PE}^r), \varepsilon = 0) = 0$ and $\lim_{r\to\infty} \text{HLAE}_J^A(N_{PE}^r), \varepsilon) = 0$.

We calculate HLAE of $\rho_n(N_{PE}^r, J)$ under H_ε^A for $\varepsilon = \sqrt{3}/21$, $\varepsilon = \sqrt{3}/12$, and $\varepsilon = 5\sqrt{3}/24$. In Figure 5.4.33 we present $\text{HLAE}_J^S(N_{PE}^r), \varepsilon)$ for these ε values conditional on the realization of $\mathcal{Y}$ in Figure 4.6.1

Note that with $\varepsilon = \sqrt{3}/21$, $\text{HLAE}_J^A\left(N_{PE}^{r=1}, \sqrt{3}/21\right) \approx .0009$ and $\text{argsup}_{r\in[1,\infty]} \text{HLAE}_J^A\left(N_{PE}^r, \sqrt{3}/21\right) \approx 1.5734$ with the supremum $\approx .0157$. With $\varepsilon = \sqrt{3}/12$, $\text{HLAE}_J^A\left(N_{PE}^{r=1}, \sqrt{3}/12\right) \approx .0168$ and $\text{argsup}_{r\in[1,\infty]} \text{HLAE}_J^A\left(N_{PE}^r, \sqrt{3}/12\right) \approx 1.6732$ with the supremum $\approx .1818$. With $\varepsilon = 5\sqrt{3}/24$, $\text{HLAE}_J^A\left(N_{PE}^{r=1}, 5\sqrt{3}/24\right) \approx .0017$ and $\text{argsup}_{r\in[1,\infty]} \text{HLAE}_J^A\left(N_{PE}^r\right), 5\sqrt{3}/24) \approx 3.2396$ with the supremum ≈ 5.7616. Furthermore, we observe that $\text{HLAE}_J^A\left(N_{PE}^r\right), 5\sqrt{3}/24) > \text{HLAE}_J^A\left(N_{PE}^r, \sqrt{3}/12\right) > \text{HLAE}_J^A(N_{PE}^r, \sqrt{3}/21)$ at each r. Based on the HLAE analysis for the given $\mathcal{Y}$ we suggest moderate r values for moderate association and larger r values for severe

association.

5.5 Relative Density of Random τ-Factor Central Similarity Proximity Catch Digraphs

We have defined the τ-factor central similarity proximity maps and the associated random digraphs in Section 3.3.3.

The test statistic used here is the relative density $\rho_n(N_{CS}^{\tau}) = \rho(\mathcal{X}_n; h, N_{CS}^{\tau}, M)$ where N_{CS}^{τ} defined with M-edge regions. The asymptotic distributions under both the null and the alternative hypotheses are determined in Sections 5.5.1 and 5.5.2 by using the standard U-statistic central limit theory. Pitman asymptotic efficacy is analyzed in Section 5.5.7.

5.5.1 Asymptotic Normality Under the Null Hypothesis

First, we present a "geometry invariance" result which will simplify our subsequent analysis by allowing us to consider the special case of the equilateral triangle.

Theorem 5.5.1. *Suppose $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T(\mathcal{Y}))$. Then for any $\tau \in [0,1]$ the distribution of $\rho_n(N_{CS}^{\tau})$ is independent of $\mathcal{Y}$ and hence the geometry of $T(\mathcal{Y})$.*

Proof: Similar to the proof of Theorem 4.7.3. ■

Based on Theorem 5.5.1 and our uniform null hypothesis, we may assume that $T(\mathcal{Y})$ is a standard equilateral triangle with $\mathcal{Y} = \{(0,0),(1,0),(1/2,\sqrt{3}/2)\}$ henceforth for $\rho_n(N_{CS}^{\tau})$.

For our τ-factor central similarity proximity map and uniform null hypothesis, the asymptotic null distribution of $\rho_n(N_{CS}^{\tau})$ can be derived as a function of τ. Let $\mu(N_{CS}^{\tau}) = \mathbf{E}[\rho_n(N_{CS}^{\tau})]$ and $\nu(N_{CS}^{\tau}) := \mathbf{Cov}[h_{12}(N_{CS}^{\tau}, M), h_{13}(N_{CS}^{\tau}, M)]$.

Lemma 5.5.2. *The expected relative density for $N_{CS}^{\tau}(\cdot, M)$ with $M \in T(\mathcal{Y})^o$ is $\mathbf{E}[\rho_n(N_{CS}^{\tau})] = \tau^2/6$.*

Proof: For τ-factor M-central similarity proximity regions, recall that $\mathscr{R}_S(N_{CS}^{\tau}, M) \subseteq \{M\}$ with $\mathscr{R}_S(N_{CS}^{\tau}, M) = \{M\}$ for $\tau = 1$, $\mathscr{R}_S(N_{CS}^{\tau}, M) = \emptyset$ for $\tau \in [0,1)$. Without loss of generality, assume $M = (m_1, m_2)$ with $m_1 > c_1$, then

$$P(X_2 \in N_{CS}^{\tau}(X_1, M)) = \sum_{j=1}^{3} P\big(X_2 \in N_{CS}^{\tau}(X_1, M), X_1 \in R_M(e_j)\big).$$

The geometry invariance of $\mu(N_{CS}^{\tau}, M)$ can be proved as in Theorem 4.7.3. So we consider the standard equilateral triangle T_e.

In $R_M(e_3)$,

$$P\big(X_2 \in N_{CS}^{\tau}(X_1, M), X_1 \in R_M(e_3)\big) = \int_0^{m_1} \int_0^{m_2\, x/m_1} \frac{A\big(N_{CS}^{\tau}((x,y), M)\big)}{A(T(\mathcal{Y}))^2} dydx$$
$$+ \int_{m_1}^{1} \int_0^{\frac{m_2\,(1-x)}{(1-m_1)}} \frac{A\big(N_{CS}^{\tau}((x,y), M)\big)}{A(T(\mathcal{Y}))^2} dydx = \sqrt{3}\,\tau^2\, m_2/9.$$

where $A\big(N_{CS}^{\tau}((x,y), M)\big) = \frac{\sqrt{3}\,\tau^2\, y^2}{4\, m_2^2}$.

In $R_M(e_1)$,

$$P\big(X_2 \in N_{CS}^{\tau}(X_1, M), X_1 \in R_M(e_1)\big) = \int_{1/2}^{m_1} \int_{\ell_{cm}(x)}^{\sqrt{3}\,(1-x)} \frac{A\big(N_{CS}^{\tau}((x,y), M)\big)}{A(T(\mathcal{Y}))^2} dydx +$$
$$\int_{m_1}^{1} \int_{\frac{m_2\,(1-x)}{(1-m_1)}}^{\sqrt{3}\,(1-x)} \frac{A\big(N_{CS}^{\tau}((x,y), M)\big)}{A(T(\mathcal{Y}))^2} dydx = \sqrt{3}\,\tau^2 \left(\sqrt{3}\,(1-m_1) - m_2\right)/18.$$

where $A\big(N_{CS}^{\tau}((x,y), M)\big) = \frac{4\,\tau^2\,\sqrt{3}\left(3\,(1-x)-\sqrt{3}\,y\right)^2}{9\left(m_2+\sqrt{3}\,(m_1-1)\right)^2}$ and $\ell_{cm}(x) = \frac{\left(2\,m_2-\sqrt{3}\right),x+m_1\sqrt{3}-m_2}{2\,m_1-1}$.

In $R_M(e_2)$,

$$P\big(X_2 \in N_{CS}^{\tau}(X_1, M), X_1 \in R_M(e_2)\big) = \int_0^{1/2} \int_{m_2\, x/m_1}^{\sqrt{3}\,x} \frac{A\big(N_{CS}^{\tau}((x,y), M)\big)}{A(T(\mathcal{Y}))^2} dydx +$$
$$\int_{1/2}^{m_1} \int_{m_2\, x/m_1}^{\ell_{cm}(x)} \frac{A\big(N_{CS}^{\tau}((x,y), M)\big)}{A(T(\mathcal{Y}))^2} dydx = \sqrt{3}\,\tau^2 \left(\sqrt{3}\, m_1 - m_2\right)/18.$$

where $A\big(N_{CS}^{\tau}((x,y), M)\big) = \frac{\sqrt{3}\,\tau^2\left(y-\sqrt{3}\,x\right)^2}{4\,\left(\sqrt{3}\,m_1-m_2\right)^2}$.

Therefore

$$\mu\big(N_{CS}^{\tau}\big) = \mu\,(N_{CS}^{\tau}, M) = \tau^2/6. \ \blacksquare$$

Notice that $\mu(N_{CS}^{\tau}, M)$ is not only independent of the geometry of $T(\mathcal{Y})$ but also of the choice of M. See Figure 5.5.1 for the plot of $\mu\big(N_{CS}^{\tau}\big)$ for $\tau \in [0, 1]$.

By detailed geometric probability calculations, the asymptotic variance of the relative density of the τ-factor central similarity proximity catch digraph can be explicitly computed. The central limit theorem for U-statistics then establishes the asymptotic normality under the uniform null hypothesis. These results are summarized in the following theorem.

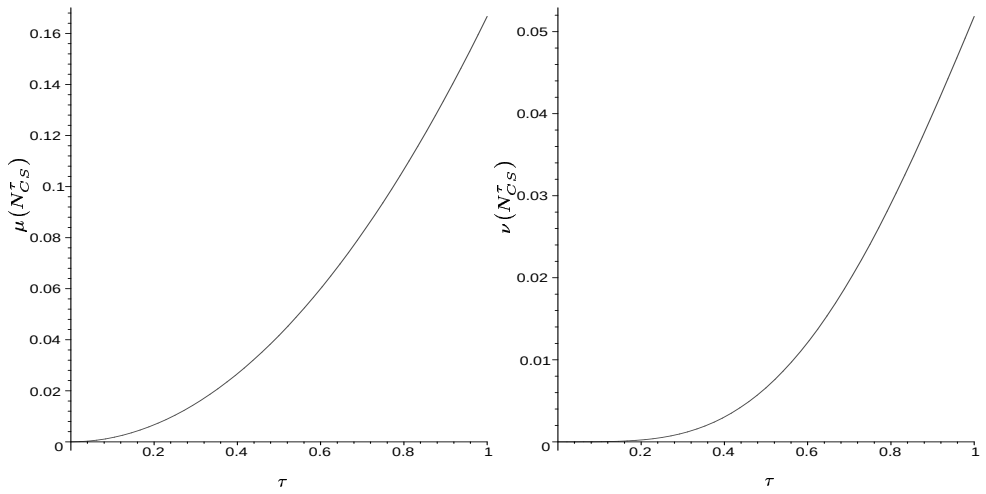


Figure 5.5.1: The asymptotic null mean $\mu(N_{CS}^{\tau})$ for $N_{CS}^{\tau}(\cdot, M)$ (left) and asymptotic null variance $\nu(N_{CS}^{\tau})$ from Equation 5.5.2 in Theorem 5.5.3 (right).

Theorem 5.5.3. *For $\tau \in (0,1]$, and $N_{CS}^{\tau}(\cdot, M_C)$*

$$\sqrt{n}\left(\rho_n(N_{CS}^{\tau}) - \mu(N_{CS}^{\tau})\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sqrt{\nu(N_{CS}^{\tau})}\right) \tag{5.5.1}$$

where

$$\mu(N_{CS}^{\tau}) = \tau^2/6 \quad \textit{and} \quad \nu(N_{CS}^{\tau}) = \frac{\tau^4(6\,\tau^5 - 3\,\tau^4 - 25\,\tau^3 + \tau^2 + 49\,\tau + 14)}{45\,(\tau+1)(2\,\tau+1)(\tau+2)}. \tag{5.5.2}$$

For $\tau = 0$, $\rho_n(N_{CS}^{\tau})$ is degenerate.

See Lemma 5.5.2 for the derivation of the mean and Appendix Section C.2 for the derivation of the variance.

Consider the form of the mean and the variance functions, which are depicted in Figure 5.5.1. Note that $\mu(N_{CS}^{\tau})$ is monotonically increasing in τ, since $N_{CS}^{\tau}(x, M)$ increases with τ for all $x \in T(\mathcal{Y})^o$. In addition, $\mu(N_{CS}^{\tau=1}) = 1/6$ and $\mu(N_{CS}^{\tau=0}) = 0$. Note also that $\mu(N_{CS}^{\tau})$ is continuous in τ.

Regarding the asymptotic variance, note that $\nu(N_{CS}^{\tau})$ is continuous in τ and $\nu(N_{CS}^{\tau=1}) = 7/135$ and $\nu(N_{CS}^{\tau=0}) = 0$ — there are no arcs when $\tau = 0$ a.s.— which explains why $\rho_n(N_{CS}^{\tau=0})$ is degenerate.

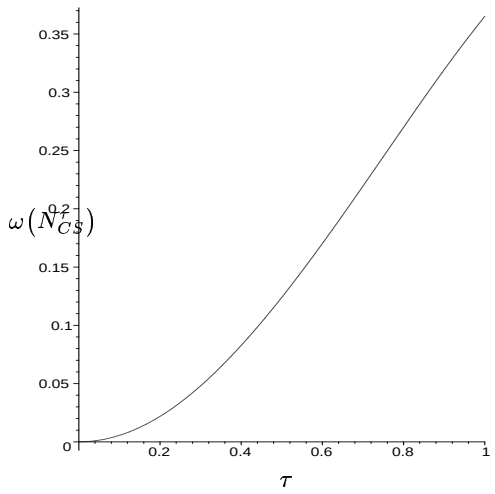


Figure 5.5.2: The graph of $\omega(N_{CS}^{\tau}) = \mathbf{Var}\left[h_{12}(N_{CS}^{\tau}, M_C)\right]$ as a function of τ.

To illustrate the limiting distribution, $\tau = 1/2$ yields

$$\frac{\sqrt{n}\left(\rho_n(N_{CS}^{1/2}) - \mu(N_{CS}^{1/2})\right)}{\sqrt{\nu(N_{CS}^{1/2})}} = \sqrt{\frac{2880\,n}{19}}\left(\rho_n(N_{CS}^{1/2}) - 1/24\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

or equivalently,

$$\rho_n(N_{CS}^{1/2}) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{1}{24}, \frac{19}{2880\,n}\right).$$

The finite sample variance and skewness may be derived analytically in much the same way as was $\mathbf{Cov}\left[h_{12}(N_{CS}^{\tau}, M_C), h_{13}(N_{CS}^{\tau}, M_C)\right]$ for the asymptotic variance. In particular, the variance of $h_{12}(N_{CS}^{\tau}, M_C)$ is

$$\omega(N_{CS}^{\tau}) = \begin{cases} \frac{\tau^2(33\,\tau+63-9\,\tau^2-2\,\tau^4-11\,\tau^3)}{9\,(\tau+3)(2\,\tau+5)}, & \text{for} \quad \tau \in [0, 1/2), \\ \frac{\tau^2(4\,\tau+12-5\,\tau^2-\tau^3)}{9\,(\tau+3)}, & \text{for} \quad \tau \in [1/2, 1]. \end{cases}$$

In Figure 5.5.2 is the graph of $\omega(N_{CS}^{\tau})$. Note that $\omega(\tau = 0) = 0$ and $\omega(\tau = 1) = 23/63 \approx .3651$. Furthermore, $\omega(N_{CS}^{\tau})$ is an increasing continous function of τ.

In fact, the exact distribution of $\rho_n(N_{CS}^{\tau})$ is, in principle, available by successively conditioning on the values of the X_i. Alas, while the joint distribution of $h_{12}(N_{CS}^{\tau}, M_C), h_{13}(N_{CS}^{\tau}, M_C)$ is available, the joint distribution of $\{h_{ij}(N_{CS}^{\tau}, M_C)\}_{1 \le i < j \le n}$, and hence the calculation for the exact distribution of $\rho_n(N_{CS}^{\tau})$, is extraordinarily tedious and lengthy for even small values of n.

Figure 5.5.3 indicates that, for $\tau = 1/2$, the normal approximation is accurate even for small

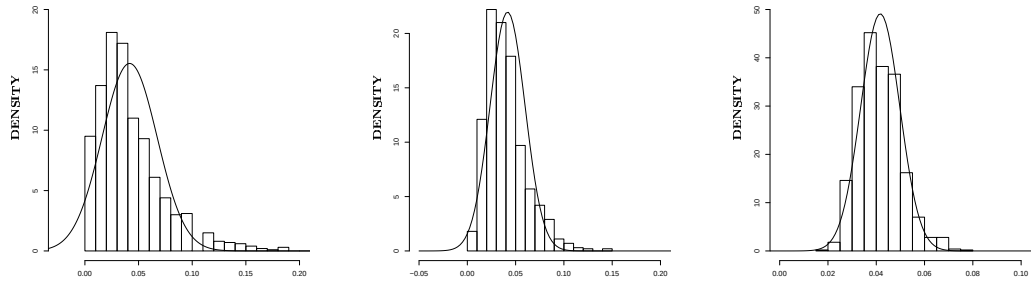


Figure 5.5.3: Depicted are $\rho_n(N_{CS}^{1/2}) \overset{\text{approx}}{\sim} \mathcal{N}\left(\frac{1}{24}, \frac{19}{2880\,n}\right)$ for 10, 20, 100 (left to right). Histograms are based on 1000 Monte Carlo replicates. Solid curves represent the approximating normal densities given in Theorem 5.4.2. Note that the vertical axes are differently scaled.

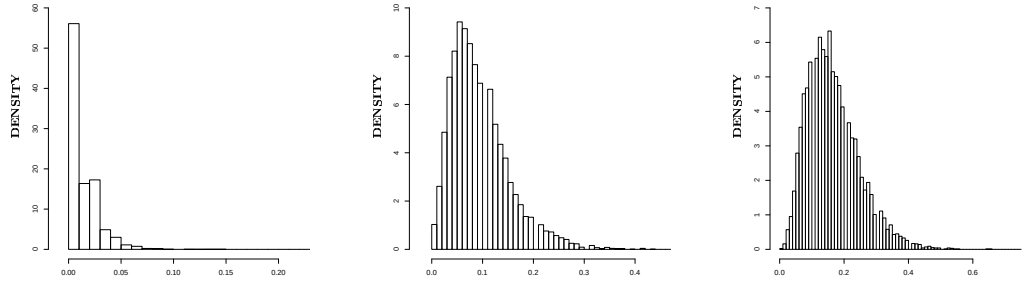


Figure 5.5.4: Depicted are the histograms for 10000 Monte Carlo replicates of $\rho_{10}(N_{CS}^{1/4})$ (left), $\rho_{10}(N_{CS}^{3/4})$ (middle), and $\rho_{10}(N_{CS}^{1})$ (right) indicating severe small sample skewness for small values of τ.

n (although kurtosis and skewness may be indicated for 10, 20). Figure 5.5.4 demonstrates, however, that the smaller the value of τ more severe the skewness of the probability density.

5.5.2 Asymptotic Normality Under the Alternatives

Asymptotic normality of relative density of the proximity catch digraph can be established under the alternative hypothesis of segregation and association by the same method as under the null hypothesis.

Theorem 5.5.4. *Let $\mu_S(N_{CS}^{\tau}, \varepsilon)$ be the mean, $\mathbf{E}_{\varepsilon}^{S}[\rho_n(N_{CS}^{\tau})]$ and $\nu_S(N_{CS}^{\tau}, \varepsilon)$ be the covariance, $\mathbf{Cov}_{\varepsilon}^{S}[h_{12}(N_{CS}^{\tau}, M_C), h_{13}(N_{CS}^{\tau}, M_C)]$ for $\tau \in (0, 1]$ and $\varepsilon \in \left(0, \sqrt{3}/3\right)$ under H_{ε}^{S}, then $\sqrt{n}\left(\rho_n(N_{CS}^{\tau}) - \mu_S(N_{CS}^{\tau}, \varepsilon)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \nu_S(N_{CS}^{\tau}, \varepsilon)\right)$ for the values of the pair (τ, ε) for which $\nu_S(N_{CS}^{\tau}, \varepsilon) > 0$. Likewise for H_{ε}^{A}.*

Sketch of Proof: Under the alternatives, i.e., $\varepsilon > 0$, $\rho_n(N_{CS}^\tau)$ is a U-statistic with the same symmetric kernel $h_{ij}(N_{CS}^\tau, M_C)$ as in the null case. Under H_ε^S, the mean $\mu_S(N_{CS}^\tau, \varepsilon) = \mathbf{E}_\varepsilon^S[\rho_n(N_{CS}^\tau)] = \mathbf{E}_\varepsilon^S[h_{12}(N_{CS}^\tau, M_C)]/2$, now a function of both τ and ε, is again in $[0,1]$. $\nu_S(N_{CS}^\tau, \varepsilon) = \mathbf{Cov}_\varepsilon^S[h_{12}(N_{CS}^\tau, M_C), h_{13}(N_{CS}^\tau, M_C)]$, also a function of both τ and ε, is bounded above by $1/4$, as before. Thus asymptotic normality obtains provided that $\nu_S(N_{CS}^\tau, \varepsilon) > 0$; otherwise $\rho_n(N_{CS}^\tau)$ is degenerate. Likewise for H_ε^A. The explicit forms of $\mu_S(N_{CS}^\tau, \varepsilon)$ and $\mu_A(N_{CS}^\tau, \varepsilon)$ are given, defined piecewise, in Appendix Section C.3. Note that under H_ε^S,

$$\nu_S(N_{CS}^\tau, \varepsilon) > 0 \text{ iff } (\tau, \varepsilon) \in (0,1] \times \left(0, 3\sqrt{3}/10\right] \cup \left(\frac{2(\sqrt{3} - 3\varepsilon)}{4\varepsilon - \sqrt{3}}, 1\right] \times \left(3\sqrt{3}/10, \sqrt{3}/3\right),$$

and under H_ε^A,

$$\nu_A(N_{CS}^\tau, \varepsilon) > 0 \text{ iff } (\tau, \varepsilon) \in (0,1] \times \left(0, \sqrt{3}/3\right). \blacksquare$$

Notice that under the association alternatives any $\tau \in (0,1]$ yields asymptotic normality for all $\varepsilon \in (0, \sqrt{3}/3)$, while under the segregation alternatives only $\tau = 1$ yields this universal asymptotic normality.

5.5.3 The Test and Analysis

The relative density of the central similarity proximity catch digraph is a test statistic for the segregation/association alternative; rejecting for extreme values of $\rho_n(N_{CS}^\tau)$ is appropriate since under segregation we expect $\rho_n(N_{CS}^\tau)$ to be large, while under association we expect $\rho_n(N_{CS}^\tau)$ to be small. Using the test statistic

$$R(N_{CS}^\tau) = \frac{\sqrt{n}(\rho_n(N_{CS}^\tau) - \mu(N_{CS}^\tau))}{\sqrt{\nu(N_{CS}^\tau)}}, \tag{5.5.3}$$

the asymptotic critical value for the one-sided level α test against segregation is given by

$$z_\alpha = \Phi^{-1}(1 - \alpha).$$

Against segregation, the test rejects for $R(N_{CS}^\tau) > z_\alpha$ and against association, the test rejects for $R(N_{CS}^\tau) < z_{1-\alpha}$.

5.5.4 Consistency

Theorem 5.5.5. *The test against H_ε^S which rejects for $R(N_{CS}^\tau) > z_\alpha$ and the test against H_ε^A which rejects for $R(N_{CS}^\tau) < z_{1-\alpha}$ are consistent for $\tau \in (0,1]$ and $\varepsilon \in (0, \sqrt{3}/3)$.*

Proof: Since the variance of the asymptotically normal test statistic, under both the null and the alternatives, converges to 0 as $n \to \infty$ (or is degenerate), it remains to show that the mean under the null, $\mu(N_{CS}^\tau) = \mathbf{E}[\rho_n(N_{CS}^\tau)]$, is less than (greater than) the mean under the alternative, $\mu_S(N_{CS}^\tau, \varepsilon) = \mathbf{E}_\varepsilon^S[\rho_n(N_{CS}^\tau)]$ against segregation ($\mu_A(N_{CS}^\tau, \varepsilon) = \mathbf{E}_\varepsilon^A[\rho_n(N_{CS}^\tau)]$ against association) for $\varepsilon > 0$. Whence it will follow that power converges to 1 as $n \to \infty$.

Detailed analysis of $\mu_S(N_{CS}^\tau, \varepsilon)$ in Appendix Section C.3.1 indicates that under segregation $\mu_S(N_{CS}^\tau, \varepsilon) > \mu(N_{CS}^\tau)$ for all $\varepsilon > 0$ and $\tau \in (0,1]$. Likewise, detailed analysis of $\mu_A(N_{CS}^\tau, \varepsilon)$ in Appendix Section C.3.2 indicates that under association $\mu_A(N_{CS}^\tau, \varepsilon) < \mu(N_{CS}^\tau)$ for all $\varepsilon > 0$ and $\tau \in (0,1]$. Hence the desired result follows for both alternatives. ■

The Remarks 5.4.5 and 5.4.6 also hold for $\mu(\tau, \varepsilon)$.

5.5.5 Monte Carlo Power Analysis Under Segregation

In segregation alternatives with $\varepsilon > 0$, we implement the above described Monte Carlo experiment for $\tau \in \{.1, .2, .3, .4, .5, .6, .7, .8, .9, 1.0\}$. For each τ value, we estimate the empirical critical value $\widehat{c}_n^S(\tau)$ and the empirical significance level $\widehat{\alpha}_{mc}^S(\tau, n)$ under H_0 and the empirical power $\widehat{\beta}_{mc}^S(\tau, n, \varepsilon)$ under H_ε^S with $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/7\}$.

The kernel (probability) density estimates of the relative density under the null case and the segregation alternatives under H_ε^S with $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/8, 2\sqrt{3}/7\}$, $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$ and $N = 10000$ are plotted in Figures 5.5.5, 5.5.6, and 5.5.7, respectively. Observe that under H_0 kernel density estimates are skewed right for small τ values, (with skewness increasing as τ gets smaller) and almost symmetric for large τ values, with most symmetry occurring at $\tau = 1$. Kernel density estimates are skewed right under $H_{\sqrt{3}/8}^S$, symmetric under $H_{\sqrt{3}/4}^S$, and skewed left under $H_{2\sqrt{3}/7}^S$. The empirical critical values, empirical significance levels, and empirical power estimates under $H_{\sqrt{3}/8}^S$ with 10000 Monte Carlo replicates are presented in Table 5.5.1. Notice that the estimated significance levels are all below .05, and about .05 at $\tau \in \{.2, .8, .9\}$. Under $H_{\sqrt{3}/8}^S$, with $n = 10$, the null and alternative probability density functions for $\rho_{10}(\tau)$ are very similar, implying small power. Notice also that more severe the segregation, larger the empirical power estimates at each τ value. These empirical power estimates are

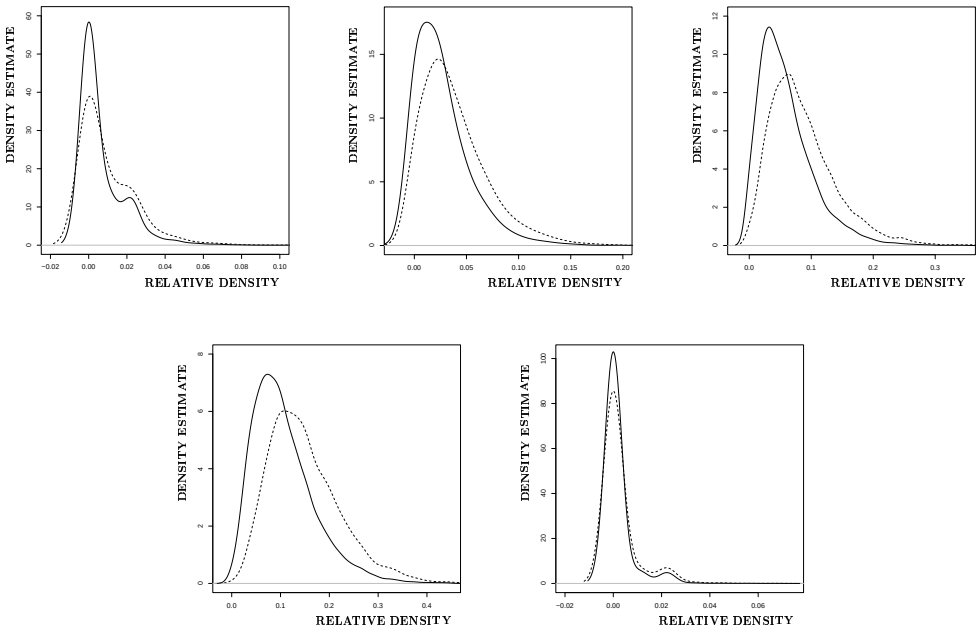


Figure 5.5.5: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/8}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$, and $N = 10000$ (left-to-right).

plotted in Figure 5.5.8. Maximum Monte Carlo power estimate occurs at $\tau = .8$ for $\varepsilon = \sqrt{3}/8$ and $\varepsilon = \sqrt{3}/4$. Furthermore, Monte Carlo power estimate tends to increase as τ gets larger.

With $n = 100$, there is more separation between null and alternative probability density functions. See Figure 5.5.9 for the kernel density estimates under H_0 and $H^S_{\sqrt{3}/8}$ with $N = 1000$ Monte Carlo replicates. Notice that the probability density functions are almost symmetric for all τ values. The corresponding empirical critical values, empirical significance levels, and empirical powers are presented in Table 5.5.2. The estimated significance levels are all about .05 and under $H^S_{\sqrt{3}/4}$ and $H^S_{2\sqrt{3}/7}$ the empirical power estimates are all 1 at each τ value. As n gets larger, the normal approximation gets more accurate and the separation between the null and alternative density estimates increase. See for example Figure 5.5.10 for $\tau = .5$ with $n = 10$, $N = 10000$ (left) and $n = 100$, $N = 1000$ (right). The empirical power estimates are plotted in Figure 5.5.11. Note that for $\varepsilon = \sqrt{3}/8$, maximum Monte Carlo power estimate occurs at $\tau = .8$.

For a given alternative and sample size, we may consider analyzing the power of the test — using the asymptotic critical value— as a function of the proximity factor τ. For each τ value, the α-level asymptotic critical value is $\mu(N^\tau_{CS}) + z_{(1-\alpha)} \cdot \sqrt{\nu(N^\tau_{CS})/n}$. In Figure 5.5.12,

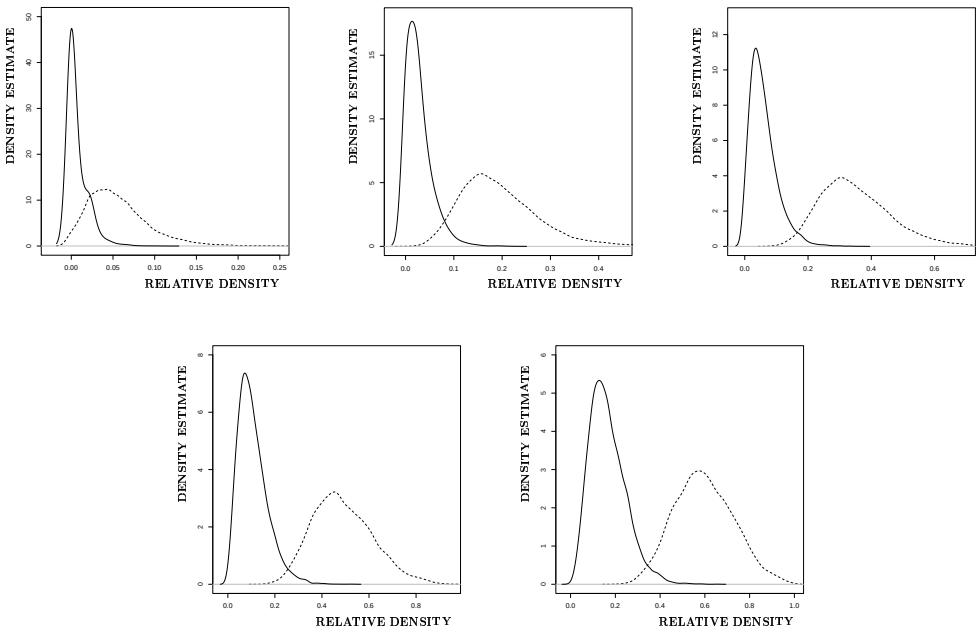


Figure 5.5.6: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/4}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$, and $N = 10000$ (left-to-right).

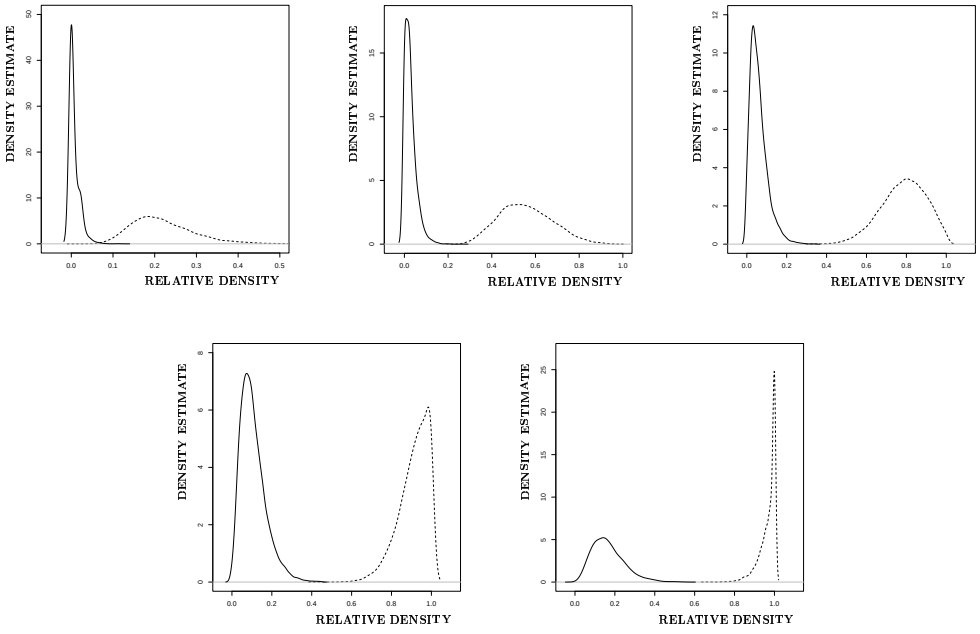


Figure 5.5.7: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{2\sqrt{3}/7}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$, and $N = 10000$ (left-to-right).

τ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\widehat{c}_n^S(\tau)$	$.0\overline{1}$	$.0\overline{2}$	$.0\overline{5}$	$.0\overline{7}$	$.\overline{1}$	$.1\overline{4}$	$.1\overline{8}$	$.\overline{2}$	$.2\overline{6}$	$.3\overline{2}$
$\widehat{\alpha}_{mc}^S(\tau, n)$	.0456	.0498	.0324	.0384	.0387	.0452	.041	.0498	.0483	.0422
$\widehat{\beta}_{mc}^S\left(\tau, n, \sqrt{3}/8\right)$	.0777	.0876	.0647	.0912	.0994	.1172	.1139	.14	.144	.1234
$\widehat{\beta}_{mc}^S\left(\tau, n, \sqrt{3}/4\right)$	.4044	.635	.8556	.9538	.9777	.9887	.9939	.9943	.9939	.991
$\widehat{\beta}_{mc}^S\left(\tau, n, 2\sqrt{3}/7\right)$	.9486	.9998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.5.1: The empirical critical values, empirical significance levels, and empirical power estimates under $H^S_{\sqrt{3}/8}$, $N = 10000$, and $n = 10$ at $\alpha = .05$.

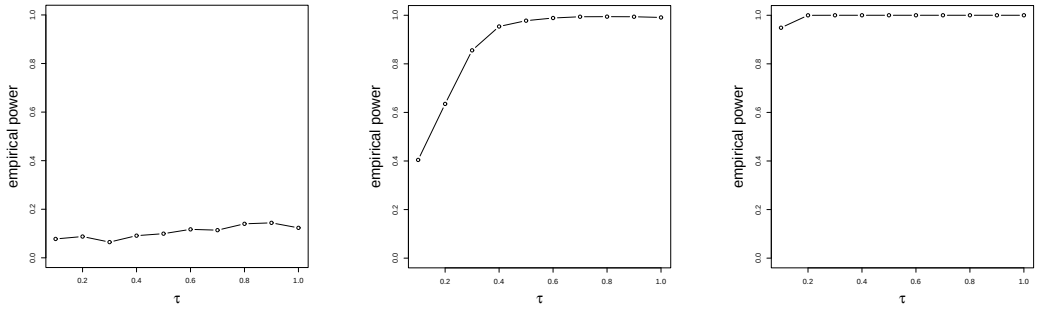


Figure 5.5.8: Monte Carlo power using the empirical critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of τ for $n = 10$.

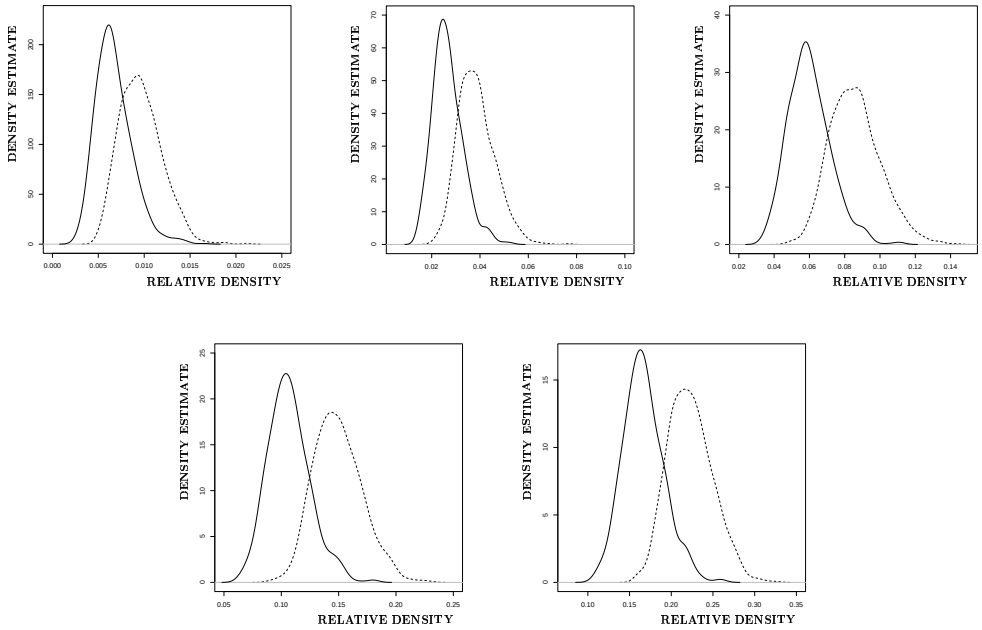


Figure 5.5.9: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/8}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 100$, and $N = 1000$ (left-to-right).

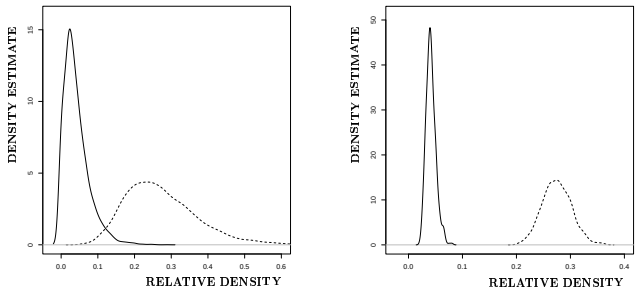


Figure 5.5.10: Kernel density estimates for the null (solid) and the segregation alternative $H^S_{\sqrt{3}/4}$ (dashed) for $\tau = .5$ with $n = 10$ and $N = 10000$ (left) and $n = 100$, $N = 1000$ (right).

τ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\widehat{c}^S_n(\tau)$	$.00\overline{29}$	$.01\overline{02}$	$.02\overline{16}$	$.03\overline{76}$	$.05\overline{78}$	$.08\overline{09}$	$.10\overline{92}$	$.14\,\overline{05}$	$.17\overline{51}$	$.21\overline{32}$
$\widehat{\alpha}^S_{mc}(\tau, n)$	.049	.05	.049	.05	.05	.05	.05	.05	.049	.05
$\widehat{\beta}^S_{mc}\left(\tau, n, \sqrt{3}/8\right)$	.225	.363	.461	.501	.544	.606	.625	.639	.628	.609
$\widehat{\beta}^S_{mc}\left(\tau, n, \sqrt{3}/4\right)$	1	1	1	1	1	1	1	1	1	1
$\widehat{\beta}^S_{mc}\left(\tau, n, 2\sqrt{3}/7\right)$	1	1	1	1	1	1	1	1	1	1

Table 5.5.2: The empirical critical values, empirical significance levels, and empirical power estimates under $H^S_{\sqrt{3}/8}$, $N = 1000$, and $n = 100$ at $\alpha = .05$.

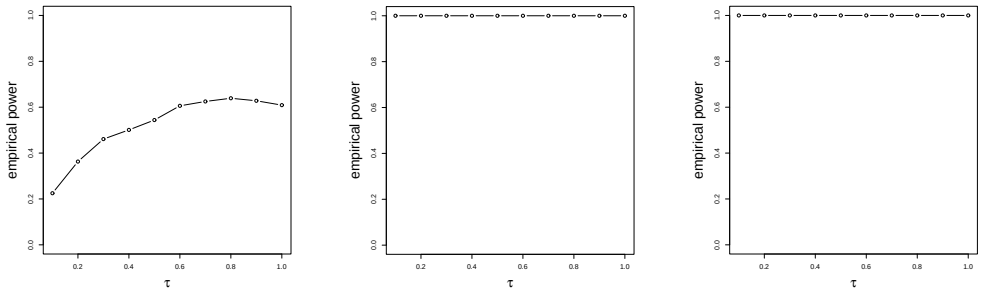


Figure 5.5.11: Monte Carlo power using the empirical critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of τ for $n = 100$.

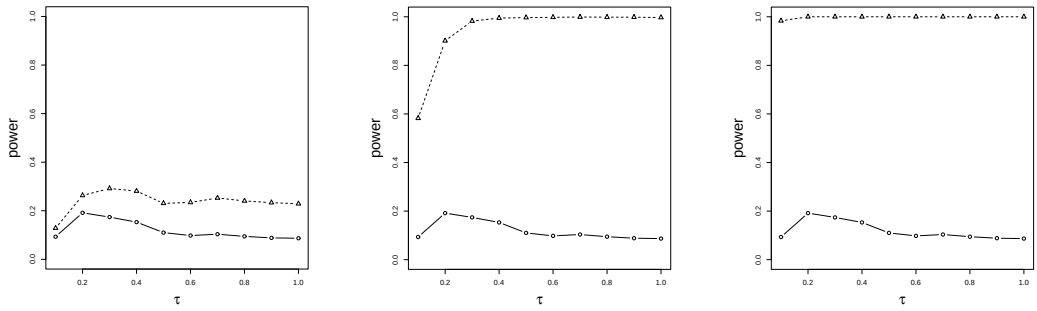


Figure 5.5.12: Monte Carlo power using the asymptotic critical value against segregation alternatives $H^S_{\sqrt{3}/8}$ (left), $H^S_{\sqrt{3}/4}$ (middle), and $H^S_{2\sqrt{3}/7}$ (right) as a function of τ for $n = 10$ and $N = 10000$. The circles represent the empirical significance levels while triangles represent the empirical power values.

we present a Monte Carlo investigation of power $\widehat{\beta}^S_n(\tau, \varepsilon)$ against $H^S_{\sqrt{3}/8}$, $H^S_{\sqrt{3}/4}$, and $H^S_{2\sqrt{3}/7}$ as a function of τ for $n = 10$. The empirical significance level, $\widehat{\alpha}_S(\tau, n)$, is about .10 for all τ values which indicate that $n = 10$ is not large enough for normal approximation. The empirical significance levels and empirical power $\widehat{\beta}^S_n(\tau, \varepsilon)$ values under H^S_ε for $\varepsilon \in \{\sqrt{3}/8,\ \sqrt{3}/4,\ 2\sqrt{3}/21\}$ are presented in Table 5.5.3. Notice that the smallest estimated significance level is .0866 at $\tau = 1$ with estimated power values $\widehat{\beta}^S_n(\tau = 1, \sqrt{3}/8) = .2289$, $\widehat{\beta}^S_n\left(\tau = 1, \sqrt{3}/4\right) = .9969$, and $\widehat{\beta}^S_n\left(\tau = 1, 2\sqrt{3}/7\right) = 1.000$. With $n = 20$ and $N = 10000$ the estimated significance levels get smaller (except for $\tau = .1$), but still are larger than .05 with smallest being .0763 at $\tau = 1$ with the corresponding power estimate $\widehat{\beta}^S_{20}\left(\tau = 1, \sqrt{3}/4\right) = .3038$. With $n = 100$, $N = 10000$ the estimated significance levels get even smaller, but still are larger than .05 with smallest being .063 at $\tau = .8$ with the corresponding power estimate $\widehat{\beta}^S_{100}\left(\tau = .8, \sqrt{3}/4\right) = .786$. In Figure 5.5.13, are the plots of the estimated significance levels and the empirical power estimates for $n \in \{10, 20, 100\}$ under $H^S_{\sqrt{3}/8}$. Observe that as n increases the estimated significance levels decrease, while the power estimates increase.

Based on the Monte Carlo power analysis, we see that for small n moderate τ values, $\tau \in [.4, .6]$, and for large n, large τ values $\tau = 1.0$ yield better performance for the normal approximation and the empirical power against association.

5.5.6 Monte Carlo Power Analysis Under Association

In association alternatives with $\varepsilon > 0$, we implement the Monte Carlo experiment for $\varepsilon \in \left\{\sqrt{3}/21,\ \sqrt{3}/12,\ 5\sqrt{3}/24\right\}$ and $\tau \in \left\{.1,\ .2,\ .3,\ .4,\ .5,\ .6,\ .7,\ .8,\ .9,\ 1.0\right\}$. For each τ value, we

$n = 10$ and $N = 10000$										
τ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\widehat{\alpha}_S(\tau, n)$	.0932	.1916	.1740	.1533	.1101	.0979	.1035	.0945	.0883	.0868
$\widehat{\beta}_n^S\left(\tau, \frac{\sqrt{3}}{8}\right)$	.1286	.2630	.2917	.2811	.2305	.2342	.2526	.2405	.2334	.2289
$\widehat{\beta}_n^S\left(\tau, \frac{\sqrt{3}}{4}\right)$	.5821	.9011	.9824	.9945	.9967	.9979	.9990	.9985	.9983	.9969
$\widehat{\beta}_n^S\left(\tau, 2\frac{\sqrt{3}}{7}\right)$	.9834	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$n = 20$ and $N = 10000$										
$\widehat{\alpha}_S(\tau, n)$	.2018	.1707	.1151	.1099	.0898	.0864	.0866	.0800	.0786	.0763
$\widehat{\beta}_n^S\left(\tau, \frac{\sqrt{3}}{8}\right)$	.2931	.3245	.2744	.3021	.2844	.2926	.3117	.3113	.3119	.3038
100, $N = 1000$										
$\widehat{\alpha}_S(\tau, n)$	0.155	0.101	0.080	0.077	0.075	0.066	0.065	0.063	0.066	0.069
$\widehat{\beta}_n^S\left(\tau, \frac{\sqrt{3}}{8}\right)$	0.574	0.574	0.612	0.655	0.709	0.742	0.774	0.786	0.793	0.793

Table 5.5.3: The empirical significance level and empirical power values under H_ε^S for $\varepsilon \in \{\sqrt{3}/8, \sqrt{3}/4, 2\sqrt{3}/21\}$ at $\alpha = .05$.

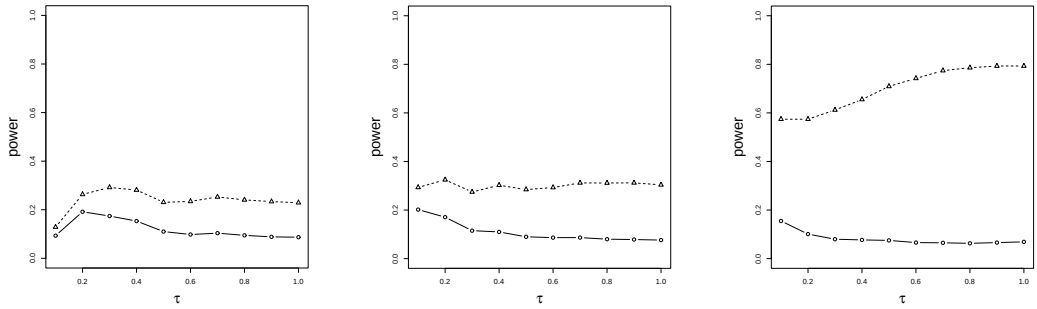


Figure 5.5.13: Monte Carlo power using the asymptotic critical value against segregation alternative $H^S_{\sqrt{3}/8}$, as a function of τ for $n = 10$ and $N = 10000$ (left), $n = 20$ and $N = 10000$ (middle), and $n = 100$ and $N = 1000$ (right). The circles represent the empirical significance levels while triangles represent the empirical power values.

estimate the empirical critical value $\widehat{c}_n^A(\tau)$ and the empirical significance level $\widehat{\alpha}_{mc}^A(\tau, n)$ under H_0 and the empirical power $\widehat{\beta}_{mc}^A(\tau, n, \varepsilon)$.

The kernel density estimates under the null case and under the association alternatives H_ε^A with $\varepsilon \in \{\sqrt{3}/21, \sqrt{3}/12, 5\sqrt{3}/24\}$, $\tau \in \{.2, .4, .6, .8, 1.0\}$, $n = 10$ and $N = 10000$ are plotted in Figures 5.5.14, 5.5.15, and 5.5.16, respectively. Observe that under H_0 kernel density estimates are skewed right for small τ values, (with skewness increasing as τ gets smaller) and almost symmetric for large τ values, with most symmetry occurring at $\tau = 1$. Kernel density estimates are skewed right under $H^A_{\sqrt{3}/21}$, $H^A_{\sqrt{3}/12}$, and $H^A_{5\sqrt{3}/24}$. The empirical critical values, empirical significance levels, and empirical power estimates under H_ε^A with 10000 Monte Carlo replicates are presented in Table 5.5.4. Notice that the estimated significance levels are all

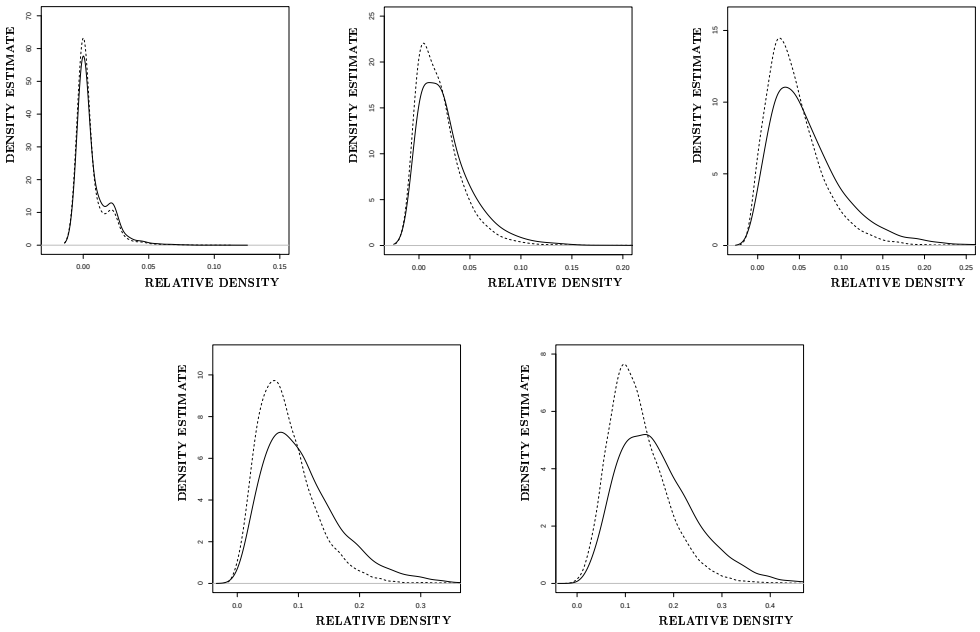


Figure 5.5.14: Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/21}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ with $n = 10$ and $N = 10000$ (left-to-right).

below .05, and about .05 at $\tau \in \{.6, .9\}$. Under $H^A_{\sqrt{3}/21}$ and $H^A_{\sqrt{3}/12}$, with $n = 10$, the null and alternative probability density functions for $\rho_{10}(\tau)$ are very similar, implying small power. Notice also that more severe the segregation, larger the empirical power estimates at each τ value. These empirical power estimates are plotted in Figure 5.5.17. Note that maximum Monte Carlo power estimate occurs at $\tau = .9$ for all ε values. estimate increases as τ gets larger.

With $n = 100$, there is more separation between null and alternative probability density functions. See Figure 5.5.18 for the kernel density estimates under H_0 and $H^A_{\sqrt{3}/21}$ with $N = 1000$ Monte Carlo replicates. Notice that the probability density functions are almost symmetric for all τ values. The corresponding empirical critical values, empirical significance levels, and empirical powers are presented in Table 5.5.5. The estimated critical values are all about .05 for $\tau \geq .3$. As n gets larger, the normal approximation gets more accurate and the separation between the null and alternative density estimates increase. See for example Figure 5.5.19 for $\tau = .5$ with $n = 10$ and $N = 10000$ (left) and $n = 100$, $N = 1000$ (right). The empirical power estimates are plotted in Figure 5.5.20. Note that maximum Monte Carlo power estimate occurs at $\tau = 1$ for all ε values.

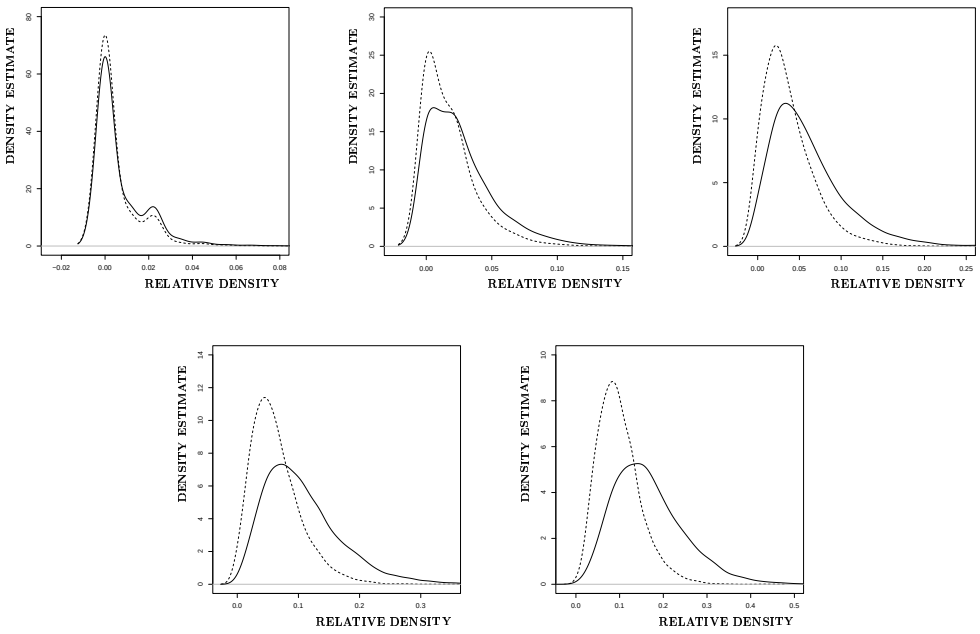


Figure 5.5.15: Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/12}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ (left-to-right).

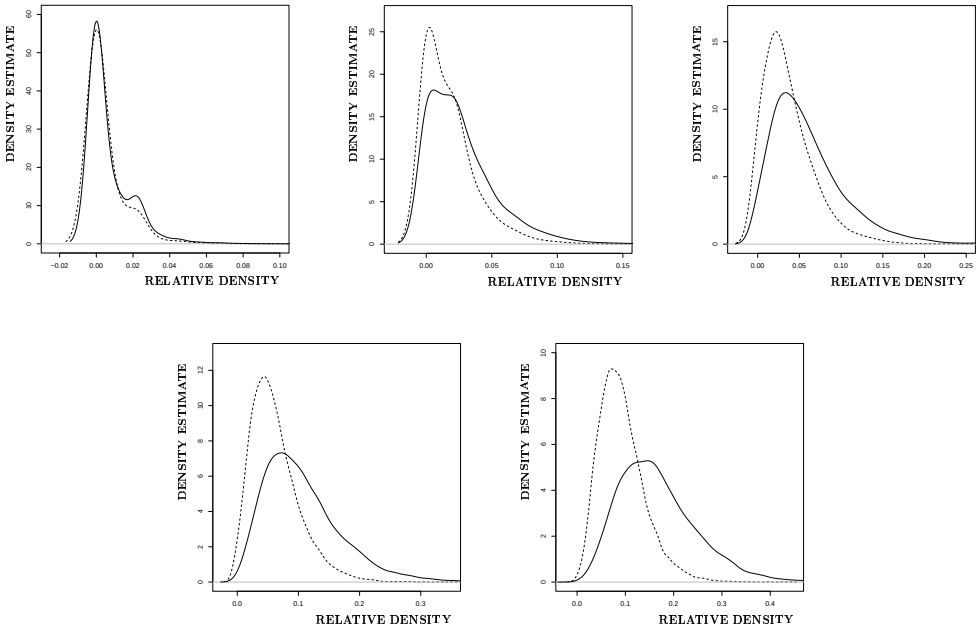


Figure 5.5.16: Kernel density estimates for the null (solid) and the association alternative $H^A_{5\sqrt{3}/24}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ (left-to-right).

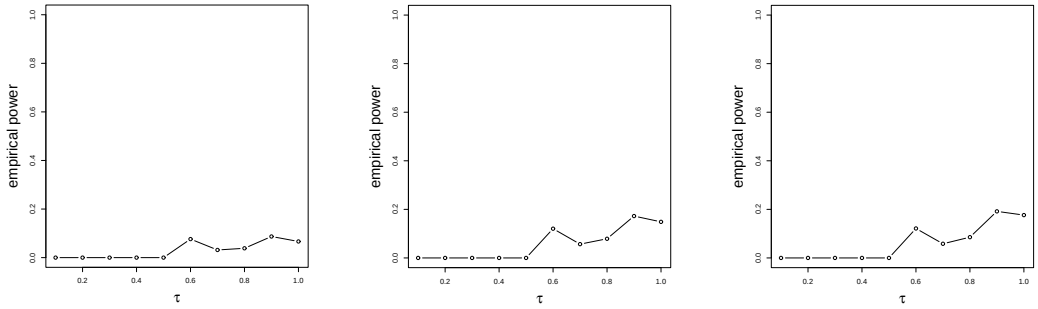


Figure 5.5.17: Monte Carlo power using the empirical critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of τ with $n = 10$ and $N = 10000$.

τ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\widehat{c}^A_n(\tau)$	0	0	0	0	0	$.1\bar{1}$	$.1\bar{1}$	$.0\bar{2}$	$.0\bar{4}$	$.0\bar{5}$
$\widehat{\alpha}^A_{mc}(\tau, n)$	0	0	0	0	0	.0471	.0163	.0217	.0451	.0343
$\widehat{\beta}^A_{mc}\left(\tau, n, \sqrt{3}/21\right)$	0	0	0	0	0	.0763	.0315	.0384	.087	.0665
$\widehat{\beta}^A_{mc}\left(\tau, n, \sqrt{3}/12\right)$	0	0	0	0	0	.1204	.0568	.0789	.1724	.1486
$\widehat{\beta}^A_{mc}\left(5\sqrt{3}/24\right)$	0	0	0	0	0	.1213	.0583	.0855	.1918	.1764

Table 5.5.4: The empirical critical values, empirical significance levels, and empirical power estimates under H^A_ε with $n = 10$ at $\alpha = .05$.

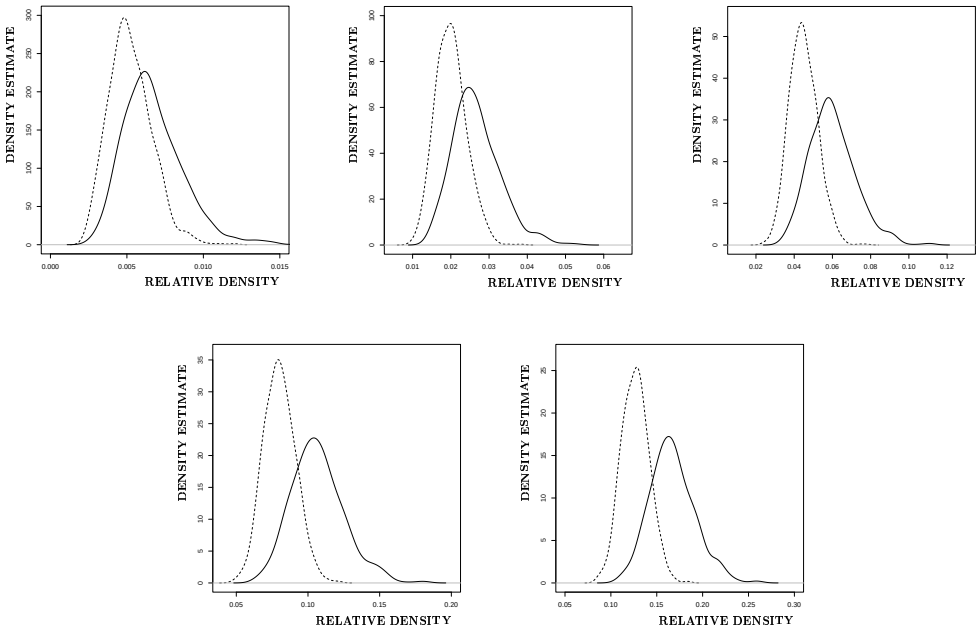


Figure 5.5.18: Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/21}$ (dashed) for $\tau \in \{.2, .4, .6, .8, 1.0\}$ with $n = 100$, $N = 1000$ (left-to-right).

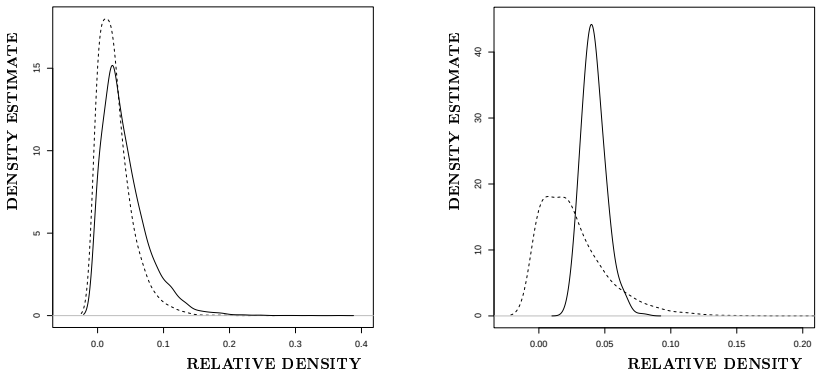


Figure 5.5.19: Kernel density estimates for the null (solid) and the association alternative $H^A_{\sqrt{3}/12}$ (dashed) for $\tau = .5$ with $n = 10$ and $N = 10000$ (left) and $n = 100$, $N = 1000$ (right).

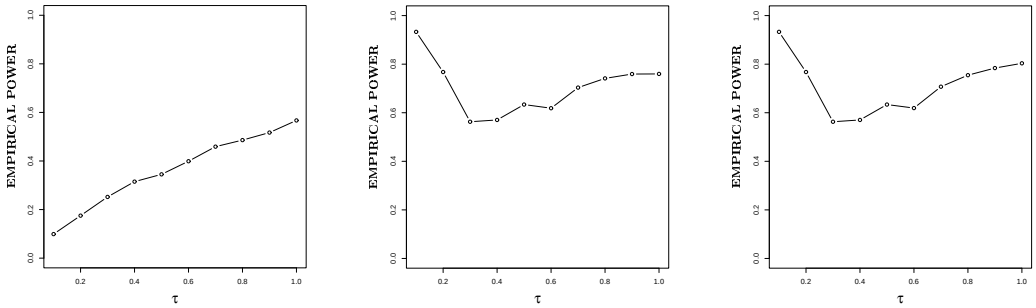


Figure 5.5.20: Monte Carlo power using the empirical critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of τ with $n = 100$ and $N = 1000$.

τ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\widehat{c}^A_n(\tau)$	$.00\overline{70}$	$.00\overline{40}$	$.00\overline{92}$	$.01\overline{75}$	$.02\overline{84}$	$.04\,\overline{2}$	$.05\overline{92}$	$.07\overline{952}$	$.10\overline{25}$	$.13\overline{03}$
$\widehat{\alpha}^A_{mc}(\tau, n)$	.036	.045	.048	.049	.049	.048	.049	.049	.049	.049
$\widehat{\beta}^A_{mc}\left(\tau, n, \sqrt{3}/21\right)$	.099	.175	.252	.315	.345	.399	.459	.486	.517	.567
$\widehat{\beta}^A_{mc}\left(\tau, n, \sqrt{3}/12\right)$	.933	.768	.563	.570	.634	.619	.704	.741	.759	.760
$\widehat{\beta}^A_{mc}\left(\tau, n, 5\sqrt{3}/24\right)$	.933	.768	.563	.570	.634	.619	.707	.755	.784	.804

Table 5.5.5: The empirical critical values, empirical significance levels, and empirical power estimates under $H^A_{\sqrt{3}/21}$ and $n = 100$ at $\alpha = .05$.

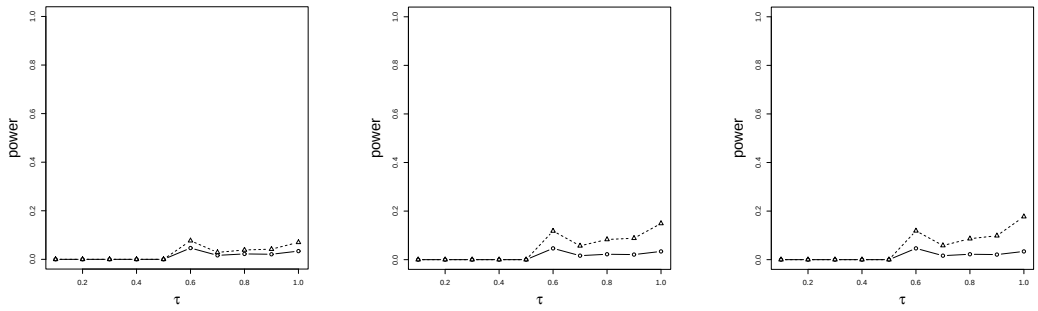


Figure 5.5.21: Monte Carlo power using the asymptotic critical value against association alternatives $H^A_{\sqrt{3}/21}$ (left), $H^A_{\sqrt{3}/12}$ (middle), and $H^A_{5\sqrt{3}/24}$ (right) as a function of τ for $n = 10$. The circles represent the empirical significance levels while triangles represent the empirical power values.

We also estimate the power — using the asymptotic critical value— as a function of the proximity factor τ. For each τ value, the α-level asymptotic critical value is $\mu\big(N^\tau_{CS}\big) + z_\alpha \cdot \sqrt{\nu\big(N^\tau_{CS}\big)/n}$. In Figure 5.5.21, we present a Monte Carlo investigation of power $\widehat{\beta}^A_n(\tau, \varepsilon)$ against $H^A_{\sqrt{3}/21}$, $H^A_{\sqrt{3}/12}$, and $H^A_{5\sqrt{3}/24}$ as a function of τ for $n = 10$. The empirical significance level $\widehat{\alpha}_A(\tau, n)$ is about .05 for $\tau = .6$ and much smaller for other τ values which indicate that $n = 10$ is not large enough for normal approximation. The empirical significance levels and empirical power $\widehat{\beta}^A_n(\tau, \varepsilon)$ values under H^A_ε for $\varepsilon \in \{\sqrt{3}/21,\ \sqrt{3}/12,\ 5\sqrt{3}/24\}$ are presented in Table 5.5.6. Notice that the estimated power values $\widehat{\beta}^A_n\left(\tau = .6, \sqrt{3}/21\right) = .0767$, $\widehat{\beta}^A_n\left(\tau = .6, \sqrt{3}/21\right) = .1181$, and $\widehat{\beta}^A_n\left(\tau = .6, 5\sqrt{3}/24\right) = .1187$.

In Figure 5.5.21, we present a Monte Carlo investigation of power against $H^A_{\sqrt{3}/21}$, $H^A_{\sqrt{3}/12}$, and $H^A_{5\sqrt{3}/24}$ as a function of τ for $n = 10$. The empirical significance level, $\widehat{\alpha}_A(n)$, is about .05 for $\tau = .6$ which have the empirical power $\widehat{\beta}^A_{10}\left(\tau, \sqrt{3}/12\right) = .1181$ with maximum power at also $\tau = .6$, and $\widehat{\beta}^A_{10}\left(\tau, 5\sqrt{3}/24\right) = .1187$. So, for small sample sizes, moderate values of τ is more appropriate for normal approximation, as they yield the desired significance level and have the property that more severe association yields higher power.

The empirical significance levels and empirical power $\widehat{\beta}^S_n(\tau, \varepsilon)$ values under H^A_ε for $\varepsilon \in \{5\sqrt{3}/24,\ \sqrt{3}/12,\ \sqrt{3}/21\}$ are presented in Table 5.5.6.

With $n = 20$ and $N = 10000$ the estimated significance levels gets closer to .05 for $\tau \in \{.4, .7, .8, .9, 1.0\}$, with closest being .0496 at $\tau = .6$ with the corresponding power estimate $\widehat{\beta}^A_n\left(\tau = .6, \sqrt{3}/12\right) = .1497$. With $n = 100$, $N = 1000$ the estimated significance levels gets even closer to .05 for all τ values, with closest being .049 at $\tau \in \{.5, .9, 1\}$ with the corresponding

$n = 10$ and $N = 10000$										
τ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\widehat{\alpha}_A(\tau, n)$	0	0	0	0	0	.0465	.0164	.0223	.0209	.0339
$\widehat{\alpha}_A(n)$	0	0	0	0	0	.0465	.0164	.0223	.0209	.0339
$\widehat{\beta}_n^A\left(\tau, \sqrt{3}/21\right)$	0	0	0	0	0	.0767	.0286	.0386	.0419	.0700
$\widehat{\beta}_n^A\left(\tau, \sqrt{3}/12\right)$	0	0	0	0	0	.1181	.0569	.0831	.0882	.1490
$\widehat{\beta}_n^A\left(\tau, 5\sqrt{3}/24\right)$	0	0	0	0	0	.1187	.0581	.0863	.0985	.1771
$n = 20$ and $N = 10000$										
$\widehat{\alpha}_A(\tau, n)$	.6603	.2203	.1069	.0496	.0338	.0301	.0290	.0267	.0333	.0372
$\widehat{\beta}_n^A\left(\tau, \sqrt{3}/12\right)$	.7398	.3326	.2154	.1497	.1442	.1608	.1818	.2084	.2663	.3167
$n = 100$, $N = 1000$										
$\widehat{\alpha}_A(\tau, n)$	.169	.075	.053	.047	.049	.044	.040	.044	.049	.049
$\widehat{\beta}_n^A\left(\tau, \sqrt{3}/12\right)$	.433	.399	.460	.559	.687	.789	.887	.938	.977	.997

Table 5.5.6: The empirical significance level and empirical power values under H_ε^A for $\varepsilon \in \left\{5\sqrt{3}/24, \sqrt{3}/12, \sqrt{3}/21\right\}$ with $N = 10000$, and $n \in \{10, 20, 100\}$ at $\alpha = .05$.

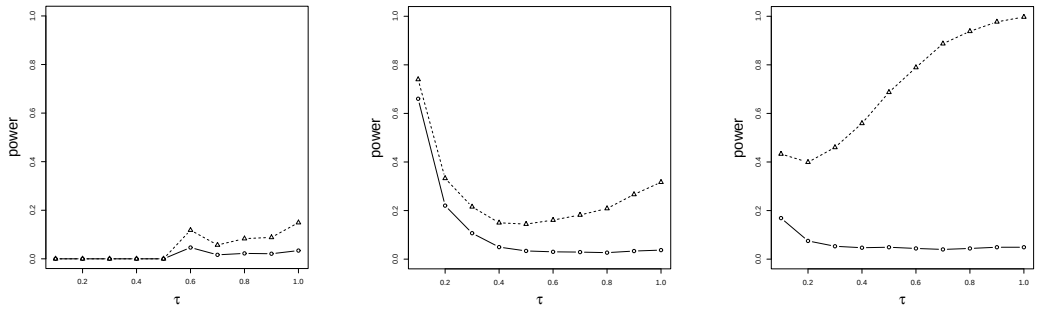


Figure 5.5.22: Monte Carlo power using the asymptotic critical value against segregation alternative $H^A_{\sqrt{3}/12}$, as a function of τ for $n = 10$ and $N = 10000$ (left), $n = 20$ and $N = 10000$ (middle), and $n = 100$, $N = 1000$ (right). The circles represent the empirical significance levels while triangles represent the empirical power values.

power estimates $\widehat{\beta}_n^A\left(\tau, \sqrt{3}/12\right) \in \{.687, .977, .997\}$, respectively. See Figure 5.5.22 for the plots of the estimated significance levels and the power estimates for $n \in \{10, 20, 100\}$ under $H^A_{\sqrt{3}/12}$. Observe that high power estimates with small significance levels occur for $\tau \geq .5$, and for these τ values, as n increases, the power estimates increase, while the significance levels decrease.

Based on the Monte Carlo power analysis, for both small and large n, we see that large τ values, e.g.,$\tau = .8$, yield the better performance for the normal approximation and the empirical power against association.

5.5.7 Pitman Asymptotic Efficacy

For segregation or association alternatives the PAE of $\rho_n\left(N_{CS}^{\tau}\right)$ is given by $\text{PAE}(\tau) = \frac{\left(\mu^{''}\left(N_{CS}^{\tau},\varepsilon=0\right)\right)^2}{\nu\left(N_{CS}^{\tau}\right)}$, since $\mu^{''}\left(N_{CS}^{\tau},\varepsilon=0\right) \neq 0$ but $\mu^{'}\left(N_{CS}^{\tau},\varepsilon=0\right) = 0$.

5.5.7.1 Pitman Asymptotic Efficacy Under Segregation Alternatives

Consider the test sequences $\rho(\tau) = \left\{\rho_n\left(N_{CS}^{\tau}\right)\right\}$ for sufficiently small $\varepsilon > 0$ and $\tau \in (0,1]$.

In the PAE framework of Section 2.7.2.1, $\theta = \varepsilon$ and $\theta_0 = 0$. Suppose, $\mu_n(\varepsilon) = E_{\varepsilon}^{S}[\rho_n\left(N_{CS}^{\tau}\right)] = \mu_S\left(N_{CS}^{\tau},\varepsilon\right)$. For $\varepsilon \in \left[0,\sqrt{3}/5\right)$,

$$\mu_S\left(N_{CS}^{\tau},\varepsilon\right) = \sum_{j=1}^{2} \varpi_{1,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j)$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 1-\sqrt{3}\,\varepsilon\right)$, $\mathcal{I}_2 = \left[1-\sqrt{3}\,\varepsilon, 1\right)$. See Section C.3.1 for the explicit form of $\mu_S\left(N_{CS}^{\tau},\varepsilon\right)$. Notice that as $\varepsilon \to 0$, only $\mathcal{I}_1 = \left[1, 3/2-\sqrt{3}\,\varepsilon\right)$ does not vanish, so we only keep the component of $\mu_S\left(N_{CS}^{\tau},\varepsilon\right)$ on $\mathcal{I}_1$.

Furthermore,

$$\begin{aligned}\sigma_n^2(\varepsilon) &= \mathbf{Var}_{\varepsilon}^{S}(\rho_n\left(N_{CS}^{\tau}\right)) \\ &= \frac{1}{2\,n\,(n-1)}\mathbf{Var}_{\varepsilon}^{S}[h_{12}(N_{CS}^{\tau},M_C)] + \frac{(n-2)}{n\,(n-1)}\mathbf{Cov}_{\varepsilon}^{S}[h_{12}(N_{CS}^{\tau},M_C),h_{13}(N_{CS}^{\tau},M_C)].\end{aligned}$$

The explicit forms of $\mathbf{Var}_{\varepsilon}^{S}[h_{12}(N_{CS}^{\tau},M_C)]$ and $\mathbf{Cov}_{\varepsilon}^{S}[h_{12}(N_{CS}^{\tau},M_C),h_{13}(N_{CS}^{\tau},M_C)]$ are not calculated, since we only need $\lim_{n\to\infty}\sigma_n^2(\varepsilon=0) = \nu\left(N_{CS}^{\tau}\right)$ which is given in Equation 5.5.2.

Notice that $\mathbf{E}_{\varepsilon}^{S}|h_{12}(N_{CS}^{\tau},M_C)|^3 \leq 8 < \infty$ and $\mathbf{E}_{\varepsilon}^{S}[h_{12}(N_{CS}^{\tau},M_C)\,h_{13}(N_{CS}^{\tau},M_C)] - \mathbf{E}_{\varepsilon}^{S}[h_{12}]^2 = \mathbf{Cov}_{\varepsilon}^{S}[h_{12}(N_{CS}^{\tau},M_C),h_{13}(N_{CS}^{\tau},M_C)] > 0$ then (PC1) follows for each $\tau \in (0,1]$ and $\varepsilon \in \left[0,\sqrt{3}/3\right)$.

Differentiating $\mu_S\left(N_{CS}^{\tau},\varepsilon\right)$ with respect to ε yields

$$\mu_S'\left(N_{CS}^{\tau},\varepsilon\right) = \frac{8\,\varepsilon\,\tau^2(5\,\varepsilon^2\tau - 9\,\varepsilon^2 - 3\,\tau + 3)}{9(1-2\,\varepsilon)^3(2\,\varepsilon+1)^3(1-\tau)}$$

hence $\mu_S'(\tau,\varepsilon=0) = 0$, so we need higher order derivatives for (PC2). Differentiating $\mu_S'\left(N_{CS}^{\tau},\varepsilon\right)$

with respect to ε, we get

$$\mu_S''\left(N_{CS}^{\tau},\varepsilon\right)=\frac{8\,\tau^2(20\,\varepsilon^4\tau-15\,\varepsilon^2\tau-36\,\varepsilon^4+11\,\varepsilon^2-\tau+1)}{3\,(2\,\varepsilon-1)^4(2\,\varepsilon+1)^4(1-\tau)}.$$

Hence

$$\mu_S''\left(N_{CS}^{\tau},\varepsilon=0\right)=8\,\tau^2/3. \tag{5.5.4}$$

Observe that $\mu_S''\left(N_{CS}^{\tau},\varepsilon=0\right)>0$ for all $\tau\in(0,1]$, so (PC2) holds with the second derivative. (PC3) in the second derivative form follows from continuity of $\mu_S''\left(N_{CS}^{\tau},\varepsilon\right)$ in ε and (PC4) follows from continuity of $\sigma_n^2(\tau,\varepsilon)$ in ε.

Next, we find $C_S(\rho(\tau))=\lim_{n\to\infty}\frac{\mu_S''\left(N_{CS}^{\tau},\varepsilon=0\right)}{\sqrt{n}\,\sigma_n\,(\tau,\varepsilon=0)}=\frac{\mu_S''\left(N_{CS}^{\tau},\varepsilon=0\right)}{\sqrt{\nu\left(N_{CS}^{\tau}\right)}}$, where numerator is given in Equation 5.5.4 and denominator is given in Equation 5.5.2. We can easily see that $C_S(\rho(\tau))>0$, then (PC5) follows.

So under segregation alternatives H_{ε}^{S}, the PAE of $\rho_n\left(N_{CS}^{\tau}\right)$ is given by

$$\mathrm{PAE}^{S}\left(N_{CS}^{\tau}\right)=C_S^2(\rho(\tau))=\frac{\left(\mu_S''\left(N_{CS}^{\tau},\varepsilon=0\right)\right)^2}{\nu\left(N_{CS}^{\tau}\right)}.$$

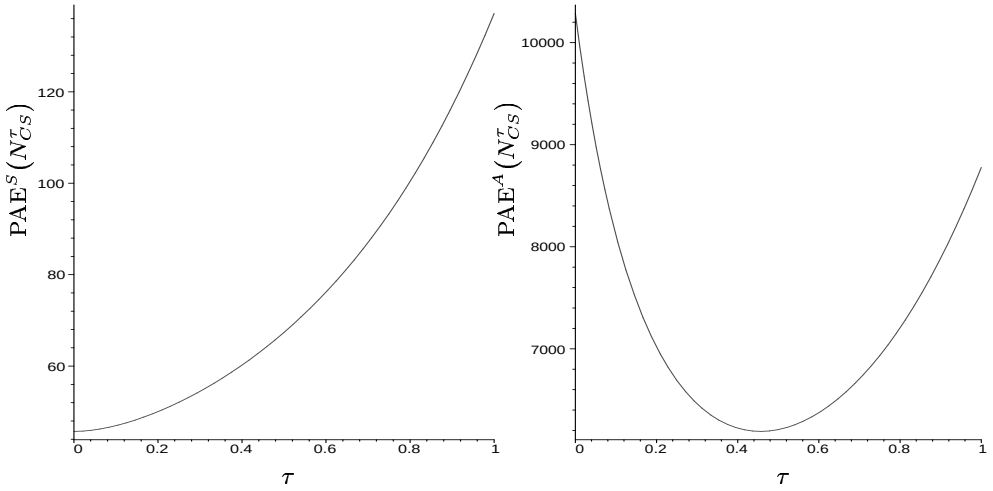


Figure 5.5.23: Pitman asymptotic efficacy against segregation (left) and against association (right) as a function of τ.

In Figure 5.5.23 (left), we present the PAE as a function of τ for segregation. Notice that $\lim_{\tau\to 0}\mathrm{PAE}^{S}\left(N_{CS}^{\tau}\right)=320/7\approx 45.7143$, $\mathrm{argsup}_{\tau\in(0,1]}\,\mathrm{PAE}^{S}\left(N_{CS}^{\tau}\right)=1.0$, and $\mathrm{PAE}^{S}\left(N_{CS}^{\tau=1}\right)=960/7\approx 137.1429$. Based on the PAE analysis, we suggest, for large n and small ε, choosing τ

large for testing against segregation.

5.5.7.2 Pitman Asymptotic Efficacy Under Association Alternatives

Consider the test sequences $\rho(\tau) = \{\rho_n(N_{CS}^\tau)\}$ for sufficiently small $\varepsilon > 0$ and $\tau \in (0,1]$.

In the PAE framework of Section 2.7.2.1, $\theta = \varepsilon$ and $\theta_0 = 0$. Suppose, $\mu_n(\varepsilon) = E_\varepsilon^A[\rho_n(N_{CS}^\tau)] = \mu_A(N_{CS}^\tau, \varepsilon)$. For $\varepsilon \in [0, \sqrt{3}/21)$,

$$\mu_A(N_{CS}^\tau, \varepsilon) = \sum_{j=1}^{7} \varpi_{1,j}(\tau, \varepsilon)\, \mathbf{I}(\tau \in \mathcal{I}_j)$$

with the corresponding intervals

$\mathcal{I}_1 = \left[0, \frac{3\sqrt{3}\,\varepsilon}{2\,(1-\sqrt{3}\,\varepsilon)}\right)$, $\mathcal{I}_2 = \left[\frac{3\sqrt{3}\,\varepsilon}{2\,(1-\sqrt{3}\,\varepsilon)}, \frac{2\sqrt{3}\,\varepsilon}{1-2\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_3 = \left[\frac{2\sqrt{3}\,\varepsilon}{1-2\sqrt{3}\,\varepsilon}, \frac{3\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_4 = \left[\frac{3\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}, \frac{3\sqrt{3}\,\varepsilon}{1-4\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_5 = \left[\frac{3\sqrt{3}\,\varepsilon}{1-4\sqrt{3}\,\varepsilon}, \frac{6\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_6 = \left[\frac{6\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}, \frac{6\sqrt{3}\,\varepsilon}{1-4\sqrt{3}\,\varepsilon}\right)$, and $\mathcal{I}_7 = \left[\frac{6\sqrt{3}\,\varepsilon}{1-4\sqrt{3}\,\varepsilon}, 1\right)$. Notice that as $\varepsilon \to 0$, only $\mathcal{I}_7$ does not vanish, so we only keep the component of $\mu_A(N_{CS}^\tau, \varepsilon)$ on $\mathcal{I}_7$. See Section C.3.2 for the explicit form of $\mu(\tau, \varepsilon)$.

Furthermore,

$$\begin{aligned}\sigma_n^2(\tau,\varepsilon) &= \mathbf{Var}_\varepsilon^A(\rho_n(N_{CS}^\tau)) \\ &= \frac{1}{2\,n\,(n-1)}\mathbf{Var}_\varepsilon^A[h_{12}(N_{CS}^\tau, M_C)] + \frac{(n-2)}{n\,(n-1)}\,\mathbf{Cov}_\varepsilon^A[h_{12}(N_{CS}^\tau, M_C), h_{13}(N_{CS}^\tau, M_C)]\end{aligned}$$

whose explicit form is not calculated, since we only need $\lim_{n\to\infty}\sqrt{n}\,\sigma_n(\varepsilon = 0) = \nu(N_{CS}^\tau)$ which is given Equation 5.5.2.

(PC1) follows for each $\tau \in (0,1]$ and $\varepsilon \in [0, \sqrt{3}/3)$ as in the segregation case.

Differentiating $\mu_A(N_{CS}^\tau, \varepsilon)$ with respect to ε, we get

$$\begin{aligned}\mu_A'(N_{CS}^\tau, \varepsilon) = -72\,\Big[\sqrt{3}\Big(10\,\sqrt{3}\tau^2 + 10\,\tau^3\sqrt{3} - 90\,\varepsilon\tau^4 + 198\,\tau^4\sqrt{3}\,\varepsilon^2 - 360\,\tau^4\,\varepsilon^3 - 90\,\tau^2\,\varepsilon + 126\,\sqrt{3}\,\varepsilon^2\tau^2 - \\ 360\,\tau^2\,\varepsilon^3 - 225\,\varepsilon\tau^3 + 495\,\tau^3\sqrt{3}\,\varepsilon^2 - 900\,\tau^3\,\varepsilon^3 - 144\,\sqrt{3}\,\varepsilon^2 - 396\,\sqrt{3}\,\varepsilon^2\tau + 4\,\tau^4\sqrt{3}\Big)\varepsilon\Big] \Big/ \Big[\Big(\sqrt{3} + 6\,\varepsilon\Big)^3 \\ \Big(\sqrt{3} - 6\,\varepsilon\Big)^3 (1 + 2\,\tau)(\tau + 2)\Big].\end{aligned}$$

Hence $\mu'_A\big(N^\tau_{CS}, \varepsilon = 0\big) = 0$, so we differentiate $\mu'_A\big(N^\tau_{CS}, \varepsilon\big)$ with respect to ε and get

$$\mu''_A\big(N^\tau_{CS}, \varepsilon\big) = 648\Big[1920\sqrt{3}\varepsilon^3\tau^2 + 1920\sqrt{3}\varepsilon^3\tau^4 - 10\,\tau^2 - 17820\,\tau^3\varepsilon^4 + 5184\,\varepsilon^4 - 10\,\tau^3 - 4\,\tau^4 -$$
$$978\,\varepsilon^2\tau^2 - 2085\,\varepsilon^2\tau^3 + 14256\,\varepsilon^4\tau - 4536\,\varepsilon^4\tau^2 + 4800\sqrt{3}\varepsilon^3\tau^3 + 150\sqrt{3}\,\varepsilon\,\tau^3 + 60\sqrt{3}\,\varepsilon\,\tau^2 +$$
$$2880\sqrt{3}\varepsilon^5\tau^2 + 7200\sqrt{3}\varepsilon^5\tau^3 + 2880\sqrt{3}\varepsilon^5\tau^4 - 7128\,\varepsilon^4\tau^4 + 432\,\varepsilon^2 + 1188\,\varepsilon^2\tau - 834\,\varepsilon^2\tau^4 +$$
$$60\,\varepsilon\sqrt{3}\tau^4\Big]\Big/\Big[\Big(\sqrt{3}+6\,\varepsilon\Big)^4\Big(\sqrt{3}-6\,\varepsilon\Big)^4(1+2\,\tau)(\tau+2)\Big].$$

Thus,

$$\mu''_A\big(N^\tau_{CS}, \varepsilon = 0\big) = -16\,\frac{\tau^2(5+5\,\tau+2\,\tau^2)}{(1+2\,\tau)(\tau+2)}. \tag{5.5.5}$$

Note that $\mu''(\tau, \varepsilon = 0) < 0$ for all $\tau \in (0,1]$, so (PC2) follows with the second derivative. (PC3) and (PC4) follow from continuity of $\mu''\big(N^\tau_{CS}, \varepsilon\big)$ and $\sigma_n^2(\tau, \varepsilon)$ in ε.

Next, we find $C_A(\rho(\tau)) = \lim_{n\to\infty} \frac{\mu''_A\left(N^\tau_{CS}, \varepsilon=0\right)}{\sqrt{n}\,\sigma_n(\tau,\varepsilon=0)} = \frac{\mu''_A\left(N^\tau_{CS}, \varepsilon=0\right)}{\sqrt{\nu\left(N^\tau_{CS}\right)}}$, by substituting the numerator from Equation 5.5.5 and denominator from Equation 5.5.2. We can easily see that $C_A(\rho(\tau)) < 0$ for all $\tau \in (0,1]$. Then (PC5) holds, so under association alternatives H^A_ε, the PAE of $\rho_n\big(N^\tau_{CS}\big)$ is

$$\mathrm{PAE}^A\big(N^\tau_{CS}\big) = C_A^2(\rho(\tau)) = \frac{\big(\mu''_A\big(N^\tau_{CS}, \varepsilon = 0\big)\big)^2}{\nu\big(N^\tau_{CS}\big)}.$$

In Figure 5.5.23 (right), we present the PAE as a function of τ for association. Notice that $\lim_{\tau\to 0}\mathrm{PAE}^A\big(N^\tau_{CS}\big) = 72000/7 \approx 10285.7143$, $\mathrm{PAE}^A\big(N^{\tau=1}_{CS}\big) = 61440/7 \approx 8777.1429$, $\mathrm{arginf}_{\tau\in(0,1]}\mathrm{PAE}^A\big(N^\tau_{CS}\big) \approx .4566$ with $\mathrm{PAE}^A\big(N^{\tau\approx.4566}_{CS}\big) \approx 6191.0939$. Based on the PAE analysis, we suggest, for large n and small ε, choosing τ small for testing against association. However, for small and moderate values of n normal approximation is not appropriate due to the skewness in the density of $\rho_n\big(N^\tau_{CS}\big)$. Therefore, for small n, we suggest large τ values.

Remark 5.5.6. Hodges-Lehmann asymptotic efficacy HLAE of $\rho_n\big(N^\tau_{CS}\big)$ is given by

$$\mathrm{HLAE}\big(N^\tau_{CS}, \varepsilon\big) := \frac{\big(\mu\big(N^\tau_{CS}, \varepsilon\big) - \mu\big(N^\tau_{CS}\big)\big)^2}{\nu\big(N^\tau_{CS}, \varepsilon\big)}.$$

Unlike PAE, HLAE does only involve $n \to \infty$ at a fixed $\varepsilon > 0$. Hence HLAE requires the mean and, especially, the asymptotic variance of $\rho_n\big(N^\tau_{CS}\big)$ *under a fixed alternative.* So, one can

investigate HLAE for specific values of ε, if not for all $\varepsilon \in \left(0, \sqrt{3}/3\right)$. $\square$

Remark 5.5.7. The asymptotic power function allows investigation as a function of τ, n, and ε using the asymptotic critical value and an appeal to normality. Under a specific segregation alternative H_ε^S, the asymptotic power function is given by

$$\Pi_S\left(N_{CS}^\tau, n, \varepsilon\right) = 1 - \Phi\left(\frac{z_{(1-\alpha)} \cdot \sqrt{\nu\left(N_{CS}^\tau\right)} + \sqrt{n}\ \left(\mu\left(N_{CS}^\tau\right) - \mu_S\left(N_{CS}^\tau, \varepsilon\right)\right)}{\sqrt{\nu_S\left(N_{CS}^\tau, \varepsilon\right)}}\right).$$

Under H_ε^A, we have

$$\Pi_A\left(N_{CS}^\tau, n, \varepsilon\right) = \Phi\left(\frac{z_\alpha \sqrt{\nu\left(N_{CS}^\tau\right)} + \sqrt{n} \cdot \left(\mu\left(N_{CS}^\tau\right) - \mu_A\left(N_{CS}^\tau, \varepsilon\right)\right)}{\sqrt{\nu_A\left(N_{CS}^\tau, \varepsilon\right)}}\right). \square$$

5.5.8 Multiple Triangle Case

As in Section 5.4.11, we wish to test

$$H_0 : X_i \overset{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}))$$

against segregation and association alternatives.

Figure 5.4.28 and Figure 4.6.1 are graphs of realizations of $n = 100$ and $n = 1000$ observations which are independent and identically distributed according to $\mathcal{U}(C_H(\mathcal{Y}))$ for $|\mathcal{Y}| = 10$ and $J = 13$ and under segregation and association for the same $\mathcal{Y}$.

Let $\rho_n\left(N_{CS}^\tau, J\right)$ be the relative density of the digraph based on $\mathcal{X}_n$ and $\mathcal{Y}$ which yields J Delaunay triangles, and let $w_j = A(T_j)/A(C_H(\mathcal{Y}))$ for $j \in \{1, \ldots, J\}$. Then we obtain the following as a corollary to Theorem 5.5.3.

Corollary 5.5.8. *The asymptotic null distribution for $\rho_n\left(N_{CS}^\tau, J\right)$ conditional on $\mathcal{W} = \{w_1, \ldots, w_J\}$ for $\tau \in (0, 1]$ is given by $\mathcal{N}(\mu\left(N_{CS}^\tau, J\right), \nu\left(N_{CS}^\tau, J\right)/n)$ provided that $\nu\left(N_{CS}^\tau, J\right) > 0$ with*

$$\begin{aligned}
\mu\left(N_{CS}^\tau, J\right) &:= \mu\left(N_{CS}^\tau\right) \sum_{j=1}^{J} w_j^2 \quad \textit{and} \\
\nu\left(N_{CS}^\tau, J\right) &:= \nu\left(N_{CS}^\tau\right) \sum_{j=1}^{J} w_j^3 + 4\,\mu\left(N_{CS}^\tau\right)^2 \left[\sum_{j=1}^{J} w_j^3 - \left(\sum_{j=1}^{J} w_j^2\right)^2\right], \qquad (5.5.6)
\end{aligned}$$

where $\mu\left(N_{CS}^{\tau}\right)$ *and* $\nu\left(N_{CS}^{\tau}\right)$ *are given by equation (5.5.2).*

Proof: Similar to the proof of Corollary 5.4.9. ■

By an appropriate application of Jensen's inequality, we see that $\sum_{j=1}^{J} w_j^3 \geq \left(\sum_{j=1}^{J} w_j^2\right)^2$. Therefore, the covariance $\nu\left(N_{CS}^{\tau}, J\right) = 0$ iff both $\nu\left(N_{CS}^{\tau}\right) = 0$ and $\sum_{j=1}^{J} w_j^3 = \left(\sum_{j=1}^{J} w_j^2\right)^2$ hold, so asymptotic normality may hold even when $\nu\left(N_{CS}^{\tau}\right) = 0$ (provided that $\mu\left(N_{CS}^{\tau}\right) > 0$).

Similarly, for the segregation (association) alternatives with $4\,\varepsilon^2/3 \times 100\%$ of the triangles around the vertices of each triangle is forbidden (allowed), we obtain the above asymptotic distribution of $\rho_n\left(N_{CS}^{\tau}, J\right)$ with $\mu\left(N_{CS}^{\tau}, J\right)$ being replaced by $\mu_J\left(N_{CS}^{\tau}, \varepsilon\right)$, $\nu\left(N_{CS}^{\tau}, J\right)$ by $\nu(\tau, J, \varepsilon)$, $\mu\left(N_{CS}^{\tau}\right)$ by $\mu\left(N_{CS}^{\tau}, \varepsilon\right)$, and $\nu\left(N_{CS}^{\tau}\right)$ by $\nu\left(N_{CS}^{\tau}, \varepsilon\right)$.

Thus in the case of $J > 1$, we have a (conditional) test of $H_0 : X_i \overset{iid}{\sim} \mathcal{U}(C_H(\mathcal{Y}))$ which once again rejects against segregation for large values of $\rho_n\left(N_{CS}^{\tau}, J\right)$ and rejects against association for small values of $\rho_n\left(N_{CS}^{\tau}, J\right)$.

The segregation (with $\delta = 1/16$, i.e., $\varepsilon = \sqrt{3}/8$), null, and association (with $\delta = 1/4$, i.e., $\varepsilon = \sqrt{3}/12$) realizations (from left to right) are depicted in Figure 5.4.28 with $n = 100$ and in Figure 4.6.1 with $n = 1000$. With $n = 100$, for the null realization, the p-value is zero for all τ values relative to the segregation alternative, so p-value is one relative to the association alternative. For the segregation realization, we obtain $p \approx 0.0$ for all τ values. For the association realization, we obtain $p \approx 1.0$ for all τ values. Note that this is only for one realization of $\mathcal{X}_n$ with $n = 100$.

With $n = 1000$, for the null realization, the $p \geq .34$ for all τ values relative to the segregation alternative, so $p \geq .32$ for all τ values relative to the association alternative. For the segregation realization, we obtain $p \leq .021$ for all $\tau \geq .2$. For the association realization, we obtain $p \leq .02$ for all $\tau \geq .2$ and $p = .07$ at $\tau = .1$. Note that this is only for one realization of $\mathcal{X}_n$ with $n = 1000$.

We repeat the null and alternative realizations $n = 1000$ times. The estimated significance levels and the empirical power values are presented in Table 5.5.7. In the null case, we find the estimated significance level $\widehat{\alpha}_{10}(\tau, J)$ to be less than .10 for all $\tau \geq .6$ with respect to both alternatives, with smallest being .09 for segregation and .07 for association. This indicates that $n = 100$ is not large enough for normal approximation. In the segregation alternative with $\delta = 1/16$ we find the empirical power $\widehat{\beta}_{10}^{S}(\tau, J)$ to be greater than .6 for all $\tau \geq .7$ with highest being .613 at $\tau = .8$. In the association alternative with $\delta = 1/4$ we find the empirical power

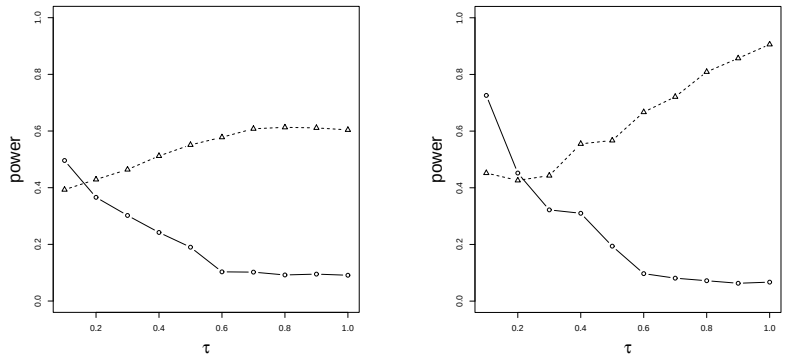


Figure 5.5.24: Monte Carlo power using the asymptotic critical value against $H^S_{\sqrt{3}/8}$ (left), $H^A_{\sqrt{3}/12}$ (right) as a function of τ for $n = 100$ conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. The circles represent the empirical significance levels while triangles represent the empirical power values.

$\widehat{\beta}^A_{10}(\tau, J)$ to be greater than .7 for all $\tau \geq .7$ with highest being .91 at $\tau = 1$. This analysis suggests that, conditional on these $\mathcal{Y}$ values, $\tau = .8$ or 1, (higher values of τ) should be used for small n. See Figure 5.5.24 for the plots of the power estimates with $n = 100$.

With $n = 500$, the corresponding power estimates are plotted in Figure 5.5.25. Note that, relative to segregation the empirical significance levels are about .05 for $\tau \geq .3$ and relative to association about .07 for $\tau \geq .5$ (larger for other τ values). This indicates that for moderate n, approximate normality yields an approximate level $\alpha = .05$ test against segregation, but the test is liberal against association for $n \in \{100, 500\}$. Based on the analysis of the empirical power curves, we suggest moderate or large τ values for both alternatives.

Remark 5.4.10 applies for $\rho_n\big(N^\tau_{CS}\big)$ also.

Remark 5.5.9. We can derive related test statistics in multiple triangle case similar to the ones in Section 5.4.11.1, by replacing r with τ.

5.5.8.1 Asymptotic Efficacy Analysis for $J > 1$

The PAE, HLAE, and asymptotic power function analysis are given for $J = 1$. For $J > 1$, the analyses will depend both the number of triangles as well as the size of the triangles. So the optimal τ values with respect to these efficacy criteria for $J = 1$ are not necessarily optimal for $J > 1$, so the analyses need to be updated, conditional on the values of J and $\mathcal{W}$.

Under segregation alternative H^S_ε, the PAE of $\rho_n\big(N^\tau_{CS}\big)$ is as in Equation 5.4.8 with r be-

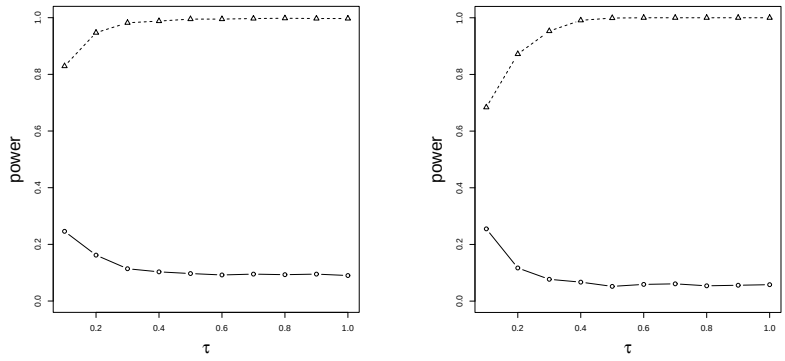


Figure 5.5.25: Monte Carlo power using the asymptotic critical value against $H^S_{\sqrt{3}/8}$ (left), $H^A_{\sqrt{3}/12}$ (right) as a function of τ for $n = 500$ conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. The circles represent the empirical significance levels while triangles represent the empirical power values.

τ	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$n = 100$, $N = 1000$										
$\widehat{\alpha}_S(\tau, n, J)$	.496	.366	.302	.242	.190	.103	.102	.092	.095	.091
$\widehat{\beta}^S_n(\tau, \sqrt{3}/8, J)$	.393	.429	.464	.512	.551	.578	.608	.613	.611	.604
$\widehat{\alpha}_A(\tau, n, J)$	.726	.452	.322	.310	.194	.097	.081	.072	.063	.067
$\widehat{\beta}^A_n(\tau, \sqrt{3}/12, J)$	.452	.426	.443	.555	.567	.667	.721	.809	.857	.906
$n = 500$, $N = 1000$										
$\widehat{\alpha}_S(\tau, n, J)$	.246	.162	.114	.103	.097	.092	.095	.093	.095	.090
$\widehat{\beta}^S_n(\tau, \sqrt{3}/8, J)$	.829	.947	.982	.988	.995	.995	.997	.998	.997	.997
$\widehat{\alpha}_A(\tau, n, J)$	.255	.117	.077	.067	.052	.059	.061	.054	.056	.058
$\widehat{\beta}^A_n(\tau, \sqrt{3}/12, J)$	.684	.872	.953	.991	.999	1.000	1.000	1.000	1.000	1.000

Table 5.5.7: The empirical significance level and empirical power values under $H^S_{\sqrt{3}/8}$ and $H^A_{\sqrt{3}/12}$, $N = 1000$, $n = 100$, and $J = 13$, at $\alpha = .05$ for the realization of $\mathcal{Y}$ in Figure 4.6.1.

ing replaced by τ. Under association alternative H^A_ε the PAE of $\rho_n(N^\tau_{CS})$ is similar. In Figure 5.5.26, we present the PAE as a function of τ for both segregation and association conditional on the realization of $\mathcal{Y}$ in Figure 5.4.28. Notice that PAE curves for $J > 1$ are similar to the ones for $J = 1$ case. See Figure 5.5.23. Some values of interest are $\lim_{\tau\to 0} \mathrm{PAE}^S_J(N^\tau_{CS}) \approx 38.1954$, $\mathrm{argsup}_{\tau\in(0,1]} \mathrm{PAE}^S_J(N^\tau_{CS}) = 1$ with supremum ≈ 100.7740. As for association, $\lim_{\tau\to 0} \mathrm{PAE}^A_J(N^\tau_{CS}) \approx 8593.9734$, $\mathrm{PAE}^A_J(N^{\tau=1}_{CS}) \approx 6449.5356$, $\mathrm{arginf}_{\tau\in(0,1]} \mathrm{PAE}^A_J(N^\tau_{CS}) \approx .4948$ with infimum $\approx .4948) \approx 5024.2236$. Based on the PAE analysis, we suggest, for large n and small ε, choosing large τ for testing against segregation and small τ against association. However, for moderate and small n, we suggest large τ values for association due

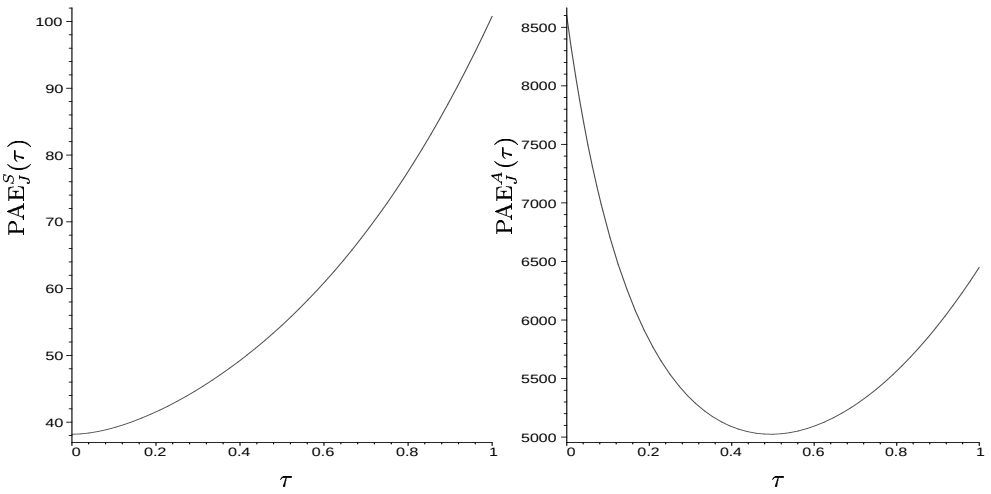


Figure 5.5.26: Pitman asymptotic efficacy against segregation (left) and association (right) as a function of τ with $J = 13$. Notice that vertical axes are differently scaled.

to the skewness of the density of $\rho_n\left(N_{CS}^{\tau}\right)$.

Under segregation, the HLAE is as in Equation 5.4.9 with r being replaced by τ. Notice that $\mathrm{HLAE}_J^S\left(N_{CS}^{\tau}, \varepsilon = 0\right) = 0$. The functional form of $\mathrm{HLAE}_J^A\left(N_{CS}^{\tau}, \varepsilon\right)$ is similar which implies $\mathrm{HLAE}_J^A\left(N_{CS}^{\tau}, \varepsilon = 0\right) = 0$.

CHAPTER 6

Conclusion

In this thesis, we analyze the PCDs for data from one class $\mathcal{X}_n$ in the triangle, $T(\mathcal{Y})$ (based on $\mathcal{Y}$ with $|\mathcal{Y}| = 3$) and then extended the results to multiple triangle case. Our PCDs are defined based on the relative positions of the class $\mathcal{X}_n$ points with respect to class $\mathcal{Y}$ points. The distribution of $\mathcal{X}_n$ is conditional on some quantity related to $\mathcal{Y}$, e.g., the relative sizes or the number of the Delaunay triangles based on $\mathcal{Y}$.

The first proximity map related to our proximity maps that appeared in the literature is the *spherical proximity map* $N_S(x) := B(x, r(x))$ (which is called CCCD in the literature, see [35], [10], [29], [36], and [37]) where $r(x) = d(x, \mathcal{Y})$.

We define the *arc-slice proximity map* $N_{AS}(\cdot, M)$, which is a slight variation of $N_S(\cdot)$, the r-factor proportional-edge proximity map $N_{PE}^r(\cdot, M)$, and the τ-factor central similarity proximity map $N_{CS}^\tau(\cdot, M)$.

In this section, first, we compare the proximity maps with respect to the appealing properties in Section 3.1. The CCCDs based on $N_S(\cdot)$ for data in compact intervals in $\mathbb{R}$ satisfies all of the appealing properties. Among the appealing properties in Section 3.1, for data in triangles, $N_S(\cdot)$ satisfies properties **P1**, **P2**, **P3**, and **P8**; $N_{AS}(\cdot, M_{CC})$ satisfies properties **P1**, **P2**, **P6**, **P7**, and **P8**; $N_{PE}^r(\cdot, M)$ satisfies the properties **P1**, **P2**, **P4**, **P5**, and **P7**. Moreover, **P6** holds for $N_{PE}^r(\cdot, M)$, if $M \in \mathscr{T}^r$, and **P8** holds for $M = M_C$ and $r = 3/2$; $N_{CS}^\tau(\cdot, M)$ satisfies all of the properties **P1**—**P8** for all $M \in T(\mathcal{Y})^o$.

The proximity map $N_{PE}^r(\cdot, M)$, when compared to the others, has the advantage that the asymptotic distribution of the domination number $\gamma_n\left(N_{PE}^r\right)$ is available (see Section 4.4 for the asymptotic distribution) and the exact distribution of $\gamma_n\left(N_{PE}^r\right)$ can be obtained by numerical methods (which is open for future research). Although we have a.s. upper and lower bounds, the exact and asymptotic distribution of the domination number of the proximity catch digraphs based on $N_S(\cdot)$ and $N_{AS}(\cdot, M_{CC})$ are not tractable and that of $N_{CS}^\tau(\cdot, M)$ is an open problem

for $\tau \in (0,1]$. For $N \in \{N_{PE}^{r}(\cdot, M), N_{CS}^{\tau}(\cdot, M)\}$, the asymptotic distribution of $\gamma(\mathcal{X}_n, N)$ does not require conditioning on $\mathcal{Y}$ but the number of Delaunay triangles based on $\mathcal{Y}$ should be large, whereas the exact distribution of $\gamma(\mathcal{X}_n, N)$ requires conditioning on $\mathcal{Y}$ (in fact, conditioning on the relative sizes of Delaunay triangles based on $\mathcal{Y}$). Furthermore, $N_{CS}^{\tau}(\cdot, M)$ and $N_{PE}^{r}(\cdot, M)$ enjoy the geometry invariance property for uniform data on triangles. While finding an exact minimum dominating set is an NP-Hard problem for $N_S(\cdot)$, $N_{AS}(\cdot, M)$, and $N_{CS}^{\tau}(\cdot, M)$, a particular exact minimum dominating set can be found in polynomial time for $N_{PE}^{r}(\cdot, M)$. The mean and variance of the relative density $\rho_n(N)$ is not analytically tractable for PCDs based on $N \in \{N_S(\cdot), N_{AS}(\cdot, M)\}$, but —in principle— can be obtained by numerical methods, conditional on the inner angles of triangle (or equivalently on (c_1, c_2)). The asymptotic distribution of $\rho_n(N)$ for PCDs based on $N \in \{N_{PE}^{r}(\cdot, M), N_{CS}^{\tau}(\cdot, M)\}$ is available for both one-triangle and multiple-triangle cases. However, removing the conditioning on the set $\mathcal{Y}$ (more explicitly conditioning on the relative sizes of the Delaunay triangles) is an open problem. Additionally, $N_{AS}(x, M)$, $N_{PE}^{r}(x, M)$, and $N_{CS}^{\tau}(x, M)$ are well defined only for $x \in C_H(\mathcal{Y})$, the convex hull of $\mathcal{Y}$, while $N_S(x)$ is well defined for all $x \in \mathbb{R}^d$. The extension of proximity maps $N_S(\cdot)$ and $N_{AS}(\cdot, M)$ to higher dimensions is straightforward, which is not the case for $N_{PE}^{r}(x, M)$ and $N_{CS}^{\tau}(x, M)$.

We develop a technique to test the spatial patterns of segregation or association against some form of complete spatial randomness. There are many tests available for segregation and association in the literature. See [12] for a survey on these tests and relevant references. Two of the most commonly used tests are Pielou's χ^2 test of independence and Ripley's test based on $K(t)$ and $L(t)$ functions. The null hypotheses we consider ($H_0 : X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}))$ or $H_0 : X_i \overset{iid}{\sim} \mathcal{C}_H(\mathsf{y})$) are considerably more restrictive than current approaches, which can be used much more generally. The null hypothesis for testing segregation or association can be described in two slightly different forms (see [12]):

(i) complete spatial randomness, that is, each class is distributed randomly throughout the area of interest. It describes both the arrangement of the locations and the association between classes.

(ii) random labeling of locations, which is less restrictive than spatial randomness, in the sense that arrangement of the locations can either be random or non-random.

Our null hypotheses are closer to the latter in this regard. Pielou's test provides insight only on the nearest neighbor type relationships between classes, hence there is no assumption on the allocation of the observations, which makes it more appropriate for testing the null hypothesis of random labeling. Ripley's tests can be used for both types of null hypotheses, in particular, it can be used to test a type of (random) spatial pattern against another type of (random) spatial pattern.

We use the mean domination number of the components of the PCD, $\overline{G}_J\left(N_{PE}^r\right)$ and the relative densities, $\rho_n\left(N_{PE}^r\right)$ and $\rho_n\left(N_{CS}^r\right)$ as our test statistics in testing the spatial patterns of segregation and association.

The test based on the mean domination number, $\overline{G}_J\left(N_{PE}^r\right)$, of the components of the PCD in Section 4.6 is not a conditional test, but requires both n and number of Delaunay triangles J to be large (with n being much larger than J). The comparison of this test with the existing tests in the literature for large values of J is possible. Furthermore, under the segregation alternatives we consider, PAE is not applicable to the mean domination number (of components) case, however, for large n and J, we suggest the use of it over relative density since for each $\varepsilon > 0$, Hodges-Lehmann asymptotic efficacy is arbitrarily large for the mean domination number case; while HLAE is bounded for relative density case with $J > 1$. For the association alternative, HLAE suggests moderate r values for which $\rho_n\left(N_{PE}^r\right)$ has finite HLAE. So again, for large J and n mean domination number is preferable. The basic advantage of $\rho_n\left(N_{PE}^r\right)$ and $\rho_n\left(N_{CS}^r\right)$ is that they do not require J to be large, so for small J, they are preferable.

The tests based on the relative densities $\rho_n\left(N_{PE}^r\right)$ and $\rho_n\left(N_{CS}^r\right)$ are conditional tests — conditional on $\mathcal{W}$ (the set of relative areas of the Delaunay triangles) and we require the number of triangles J is fixed and relatively small compared to $n = |\mathcal{X}_n|$. Furthermore, our method deals with a slightly different type of data than most methods to examine spatial patterns. The sample size for one type of point (type $\mathcal{X}_n$ points) is much larger compared to the the other (type $\mathcal{Y}$ points).

There are two major types of asymptotic structures for spatial data (see, e.g., [24]). In the first, any two observations are required to be at least a fixed distance apart, hence as the number of observations increase, the region on which the process is observed eventually becomes unbounded. This type of sampling structure is called "increasing domain asymptotics". In the second type, the region of interest is a fixed bounded region and more or more points are

observed in this region. Hence the minimum distance between data points tends to zero as the sample size tends to infinity. This type of structure is called "infill asymptotics" due to Cressie [7]. The sampling structure for our asymptotic analysis for the mean domination number case can be viewed as increasing domain asymptotics or infill asymptotics for $\mathcal{Y}$ since we can assume $\mathcal{Y}$ is from a distribution on a compact region or from a process for which minimum distance between $\mathcal{Y}$ points is fixed and infill asymptotics for $\mathcal{X}_n$ since in the convex hull of $\mathcal{Y}$, the size of $\mathcal{X}_n$ tends to infinity (faster than the size of $\mathcal{Y}$), while for the relative density case, the sampling structure is infill since only the size of $\mathcal{X}_n$ tends to infinity, while the support, the convex hull of $\mathcal{Y}$, $C_H(\mathcal{Y})$ is a fixed bounded region.

Moreover, our statistic can be written as a U-statistic based on the locations of $\mathcal{X}_n$ points with respect to $\mathcal{Y}$ points. This is one advantage of the proposed method; most statistics for spatial patterns can not be written as U-statistics. The U-statistic form avails us the asymptotic normality, once the mean and variance is obtained by tedious detailed geometric calculations.

The expressions of the mean and the variance of $\rho_n\big(N_{CS}^{\tau}\big)$ have a simpler form than those of $\rho_n\big(N_{PE}^{r}\big)$. The means of relative densities have the relation $\mu\big(N_{PE}^{r}\big) > \mu\big(N_{CS}^{\tau}\big)$ for all $r \geq 1$ and $\tau \in [0,1]$; the asymptotic variances $\nu\big(N_{PE}^{r}\big)$ and $\nu\big(N_{CS}^{\tau}\big)$ are both less than $7/135$ for all $\tau \in [0,1]$ and for extreme r values ($r \lessapprox 1.443$ and $r \gtrapprox 4.215$). Note that when $J = 1$ (one triangle case), $\rho_n\big(N_{CS}^{\tau}\big)$ is degenerate at $\tau = 0$, while $\rho_n\big(N_{PE}^{r}\big)$ is degenerate at $r = \infty$.

The comparison of the empirical significance levels (see Sections 5.4.6, 5.4.7, 5.5.5, and 5.5.6) in the case of $J = 1$ (one triangle case) indicates that $\rho_n\big(N_{PE}^{r}\big)$ is appropriate for moderate r values ($2 \leq r \leq 3$ against segregation and $\sqrt{2} \leq r \leq 5$ against association) while $\rho_n\big(N_{CS}^{\tau}\big)$ is liberal for all $\tau \in (0,1]$ for moderate and small n. Moreover, a larger n is required to attain normal approximation for $\rho_n\big(N_{CS}^{\tau}\big)$ compared to $\rho_n\big(N_{PE}^{r}\big)$ with moderate r values. Additionally, for small n the kernel density estimate of $\rho_n\big(N_{CS}^{\tau}\big)$ is skewed right, while $\rho_n\big(N_{PE}^{r}\big)$ is skewed right for values of r close to 1 and skewed left for values of r larger than 5.

For $J = 1$, PAE compares as $\mathrm{PAE}^{S}\big(N_{PE}^{r}\big) \geq \sup_{\tau\in[0,1]} \mathrm{PAE}^{S}\big(N_{CS}^{\tau}\big)$ for $r \gtrapprox 1.1316$ and $\mathrm{PAE}^{A}\big(N_{PE}^{r}\big) \leq \inf_{\tau\in[0,1]} \mathrm{PAE}^{A}\big(N_{CS}^{\tau}\big)$ for $r \gtrapprox 1.057$. Hence based on the PAE analyses in the one-triangle case, for large n and small ε, we suggest the use of $\rho_n\big(N_{PE}^{r}\big)$ with moderate r values ($r \in [3/2, 3]$) against segregation and the use of $\rho_n\big(N_{CS}^{\tau}\big)$ with large τ values ($\tau \gtrapprox 1$) against association.

For the realization of $\mathcal{Y}$ given in Figure 4.6.1, against segregation, $\rho_n\big(N_{PE}^{r}\big)$ and $\rho_n\big(N_{CS}^{\tau}\big)$

are both liberal with $\rho_n\left(N_{PE}^r\right)$ being less liberal in rejecting H_0 for small and moderate n. Against association, $\rho_n\left(N_{CS}^\tau\right)$ is appropriate for large τ values, while $\rho_n\left(N_{PE}^r\right)$ is again liberal in rejecting H_0. In terms of PAE analysis, $\mathrm{PAE}_J^S\left(N_{PE}^r\right) \geq \sup_{\tau\in[0,1]} \mathrm{PAE}_J^S\left(N_{CS}^\tau\right)$ for $r \gtrapprox 1.6387$ and $\mathrm{PAE}_J^A\left(N_{PE}^r\right) \leq \inf_{\tau\in[0,1]} \mathrm{PAE}_J^A\left(N_{CS}^\tau\right)$ for all $r \geq 1$. Hence in the multiple-triangle case, for large n and small ε, we suggest the use of $\rho_n\left(N_{PE}^r\right)$ with moderate r values ($r \in [2,3]$) against segregation and the use of $\rho_n\left(N_{CS}^\tau\right)$ with large τ values ($\tau \lessapprox 1$) against association.

In addition to the open problems mentioned so far, we suggest the investigation of the underlying graphs of the PCDs we have defined for further research. For any given digraph, there are two associated underlying graphs; namely AND and OR underlying graphs. For digraph $D = (\mathcal{V}, \mathcal{A})$, the AND underlying graph is constructed with the same vertex set $\mathcal{V}$ and the edge set $\mathcal{E}_{\mathrm{AND}}$ defined by $(u, v) \in \mathcal{E}_{\mathrm{AND}}$ iff both uv and vu are arcs in $\mathcal{A}$. Similarly, the OR underlying graph is constructed with the same vertex set $\mathcal{V}$ and the edge set $\mathcal{E}_{\mathrm{OR}}$ defined by $(u, v) \in \mathcal{E}_{\mathrm{OR}}$ iff either uv or vu are arcs in $\mathcal{A}$. One can construct these underlying graphs based on the PCDs and investigate their properties (such as domination number, relative density, etc.) as well as possible applications.

Part III

Appendices and Bibliography

Appendix A

Proofs of Some of the Theorems in Chapters 3 and 4

A.1 Proof of Theorem 3.4.30

First, we begin the proof with a remark that introduces some terminology which we will use for asymptotics throughout this dissertation.

Remark A.1.1. Suppose $\mathcal{X}_n$ is a random sample from F with support $\mathcal{S}(F) \subseteq \Omega$. If over a sequence $\Omega_n \subseteq \Omega,\ n = 1, 2, 3, \ldots$, X restricted to Ω_n, $X|_{\Omega_n}$, has distribution F_n with $F_n(x) = F(x)/P_F(X \in \Omega_n)$ and $P_F(X \in \Omega_n) \to 1$ as $n \to \infty$, then we call F_n the *asymptotically accurate distribution* of X and Ω_n the *asymptotically accurate support* of F . If F has density f, $f_n = f(x)/P_F(X \in \Omega_n)$ is called the *asymptotically accurate pdf* of X. In both cases, if we are concerned with asymptotic results, for simplicity we will use F and f for asymptotically accurate distribution and pdf, respectively. Conditioning will be alluded to by stating that $X \in \Omega_n$ with probability 1, as $n \to \infty$ or for sufficiently large n. □

A.1.1 Proof of Theorem 3.4.30

For $r = 3/2$ and $M = M_C$, and sufficiently large n, $\Gamma_1\left(X_{e_j}(n), N_{PE}^r, M_C\right) \cap R_{CM}(\mathsf{y}_j)$ is a triangle for each $j \in \{1, 2, 3\}$ w.p. 1. See Figure A.1.1 (left). Let $X_{e_j}(n) \in \operatorname{argmin}_{X \in \mathcal{X}_n} d(X, e_j)$ be the closest edge extrema for edge e_j. Note that for $n \geq 3$, $X_{e_j}(n)$ uniquely exist for each edge e_j w.p. 1. Suppose in a realization of $\mathcal{X}_n$, the edge extrema are denoted as $x_{e_j}(n) = (x_j, y_j)$ and are close enough to e_j for $j \in \{1, 2, 3\}$ that $\Gamma_1\left(X_{e_j}(n), N_{PE}^r, M_C\right) \cap R_{CM}(\mathsf{y}_j)$ is a triangle. Then $\Gamma_1\left(x_{e_1}, N_{PE}^{3/2}, M_C\right) \cap R_{CM}(\mathsf{y}_1)$ is the triangle with vertices

$$\left(\frac{\sqrt{3}}{3} y_2 + x_2 - \frac{1}{2}, -\frac{1}{6}\left(-3\sqrt{3} + 2\, y_2 + 2\sqrt{3}\, x_2\right)\right), \left(\frac{1}{2}, \frac{1}{3}\left(-\frac{3\sqrt{3}}{2} + 2\, y_2 + 2\sqrt{3}\, x_2\right)\right), \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right);$$

$\Gamma_1\left(x_{e_2}, N_{PE}^{3/2}, M_C\right) \cap R_{CM}(\mathsf{y}_2)$ is the triangle with vertices

$$\left(\frac{1}{2}, \frac{1}{6}\left(\sqrt{3}+4\,y_3-4\sqrt{3}\,x_3\right)\right), \left(\frac{1}{2}-\frac{\sqrt{3}}{3}\,y_3+x_3, -\frac{1}{6}\left(-\sqrt{3}+2\,y_3-2\sqrt{3}\,x_3\right)\right), \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right);$$

and $\Gamma_1\left(x_{e_3}, N_{PE}^{3/2}, M_C\right) \cap R_{CM}(\mathsf{y}_3)$ is the triangle with vertices

$$\left(-\frac{\sqrt{3}}{6}\left(-\sqrt{3}+4\,y_1\right), \frac{\sqrt{3}}{6}+\frac{2}{3}\,y_1\right), \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right), \left(\frac{\sqrt{3}}{6}\left(\sqrt{3}+4\,y_1\right), \frac{\sqrt{3}}{6}+\frac{2}{3}\,y_1\right).$$

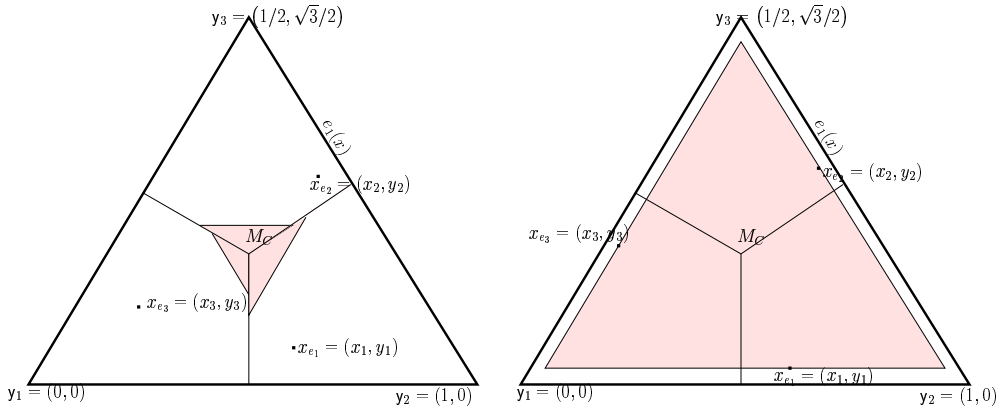


Figure A.1.1: The shaded regions are the triangular $\Gamma_1\left(x_{e_j}, N_{PE}^{3/2}, M_C\right) \cap R_{CM}(\mathsf{y}_j)$ regions for $j \in \{1,2,3\}$ (left); the figure for the description of the pdf of $X_{e_j}(n)$, the shaded region is $T\left(\vec{\zeta}\right)$ (right) given $X_{e_j}(n) = x_{e_j} = (x_j, y_j)$ for $j \in \{1,2,3\}$.

For a realization of $\mathcal{X}_n$ with $X_{e_j}(n) = x_{e_j} = (x_j, y_j)$ close enough to e_j for $j \in \{1,2,3\}$

$$\begin{aligned} A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right) &= \frac{\sqrt{3}}{9}\left(\sqrt{3}\,x_2-\sqrt{3}+y_2\right)^2 + \frac{\sqrt{3}}{9}\left(-\sqrt{3}\,x_3+y_3\right)^2 + \frac{4\sqrt{3}}{9}y_1^2 \\ &= \frac{\sqrt{3}}{9}\left(3\,x_2^2-6\,x_2+2\sqrt{3}\,y_2\,x_2-2\sqrt{3}\,y_2+y_2^2+3+y_3^2-2\sqrt{3}\,y_3\,x_3+3\,x_3^2+4\,y_1^2\right). \end{aligned}$$

To find the expected area, we need the joint pdf of the $X_{e_j}(n)$. The edge extrema are all distinct with probability 1 as $n \to \infty$ (see Theorem 3.4.12). Let $T\left(\vec{\zeta}\right)$ be the triangle formed by the lines at x_{e_j} parallel to e_j for $j \in \{1,2,3\}$ where $\vec{\zeta} = (x_1, y_1, x_2, y_2, x_3, y_3)$. See Figure A.1.1 (right).

Then the asymptotically accurate joint pdf of $X_{e_j}(n)$ is

$$\begin{aligned} f_3(\vec{\zeta}) &= \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3}\left(\frac{A(T(\vec{\zeta}))}{A(T(\mathcal{Y}))}\right)^{n-3} \\ &= \frac{n\,(n-1)\,(n-2)}{(\sqrt{3}/4)^n}\left(\sqrt{3}/36\left(-2\sqrt{3}\,y_1+\sqrt{3}\,y_3-3\,x_3+\sqrt{3}\,y_2+3\,x_2\right)^2\right)^{n-3}. \end{aligned}$$

where the support D_S of $f_3(\vec{\zeta})$ is a region in $T(\mathcal{Y})$ so that (x_j, y_j) are distinct. Here the joint pdf is asymptotically accurate in the sense that $P(X_{e_j}(n) \in D_S) \to 1$ as $n \to \infty$, so we can just use this function as a pdf for $X_{e_j}(n)$ in asymptotics.

Then for sufficiently large n,

$$\begin{aligned} \mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right)\right] &\approx \int_{D_S} A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right) f_3(\vec{\zeta})\,d\vec{\zeta} \\ &= \int_{D_S} A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right) \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3}\left(\frac{A(T(\vec{\zeta}))}{A(T(\mathcal{Y}))}\right)^{n-3} d\vec{\zeta}, \end{aligned}$$

where $d\vec{\zeta} = dx_1\,dy_1\,dx_2\,dy_2\,dx_3\,dy_3$. Let

$$G(\vec{\zeta}) = A(T(\vec{\zeta}))/A(T(\mathcal{Y})).$$

Notice that the integrand is critical when $x_{e_j} \in e_j$, since $G(\vec{\zeta}) = 1$ when $x_{e_j} \in e_j$ for $j \in \{1,2,3\}$. So we make the change of variables $y_1 = z_1$, $y_2 = \sqrt{3}\,(1-x_2) - z_2$, and $y_3 = \sqrt{3}\,x_3 - y_3$, then $G(\vec{\zeta})$ and $A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right)$ become

$$\left(2\,z_1 + z_3 - \sqrt{3} + z_2\right)^2/3 \quad \text{and} \quad \sqrt{3}\,\left(z_2^2 + z_3^2 + 4\,z_1^2\right)/9,$$

respectively. Hence the integrand does not depend on x_1, x_2, x_3 and integrating with respect to x_1, x_2, x_3 yields a constant K. Now, the integrand is critical at $(z_1, z_2, z_3) = (0,0,0)$, since $G(0,0,0) = 1$. So let E_u^ε be the event that $0 \le z_j \le \varepsilon$ for $j \in \{1,2,3\}$ for sufficiently small $\varepsilon > 0$. Then making the change of variables $z_j = w_j/n$ for $j \in \{1,2,3\}$, we get $A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right) = O\left(n^{-2}\right)$ and $G(z_1, z_2, z_3)$ becomes $G(w_1, w_2, w_3) = 1 -$

$\frac{1}{n}\left(2/\sqrt{3}\,(2\,w_1+w_2+w_3)\right)+O\left(n^{-2}\right)$. Hence, for sufficiently large n

$$\begin{aligned}
&\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right)\right] \approx \\
&K\int_0^{n\varepsilon}\int_0^{n\varepsilon}\int_0^{n\varepsilon} A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right)\frac{n\,(n-1)\,(n-2)}{n^3}G(w_1,w_2,w_3)^{n-3}dw_1dw_2dw_3, \\
&\approx K\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} O\left(n^{-2}\right)\exp\left(-2/\sqrt{3}\,(2\,w_1+w_2+w_3)\right)dw_1dw_2dw_3 = O\left(n^{-2}\right),
\end{aligned}$$

since $\int_0^{\infty}\int_0^{\infty}\int_0^{\infty}\exp\left(-2/\sqrt{3}\,(2\,w_1+w_2+w_3)\right)dw_1dw_2dw_3 = 3\sqrt{3}/16$ which is a finite constant. Hence $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right)\right] \to 0$ at rate $O\left(n^{-2}\right)$. ■

A.2 Proofs of Some of the Theorems in Chapter 4

A.2.1 Proof of Proposition 4.3.4

Recall that the circumcenter in T_b is $M_{CC} = \left(\frac{1}{2}, \frac{c_2^2 - c_1(1-c_1)}{2\,c_2}\right)$ and the circumradius is $r_{cc} = \frac{1}{2}\sqrt{1+\frac{(c_2^2-c_1(1-c_1))^2}{c_2^2}}$.

Suppose $T(\mathcal{Y})$ is an acute triangle. Then $c_2^2 - c_1(1-c_1) > 0$. Moreover, $d(M_{CC}, e_3) = \frac{c_2^2-c_1(1-c_1)}{2\,c_2}$, $d(M_{CC}, e_2) = \frac{(1-c_1)\sqrt{c_1^2+c_2^2}}{2\,c_2}$, and $d(M_{CC}, e_1) = \frac{\sqrt{(c_1-c_2)^2}\,c_1}{2\,c_2}$. By algebraic manipulations, we can see that $r_{cc} > d\left(M_{CC}, e_j\right)$ for each $j \in \{1,2,3\}$. Geometrically, this can be seen by using the appropriate right triangles. See Figure A.2.1. So let $\varepsilon > 0$ be small enough that for each $\mathsf{y} \in \mathcal{Y}$, $d(x, M_{CC}) < \varepsilon$ and $x \in R_{CC}(\mathsf{y})$ will imply $R_{CC}(\mathsf{y}) \subset N_{AS}\left(x, M_{CC}\right)$, hence $\mathcal{X}_n \cap R_{CC}(\mathsf{y}) \subset N_{AS}\left(x, M_{CC}\right)$. Such an ε exists since $r_{cc} > d\left(M_{CC}, e_j\right)$ for each $j \in \{1,2,3\}$. Then for each $\mathsf{y} \in \mathcal{Y}$, $P\left(\mathcal{X}_n \cap \left[R_{CC}(\mathsf{y}) \cap B(M_{CC}, \varepsilon)\right] \neq \emptyset\right) \to 1$ as $n \to \infty$. So at most three points are required to dominate $\mathcal{X}_n$ as $n \to \infty$.

Now suppose $T(\mathcal{Y})$ is a right triangle¿ Then in T_b the circumcenter is $M_{CC} = (1/2, 0)$ and the circumradius is $r_{cc} = 1/2$. Then

$$d(M_{CC}, e_3) = 0, \;\; d(M_{CC}, e_2) = \frac{c_2}{2\sqrt{c_1^2+c_2^2}}, \text{ and } d(M_{CC}, e_1) = \frac{c_2}{2\sqrt{(1-c_1)^2+c_2^2}}.$$

Notice that $d(M_{CC}, e_2) < r_{cc}$ and $d(M_{CC}, e_1) < r_{cc}$. Hence the desired result follows as in the acute triangle case.

Finally, suppose $T(\mathcal{Y})$ is an obtuse triangle. Then in T_b we have $c_2^2 - c_1(1-c_1) < 0$ and $M_{CC} \notin T(\mathcal{Y})$. Then $R_{CC}(\mathsf{y}_3)$ is the pentagon with vertices $\mathsf{y}_3, M_2, z_1, z_2, M_1$ where $z_1 = \left(\frac{c_1^2+c_2^2}{2\,c_1}, 0\right)$, $z_2 = \left(\frac{1-(c_1^2+c_2^2)}{2\,(1-c_1)}, 0\right)$; $R_{CC}(\mathsf{y}_1)$ is the triangle with vertices y_1, z_1, M_2; and $R_{CC}(\mathsf{y}_2)$ is the triangle

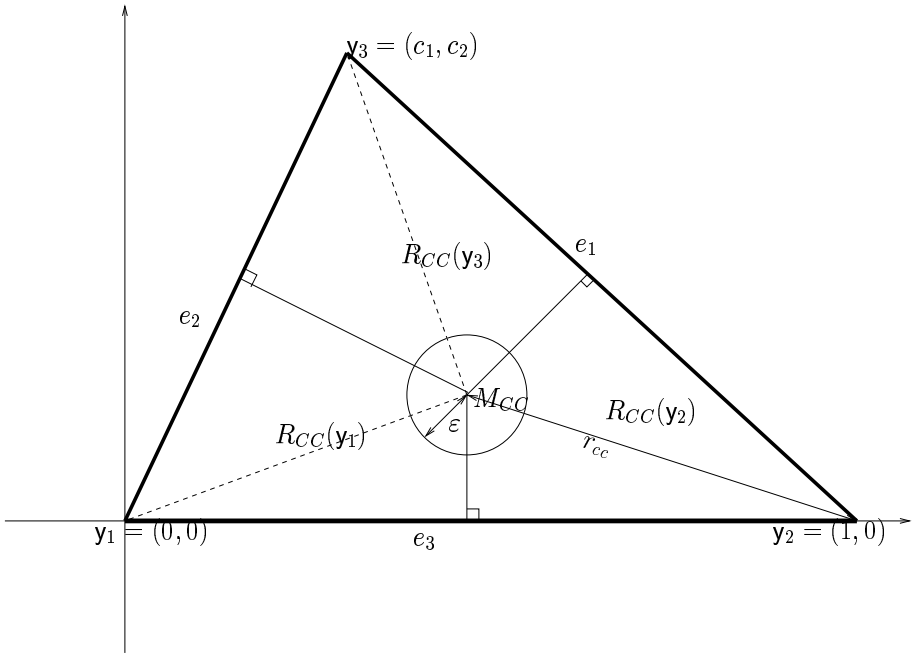


Figure A.2.1: A figure for the description of the ball $B(M_{CC}, \varepsilon)$.

with vertices y_2, M_1, z_2. See Figure A.2.2. Since, the latter two triangles are right triangles, it follows that $d(z_1, e_2) < d(\mathsf{y}_1, z_1)$ and $d(z_2, e_1) < d(z_1, \mathsf{y}_2)$. The explicit forms of these distances are $d(z_1, e_2) = \frac{c_2\sqrt{c_2^2+c_1^2}}{2\,c_1}$ and $d(z_2, e_1) = \frac{c_2\sqrt{1-2\,c_1+c_1^2+c_2^2}}{2\,(1-c_1)}$. So we need one point for each of $R_{CC}(\mathsf{y}_1)$ and $R_{CC}(\mathsf{y}_2)$ around z_1 and z_2 to cover $R_{CC}(\mathsf{y}_1)$ and $R_{CC}(\mathsf{y}_2)$, respectively. Note also that $2\,d(z_1, \mathsf{y}_1) > 1 - 2\,d(z_2, \mathsf{y}_2)$, since $c_2 > 0$. This implies that picking points sufficiently close to z_1, z_2 in $R_{CC}(\mathsf{y}_1)$ and $R_{CC}(\mathsf{y}_2)$, say x_1, x_2, $B(x_1, r(x_1))$ and $B(x_2, r(x_2))$ intersect as in Figure A.2.2. As for $R_{CC}(\mathsf{y}_3)$, in the basic triangle T_b $\operatorname{argsup}_{x \in R_{CC}(\mathsf{y}_3)} d(x, \mathsf{y}_3) = z_2$ and we see that $d(z_2, M_2) > d(z_2, \mathsf{y}_3)$. See Figure A.2.2. So it seems that two points are required for $R_{CC}(\mathsf{y}_3)$; one around z_1 and one around z_2. However, if the closest points in $R_{CC}(\mathsf{y}_1)$ and $R_{CC}(\mathsf{y}_2)$ are sufficiently close to z_1 and z_2 respectively, only a small region in $R_{CC}(\mathsf{y}_3)$ is not covered by $B(x_1, r(x_1)) \cup B(x_2, r(x_2))$ (see the solid ball and dashed ball in Figure A.2.2). To cover the remaining region, pick a point sufficiently close to $(c_1, 0)$, say x_3 as in Figure A.2.2. So three points will suffice to cover $T(\mathcal{Y})$. An ε argument as before will imply that this happens with probability 1 for three points in $\mathcal{X}_n$ as $n \to \infty$. Hence the desired result follows. ■.

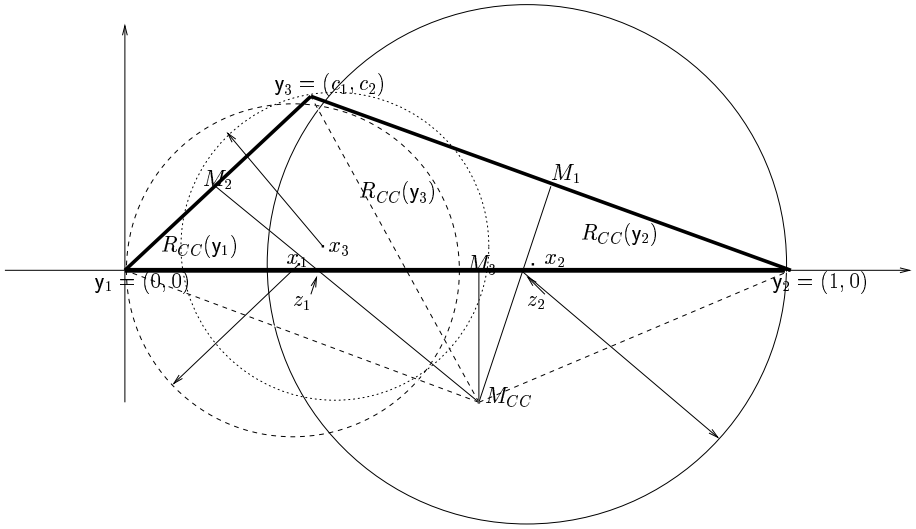


Figure A.2.2: The vertex regions $R_{CC}(\mathsf{y})$ in an obtuse triangle.

A.2.2 Proof of Theorem 4.3.5

Suppose $T(\mathcal{Y})$ is an acute triangle, then $M_{CC} \in T(\mathcal{Y})^o$, i.e., $c_2^2 - c_1(1-c_1) > 0$. Observe that, as $n \to \infty$, $\mathcal{X}_n \cap R_{CC}(\mathsf{y}_3) \subset N_{AS}\left(X_3^f(n), M_{CC}\right)$ with probability 1, where $X_j^f(n)$ —when exists— is one of the furthest points from y_j in $R_{CC}(\mathsf{y}_j)$; i.e., $X_j^f(n) \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_{CC}(\mathsf{y}_j)} d(X, \mathsf{y}_j)$ for $j \in \{1,2,3\}$ whenever $\mathcal{X}_n \cap R_{CC}(\mathsf{y}_j) \neq \emptyset$. Note that at least one of the $X_j^f(n)$ uniquele exists for some j w.p. 1 and as $n \to \infty$ all $X_j^f(n)$ are unique w.p. 1. See also Figure A.2.3. Then if we can show that as $n \to \infty$, the domination number of the induced digraph on $\mathcal{X}_n \cap \left[R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)\right]$ and $\{\mathsf{y}_1, \mathsf{y}_2\}$ is 1 with probability 1, the desired result will follow. Restricted to $R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)$, the set $\mathcal{X}_n \cap [R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)]$ is a random sample from $\mathcal{U}\big(R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)\big)$. Let $\gamma_{ind}\left(\mathcal{X}_n, \widetilde{N}_{AS}\right)$ be the domination number for this induced digraph where $\widetilde{N}_{AS}(x) := \overline{B}(x, r(x)) \cap \left[R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)\right]$. Consider the proximity region $N_1(z_1 = (x_1, y_1))$ defined for $z_1 \in R_{CC}(\mathsf{y}_1)$ as

$$N_1(z_1) := \begin{cases} \left\{(u,v) \in R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2) : \; u \leq \min(2\,x_1, s_1(z_1))\right\}, & \text{if } \overline{B}(z_1, r(z_1)) \cap e_1 \neq \emptyset, \\ \widetilde{N}_{AS}\left((x_1, y_1)\right)), & \text{otherwise,} \end{cases}$$

where $s_1(x_1, y_1) = \frac{x_1 + 3 - \sqrt{3}\,y_1 + \sqrt{\left(x_1 - \sqrt{3}\,y_1\right)^2 + 3\,(2\,x_1 - 1)}}{4}$ in the standard equilateral triangle T_e. See Figure A.2.4 (left). For $(x_1, y_1) \in R_{CC}(\mathsf{y}_2)$, $N_1((x_1, y_1))$ is defined similarly. Then for

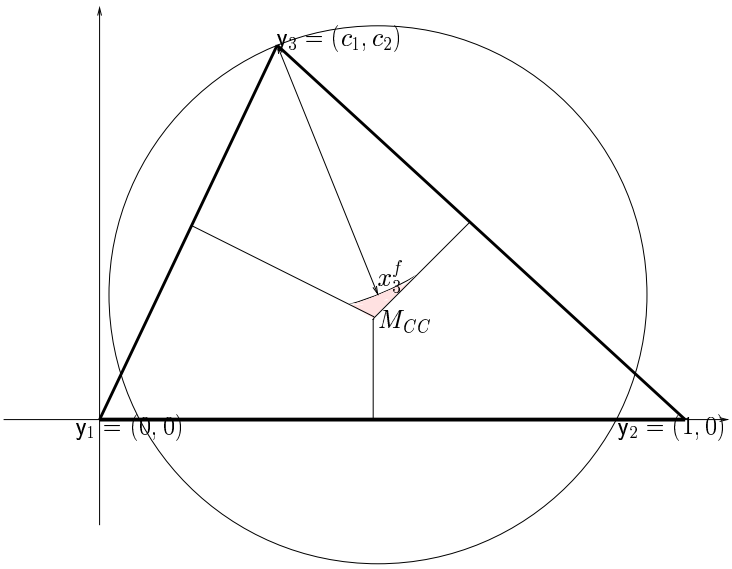


Figure A.2.3: A figure for the description of the pdf of $X_3^f(n)$ with a given $X_3^f(n) = x_3^f$.

sufficiently large n,

$$P(\gamma(\mathcal{X}_n, N_1) = 1) \leq P\left(\gamma_{ind}(\mathcal{X}_n, \widetilde{N}_{AS}) = 1\right),$$

since $N_1(x) \subsetneq \widetilde{N}_{AS}(x)$ for all x. Next, consider the transformation, $(x, y) \to x$, then the transformed variable X is distributed as F with density

$$f(x) = y/A(R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)), \text{ for } (x, y) \in \partial\left(R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)\right) \setminus e_3. \tag{A.2.1}$$

See also Figure A.2.4 (right) where $f(x)$ is multiplied by the area of $R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2)$. Let

$$N_2(x) := \left\{u \in (0, 1) : \ u \leq \min\left(2\, x_1, s_1(x_1, y_1)\right)\right\}.$$

Then $\gamma(\mathcal{X}_n, N_1) = 1$ iff $\gamma(\mathcal{X}_n, N_2) = 1$.

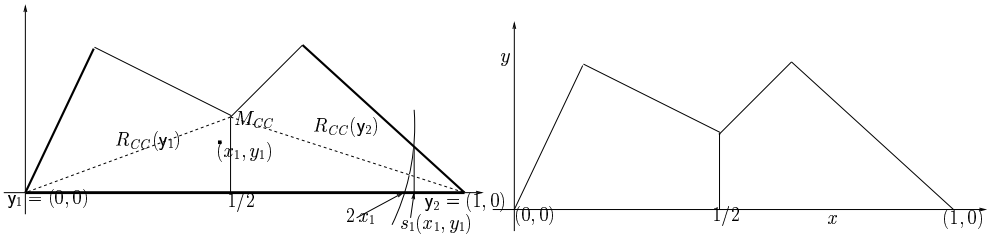


Figure A.2.4: The explanatory figure for $N_1((x_1, y_1))$ for $(x_1, y_1) \in R_{CC}(\mathsf{y}_1)$ (left); the graph of $f(x)$ in Equation A.2.1 multiplied by $A(R_{CC}(\mathsf{y}_1) \cup R_{CC}(\mathsf{y}_2))$ (right).

If x_1 sufficiently close to $1/2$, then $\min(2\,x_1, s_1(x_1, y_1)) = 2\,x_1$ will hold. Then $\min(2\,X_1, s_1(X_1, Y_1)) = 2\,X_1$ with probability 1 as $n \to \infty$. Hence $\gamma(\mathcal{X}_n, N_2) = 1$ by a Lemma in [3] since $f(0) = f(1) = 0$ and $f(1/2) = \frac{c_2^2 - c_1\,(1-c_1)}{2\,c_2}$.

Hence $\lim_{n\to\infty} P(\gamma(\mathcal{X}_n, N_2) = 1) = 1$ which in turn implies $\lim_{n\to\infty} P\left(\gamma_{ind}(\mathcal{X}_n, \widetilde{N}_{AS}) = 1\right) = 1$. Then as $n \to \infty$ two points suffice to dominate the data set $\mathcal{X}_n$ with probability 1. Hence $P(\gamma(\mathcal{X}_n, N_{AS}, M_{CC}) \leq 2) \to 1$ as $n \to \infty$. ■

A.2.3 Proof of Theorem 4.4.8

Let $M = (m_1, m_2) \in \partial(\mathscr{T}^r)$, say $M \in q_3(r, x)$ (recall that $q_j(r, x)$ are defined such that $d(\mathrm{y}_j, e_j) = r \cdot d(q_j(r, x), \mathrm{y}_j)$ for $j \in \{1, 2, 3\}$), then $m_2 = \frac{\sqrt{3}\,(2-r)}{2\,r}$ and $m_1 \in \left[\frac{3\,(r-1)}{2\,r}, \frac{3-r}{2\,r}\right]$. Let $X_{e_j}(n)$ be one of the closest point(s) to the edge e_j; i.e., $X_{e_j}(n) \in \mathrm{argmin}_{X\in\mathcal{X}_n}\, d(X, e_j)$ for $j \in \{1, 2, 3\}$. Note that $X_{e_j}(n)$ is unique a.s. for each j.

Notice that for all $j \in \{1, 2, 3\}$, $X_{e_j}(n) \notin N_{PE}^r(X)$ for all $X \in \mathcal{X}_n \cap R_M(\mathrm{y}_j)$ implies that $\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) > 1$ with probability 1. For sufficiently large n, $X_{e_j}(n) \notin N_{PE}^r(X)$ for all $X \in \mathcal{X}_n \cap R_M(\mathrm{y}_j)$ with probability 1, for $j \in \{1, 2\}$, by the choice of M. Hence we consider only $X_{e_3}(n)$. The asymptotically accurate pdf of $X_{e_3}(n)$ is

$$f_{e_3}\,(x, y) = n\left(\frac{A(S_U(x, y))}{A(T(\mathcal{Y}))}\right)^{n-1} \frac{1}{A(T(\mathcal{Y}))},$$

where $S_U(x, y)$ is the unshaded region in Figure A.2.5 (left) (for a given $X_{e_3}(n) = x_{e_3} = (x, y)$) whose area is $A(S_U(x, y)) = \sqrt{3}\,\left(2\,y - \sqrt{3}\right)^2/12$. Note that $X_{e_3}(n) \notin N_{PE}^r(X)$ for all $X \in \mathcal{X}_n \cap R_M(\mathrm{y}_3)$ iff $\mathcal{X}_n \cap [\Gamma_1\,(\mathcal{X}_n, N_{PE}^r, M) \cap R_M(\mathrm{y}_3)] = \emptyset$. Then given $X_{e_3}(n) = (x, y)$,

$$P\left(\mathcal{X}_n \cap \left[\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathrm{y}_3)\right] = \emptyset\right) = \left(\frac{A(S_U(x, y)) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathrm{y}_3)\right)}{A(S_U(x, y))}\right)^{n-1},$$

where $A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathrm{y}_3)\right) = \frac{\sqrt{3}\,y^2}{3\,(r-1)\,r}$ (see Figure A.2.5 (right) where the shaded region is $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathrm{y}_3)$ for a given $X_{e_3}(n) = (x, y)$), then for sufficiently large n

$$\begin{aligned} &P\left(\mathcal{X}_n \cap \left[\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathrm{y}_3)\right] = \emptyset\right) \approx \\ &\qquad \int \left(\frac{A(S_U(x, y)) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathrm{y}_3)\right)}{A(S_U(x, y))}\right)^{n-1} f_{e_3}\,(x, y)\,dy\,dx \\ &\qquad\quad = \int \frac{n}{A(T(\mathcal{Y}))} \left(\frac{A(S_U(x, y)) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathrm{y}_3)\right)}{A(T(\mathcal{Y}))}\right)^{n-1} dy\,dx. \end{aligned}$$

Let

$$G(x,y) = \frac{A(S_U(x,y)) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_3)\right)}{A(T(\mathcal{Y}))} = \frac{4}{\sqrt{3}}\left(\frac{\sqrt{3}\,\left(2\,y-\sqrt{3}\right)^2}{12} - \frac{\sqrt{3}\,y^2}{3\,(r-1)\,r}\right),$$

which is independent on x, so we denote it as $G(y)$.

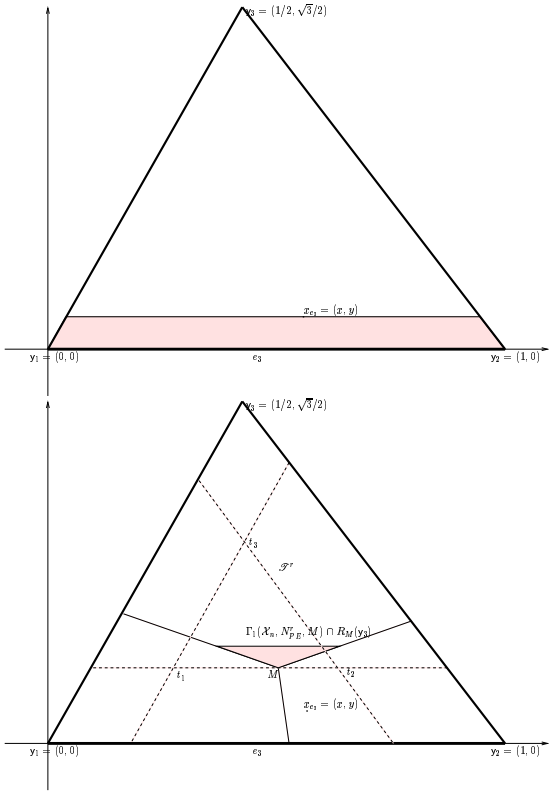


Figure A.2.5: A figure for the description of the pdf of $X_{e_3}(n)$ (left) and $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right)$ (right) given $X_{e_3}(n) = x_{e_3} = (x, y)$.

Let $\varepsilon > 0$ be sufficiently small, then for sufficiently large n,

$$P\left(\mathcal{X}_n \cap \left[\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_3)\right] = \emptyset\right) \approx$$
$$\int_0^{\varepsilon} \int_{y/\sqrt{3}}^{1-y/\sqrt{3}} n\,G(y)^{n-1}\,4/\sqrt{3}\,dy\,dx = \left(1 - 2\,y/\sqrt{3}\right) \int_0^{\varepsilon} n\,G(y)^{n-1}\,4/\sqrt{3}\,dy.$$

The integrand is critical at $y = 0$, since $G(0) = 1$ (i.e., when $x_{e_3} \in e_3$). Furthermore, $G(y) =$

$1 - 4\,y/\sqrt{3} + O\left(y^2\right)$ around $y = 0$. Then letting $y = w/n$, we get

$$P\left(\mathcal{X}_n \cap \left[\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_3)\right] = \emptyset\right) \approx \left(1 - \frac{2\,w}{\sqrt{3}\,n}\right) \frac{4}{\sqrt{3}} \int_0^{n\varepsilon} \left(1 - \frac{4}{\sqrt{3}\,n}\,w + O\left(n^{-2}\right)\right)^{n-1} dw.$$

$$\text{letting } n \to \infty, \qquad \approx 4/\sqrt{3} \int_0^{\infty} \exp\left(-4\,w/\sqrt{3}\right) dw = 1.$$

Hence $\lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) > 1\right) = 1$. For $M \in q_j(r,x) \cap \mathscr{T}^r$ with $j \in \{1,2\}$ the result follows similarly. ■

A.2.4 Proof of Theorem 4.4.9

Let $M = (m_1, m_2) \in \partial(\mathscr{T}^r) \setminus \{t_1, t_2, t_3\}$, say $M \in q_3(r,x)$. Then $m_2 = \frac{\sqrt{3}\,(r-1)}{2\,r}$. Without loss of generality, assume $\frac{1}{2} \le m_1 < \frac{3-r}{2\,r}$. See also Figure A.2.6.

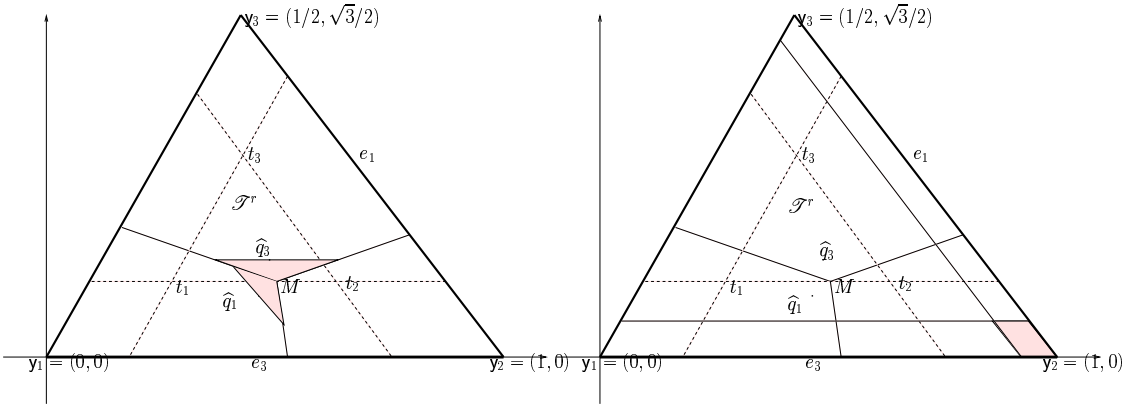


Figure A.2.6: A figure for the description of the pdf of $\hat{Q}_1(n)$ and $\hat{Q}_3(n)$ (left) and the unshaded region is $N_{PE}^r(\hat{q}_1, M) \cup N_{PE}^r(\hat{q}_3, M)$ (right).

Whenever $\mathcal{X}_n \cap R_M(\mathsf{y}_j) \neq \emptyset$, let

$$\widehat{Q}_j(n) \in \operatorname{argmin}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_j)} d\left(X, e_j\right) = \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_j)} d(\ell(\mathsf{y}_j, X), \mathsf{y}_j) \text{ for } j \in \{1,2,3\}.$$

Note that at least one of the $\widehat{Q}_j(n)$ uniquely exists w.p. 1 for finite n and as $n \to \infty$, $\widehat{Q}_j(n)$ are unique w.p. 1. Then

$$\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) \le 2 \text{ iff } \mathcal{X}_n \subset \left[N_{PE}^r\left(\widehat{Q}_1(n), M\right) \cup N_{PE}^r\left(\widehat{Q}_2(n), M\right)\right] \text{ or}$$
$$\mathcal{X}_n \subset \left[N_{PE}^r\left(\widehat{Q}_2(n), M\right) \cup N_{PE}^r\left(\widehat{Q}_3(n), M\right)\right] \text{ or } \mathcal{X}_n \subset \left[N_{PE}^r\left(\widehat{Q}_1(n), M\right) \cup N_{PE}^r\left(\widehat{Q}_3(n), M\right)\right].$$

Let $E_n^{i,j}$ be the event that $\mathcal{X}_n \subset N_{PE}^r\left(\widehat{Q}_i, M\right) \cup \left[N_{PE}^r\left(\widehat{Q}_j(n), M\right)\right]$ for $(i,j) \in \{(1,2),(1,3),(2,3)\}$.

Then

$$P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) \leq 2\right) = P\left(E_n^{1,2}\right) + P\left(E_n^{2,3}\right) + P\left(E_n^{1,3}\right) - P\left(E_n^{1,2} \cap E_n^{2,3}\right) - P\left(E_n^{1,2} \cap E_n^{1,3}\right) - P\left(E_n^{1,3} \cap E_n^{2,3}\right) + P\left(E_n^{1,2} \cap E_n^{2,3} \cap E_n^{1,3}\right).$$

But note that $P\left(E_n^{1,2}\right) \to 0$ as $n \to \infty$ by the choice of M since

$$\sup_{\substack{u \in R_M(\mathrm{y}_1) \\ v \in R_M(\mathrm{y}_2)}} N_{PE}^r(u, M) \cup N_{PE}^r(v, M) \subsetneq T(\mathcal{Y}),$$

and

$$P\left(\mathcal{X}_n \cap T(\mathcal{Y}) \setminus \left[\sup_{\substack{u \in R_M(\mathrm{y}_1) \\ v \in R_M(\mathrm{y}_2)}} N_{PE}^r(u, M) \cup N_{PE}^r(v, M)\right] \neq \emptyset\right) \to 1 \text{ as } n \to \infty.$$

Then,

$$P\left(E_n^{1,2}\right) - P\left(E_n^{1,2} \cap E_n^{2,3}\right) - P\left(E_n^{1,2} \cap E_n^{1,3}\right) + P\left(E_n^{1,2} \cap E_n^{2,3} \cap E_n^{1,3}\right) \leq 4\, P\left(E_n^{1,2}\right) \to 0 \text{ as } n \to \infty.$$

Therefore,

$$\lim_{n \to \infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) \leq 2\right) = \lim_{n \to \infty} \left(P\left(E_n^{2,3}\right) + P\left(E_n^{1,3}\right)\right).$$

Furthermore, observe that $P\left(E_n^{1,3}\right) \geq P\left(E_n^{2,3}\right)$ by the choice of M. Then we first find $\lim_{n \to \infty} P\left(E_n^{1,3}\right)$. Given a realization of $\mathcal{X}_n$ with $\widehat{Q}_1(n) = \widehat{q}_1 = (x_1, y_1)$ and $\widehat{Q}_3(n) = \widehat{q}_3 = (x_3, y_3)$, the remaining $n-2$ points should fall, for example, in the undshaded region in Figure A.2.6 (left). Then the asymptotically accurate joint pdf of $\widehat{Q}_1(n), \widehat{Q}_3(n)$ is

$$f_{13}\left(\vec{\zeta}\right) = \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} \left(\frac{A(T(\mathcal{Y})) - A(S_R\left(\vec{\zeta}\right)\left(\vec{\zeta}\right))}{A(T(\mathcal{Y}))}\right)^{n-2}$$

where $\vec{\zeta} = (x_1, y_1, x_3, y_3)$, $S_R\left(\vec{\zeta}\right)$ is the shaded region in Figure A.2.6 (left) whose area is $A(S_R\left(\vec{\zeta}\right)) = \frac{\sqrt{3}\left(2\,r\,y_3 - \sqrt{3}\,(r-1)\right)^2}{12\,r\,(r-1)} + \frac{\sqrt{3}\left[2\,\sqrt{3}\,r\,y_1 - 3\,(r-1) + 6\,r\,(x_1 - m_1)\right]^2}{72\,r\,(1 - r\,(2\,m_1 - 1))}$.

Given $\widehat{Q}_j(n) = \widehat{q}_j = (x_j, y_j)$ for $j \in \{1, 3\}$,

$$P\left(E_n^{1,3}\right) = \left(\frac{A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R\left(\vec{\zeta}\right))}{A(T(\mathcal{Y})) - A(S_R\left(\vec{\zeta}\right))}\right)^{n-2}$$

then for sufficiently large n

$$\begin{aligned} P\left(E_n^{1,3}\right) &\approx \int \left(\frac{A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y})) - A(S_R(\vec{\zeta}))} \right)^{n-2} f_{13}\left(\vec{\zeta}\right) d\vec{\zeta}, \\ &= \int \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} \left(\frac{A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y}))} \right)^{n-2} d\vec{\zeta} \end{aligned}$$

where

$$A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) = \frac{\sqrt{3}}{4} - \left(\frac{\left(\sqrt{3}\,r\,y_1 + 3\,r\,x_1 - 3\right)\,\left(\sqrt{3}\,(r-1) - 2\,r\,y_3\right)}{6} \right).$$

See Figure A.2.6 (right) for $N_{PE}^r(\widehat{q}_1, M) \cup N_{PE}^r(\widehat{q}_3, M)$. Let

$$G(\vec{\zeta}) = \frac{A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y}))}.$$

Note that the integral is critical at $x_1 = x_3 = m_1$ and $y_1 = y_3 = m_2$, since $G(\vec{\zeta}) = 1$. Since $N_{PE}^r(x, M_C)$ depends on the distance $d(x, e_j)$ for $x \in R_M(\mathsf{y}_j)$, we make the change of variables $(x_1, y_1) \to (d(M, e_1) + z_1, y_1)$ where $d(M, e_1) = \frac{\sqrt{3}\,(r+1-2\,r\,m_1)}{4\,r}$ and $(x_3, y_3) \to (x_3, m_2 + z_3)$ then $G(\vec{\zeta})$ depends only on z_1, z_3, we denote it $G(z_1, z_3)$ which is

$$G(z_1, z_3) = 1 - \frac{8\,r\,z_1^2}{3\,(1 + r\,(1 - 2\,m_1))} - \frac{4\,r\,z_3^2}{3\,(r-1)} - \frac{2\,r\,z_3\,\left(\sqrt{3}\,(3-r)\right) + r\,\left(4\,z_1 - 2\,\sqrt{3}\,m_1\right)}{3}.$$

The new integrand is $\frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} G(z_1, z_3)^{n-2}$. Integrating with respect to x_3 and y_1 yields $\frac{2\,\sqrt{3}\,z_3\,r}{3\,(r-1)}$ and $\frac{4\,\sqrt{3}\,r\,z_1}{3\,(2\,r\,m_1 - r - 1)}$, respectively. Hence for sufficiently large n

$$P\left(E_n^{1,3}\right) \approx \int_0^{\varepsilon} \int_0^{\varepsilon} \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} \left(\frac{2\,\sqrt{3}\,z_3\,r}{3\,(r-1)} \right) \left(\frac{4\,\sqrt{3}\,r\,z_1}{3\,(2\,r\,m_1 - r - 1)} \right) G(z_1, z_3)^{n-2} dz_1\,dz_3.$$

Note that the new integral is critical when $z_1 = z_3 = 0$, so we make the change of variables $z_1 = w_1/\sqrt{n}$ and $z_3 = w_3/n$ then $G(z_1, z_3)$ becomes

$$G(w_1, w_3) = 1 + \frac{1}{n} \left(\frac{2\,\sqrt{3}\,r\,(r - 3 + 2\,r\,m_1)}{3} w_3 + \frac{8\,r}{3\,(r + 1 - 2\,r\,m_1)}\,w_1^2 \right) + O\left(n^{-3/2}\right),$$

so for sufficiently large n

$$P\left(E_n^{1,3}\right) \approx \int_0^{\sqrt{n}\,\varepsilon} \int_0^{n\,\varepsilon} \frac{(n-1)}{n^3}\frac{16}{3}\left(\frac{2\sqrt{3}\,r}{3\,(r-1)}\right)\left(\frac{4\sqrt{3}\,r}{3\,(2\,r\,m_1-r-1)}\right)(-4\,m_1+2+\sqrt{2})w_1\,w_3$$
$$\left[1-\frac{1}{n}\left(\frac{2\sqrt{3}\,r\,(r-3+2\,r\,m_1)}{3}w_3+\frac{8\,r}{3\,(r+1-2\,r\,m_1)}\,w_1^2\right)+O\left(n^{-3/2}\right)\right]^{n-2} dw_3 w_1,$$
$$\approx O\left(n^{-1}\right)\int_0^{\infty}\int_0^{\infty} w_1\,w_3\,\exp\left(-\frac{2\sqrt{3}\,r\,(r-3+2\,r\,m_1)}{3}w_3-\frac{8\,r}{3\,(r+1-2\,r\,m_1)}\,w_1^2\right)dw_3 w_1$$
$$= O\left(n^{-1}\right)$$

since $\int_0^{\infty}\int_0^{\infty} w_1\,w_3\,\exp\left(-\frac{2\sqrt{3}\,r\,(r-3+2\,r\,m_1)}{3}w_3-\frac{8\,r}{3\,(r+1-2\,r\,m_1)}\,w_1^2\right)dw_3 w_1 = \frac{3}{8\,r\,(3-r\,(2\,m_1+1))}$, which is a finite constant. Then $P\left(E_n^{1,3}\right)\to 0$ as $n\to\infty$, which also implies $P\left(E_n^{2,3}\right)\to 0$ as $n\to\infty$. Then $P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right)\le 2\right)\to 0$. Hence the desired result follows. ■

A.2.5 Proof of Theorem 4.4.10

Let $M=(m_1,m_2)\in\{t_1,t_2,t_3\}$. Without loss of generality, assume $M=t_2$ then $m_1=\frac{2-r+c_1\,(r-1)}{r}$ and $m_2=\frac{c_2\,(r-1)}{r}$. See Figure A.2.7.

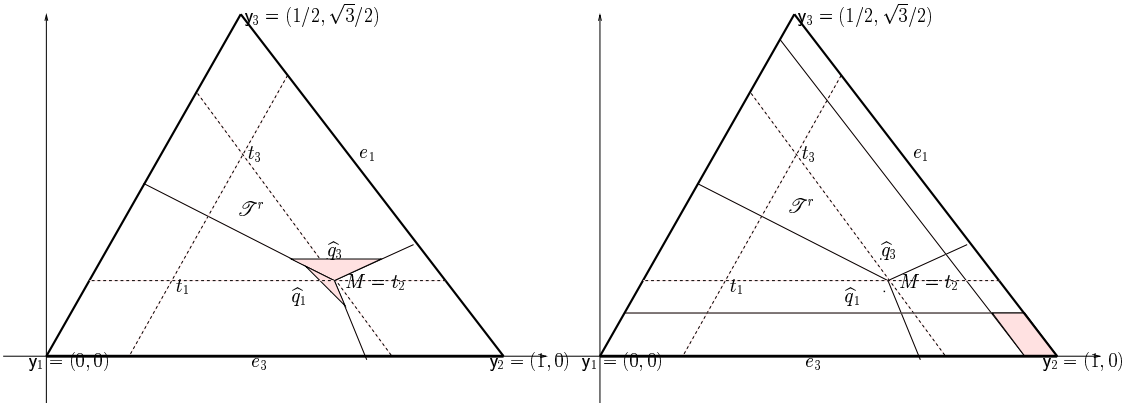


Figure A.2.7: A figure for the description of the pdf of $\hat{Q}_1(n)$ and $\hat{Q}_3(n)$ (left) and the unshaded region is $N_{PE}^r(\hat{q}_1,M)\cup N_{PE}^r(\hat{q}_3,M)$ (right) given $\hat{Q}_j(n)=\hat{q}_j$ for $j\in\{1,3\}$.

Let $\widehat{Q}_j(n)$ and the events $E_n^{i,j}$ be defined as in the proof of Theorem 4.4.9 for $(i,j)\in\left\{(1,2),(1,3),(2,3)\right\}$. Then as in the proof of Theorem 4.4.9,

$$P\left(\gamma\left(\mathcal{X}_n,N_{PE}^r,M\right)\le 2\right)=P\left(E_n^{1,2}\right)+P\left(E_n^{2,3}\right)+P\left(E_n^{1,3}\right)-P\left(E_n^{1,2}\cap E_n^{2,3}\right)-$$
$$P\left(E_n^{1,2}\cap E_n^{1,3}\right)-P\left(E_n^{1,3}\cap E_n^{2,3}\right)+P\left(E_n^{1,2}\cap E_n^{2,3}\cap E_n^{1,3}\right).$$

Observe that the choice of M implies that $P\left(E_n^{1,3}\right) \geq P\left(E_n^{2,3}\right)$ and by symmetry (in T_e) $P\left(E_n^{1,2}\right) = P\left(E_n^{2,3}\right)$. So first we find $P\left(E_n^{1,3}\right)$. As in the proof of Theorem 4.4.9 asymptotically accurate joint pdf of $\widehat{Q}_1(n), \widehat{Q}_3(n)$ is

$$f_{13}\left(\vec{\zeta}\right) = \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\left(\frac{A(T(\mathcal{Y})) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y}))}\right)^{n-2}$$

where $\vec{\zeta} = (x_1, \mathsf{y}_1, x_3, \mathsf{y}_3)$ and $S_R\left(\vec{\zeta}\right)$ is the shaded region in Figure A.2.7 (left) whose area is

$$A(S_R(\vec{\zeta})) = \frac{\sqrt{3}\,\left(2\,r\,y_3 - \sqrt{3}\,(r-1)^2\right)}{12\,(r-1)\,r} + \frac{\sqrt{3}\,\left(\sqrt{3}\,r\,y_1 + 3\,x_1\,r - 3\right)^2}{36\,(r-1)\,r}.$$

Given $\widehat{Q}_j(n) = \widehat{q}_j = (x_j, y_j)$ for $j \in \{1,3\}$,

$$P\left(E_n^{1,3}\right) = \left(\frac{A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y})) - A(S_R(\vec{\zeta}))}\right)^{n-2},$$

then for sufficiently large n

$$\begin{aligned} P\left(E_n^{1,3}\right) &\approx \int \left(\frac{A\left(N_{PE}^r\left(\widehat{Q=q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y})) - A(S_R(\vec{\zeta}))}\right)^{n-2} f_{13}\left(\vec{\zeta}\right) d\vec{\zeta}, \\ &= \int \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\left(\frac{A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y}))}\right)^{n-2} d\vec{\zeta} \end{aligned}$$

where

$$A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) = \frac{\sqrt{3}}{4} - \frac{\left(2\,r\,y_3 - \sqrt{3}\,(r-1)\right)\,\left(3 - \sqrt{3}\,r\,y_1 - 3\,r\,x_1\right)}{6}.$$

See Figure A.2.7 (right) for $N_{PE}^r(\widehat{q}_1, M) \cup N_{PE}^r(\widehat{q}_3, M)$. Let

$$G(\vec{\zeta}) = \frac{A\left(N_{PE}^r\left(\widehat{q}_1, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y}))}.$$

Note that the integral is critical when $x_1 = x_3 = m_1$ and $y_1 = y_3 = m_2$, since $G\left(\vec{\zeta}\right) = 1$.

As in the proof of Theorem 4.4.9, we make the change of variables $(x_1, y_1) \to (d(M, e_1) +$

$z_1, y_1)$ where $d(M, e_1) = \frac{\sqrt{3}\,(r-1)}{2\,r}$ and $(x_3, y_3) \to (x_3, m_2 + z_3)$. Then $G(\vec{\zeta})$ becomes

$$G(z_1, z_3) = 1 - \frac{4\,r}{3\,(r-1)}\,z_1^2 - \frac{4\,r}{3\,(r-1)}\,z_3^2 - \frac{8\,r^2}{3}\;z_1\,z_3.$$

The new integral is

$$\int \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} G(z_1, z_3)^{n-2} dx_3 dy_1 dz_3 dz_1.$$

Note that $G(z_1, z_3)$ is independent of y_1, x_3, so integrating with respect to x_3 and y_1 yields $\frac{2\sqrt{3}\,r\,z_1}{3\,(r-1)}$ and $\frac{2\sqrt{3}\,r\,z_3}{3\,(r-1)}$, respectively. The new integral is critical at $z_1 = z_3 = 0$. Hence, for sufficiently large n and sufficiently small $\varepsilon > 0$, the integral becomes,

$$P\left(E_n^{1,3}\right) \approx \int_0^{\varepsilon}\int_0^{\varepsilon} \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\left(\frac{12\,r^2}{9\,(r-1)^2}\right) z_1\,z_3\,G(z_1, z_3)^{n-2}\;dz_1 dz_3.$$

Since the new integral is critical when $z_1 = z_2 = 0$ we make the change of variables $z_j = w_j/\sqrt{n}$ for $j \in \{1, 3\}$ then $G(z_1, z_3)$ becomes

$$G(w_1, w_3) = 1 - \frac{4\,r}{3\,n\,(r-1)}\left(w_1^2 + w_3^2 + 2\,r\,(r-1)\,w_1\,w_3)\right),$$

so

$$\begin{aligned} p_r := P\left(E_n^{1,3}\right) \approx & \int_0^{\sqrt{n}\,\varepsilon}\int_0^{\sqrt{n}\,\varepsilon} \frac{(n-1)}{n}\,\frac{16}{3}\left(2\left(\frac{12\,r^2}{9\,(r-1)^2}\right) w_1\,w_3\right) \\ & \left[1 - \frac{4\,r}{3\,n\,(r-1)}\left(u_1^2 + u_3^2 + 2\,r\,(r-1)\,u_1\,u_3)\right)\right]^{n-2} dw_3 w_1, \\ & \text{letting } n \to \infty, \\ \approx & \int_0^{\infty}\int_0^{\infty} \frac{64}{9}\left(\frac{r}{r-1}\right)^2 w_1\,w_3\;\exp\left(\frac{4\,r}{3\,(r-1)}\left(u_1^2 + u_3^2 + 2\,r\,(r-1)\,u_1\,u_3\right)\right) dw_3 w_1 \end{aligned}$$

which is not analytically integrable, but p_r can be obtained by numerical integration, e.g., $p_{r=\sqrt{2}} \approx .4826$ and $p_{r=5/4} \approx .6514$.

Next, we find $\lim_{n\to\infty} P\left(E_n^{2,3}\right)$. The asymptotically accurate joint pdf of $\widehat{Q}_2(n), \widehat{Q}_3(n)$ is

$$f_{23}\left(\vec{\zeta}\right) = \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\left(\frac{A(T(\mathcal{Y})) - A(S_R^2(\vec{\zeta}))}{A(T(\mathcal{Y}))}\right)^{n-2}$$

where $\vec{\zeta} = (x_2, y_2, x_3, y_3)$ and $S_R^2(\vec{\zeta})$ is the shaded region in Figure A.2.8 (left) whose area is

$$A\left(S_R^2(\vec{\zeta})\right) = \frac{\sqrt{3}\ \left(2\,r\,\mathsf{y}_3 + \sqrt{3}\,(1-r)\right)}{12\,r\,(r-1)} + \frac{\sqrt{3}\ \left(\sqrt{3}\,r\,\mathsf{y}_2 - 3\,r\,x_2 - 3r + 6\right)}{36\,(2-r)\,r}.$$

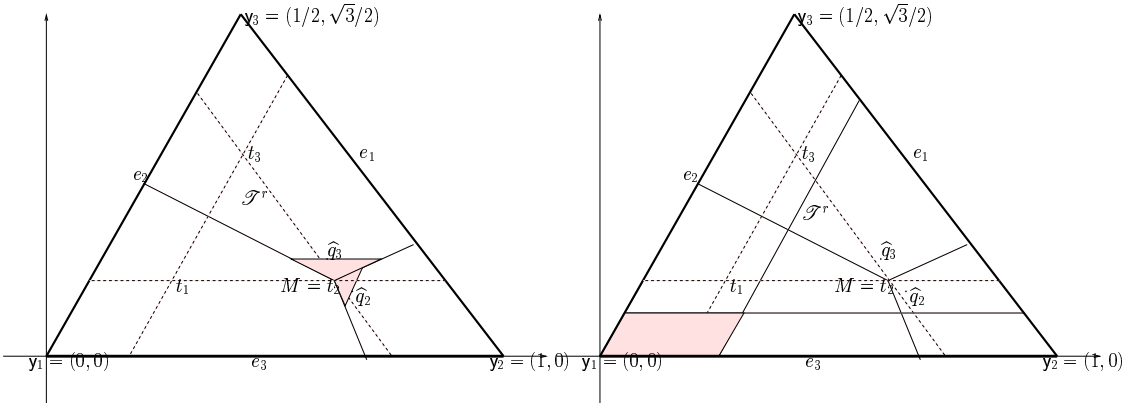


Figure A.2.8: A figure for the description of the pdf of $\hat{Q}_2(n)$ and $\hat{Q}_3(n)$ (left) and unshaded region is $N_{PE}^r(\hat{q}_2) \cup N_{PE}^r(\hat{q}_3)$ (right) given $\hat{Q}_j(n) = \hat{q}_j$ for $j \in \{2,3\}$.

As before,

$$\begin{aligned} P\left(E_n^{2,3}\right) &= \int \left(\frac{A\left(N_{PE}^r\left(\widehat{q}_2, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y})) - A(S_R^2(\vec{\zeta}))}\right)^{n-2} f_{23}\left(\vec{\zeta}\right) d\vec{\zeta} \\ &= \int \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} \left(\frac{A\left(N_{PE}^r\left(\widehat{q}_2, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y}))}\right)^{n-2} d\vec{\zeta} \end{aligned}$$

where

$$A\left(N_{PE}^r\left(\widehat{q}_2, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) = \frac{\sqrt{3}}{4} - \frac{\left(2\,r\,y_3 - \sqrt{3}\,(r-1)\right)\ \left(3 - \sqrt{3}\,r\,y_2 + 3\,r\,x_2 - 3\,r\right)}{6}.$$

See Figure A.2.8 (right) for $N_{PE}^r(\widehat{q}_2) \cup N_{PE}^r(\widehat{q}_3, M)$. Let

$$G(\vec{\zeta}) = \frac{A\left(N_{PE}^r\left(\widehat{q}_2, M\right) \cup N_{PE}^r\left(\widehat{q}_3, M\right)\right) - A(S_R(\vec{\zeta}))}{A(T(\mathcal{Y}))}.$$

Note that the integral is critical when $x_2 = x_3 = m_1$ and $y_2 = y_3 = m_2$, since $G(\vec{\zeta}) = 1$.

We make the change of variables $(x_3, y_3) \to (x_3, m_2 + z_3)$ and $(x_2, y_2) \to (d(M, e_2) + z_2, y_2)$

where $d(M,e_2)=\frac{\sqrt{3}\,(2-r)}{2\,r}$. Then $G\left(\vec{\zeta}\right)$ becomes

$$G(z_2,z_3)=1-\frac{4\,r\,z_2^2}{3\,(2-r)}-\frac{4\,r\,z_2^2}{3\,(r-2)}-\frac{4\sqrt{3}\,r\,z_3\,(3-2\,r)}{3}-\frac{8\,r^2\,z_2\,z_3}{3}.$$

The new integral is

$$\int\frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}G(z_2,z_3)^{n-2}dx_3dy_2dz_3dz_2.$$

The integrand is independent of x_3 and y_2, so integrating with respect to x_3 and y_2 yields $\frac{2\sqrt{3}\,r\,z_3}{3\,(r-1)}$ and $\frac{2\sqrt{3}\,r\,z_2}{3\,(2-r)}$, respectively. Hence, for sufficiently large n

$$P\left(E_n^{2,3}\right)\approx\int_0^{\varepsilon}\int_0^{\varepsilon}\frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\left(\frac{4\,r^2}{3\,(r-1)\,(2-r)}\right)z_3\,z_2\,G(z_2,z_3)^{n-2}dz_2dz_3.$$

Note that the new integral is critical when $z_2=z_3=0$, so we make the change of variables $z_2=w_2/\sqrt{n}$ and $z_3=w_3/n$ then $G(z_2,z_3)$ becomes

$$G(w_2,w_3)=1-\frac{1}{n}\left[\frac{4\,r\,w_2^2}{3\,(2-r)}-\frac{4\sqrt{3}\,r\,w_3\,(3-2\,r)}{3}\right]+O\left(n^{-\frac{3}{2}}\right),$$

so for sufficiently large n

$$\begin{aligned}P\left(E_n^{2,3}\right)&\approx\int_0^{\sqrt{n}\,\varepsilon}\int_0^{n\,\varepsilon}\frac{(n-1)}{n^2}\,\frac{64\,r^2}{9\,(r-1)\,(2-r)}\,w_2\,w_3\\&\left[1-\frac{1}{n}\left(\frac{4\,r\,w_2^2}{3\,(2-r)}-\frac{4\sqrt{3}\,r\,w_3\,(3-2\,r)}{3}\right)+O\left(n^{-\frac{3}{2}}\right)\right]^{n-2}dw_3w_2,\\&\approx O\left(n^{-1}\right)\int_0^{\infty}\int_0^{\infty}w_2\,w_3\;\exp\left(-\frac{4\,r\,w_2^2}{3\,(2-r)}-\frac{4\sqrt{3}\,r\,u_3\,(3-2\,r)}{3}\right)dw_3w_2=O\left(n^{-1}\right)\end{aligned}$$

since

$$\int_0^{\infty}\int_0^{\infty}w_2\,w_3\;\exp\left(-\frac{4\,r\,w_2^2}{3\,(2-r)}-\frac{4\sqrt{3}\,r\,u_3\,(3-2\,r)}{3}\right)dw_3w_2=\frac{27\,(2-r)}{384\,r^3\,(3-2\,r)^2}$$

which is a finite constant.

Thus we have shown that $P\left(E_n^{2,3}\right) \to 0$ as $n \to \infty$, which implies that as $n \to \infty$,

$$P\left(E_n^{2,3}\right) + P\left(E_n^{1,2}\right) - P\left(E_n^{1,2} \cap E_n^{2,3}\right) - P\left(E_n^{1,2} \cap E_n^{1,3}\right) \\ - P\left(E_n^{1,3} \cap E_n^{2,3}\right) + P\left(E_n^{1,2} \cap E_n^{2,3} \cap E_n^{1,3}\right) \leq 5\, P\left(E_n^{2,3}\right) \to 0.$$

Hence $\lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) \leq 2\right) = \lim_{n\to\infty} P\left(E_n^{1,3}\right)$ and $\lim_{n\to\infty} P(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) > 1) = 1$ implies that

$$\lim_{n\to\infty} P(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = 2) = p_r.$$

■

A.2.6 Proof of Theorem 4.4.11

Whenever $\mathcal{X}_n \cap R_{CM}(\mathsf{y}_j) \neq \emptyset$, let $\widetilde{Q}_j(n) \in \operatorname{argmin}_{X \in \mathcal{X}_n \cap R_{CM}(\mathsf{y}_j)} d(X, e_j) = \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_{CM}(\mathsf{y}_j)} d(\ell(\mathsf{y}_j, X), e_j)$, where e_j is the edge opposite vertex y_j for $j \in \{1,2,3\}$ and let $\widetilde{q}_j$ be a realization of $\widetilde{Q}_j(n)$ for $j \in \{1,2,3\}$. Note that for sufficiently large n, $\widetilde{Q}_j(n)$ uniquely exist w.p. 1 for each j. Then

$$\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) = 1 \text{ iff} \\ \mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_1(n), M_C\right) \text{ or } \mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_2(n), M_C\right) \text{ or } \mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_3(n), M_C\right).$$

Let $E_n^j := \left\{\mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_j(n), M_C\right)\right\}$ for $j \in \{1,2,3\}$. The events E_n^j are not necessarily disjoint. Then

$$P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) = 1\right) = P\left(E_n^1\right) + P\left(E_n^2\right) + P\left(E_n^3\right) - P\left(E_n^1 \cap E_n^2\right) - P\left(E_n^1 \cap E_n^3\right) \\ - P\left(E_n^2 \cap E_n^3\right) + P\left(E_n^1 \cap E_n^2 \cap E_n^3\right) \leq 4\, P\left(E_n^2\right).$$

Let $T(\widetilde{Q}_j(n))$ be the triangle bounded by the median lines at y_k and y_l for $k, l \neq j$ and $\ell(\mathsf{y}_j, \widetilde{Q}_j(n))$, then $T(\widetilde{q}_2)$ has vertices

$$\left(1/2, \sqrt{3}/6\right), \left(1/2, y_2 + \sqrt{3}/2 - \sqrt{3}\, x_2\right), \left(\sqrt{3}/2\, \left(-y_2 + \sqrt{3}\, x_2\right), -y_2/2 + \sqrt{3}\, x_2/2\right).$$

Then the asymptotically accurate pdf of $\widetilde{Q}_2(n)$ is

$$f_{\widetilde{Q}_2}(x_2, y_2) = \frac{n}{A(T(\mathcal{Y}))} \left(\frac{A(T(\mathcal{Y})) - A(T(\widetilde{q}_2))}{A(T(\mathcal{Y}))} \right)^{n-1},$$

where $A(T(\widetilde{q}_2)) = \left(-\sqrt{3}\, y_2 - 1 + 3\, x_2\right)^2 / 12$ with domain $D_I = \left\{(x_2, y_2) \in R_{CM}(\mathrm{y}_2) : \; y_2 \geq -(1 + x_2 - \varepsilon)/\sqrt{3}\right\}$ for $\varepsilon > 0$ small enough that $T(\widetilde{q}_2) \subsetneq R_{CM}(\mathrm{y}_2)$. See Figure A.2.9.

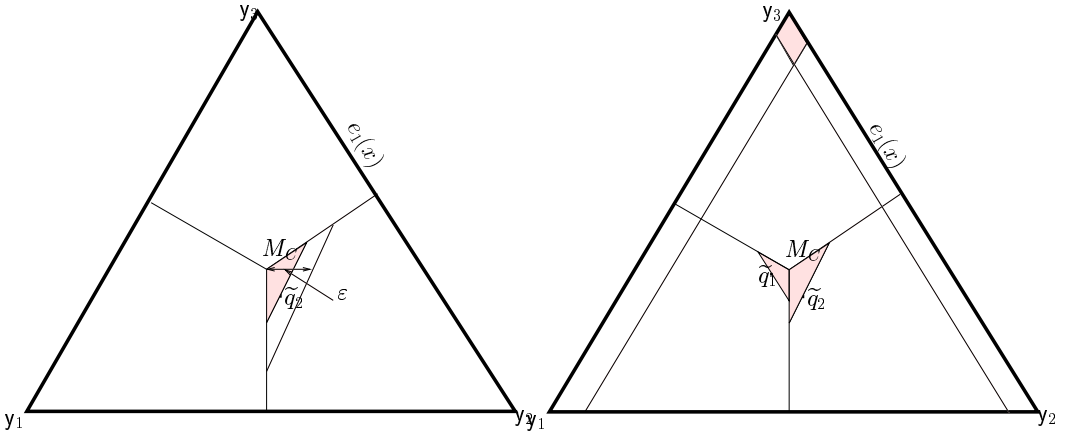


Figure A.2.9: A figure for the description of the pdf of $\tilde{Q}_2(n)$ given $\tilde{Q}_2(n) = \tilde{q}_2$. The shaded region is $T(\tilde{q}_2)$ (left). A figure for the description of the joint pdf of $\tilde{Q}_1(n), \tilde{Q}_2(n)$ given $\tilde{Q}_j(n) = \tilde{q}_j$ for $j \in \{1, 2\}$, and the event $E_n^{1,2}$. The unshaded region is $N_{PE}^{3/2}(\tilde{q}_1) \cup N_{PE}^{3/2}(\tilde{q}_2) \setminus [T(\tilde{q}_1) \cup T(\tilde{q}_2)]$ (right).

Then for a given realization of $\mathcal{X}_n$ with $\widetilde{q}_2 = (x_2, y_2)$, $N_{PE}^{3/2}(\widetilde{q}_2, M_C)$ is a triangle with vertices $\left(-1/2\, \left(\sqrt{3}\, y_2 + 1 - 3\, x_2\right), 0\right), (1, 0), \left(-1/4\, \left(\sqrt{3}\, y_2 - 1 - 3\, x_2\right), 3/4\, \left(\sqrt{3} + y_2 - \sqrt{3}\, x_2\right)\right)$. So for sufficiently large n

$$\begin{aligned} P\left(E_n^2\right) &\approx \int_{D_I} \left(\frac{A\left(N_{PE}^{3/2}(\widetilde{q}_2, M_C)\right) - T(\widetilde{q}_2)}{A\left(N_{PE}^{3/2}(\widetilde{q}_2, M_C)\right)} \right)^{n-1} f_{\widetilde{Q}_2}(x_2, y_2)\, dy_2 dx_2 \\ &= \int_{D_I} \frac{n}{A(T(\mathcal{Y}))} \left(\frac{A\left(N_{PE}^{3/2}(\widetilde{q}_2, M_C)\right) - A(T(\widetilde{q}_2))}{A(T(\mathcal{Y}))} \right)^{n-1} dy_2 dx_2 \end{aligned}$$

where $A\left(N_{PE}^{3/2}(\widetilde{q}_2, M_C)\right) = 3\sqrt{3}\, \left(-\sqrt{3} - y_2 + \sqrt{3}\, x_2\right)^2 / 16$.

Let

$$G(x_2, y_2) = \frac{A\left(N_{PE}^{3/2}(\widetilde{q}_2, M_C)\right) - A(T(\widetilde{q}_2))}{A(T(\mathcal{Y})}.$$

The integral is critical at $(x_2, y_2) = (1/2, \sqrt{3}/6)$, i.e., when $\widetilde{q}_2 = M_C$ since $G(1/2, \sqrt{3}/6) = 1$.

So we make the change of variables $x_2 = y_2/\sqrt{3} + 1/3 + z_2$, then $G(x_2, y_2)$ becomes $G(z_2, y_2) = (48 - 3(2 + z_2)^2)/4$, which does not depend on y_2, so we denote it as $G(z_2)$. Hence the integral becomes

$$P\left(E_n^2\right) \approx \int\int \frac{n}{A(T(\mathcal{Y}))} G(z_2)^{n-1} dy_2 dz_2,$$

where the limits of integration for y_2 are $\sqrt{3}/2\,(1/3 + z_2)$ and $\sqrt{3}/2\,(1/3 - 2\,z_2)$, so integrating over y_2 we get

$$P\left(E_n^2\right) \approx \int_0^{\varepsilon} \frac{n}{A(T(\mathcal{Y}))} \frac{3\sqrt{3}}{2} z_2\, G(z_2)^{n-1} dz_2.$$

The integral is critical when $z_2 = 0$, since $G(z_2 = 0) = 1$. Let $z_2 = w_2/n$. Then for sufficiently large n the integral becomes

$$\begin{aligned} P\left(E_n^2\right) &\approx \int_0^{n\,\varepsilon} 6\,\frac{w_2}{n} \left[\left(1 - 3\frac{w_2}{n} + O\left(n^{-2}\right)\right)\right]^{n-1} dw_2 \\ &\approx O\left(n^{-1}\right) \int_0^{\infty} 6\,w_2\, \exp(-3w_2)\, dw_2 = O\left(n^{-1}\right) \end{aligned}$$

since $\int_0^{\infty} 6\,w_2\, \exp(-3w_2)\, dw_2 = 2/3$. Thus as $n \to \infty$, $P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M_C\right) = 1\right) \to 0$ at rate $O\left(n^{-1}\right)$. ■

A.2.7 Proof of Theorem 4.4.12

Note that

$$\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) \leq 2 \text{ iff } \mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_1(n), M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_2(n), M_C\right) \text{ or}$$
$$\mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_1(n), M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_3(n), M_C\right) \text{ or } \mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_2(n), M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_3(n), M_C\right).$$

We have shown in Theorem 4.4.11 $\lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) > 1\right) = 1$, then

$$\lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) \leq 2\right) = \lim_{n\to\infty} P(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) = 2).$$

Let the events $E_n^{i,j} := \mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_i, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_j(n), M_C\right)$ for $(i, j) = \{(1,2), (1,3), (2,3)\}$.

Then

$$P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) \leq 2\right) = P\left(E_n^{1,2}\right) + P\left(E_n^{1,3}\right) + P\left(E_n^{2,3}\right) - P\left(E_n^{1,2} \cap E_n^{1,3}\right) - P\left(E_n^{1,2} \cap E_n^{2,3}\right) - P\left(E_n^{1,3} \cap E_n^{2,3}\right) + P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right).$$

By symmetry, $P\left(E_n^{1,2}\right) = P\left(E_n^{1,3}\right) = P\left(E_n^{2,3}\right)$ and $P\left(E_n^{1,2} \cap E_n^{1,3}\right) = P\left(E_n^{1,2} \cap E_n^{2,3}\right) = P\left(E_n^{1,3} \cap E_n^{2,3}\right)$. Hence

$$P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) \leq 2\right) = 3\left[P\left(E_n^{1,2}\right) - P\left(E_n^{1,2} \cap E_n^{1,3}\right)\right] + P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right).$$

To find $P\left(E_n^{1,2}\right)$, we need the asymptotically accurate joint pdf of $\widetilde{Q}_1(n)$, $\widetilde{Q}_2(n)$. Given $\widetilde{Q}_j(n) = \widetilde{q}_j$ for $j \in \{1,2\}$, let $\varepsilon > 0$ be small enough such that $T(\widetilde{q}_j(n)) \subsetneq R_{CM}(\mathsf{y}_j)$ for $j \in \{1,2\}$, then the asymptotically accurate joint pdf of $\widetilde{Q}_1(n)$, $\widetilde{Q}_2(n)$ is

$$f_{12}\left(x_1, y_1, x_2, y_2\right) = \frac{n\left(n-1\right)}{A(T(\mathcal{Y}))^2}\left(\frac{A(T(\mathcal{Y})) - A(T(\widetilde{q}_1)) - A(T(\widetilde{q}_2))}{A(T(\mathcal{Y}))}\right)^{n-2}$$

where $A(T(\widetilde{q}_1)) = \sqrt{3}/36\left(-2\sqrt{3} + 3\,y_1 + 3\sqrt{3}\,x_1\right)^2$ and $A(T(\widetilde{q}_2)) = \sqrt{3}/36\left(-3\,y_2 - \sqrt{3} + 3\sqrt{3}\,x_2\right)^2$ with domain $D_I = \left\{(x_1, y_1) \in R_{CM}(\mathsf{y}_1) :\ y_1 \geq -\sqrt{3}/3 + \sqrt{3}\,x_1 + \sqrt{3}\,\varepsilon,\ (x_2, y_2) \in R_{CM}(\mathsf{y}_2) :\ y_2 \leq -\sqrt{3}/3 + \sqrt{3}\,x_2 - \sqrt{3}\,\varepsilon\right\}$. See Figure A.2.9.

Note that $P\left(E_n^{1,2}\right) = P\left(\mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_1(n), M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_2(n), M_C\right)\right)$, then

$$P\left(E_n^{1,2}\right) = \int_{D_I}\left(\frac{A\left(N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right) - A(T(\widetilde{q}_1)) - A(T(\widetilde{q}_2)}{A\left(N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right)}\right)^{n-2} f_{12}\left(\vec{\zeta}\right) d\vec{\zeta}$$

where $\vec{\zeta} = (x_1, y_1, x_2, y_2)$ and $A\left(N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right) = \sqrt{3}/2 + 3\,y_2/4 - 3\sqrt{3}\,x_2/4 - 3\sqrt{3}\,y_1\,y_2/8 - 9\,x_1\,y_2/8 - 3\,y_1/8 - 3\sqrt{3}\,x_1/8 + 9\,x_2\,y_1/8 + 9\sqrt{3}\,x_1\,x_2/8$. The integrand simplifies to

$$\frac{n\left(n-1\right)}{A(T(\mathcal{Y}))^2}\left(\frac{A\left(N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right) - A(T(\widetilde{q}_1)) - A(T(\widetilde{q}_2))}{A(T(\mathcal{Y}))}\right)^{n-2}.$$

Let

$$G\left(\vec{\zeta}\right) = \frac{A\left(N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right) - A(T(\widetilde{q}_1)) - A(T(\widetilde{q}_2))}{A(T(\mathcal{Y}))}.$$

The integral is critical at $x_1 = x_2 = 1/2$ and $y_1 = y_2 = \sqrt{3}/6$, i.e., when $\widetilde{q}_1 = \widetilde{q}_2 = M_C$ since $G(\vec{\zeta}) = 1$.

Next, we make the change of variables $x_1 = 1 - y_1/\sqrt{3} - 1/3 - z_1$ and $x_2 = y_2/\sqrt{3} + 1/3 + z_2$, then $G(\vec{\zeta})$ becomes $G(z_1, z_2) = 1 - 9\,z_1\,z_2/2 - 3\,z_1^2 - 3\,z_2^2$, which does not depend on y_1, y_2. Hence for sufficiently large n the integral becomes

$$P\left(E_n^{1,2}\right) \approx \int \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} G(z_1, z_2)^{n-2} dy_1 dy_2 dz_1 dz_2,$$

where the limits of integration for y_1 are $\sqrt{3}/6 + \sqrt{3}\,z_1/2$, and $\sqrt{3}/6 - \sqrt{3}\,z_1$ and for y_2 are $\sqrt{3}/6 + \sqrt{3}\,z_2/2$, and $\sqrt{3}/6 - \sqrt{3}\,z_2$, so integrating for y_1 and y_2 yields

$$P\left(E_n^{1,2}\right) \approx \int_0^{\varepsilon} \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2} \left(3\sqrt{3}\,z_1/2\right) \left(3\sqrt{3}\,z_2/2\right) G(z_1, z_2)^{n-2} dz_1 dz_2.$$

The integral is critical when $z_1 = z_2 = 0$, since $G(0,0) = 1$. Let $z_j = w_j/\sqrt{n}$ for $j \in \{1,2\}$. Then for sufficiently large n the integral becomes

$$P\left(E_n^{1,2}\right) \approx \int_0^{\sqrt{n}\,\varepsilon} \int_0^{\sqrt{n}\,\varepsilon} \frac{1}{A(T(\mathcal{Y}))^2} \frac{n-1}{n} \frac{27}{4} w_1\,w_2 \left[1 - \frac{1}{n}\left(9\,w_1 w_2/2 + 3\,w_1^2 + 3\,w_2^2\right)\right]^{n-2} dw_1 dw_2,$$
$$\approx \int_0^{\infty} \int_0^{\infty} \frac{16}{3} \frac{27}{4} \exp\left(-9\,w_1 w_2/2 - 3\,w_1^2 - 3\,w_2^2\right)\, dw_1 dw_2 \approx .4126,$$

which is obtained by numerical integration.

Furthermore, to find $P\left(E_n^{1,2} \cap E_n^{1,3}\right)$, we need the asymptotically accurate joint pdf of $\widetilde{Q}_1(n)$, $\widetilde{Q}_2(n)$, $\widetilde{Q}_3(n)$. Given $\widetilde{Q}_j(n) = \widetilde{q}_j = (x_j, y_j)$ for $j \in \{1,2,3\}$, let $\varepsilon > 0$ be small enough such that $T(\widetilde{q}_j) \subsetneq R_{CM}(\mathrm{y}_j)$, for $j \in \{1,2,3\}$ where $T(\widetilde{q}_3)$ is the triangle with vertices $\left(\left(\sqrt{3} - 3\,y_3\right)/\sqrt{3}, y_3\right)$, $\left(1/2, \sqrt{3}/6\right)$, $\left(\sqrt{3}\,y_3, y_3\right)$. Then the asymptotically accurate joint pdf of $\widetilde{Q}_1(n)$, $\widetilde{Q}_2(n)$, $\widetilde{Q}_3(n)$ is

$$f_{123}\left(\vec{\zeta}\right) = \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} \left(\frac{A(T(\mathcal{Y})) - A(T(\widetilde{q}_1)) - A(T(\widetilde{q}_2)) - A(T(\widetilde{q}_3))}{A(T(\mathcal{Y}))}\right)^{n-3}$$

where $\vec{\zeta} = (x_1, y_1, x_2, y_2, x_3, y_3)$ and $A(T(\widetilde{q}_3)) = \frac{\sqrt{3}}{36}\left(-\sqrt{3} + 6\,y_3\right)^2$ with domain $D_I = \Big\{(x_1, y_1) \in R_{CM}(\mathrm{y}_1) : \; y_1 \geq -\sqrt{3}/3 + \sqrt{3}\,x_1 + \sqrt{3}\,\varepsilon, \; (x_2, y_2) \in R_{CM}(\mathrm{y}_2) : \; y_2 \geq -\sqrt{3}/3 + \sqrt{3}\,x_2 - \sqrt{3}\,\varepsilon, \; (x_3, y_3) \in R_{CM}(\mathrm{y}_3) : \; y_3 \leq \sqrt{3}/6 + \varepsilon\Big\}$. See Figure A.2.10.

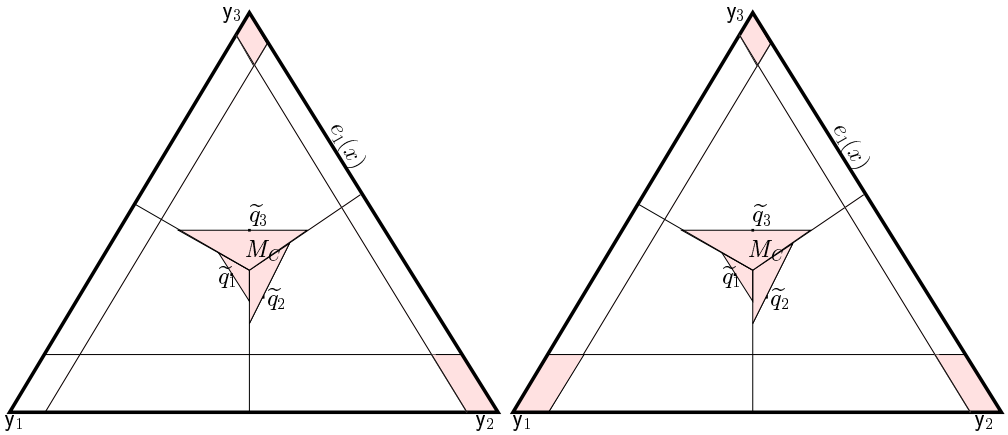


Figure A.2.10: A figure for the description of the pdf of $\tilde{Q}_1(n), \tilde{Q}_2(n), \tilde{Q}_3(n)$ and the event $E_n^{1,2} \cap E_n^{1,3}$. The unshaded region is $N_{PE}^{3/2}(\tilde{q}_1) \cup [N_{PE}^{3/2}(\tilde{q}_2) \cap N_{PE}^{3/2}(\tilde{q}_2)] \setminus [T(\tilde{q}_1) \cup T(\tilde{q}_2)]$ (left). A figure for the description of the event $E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}$. The unshaded region is $([N_{PE}^{3/2}(\tilde{q}_1) \cup N_{PE}^{3/2}(\tilde{q}_2)] \cap [N_{PE}^{3/2}(\tilde{q}_1) \cup N_{PE}^{3/2}(\tilde{q}_3)] \cap [N_{PE}^{3/2}(\tilde{q}_2) \cup N_{PE}^{3/2}(\tilde{q}_3)]) \setminus [T(\tilde{q}_1) \cup T(\tilde{q}_2) \cup T(\tilde{q}_3)]$ (right).

Note that

$$P\left(E_n^{1,2} \cap E_n^{1,3}\right) = P\left(\mathcal{X}_n \subset N_{PE}^{3/2}\left(\widetilde{Q}_1(n), M_C\right) \cup \left[N_{PE}^{3/2}\left(\widetilde{Q}_2(n), M_C\right) \cap N_{PE}^{3/2}\left(\widetilde{Q}_3(n), M_C\right)\right]\right),$$

then for sufficiently large n

$$P\left(E_n^{1,2} \cap E_n^{1,3}\right) \approx \int_{R_I} \left(\frac{A\left(N_{PE}^{3/2}(\widetilde{q}_1, M_C) \cup \left[N_{PE}^{3/2}(\widetilde{q}_2, M_C) \cap N_{PE}^{3/2}(\widetilde{q}_3, M_C)\right]\right)}{A(T(\mathcal{Y})) - A(T(\widetilde{q}_1)) - A(T(\widetilde{q}_2)) - A(T(\widetilde{q}_3))} \right)^{n-3} f_{123}(\vec{\zeta})\, d\vec{\zeta}$$

where

$$A\left(N_{PE}^{3/2}(\widetilde{q}_1, M_C) \cup \left[N_{PE}^{3/2}(\widetilde{q}_2, M_C) \cap N_{PE}^{3/2}(\widetilde{q}_3, M_C)\right]\right) = -\frac{\sqrt{3}}{8}\Big(-6 - 2\sqrt{3}\,y_2 + 6\,x_2 +$$
$$4\sqrt{3}\,y_3 + 6\,x_1 - 6\sqrt{3}\,x_1\,y_3 - 6\,y_1\,y_3 + 3\sqrt{3}\,x_1\,y_2 - 3\sqrt{3}\,x_2\,y_1 - 9\,x_1\,x_2 + 3\,y_1\,y_2 + 2\sqrt{3}\,y_1\Big).$$

The integrand simplifies to

$$\frac{n(n-1)(n-2)}{A(T(\mathcal{Y}))^3} \left(\frac{A\left(N_{PE}^{3/2}(\widetilde{q}_1, M_C) \cup \left[N_{PE}^{3/2}(\widetilde{q}_2, M_C) \cap N_{PE}^{3/2}(\widetilde{q}_3, M_C)\right]\right)}{A(T(\mathcal{Y}))} \right)^{n-3}.$$

Let

$$G(\vec{\zeta}) = \frac{A\left(N_{PE}^{3/2}(\widetilde{q}_1, M_C) \cup \left[N_{PE}^{3/2}(\widetilde{q}_2, M_C) \cap N_{PE}^{3/2}(\widetilde{q}_3, M_C)\right]\right)}{A(T(\mathcal{Y}))}.$$

The integral is critical at $x_1 = x_2 = x_3 = 1/2$ and $y_1 = y_2 = y_3 = \sqrt{3}/6$, i.e., when $\widetilde{q}_1 = \widetilde{q}_2 = \widetilde{q}_3 = M_C$ since $G(\vec{\zeta}) = 1$. Next we make the change of variables $x_1 = 1 - y_1/\sqrt{3} - 1/3 - z_1$ and $x_2 = y_2/\sqrt{3} + 1/3 + z_2$ and $y_3 = \sqrt{3}/6 + z_3$, then $G(\vec{\zeta})$ becomes

$$G(z_1, z_2, z_3) = 1 - 3\,z_1^2 - 3\,z_2^2 - 4\,z_3^2 - 9\,z_1\,z_2/2 - 3\sqrt{3}\,z_1\,z_3,$$

which does not depend on y_1, y_2, and x_3.

Hence for sufficiently large n the integral becomes

$$P\left(E_n^{1,2} \cap E_n^{1,3}\right) \approx \int \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} G(z_1, z_2, z_3)^{n-3} dy_1 dy_2 dx_3 dz_1 dz_2 dz_3,$$

where the limits of integration for y_1 are $\sqrt{3}/6 + \sqrt{3}\,z_1/2$ and $\sqrt{3}/6 - \sqrt{3}\,z_1$; for y_2 are $\sqrt{3}/6 + \sqrt{3}\,z_2/2$ and $\sqrt{3}/6 - \sqrt{3}\,z_2$; and for x_3 are $\left(1/2 + \sqrt{3}\,z_3\right)$ and $\left(-1/2 + \sqrt{3}\,z_3\right)$, so integrating over y_1, y_2, and x_3 we get

$$P\left(E_n^{1,2} \cap E_n^{1,3}\right) \approx \int_0^{\varepsilon}\int_0^{\varepsilon}\int_0^{\varepsilon} \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} \left(\frac{3\sqrt{3}}{2} z_1\right)\left(\frac{3\sqrt{3}}{2} z_2\right) 2\sqrt{3}\,z_3 \; G(z_1, z_2, z_3)^{n-3} dz_1 dz_2 dz_3.$$

The integral is critical when $z_1 = z_2 = z_3 = 0$, since $G(0,0,0) = 1$. Hence let $z_j = w_j/\sqrt{n}$ for $j \in \{1,2,3\}$, the integral becomes

$$\begin{aligned}
&P\left(E_n^{1,2} \cap E_n^{1,3}\right) = \int_0^{\sqrt{n}\,\varepsilon}\int_0^{\sqrt{n}\,\varepsilon}\int_0^{\sqrt{n}\,\varepsilon} \frac{1}{A(T(\mathcal{Y}))^3} \frac{(n-1)(n-2)}{n^2} 288 w_1\,w_2\,w_3 \\
&\left[1 - \frac{1}{n}\left(3\,w_2^2 + 3\,w_1^2 + 4\,w_3^2 + 9\,w_1\,w_2/2 + 3\sqrt{3}\,w_1\,w_3\right)\right]^{n-2} dw_1 dw_2 dw_3, \text{ letting } n \to \infty \\
&\approx \int_0^{\infty}\int_0^{\infty}\int_0^{\infty} 288 \exp\left(-3\,w_2^2 - 3\,w_1^2 - 4\,w_3^2 - 9\,w_1\,w_2/2 - 3\sqrt{3}\,w_1\,w_3\right) dw_1 dw_2 dw_3 \\
&\approx .20087,
\end{aligned}$$

which is obtained by numerical integration.

Next, we find $P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right)$. Note that

$$P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right) = P\Bigg(\mathcal{X}_n \subset \left[N_{PE}^{3/2}\left(\widetilde{Q}_1(n), M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_2(n), M_C\right)\right]$$

$$\cap \left[N_{PE}^{3/2}\left(\widetilde{Q}_1(n), M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_3(n), M_C\right)\right] \cap \left[N_{PE}^{3/2}\left(\widetilde{Q}_2(n), M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{Q}_3(n), M_C\right)\right]\Bigg)$$

So, for sufficiently large n

$$P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right) \approx \int \Bigg(A\Big(\left[N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right] \cap \left[N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right)\right]$$

$$\cap \left[N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right)\right]\Big) \Big/ \Big(A(T(\mathcal{Y})) - A(T(\widetilde{q}_1)) - A(T(\widetilde{q}_2)) - A(T(\widetilde{q}_3))\Big)\Bigg)^{n-3} f_{123}\left(\vec{\zeta}\right) d\vec{\zeta}$$

where

$$A\Bigg(\left[N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup \ N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right] \cap \left[N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right)\right] \cap$$

$$\left[N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right]\Bigg) = \frac{\sqrt{3}}{8}\Big(-3\sqrt{3}\,x_1\,y_2 + 3\sqrt{3}\,x_2\,y_1 + 5 - 6\sqrt{3}\,x_2\,y_3 +$$

$$6\,y_2\,y_3 + 6\,y_1\,y_3 + \sqrt{3}\,y_2 + 9\,x_1\,x_2 - 3\,y_1\,y_2 - 3\,x_2 - 2\sqrt{3}\,y_3 - 6\,x_1 + 6\sqrt{3}\,x_1\,y_3 - 2\sqrt{3}\,y_1\Big).$$

The integrand simplifies to

$$\frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3}\Bigg(\Big[A(N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\Big] \cap \Big[N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup$$

$$N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right)\Big] \cap \Big[N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right))\Big] \Big/ A(T(\mathcal{Y}))\Bigg)^{n-3}.$$

Let

$$G(\vec{\zeta}) = \Big[A\Big(\left[N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right)\right] \cap \left[N_{PE}^{3/2}\left(\widetilde{q}_1, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right)\right] \cap$$

$$\left[N_{PE}^{3/2}\left(\widetilde{q}_2, M_C\right) \cup N_{PE}^{3/2}\left(\widetilde{q}_3, M_C\right)\right]\Big)\Big] \Big/ A(T(\mathcal{Y})).$$

The integral is critical at $x_1 = x_2 = x_3 = 1/2$ and $y_1 = y_2 = y_3 = \sqrt{3}/6$, since $G(\vec{\zeta}) = 1$. So, we make the change of variables $x_1 = 1 - y_1/\sqrt{3} - 1/3 - z_1$ and $x_2 = y_2/\sqrt{3} + 1/3 + z_2$ and

$y_3 = \sqrt{3}/6 + z_3$, then $G(\vec{\zeta})$ becomes

$$G(z_1, z_2, z_3) = 1 - 3\,z_1^2 - 3\,z_2^2 - 4\,z_3^2 - 9\,z_1\,z_2/2 - 3\,\sqrt{3}\,z_1\,z_3 - 3\,\sqrt{3}\,z_2\,z_3.$$

Hence for sufficiently large n the integral becomes

$$P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right) \approx \int \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} G(z_1, z_2, z_3)^{n-3} dy_1 dy_2 dx_3 dz_1 dz_2 dz_3,$$

where the limits of integration for y_1, y_2, and x_3 are as above. So integrating over y_1, y_2, and x_3 we get

$$P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right) \approx \int_0^{\varepsilon}\int_0^{\varepsilon}\int_0^{\varepsilon} \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} \left(3\sqrt{3}\,z_1/2\right)\left(3\sqrt{3}\,z_2/2\right) \left(2\,\sqrt{3}\,z_3\right) G(z_1, z_2, z_3)^{n-3} dz_1 dz_2 dz_3.$$

The integral is critical when $z_1 = z_2 = z_3 = 0$, since $G(0,0,0) = 1$. Hence let $z_j = w_j/\sqrt{n}$ for $j \in \{1,2,3\}$, the integral becomes

$$P\left(E_n^{1,2} \cap E_n^{1,3} \cap E_n^{2,3}\right) \approx \int_0^{\sqrt{n}\,\varepsilon}\int_0^{\sqrt{n}\,\varepsilon}\int_0^{\sqrt{n}\,\varepsilon} \frac{1}{A(T(\mathcal{Y}))^3} \frac{(n-1)(n-2)}{n^2} 288 w_1\,w_2\,w_3$$
$$\left[1 - \frac{1}{n}\left(3\,w_1^2 + 3\,w_2^2 + 4\,w_3^2 + 9\,w_1\,w_2/2 + 3\,\sqrt{3}\,w_1\,w_3 + 3\,\sqrt{3}\,w_2\,w_3\right)\right]^{n-2} dw_1 dw_2 dw_3,$$

letting $n \to \infty$,

$$\approx \int_0^{\infty}\int_0^{\infty}\int_0^{\infty} 288 \exp\left(-3\,(w_1^2 + w_2^2) - 4\,w_3^2 - 9\,w_1\,w_2/2 - 3\,\sqrt{3}\,(w_1\,w_3 + w_2\,w_3)\right) dw_1 dw_2 dw_3$$

$$\approx .1062,$$

which is obtained by numerical integration.

Hence we get

$$\lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) = 2\right) \approx .7413, \text{ and } \lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right) = 3\right) \approx .2587. \blacksquare$$

A.2.8 Proof of Theorem 4.7.4

Let $M = (m_1, m_2) \in T(\mathcal{Y})^o$. Recall that $\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) > 1$ iff $\mathcal{X}_n \cap \Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) = \emptyset$. The Γ_1-region is determined by the (closest) edge extrema, i.e., by $X_{e_j}(n) \in \operatorname{argmin}_{X \in \mathcal{X}_n} d(X, e_j)$

for $j \in \{1,2,3\}$. Note that $X_{e_j}(n)$ is unique a.s. for each j and $n \geq 3$.

Let $E_1(n,\varepsilon)$ be the event such that $X_{e_1}(n)$ is in the strip $\{(x,y) \in T(\mathcal{Y}) : \ 0 \leq y \leq \varepsilon\}$; $X_{e_2}(n)$ is in the strip $\{(x,y) \in T(\mathcal{Y}) : \sqrt{3}\,x - \varepsilon \leq y \leq \sqrt{3}\,x\}$; and $X_{e_3}(n)$ is in the strip $\{(x,y) \in T(\mathcal{Y}) : \sqrt{3}\,(1-x) - \varepsilon \leq y \leq \sqrt{3}\,(1-x)\}$ (that is, the edge extrema are in ε strips around the edges) and $X_{e_j}(n)$ are distinct. Note that $P(E_1(n,\varepsilon)) \to 1$ as $n \to \infty$. See Figure A.2.11.

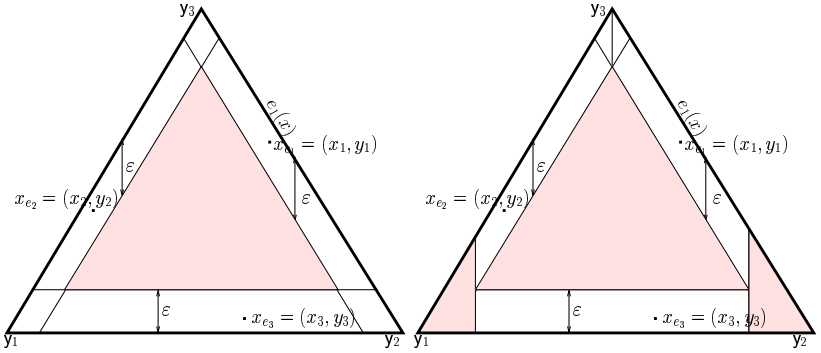


Figure A.2.11: A figure for the description of the event $E_1(n,\varepsilon)$ (left) and $E_2(n,\varepsilon)$ (right). The unshaded regions are the corresponding ε-strips around the edges given $X_{e_j}(n) = x_{e_j} = (x_j, y_j)$ for $j \in \{1,2,3\}$.

Let $X_{e_j}(n) = x_{e_j} = (x_j, y_j)$ be given and let $T(\vec{\zeta})$ be the triangle formed by the lines at x_{e_j} parallel to e_j for $j \in \{1,2,3\}$ where $\vec{\zeta} = (x_1, y_1, x_2, y_2, x_3, y_3)$. See Figure A.2.11 (left).

Then the asymptotically accurate joint pdf of $X_{e_j}(n)$ is

$$\begin{aligned} f_3(\vec{\zeta}) &= \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} \left(\frac{A(T(\vec{\zeta}))}{A(T(\mathcal{Y}))} \right)^{n-3} \qquad \text{(A.2.2)} \\ &= n\,(n-1)\,(n-2) \left(\sqrt{3}/12 \left(-2\,y_1 + y_3 - \sqrt{3}\,x_3 + y_2 + 3\,x_2 \right)^2 \right)^{n-3} \Big/ \left(\sqrt{3}/4 \right)^n . \end{aligned}$$

with the support $D_S = \{x_{e_j} = (x_j, y_j),\ j \in \{1,2,3\} :\ d(x_{e_j}, e_j) < \varepsilon\}$.

Given $X_{e_j}(n) = (x_j, y_j)$,

$$P\left(\mathcal{X}_n \cap \Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M \right) = \emptyset \right) = \left(\frac{A(T(\vec{\zeta})) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M \right) \right)}{A(T(\vec{\zeta}))} \right)^{n-3}$$

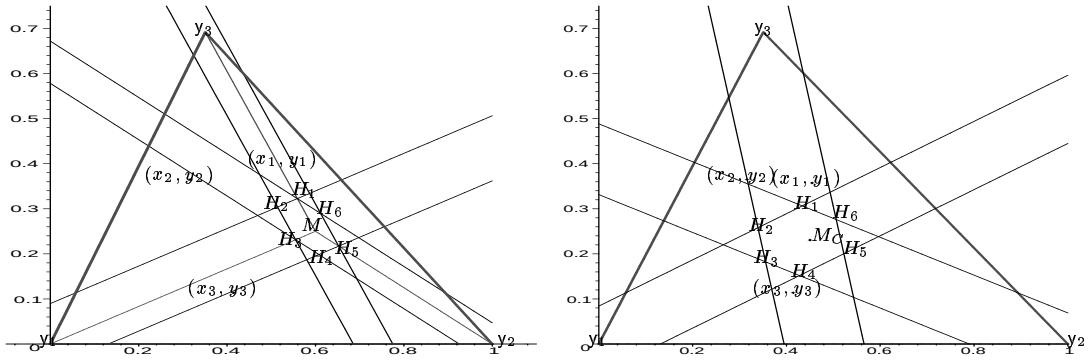


Figure A.2.12: The Γ_1-region for $N_{CS}^{\tau=1}(\cdot, M)$(left) and $N_{CS}^{\tau=1}(\cdot, M_C)$ (right) in $T(\mathcal{Y})$.

where $\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)$ is the hexagon with area

$$A\left(\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)\right) = m_2\,x_2 + y_2\,m_1\,y_3/\sqrt{3} - x_2\,y_3\,m_2/\sqrt{3} - y_3\,m_1\,x_2 + y_2\,y_3\,m_2/3 - m_2\,y_3/\sqrt{3}$$
$$- m_2\,y_2/\sqrt{3} + y_3\,m_2\,x_1/\sqrt{3} - y_3\,x_1\,m_1 + y_3\,y_1\,m_2/3 - y_3\,y_1\,m_1/\sqrt{3} - m_2\,y_1\,x_2/\sqrt{3} +$$
$$m_2\,y_1\,y_2/3 + m_2\,x_1\,y_2/\sqrt{3} + x_2\,y_3 + y_3\,m_1 - y_2\,y_3/\sqrt{3} - m_2\,x_1\,x_2$$

and

$$A(T(\vec{\zeta})) = \frac{\sqrt{3}\,\left[(-2\,y_3 + y_2 + y_1) + \sqrt{3}\,(x_1 - x_2)\right]^2}{12}.$$

See Figure A.2.12 (left) for a realization of $\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)$.

Then

$$P\left(\mathcal{X}_n \cap \Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) = \emptyset\right) = \int \left(\frac{A(T(\vec{\zeta})) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)\right)}{A(T(\vec{\zeta}))}\right)^{n-3} f_3(\vec{\zeta})\,d\vec{\zeta}$$
$$= \int \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} \left(\frac{A(T(\vec{\zeta})) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)\right)}{A(T(\mathcal{Y}))}\right)^{n-3} d\vec{\zeta},$$

where $\vec{\zeta} = (x_1, y_1, x_2, y_2, x_3, y_3)$.

Let

$$G(\vec{\zeta}) = \frac{A(T(\vec{\zeta})) - A\left(\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)\right)}{A(T(\mathcal{Y}))}.$$

The integral is critical when $(x_j, y_j) \in e_j$, i.e., when $y_3 = 0$, $y_1 = \sqrt{3}\,(1 - x_1)$, and $y_2 = \sqrt{3}\,x_2$, since $G(\vec{\zeta}) = 1$. So we make the change of variables $y_3 = z_3$, $y_1 = \sqrt{3}\,(1 - x_1) - z_1$, and

$y_2 = \sqrt{3}\, x_2 - z_2$. Hence $G(\vec{\zeta})$ becomes

$$\begin{aligned} G(x_1, x_2, x_3, z_1, z_2, z_3) &= \sqrt{3}\left(3\sqrt{3} - 6\,(z_2 + 2\,z_3) + 4\sqrt{3}\,m_1\,(z_2\,z_3 - z_1\,z_3) + 2\sqrt{3}\,(z_1\,z_2 + 2\,z_1\,z_3)\right. \\ &\quad \left. + \sqrt{3}\,(4\,z_3^2 + z_2^2 + z_1^2) + 4\,m_2\,(z_2\,z_3 + z_1\,z_3 - z_1\,z_2)\right)/9 \\ &= 1 - 2\,(z_1 + z_2 + 2\,z_3)/\sqrt{3} + \text{higher order terms in } z_j, \end{aligned}$$

which is independent of x_j, so we denote it as $G(z_1, z_2, z_3)$. So for sufficienlty large n and sufficiently small $\varepsilon > 0$, we have

$$\begin{aligned} &P\left(\mathcal{X}_n \cap \Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) = \emptyset\right) \approx \\ &\int_0^{\varepsilon}\int_0^{\varepsilon}\int_0^{\varepsilon}\int_{\varepsilon/\sqrt{3}}^{1-\varepsilon/\sqrt{3}}\int_{\varepsilon/\sqrt{3}}^{1/2}\int_{1/2}^{1-\varepsilon/\sqrt{3}} \frac{n\,(n-1)\,(n-2)}{A(T(\mathcal{Y}))^3} G(z_1, z_2, z_3)^{n-3}\, dx_3\, dx_2\, dx_1\, dz_1\, dz_2\, dz_3. \end{aligned}$$

Integrating over x_j for $j \in \{1, 2, 3\}$, yields

$$K_\varepsilon = \left(1 - 2\,\varepsilon/\sqrt{3}\right)\left(1/2 - \varepsilon/\sqrt{3}\right)\left(1/2 - \varepsilon/\sqrt{3}\right).$$

Then for sufficiently large n

$$P\left(\mathcal{X}_n \cap \Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) = \emptyset\right) \approx \int_0^{\varepsilon}\int_0^{\varepsilon}\int_0^{\varepsilon} K_\varepsilon\, \frac{64}{3\sqrt{4}}\, n\,(n-1)\,(n-3)\, G(z_1, z_2, z_3)^{n-3}\, dz_1\, dz_2\, dz_3.$$

The new integrand is critical at $(z_1, z_2, z_3) = (0, 0, 0)$, so we make the change of variables $z_j = w_j/n$ for $j \in \{1, 2, 3\}$, then

$$\begin{aligned} &P\left(\mathcal{X}_n \cap \Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) = \emptyset\right) \approx \int_0^{n\,\varepsilon}\int_0^{n\,\varepsilon}\int_0^{n\,\varepsilon} K_\varepsilon\, \frac{64}{3\sqrt{4}}\, \frac{n\,(n-1)\,(n-3)}{n^3} \\ &\left(1 - \frac{1}{n}\left(2\,(w_1 + w_2 + 2\,w_3)/\sqrt{3}\right) + O\left(n^{-2}\right)\right)^{n-3} dw_1\, dw_2\, dw_3, \\ &\text{lettting } n \to \infty \\ &\approx \int_0^{\infty}\int_0^{\infty}\int_0^{\infty} K_\varepsilon\, \frac{64}{3\sqrt{4}}\, \exp\left(-(2\,(w_1 + w_2 + 2\,w_3)/\sqrt{3}\right) dw_1\, dw_2\, dw_3 \\ &= \frac{9}{16\sqrt{3}} K_\varepsilon\, \frac{64}{3\sqrt{4}} = 4\,K_\varepsilon. \end{aligned}$$

Since $\lim_{\varepsilon \downarrow 0} K_\varepsilon = 1/4$, the desired result follows by continuity of K_ε in ε. ■

A.2.9 Proof of Theorem 4.7.6

Recall that $\gamma(\mathcal{X}_n, N_{CS}^{\tau}, M_C) = 1$ iff $\mathcal{X}_n \subset N_{CS}^{\tau}(X, M_C)$ for some $X \in \mathcal{X}_n$, or equivalently, $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{CS}^{\tau}, M_C) \neq \emptyset$. Observe that

$$P\left(X_{e_j}^f(n) \in \Gamma_1(\mathcal{X}_n, N_{CS}^{\tau}, M_C)\right) \geq P\left(X \in \Gamma_1(\mathcal{X}_n, N_{CS}^{\tau}, M_C), X \in R_{CM}(\mathsf{y}_j)\right)$$

for all $X \neq X_{e_j}^f(n)$. In fact, one can see that

$$P\left(X_{e_j}^f(n) \notin \Gamma_1(\mathcal{X}_n, N_{CS}^{\tau}, M_C), X \in \Gamma_1(\mathcal{X}_n, N_{CS}^{\tau}, M_C) \cap R_{CM}(\mathsf{y}_j)\right) \to 0 \text{ as } n \to \infty$$

for all $X \neq X_{e_j}^f(n)$. Hence for sufficiently large n, if $\gamma(\mathcal{X}_n, N_{CS}^{\tau}, M_C) = 1$ iff $\mathcal{X}_n \subset N_{CS}^{\tau}\left(X_{e_j}^f(n), M_C\right)$ for some $j \in \{1,2,3\}$ were true the conjecture above Theorem 4.7.6 would follow. Alas, proving this assertion is still an open problem. Let E_j^{τ} be the event that $\mathcal{X}_n \subset N_{CS}^{\tau}\left(X_{e_j}^f(n), M_C\right)$ for $j \in \{1,2,3\}$. By symmetry in T_e, $P\left(E_j^{\tau}\right)$ are identical for all $j \in \{1,2,3\}$. So we only consider $j = 3$.

The asymptotically accurate pdf of $X_{e_3}^f(n)$ is

$$f_3(x_3, y_3) = n\left(\frac{A(T(\mathcal{Y})) - A(S_R(x_3, y_3))}{A(T(\mathcal{Y}))}\right)^{n-1} \cdot \frac{1}{A(T(\mathcal{Y}))},$$

where $S_R(x_3, y_3)$ is the shaded region in Figure A.2.13 (left), whose area, for a given $X_{e_3}^f(n) = x_{e_3}^f = (x_3, y_3)$, is $A(S_R(x_3, y_3)) = \frac{\sqrt{3}\left(2\sqrt{3}\, y_3 - 1\right)^2}{12}$.

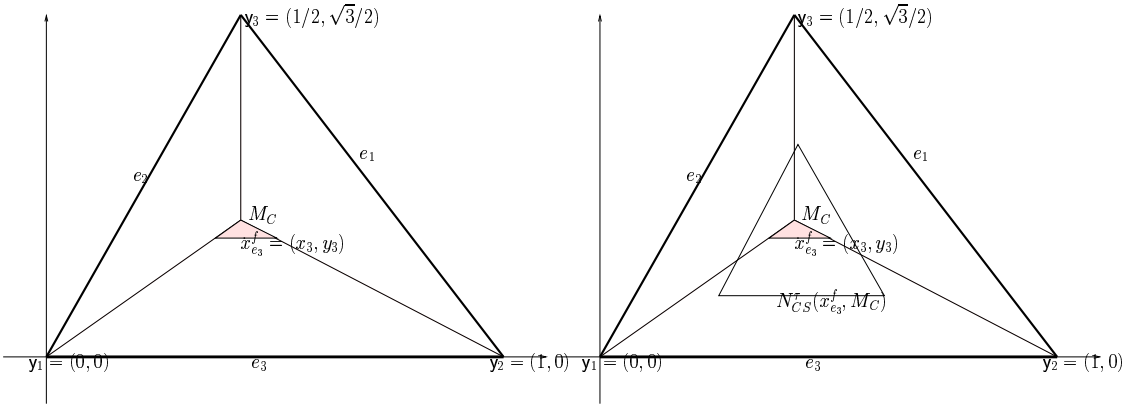


Figure A.2.13: A figure for the description of the pdf of $X_{e_3}^f(n)$ (left) and $N_{CS}^{\tau}\left(X_{e_3}^f(n), M_C\right)$ (right) given $X_{e_3}^f(n) = x_{e_3}^f = (x_3, y_3)$.

Given $X_{e_3}^f(n) = x_{e_3}^f = (x_3, y_3)$, $P(E_3^\tau) = \left(\frac{A\left(N_{CS}^\tau\left(x_{e_3}^f, M_C\right)\right) - A(S_R(x_3,y_3))}{A(T(\mathcal{Y})) - A(S_R(x_3,y_3))} \right)^{n-1}$, then

$$\begin{aligned} P(E_3^\tau) &= \int \left(\frac{A\left(N_{CS}^\tau\left(x_{e_3}^f, M_C\right)\right) - A(S_R(x_3,y_3))}{A(T(\mathcal{Y})) - A(S_R(x_3,y_3))} \right)^{n-1} f_3(x_3,y_3)\, dx_3\, dy_3 \\ &= \int n \left(\frac{A\left(N_{CS}^\tau\left(x_{e_3}^f, M_C\right)\right) - A(S_R(x_3,y_3))}{A(T(\mathcal{Y}))} \right)^{n-1} \frac{1}{A(T(\mathcal{Y}))}\, dx_3\, dy_3 \end{aligned}$$

where $A\left(N_{CS}^\tau\left(x_{e_3}^f, M_C\right)\right) = 3\sqrt{3}\,\tau^2\, y_3^2$ and we use the fact that for fixed $\tau \in (0,1]$, $S_R(x_3,y_3) \subsetneq N_{CS}^\tau\left(x_{e_3}^f, M_C\right)$ with probability 1 for sufficiently large n. See Figure A.2.13 (right) for a realization of $N_{CS}^\tau\left(X_{e_j}^f(n), M_C\right)$ given $X_{e_3}^f(n) = x_{e_3}^f = (x_3, y_3)$.

Let

$$G(x_3,y_3) = \frac{A\left(N_{CS}^\tau\left(x_{e_3}^f, M_C\right)\right) - A(S_R(x_3,y_3))}{A(T(\mathcal{Y}))} = \frac{12\,\left(\sqrt{3}\,\tau - 1\right)^2\, y_3^2 - 1}{3}$$

which is independent of x_3, so we denote it as $G(y_3)$. Then for sufficiently large n and sufficiently small $\varepsilon > 0$, we have

$$P(E_3^\tau) \approx \int_{1-\sqrt{3}\,y_3}^{\sqrt{3}\,y_3} \int_{\frac{\sqrt{3}}{6}-\varepsilon}^{\frac{\sqrt{3}}{6}} n\, G(y_3)^{n-1} \frac{1}{A(T(\mathcal{Y}))}\, dx_3\, dy_3 = \int_{\frac{\sqrt{3}}{6}-\varepsilon}^{\frac{\sqrt{3}}{6}} \frac{\left(1 - 2\sqrt{3}\,y_3\right)\, n}{A(T(\mathcal{Y}))}\, G(y_3)^{n-1}\, dy_3.$$

The integrand is maximized when $y_3 = \sqrt{3}/6$, i.e., when $(x_3, y_3) = M_C$, since $G(y_3) = \tau^2$. So we make the change of variable $y_3 = \sqrt{3}/6 - z_3$, and $G(y_3)$ becomes

$$G(z_3) = \tau^2\, \left(1 - 4\sqrt{3}\, z_3 + 12\, z_3^2\right) - 4\, z_3^2.$$

Then we have

$$P\left(\mathcal{X}_n \subset N_{CS}^\tau\left(X_{e_j}^f(n), M_C\right)\right) \approx \int_0^\varepsilon \left(2\sqrt{3}\, z_3\right)\, n\, G(z_3)^{n-1}\, \frac{1}{A(T(\mathcal{Y}))}\, dz_3.$$

The new integrand is maximized when $z_3 = 0$. So we make the change of variable $z_3 = w_3/n$,

and for sufficiently large n and sufficiently small $\varepsilon > 0$, we have

$$\begin{aligned} P\left(E_3^\tau\right) &\approx \int_0^{n\,\varepsilon} \left(\frac{2\sqrt{3}\,w_3}{n}\right) \tau^{2(n-1)} \left(1 - \frac{1}{n}\,4\sqrt{3}\,w_3 + O\left(n^{-2}\right)\right)^{n-1} \frac{1}{A(T(\mathcal{Y}))}\,dw_3 \\ &\approx O\left(\tau^{2(n-1)}\,n^{-1}\right) \int_0^{\infty} \left(2\sqrt{3}\,w_3\right) \exp\left(-4\sqrt{3}\,w_3\right) \frac{1}{A(T(\mathcal{Y}))}\,dw_3 = O\left(\tau^{2(n-1)}\,n^{-1}\right) \end{aligned}$$

since $\int_0^{\infty} \left(2\sqrt{3}\,w_3\right) \exp\left(-4\sqrt{3}\,w_3\right) \frac{1}{A(T(\mathcal{Y}))}\,dw_3 = 1/6$. Hence the desired result follows. ■

A.2.10 Proof of Theorem 4.7.7

For sufficiently large n, if

$$\gamma\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C\right) \leq 2 \text{ iff } \mathcal{X}_n \subset N_{CS}^{\tau=1}\left(X_{e_i}^f, M_C\right) \cup N_{CS}^{\tau=1}\left(X_{e_j}^f(n), M_C\right) \tag{A.2.3}$$

for some $(i,j) \in \left\{(1,2),(1,3),(2,3)\right\}$ were true then the conjecture above Theorem 4.7.7 would follow. But proving the expression in A.2.3 is still an open problem. Let E_j^τ be defined as in the proof of Theorem 4.7.6. By symmetry in T_e, $P\left(E_i^{\tau=1} \cup E_j^{\tau=1}\right)$ are identical for all $(i,j) \in \left\{(1,2),(1,3),(2,3)\right\}$. So we consider $(i,j) = (1,3)$.

The joint density of $X_{e_1}^f(n), X_{e_3}^f(n)$ is

$$f_{13}\left(\vec{\zeta}\right) = n\,(n-1) \left(\frac{A(T(\mathcal{Y})) - A\left(S_R^2(\vec{\zeta})\right)}{A(T(\mathcal{Y}))}\right)^{n-2} \frac{1}{A(T(\mathcal{Y}))^2}$$

where $S_R^2(\vec{\zeta})$ is the shaded region in Figure A.2.14 (left) for given $X_{e_j}^f(n) = x_{e_1}^f = (x_j, y_j)$ for $j \in \{1,3\}$ with area

$$A\left(S_R^2(\vec{\zeta})\right) = \sqrt{3}\left(\left(3\,x_1 + \sqrt{3}\,y_1 - 2\right)^2 + \left(2\sqrt{3}\,y_3 - 1\right)^2\right) \Big/ 12.$$

Given $X_{e_j}^f(n) = x_{e_1}^f = (x_j, y_j)$ for $j \in \{1,3\}$,

$$P\left(E_1^{\tau=1} \cup E_3^{\tau=1}\right) = \left(\frac{A\left(N_{CS}^{\tau=1}\left(x_{e_1}^f, M_C\right) \cup N_{CS}^{\tau=1}\left(x_{e_3}^f, M_C\right)\right) - A\left(S_R^2(\vec{\zeta})\right)}{A(T(\mathcal{Y})) - A\left(S_R^2(\vec{\zeta})\right)}\right)^{n-2}$$

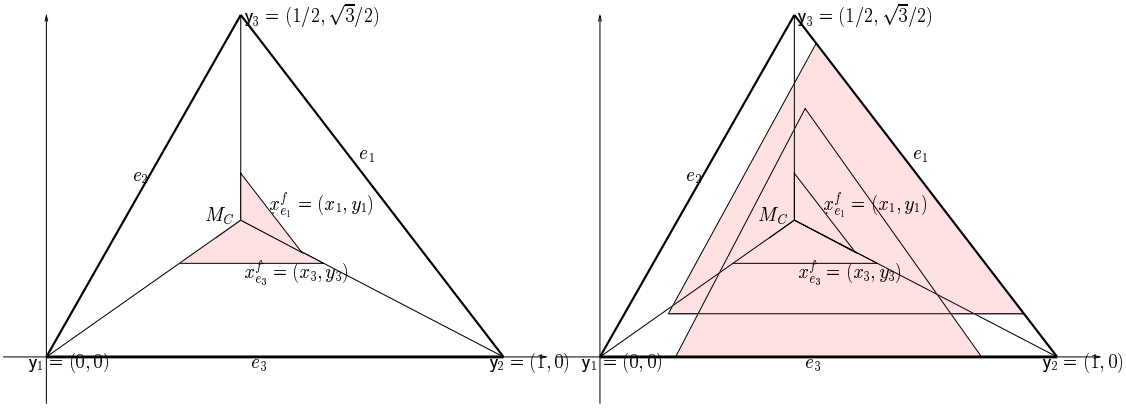


Figure A.2.14: A figure for the description of the joint pdf of $X^f_{e_1}(n), X^f_{e_3}(n)$ (left), and $N^\tau_{CS}\left(X^f_{e_1}(n), M_C\right) \cup N^\tau_{CS}\left(X^f_{e_3}(n), M_C\right)$ (right) given $X^f_{e_j}(n) = x^f_{e_1} = (x_j, y_j)$ for $j \in \{1, 3\}$.

where

$$A\left(N^{\tau=1}_{CS}\left(x^f_{e_1}, M_C\right) \cup N^{\tau=1}_{CS}\left(x^f_{e_3}, M_C\right)\right) = \sqrt{3}\left(2\,x_1^2 - 4\,x_1 + \sqrt{3}\,(y_3 + y_1)\,(x_1 - 1) + 3\,y_1\,y_3\right).$$

See Figure A.2.14 (right) for a realization of $N^{\tau=1}_{CS}\left(X^f_{e_1}(n), M_C\right) \cup N^{\tau=1}_{CS}\left(X^f_{e_3}(n), M_C\right)$. The configuration in Figure A.2.14 (right) can be assumed without loss of generality, since by symmetry it is equally likely to have the vertex of $N^{\tau=1}_{CS}\left(X^f_{e_1}(n), M_C\right)$ that is not on edge e_1 inside $N^{\tau=1}_{CS}\left(X^f_{e_3}(n), M_C\right)$.

Then

$$P\left(E_1^{\tau=1} \cup E_3^{\tau=1}\right) = \int \left(\frac{A\left(N^{\tau=1}_{CS}\left(x^f_{e_1}, M_C\right) \cup N^{\tau=1}_{CS}\left(x^f_{e_3}, M_C\right)\right) - A\left(S^2_R(\vec{\zeta})\right)}{A(T(\mathcal{Y})) - A\left(S^2_R(\vec{\zeta})\right)}\right)^{n-2} f_{13}\left(\vec{\zeta}\right) d\vec{\zeta}$$

$$= \int \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\left(\frac{A\left(N^{\tau=1}_{CS}\left(x^f_{e_1}, M_C\right) \cup N^{\tau=1}_{CS}\left(x^f_{e_3}, M_C\right)\right) - A\left(S^2_R(\vec{\zeta})\right)}{A(T(\mathcal{Y}))}\right)^{n-2} d\vec{\zeta}.$$

Let

$$G(\vec{\zeta}) = \frac{A\left(N^{\tau=1}_{CS}\left(x^f_{e_1}, M_C\right) \cup N^{\tau=1}_{CS}\left(x^f_{e_3}, M_C\right)\right) - A\left(S^2_R(\vec{\zeta})\right)}{A(T(\mathcal{Y}))}.$$

The integral is critical at $(x_1, y_1) = (x_3, y_3) = M_C$, since $G(\vec{\zeta}) = 1$. So we make the change of

variables $y_3 = \sqrt{3}/6 - z_3$ and $(x_1, y_1) \to \left(1/2 + r\ \cos(\theta), \sqrt{3}/6 + r\ \sin(\theta)\right)$. Then $G\left(\vec{\zeta}\right)$ becomes

$$G(r, z_3, \theta) = 1 - 4\,r\ \cos(t) - 4\,z_3^2 + 5\,r^2\ \cos(t)^2 - 4\,\sqrt{3}\,r\ \cos(t)\,z_3 - r^2\ \sin(t)^2 - 12\,z_3\,r\ \sin(t) + 2\,\sqrt{3}\,r^2\ \sin(t)\ \cos(t)$$

which is independent of x_3. The new integrand is critical at $r = z_3 = 0$, since $G(r, z_3, \theta) = 1$. So for sufficiently large n and sufficiently small $\varepsilon > 0$, we have

$$\begin{aligned} P\left(E_1^{\tau=1} \cup E_3^{\tau=1}\right) &\approx \int_{1/2-\sqrt{3}\,z_3}^{1/2+\sqrt{3}\,z_3} \int_0^{\varepsilon} \int_{-\pi/6}^{\pi/2} \int_0^{\varepsilon} \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\, G(r, \theta, z_3)^{n-2}\, r\, dr\, d\theta\, dz_3\, dx_3 \\ &= \int_0^{\varepsilon} \int_{-\pi/6}^{\pi/2} \int_0^{\varepsilon} 2\,\sqrt{3}\,z_3\, \frac{n\,(n-1)}{A(T(\mathcal{Y}))^2}\, G(r, \theta, z_3)^{n-2}\, r\, dr\, d\theta\, dz_3. \end{aligned}$$

Next, we make the change of variables $r = w/\sqrt{n}$, $z_3 = w_3/\sqrt{n}$, then we get

$$\begin{aligned} &P\left(E_1^{\tau=1} \cup E_3^{\tau=1}\right) \approx \int_0^{\sqrt{n}\,\varepsilon} \int_{-\pi/6}^{\pi/2} \int_0^{\sqrt{n}\,\varepsilon} 2\,\sqrt{3}\,w_3\, \frac{n\,(n-1)}{n^2} \\ &\left(1 - \frac{1}{\sqrt{n}}\,4\,w\ \cos(t) + O\left(n^{-3/2}\right)\right)^{n-2} \frac{1}{A(T(\mathcal{Y}))^2}\, w\, dw\, d\theta\, dw_3, \\ &\approx \int_0^{\infty} \int_{-\pi/6}^{\pi/2} \int_0^{\infty} 2\,\sqrt{3}\,w_3 \left[1 - \frac{1}{\sqrt{n}}\,4\,w\ \cos(t) + O\left(n^{-3/2}\right)\right]^{n-2} \frac{1}{A(T(\mathcal{Y}))^2}\, w\, dw\, d\theta\, dw_3 \to 0 \end{aligned}$$

as $n \to \infty$ since $\left(1 - \frac{1}{\sqrt{n}}\,4\,w\,cos(t) + O\left(n^{-3/2}\right)\right)^{n-2} \to 0$. Hence the desired result follows. ■

A.2.11 Proof of Theorem 4.7.8

Consider the induced digraph on the triangle $T_s := T(\mathrm{y}_1, M_3, M_C)$ based on $\mathcal{X}_n' = \mathcal{X}_m \cap T_s$ and $\mathcal{Y}$ (conditional on the event that $|\mathcal{X}_m \cap T_s| = n$). See Figure A.2.15 (left) for T_s in T_e. Let $\gamma_{ind}(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C)$ denote the domination number for the induced digraph. Then to prove Theorem 4.7.8, it will suffice to show that $P(\gamma_{ind}(\mathcal{X}_n, N_{CS}^{\tau=1}, M_C) = 1) \to 1$ as $n \to \infty$. Note that, conditional on $\mathcal{X}_n \cap T_s \neq \emptyset$, the set $\mathcal{X}_n \cap T_s$ is a random sample from $\mathcal{U}(T_s)$. So, without loss of generality, assume $\mathcal{X}_n'$ be a random sample of size n from $\mathcal{U}(T_s)$.

Consider the extremum Z_M^1 such that it is the closest data point to the line joining y_1 and M_C in $T_s \cap \left\{(x, y) \in R_{CM}(e_3) :\ y \geq \frac{\sqrt{3}\,(2-3\,x)}{9}\right\}$. See Figure A.2.15 (right). The pdf of such an

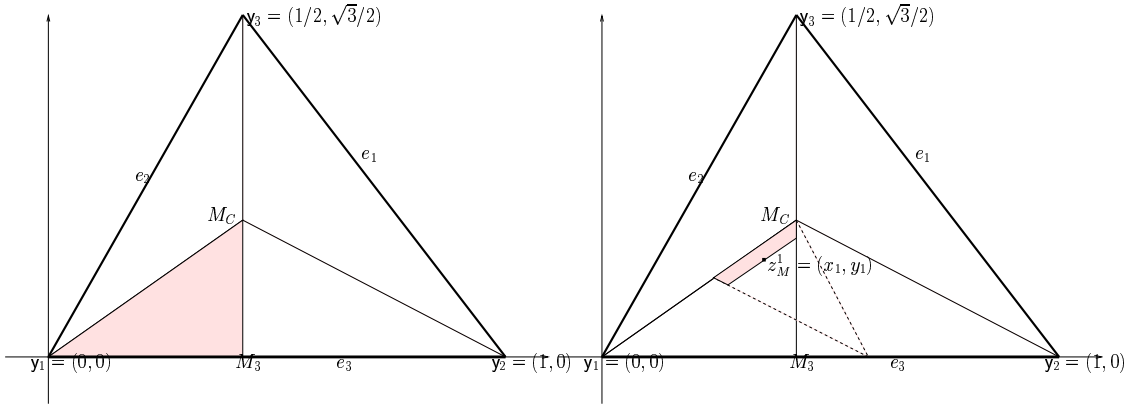


Figure A.2.15: The triangle T_s (left) and a realization of the special extremum Z_M^1 in T_s (right).

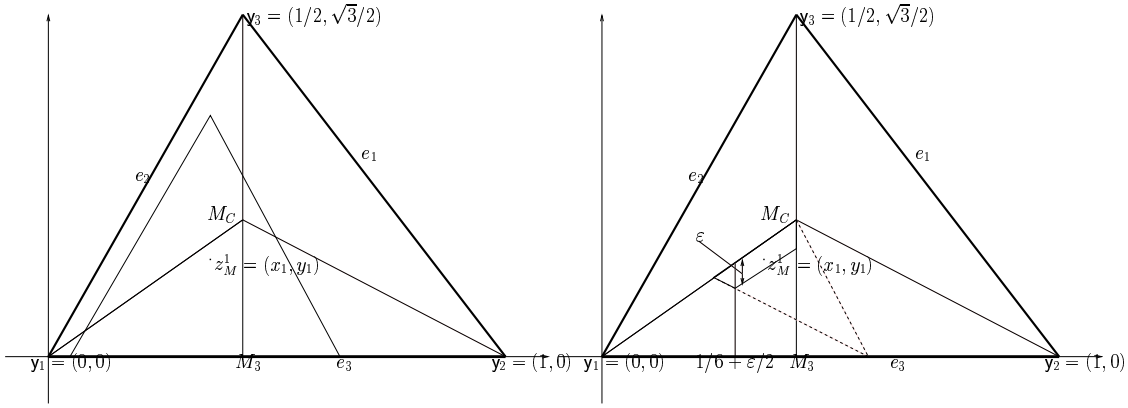


Figure A.2.16: The explanatory figure for $\widetilde{N}_{CS}^{\tau=1}(Z_M^1)$ (left) and the asymptotically accurate support for the pdf of Z_M^1 (right) given $Z_M^1 = z_M^1 = (x_1, y_1)$.

extremum point is

$$f(x_1, y_1) = n \left(\frac{A(T_s) - A(S_R(x_1, y_1))}{A(T_s)} \right)^{n-1} \frac{1}{A(T_s)}$$

where $S_R(x_1, y_1)$ is the shaded region in Figure A.2.15 (right) for a given realization of $Z_M^1 = z_M^1 = (x_1, y_1)$ whose area is $A(S_R(x_1, y_1)) = \frac{\sqrt{3}}{36} \left(\sqrt{3}\, y_1 - x_1\right) \cdot$
$\left(3\, x_3 - 3\sqrt{3}\, y_3 - 2\right)$.

Let $\widetilde{N}_{CS}^{\tau=1}(x, M_C) := N_{CS}^{\tau=1}(x, M_C) \cap T_s$ for all $x \in T_s$. Given $Z_M^1 = z_M^1 = (x_1, y_1)$,

$$P\left(\mathcal{X}_n' \subset \widetilde{N}_{CS}^{\tau=1}(Z_M^1)\right) = \left(\frac{A\left(\widetilde{N}_{CS}^{\tau=1}(z_M^1)\right) - A(S_R(x_1, y_1))}{A(T_s) - A(S_R(x_1, y_1))} \right)^{n-1}.$$

Then for sufficiently large n

$$P\left(\mathcal{X}'_n \subset \widetilde{N}_{CS}^{\tau=1}(Z_M^1)\right) \approx \int \left(\frac{A\left(\widetilde{N}_{CS}^{\tau=1}(z_M^1)\right) - A(S_R(x_1,y_1))}{A(T_s) - A(S_R(x_1,y_1))}\right)^{n-1} f(x_1,y_1)\,dy_1\,dx_1$$
$$= \int n \left(\frac{A\left(\widetilde{N}_{CS}^{\tau=1}(z_M^1)\right) - A(S_R(x_1,y_1))}{A(T_s)}\right)^{n-1} \frac{1}{A(T_s)}\,dy_1\,dx_1.$$

Let

$$G(x_1,y_1) = \frac{A\left(\widetilde{N}_{CS}^{\tau=1}(z_M^1)\right) - A(S_R(x_1,y_1))}{A(T_s)} = 1 + 4\,(y_1 - x_1)/\sqrt{3} - 4\left(x_1^2 + 3\,y_1^2 - 2\sqrt{3}\,x_1\,y_1\right)$$

where $A\left(\widetilde{N}_{CS}^{\tau=1}(z_M^1)\right) = \frac{\sqrt{3}}{24}\left(1 - 6\,\left(3 - \sqrt{3}\,y_1\right)^2\right)$. See Figure A.2.16 (left) for a realization of $\widetilde{N}_{CS}^{\tau=1}(Z_M^1)$. The integral is critical when $y_1 = x_1/\sqrt{3}$, i.e., (x_1,y_1) lies on the line joining y_1 and M_C. Then we make the change of variable $y_1 = x_1/\sqrt{3} - z_1$, then $G(x_1,y_1)$ becomes

$$G(z_1) = 1 - 4\,z_1/\sqrt{3} - 12\,z_1^2,$$

which is independent of x_1. Let ε be small enough and the asymptotically accurate support of Z_M^1 is $\left\{(x,y) \in T_s :\ 1/6 + \varepsilon/2 \le x \le 1/2;\ y \ge x/\sqrt{3} - \varepsilon\right\}$. See Figure A.2.16 (right) for this support. Then for sufficiently large n and sufficiently small $\varepsilon > 0$, we have

$$P\left(\mathcal{X}'_n \subset \widetilde{N}_{CS}^{\tau=1}(Z_M^1)\right) \approx \int_{1/6+\varepsilon/2}^{1/2}\int_0^{\varepsilon} \frac{12}{\sqrt{3}}\,n\,\left(1 - 4\,z_1/\sqrt{3} - 12\,z_1^2\right)^{n-1}\,dz_1\,dx_1$$
$$= \int_0^{\varepsilon} (1/3 - \varepsilon/2)\,\frac{12}{\sqrt{3}}\,n\,\left(1 - 4\,z_1/\sqrt{3} - 12\,z_1^2\right)^{n-1}\,dz_1.$$

The new integrand is critical at $z_1 = 0$, since $G(z_1) = 1$, so we let $z_1 = w_1/n$, then

$$P\left(\mathcal{X}'_n \subset \widetilde{N}_{CS}^{\tau=1}(Z_M^1)\right) \approx \int_0^{n\,\varepsilon} (1/3 - \varepsilon/2)\,\frac{12}{\sqrt{3}}\,\left(1 - \frac{1}{n}\,4\,w_1/\sqrt{3} + O\left(n^{-2}\right)\right)^{n-1}\,dw_1$$
$$\text{letting } n \to \infty, \quad \approx \int_0^{\infty} (1/3 - \varepsilon/2)\,\frac{12}{\sqrt{3}}\,\exp\left(-4\,w_1/\sqrt{3}\right)\,dw_1$$
$$= (1/3 - \varepsilon/2)\,\frac{12}{\sqrt{3}}\,\frac{\sqrt{3}}{4}$$

which converges to 1 as $\varepsilon \downarrow 0$. Hence the desired result follows by continuity in ε. ■

Appendix B

Derivations of Mean and Variance Terms in Section 5.4.1

B.1 Derivation of $\mu(N_{PE}^r)$ in Theorem 5.4.2

Let $T_s := T(\mathsf{y}_1, M_3, M_C)$. Then, observe that, by symmetry,

$$\mu(N_{PE}^r) = P\big(X_2 \in N_{PE}^r(X_1)\big) = 6\,P\big(X_2 \in N_{PE}^r(X_1), X_1 \in T_s\big).$$

Let $\ell_s(r,x)$ be the line such that $r\,d(\mathsf{y}_1, \ell_s(r,x)) = d(\mathsf{y}_1, e_1)$, so $\ell_s(r,x) = \sqrt{3}\,(1/r - x)$. Then if $x_1 \in T_s$ is above $\ell_s(r,x)$ then $N_{PE}^r(x_1) = T(\mathcal{Y})$, otherwise, $N_{PE}^r(x_1) \subsetneq T(\mathcal{Y})$.

For $r \in [1, 3/2)$, $\ell_s(r,x) \cap T_s = \emptyset$, so $N_{PE}^r(x) \subsetneq T(\mathcal{Y})$ for all $x \in T_s$. Then

$$\mu(N_{PE}^r) = 6\,P\big(X_2 \in N_{PE}^r(X_1), X_1 \in T_s\big) = 6\int_0^{1/2}\int_0^{x/\sqrt{3}} \frac{A(N_{PE}^r(x_1))}{A(T(\mathcal{Y}))^2} dydx = 6\cdot\frac{37}{1296}\,r^2 = \frac{37}{216}\,r^2.$$

where $A(N_{PE}^r(x_1)) = \frac{\sqrt{3}}{12}\,r^2\left(y + \sqrt{3}\,x\right)^2$ and $A(T(\mathcal{Y})) = \sqrt{3}/4$.

For $r \in [3/2, 2)$, $\ell_s(r,x)$ crosses through $\overline{M_3 M}_C$. Let the x coordinate of $\ell_s(r,x) \cap \overline{\mathsf{y}_1 M}_C$ be s_1. Then $s_1 = \frac{3}{4\,r}$. See Figure B.1.1.

Then

$$\begin{aligned} P\big(X_2 \in N_{PE}^r(X_1), X_1 \in T_s\big) &= \int_0^{s_1}\int_0^{x/\sqrt{3}} \frac{A(N_{PE}^r(x_1))}{A(T(\mathcal{Y}))^2} dydx + \int_{s_1}^{1/2}\int_0^{\ell_s(r,x)} \frac{A(N_{PE}^r(x_1))}{A(T(\mathcal{Y}))^2} dydx \\ &\quad + \int_{s_1}^{1/2}\int_{\ell_s(r,x)}^{x/\sqrt{3}} \frac{1}{A(T(\mathcal{Y}))} dydx = -\frac{-36 + r^4 + 64\,r - 32\,r^2}{48\,r^2}. \end{aligned}$$

Hence for $r \in [3/2, 2)$,

$$\mu(N_{PE}^r) = -\frac{1}{8}\,r^2 - 8\,r^{-1} + \frac{9}{2}\,r^{-2} + 4.$$

For $r \in [2, \infty)$, $\ell_s(r,x)$ crosses through $\overline{\mathsf{y}_1 M}_3$. Let the x coordinate of $\ell_s(r,x) \cap \overline{\mathsf{y}_1 M}_3$ be s_2, then $s_2 = 1/r$. See Figure B.1.1.

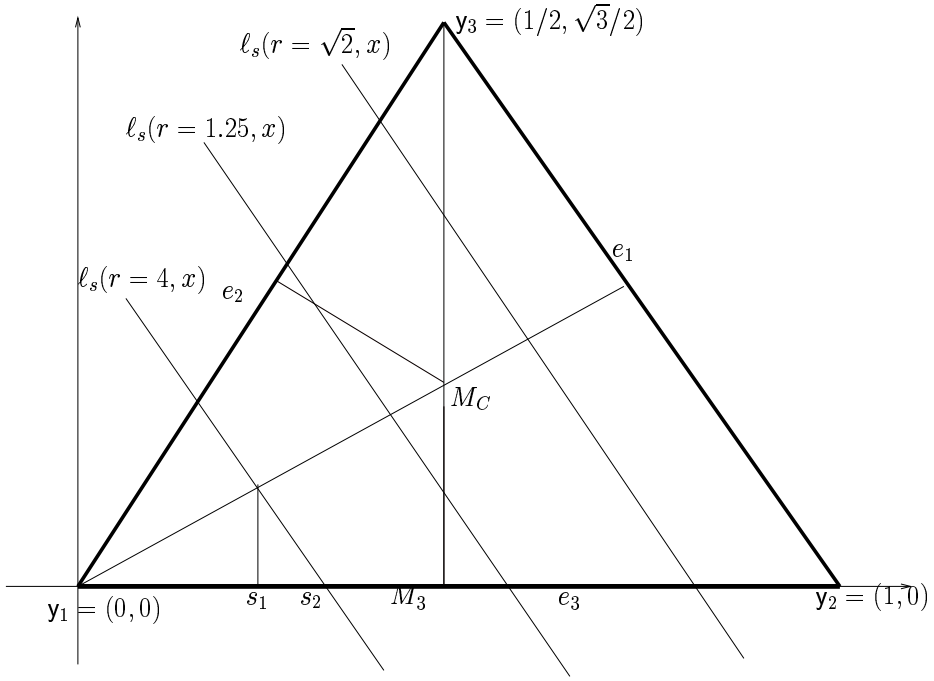


Figure B.1.1: The cases for relative position of $\ell_s(r, x)$ with various r values.

Then

$$P\big(X_2 \in N_{PE}^r(X_1), X_1 \in T_s\big) = \int_0^{s_1}\int_0^{x/\sqrt{3}} \frac{A(N_{PE}^r(x_1))}{A(T(\mathcal{Y}))^2} dydx + \int_{s_1}^{s_2}\int_0^{\ell_s(r,x)} \frac{A(N_{PE}^r(x_1))}{A(T(\mathcal{Y}))^2} dydx$$
$$+ \int_{s_1}^{s_2}\int_{\ell_s(r,x)}^{x/\sqrt{3}} \frac{1}{A(T(\mathcal{Y}))} dydx + \int_{s_2}^{1/2}\int_0^{x/\sqrt{3}} \frac{1}{A(T(\mathcal{Y}))} dydx = \frac{-3+2\,r^2}{12\,r^2}.$$

Hence for $r \in [2, \infty)$,

$$\mu(N_{PE}^r) = 1 - \frac{3}{2}\, r^{-2}.$$

For $r = \infty$, $\mu(N_{PE}^r) = 1$ follows trivially.

B.2 Derivation of $\nu(N_{PE}^r)$ in Theorem 5.4.2

Recall that $P_{2N}(N_{PE}^r) := P(\{X_2, X_3\} \subset N_{PE}^r(X_1))$, $P_{2G}(N_{PE}^r) := P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r))$, and $P_M(N_{PE}^r) := P(X_2 \in N_{PE}^r(X_1), X_3 \in \Gamma_1(X_1, N_{PE}^r))$. Then

$$\nu(N_{PE}^r) = \mathbf{Cov}\left[h_{12}(N_{PE}^r), h_{13}(N_{PE}^r)\right] = \big(P_{2N}(N_{PE}^r) + 2\,P_M(N_{PE}^r) + P_{2G}(N_{PE}^r)\big) - [2\,\mu(N_{PE}^r)]^2.$$

Recall the lines $\xi_j(r, x)$ in Section 3.4.2. To find the covariance, we need to find the possible

types of $\Gamma_1(x_1, N_{PE}^r)$ and $N_{PE}^r(x_1)$ for $r \in [1, \infty)$.

We partition $[1, \infty)$ with respect to the types of $N_{PE}^r(x_1)$ and $\Gamma_1(x_1, N_{PE}^r)$ and obtain $[1, 4/3)$, $[4/3, 3/2)$, $[3/2, 2)$, and $[2, \infty)$.

For $r \in [1, 4/3)$, there are six cases regarding $\Gamma_1(x_1, N_{PE}^r)$ and one case for $N_{PE}^r(x_1)$. See Figure B.2.1 for the prototypes of these six cases of $\Gamma_1(x_1, N_{PE}^r)$. Each case j, corresponds to the region R_j in Figure B.2.2, where $\ell_{am}(x) = x/\sqrt{3}$, $q_1(x) = (2\,r + 3\,x - 3)/\sqrt{3}$, $q_2(x) = \sqrt{3}\,(1/2 - r/3)$, $q_3(x) = \sqrt{3}\,(x - 1 + r/2)$, $q_4(x) = \sqrt{3}\,(1/2 - r/4)$, $q_{12}(x) = \sqrt{3}\,(r/2 - x)$, and $s_1 = 1 - 2\,r/3$, $s_2 = 3/2 - r$, $s_3 = 1 - r/2$, $s_4 = 3/2 - 5\,r/6$, $s_5 = 3\,r/8$. The explicit forms of R_j for $j \in \{1, \ldots, 6\}$ are given by

$$R_1 = \big\{(x,y) \in [0, s_1] \times [0, \ell_{am}(x)] \cup [s_1, s_2] \times [q_1(x), \ell_{am}(x)]\big\},$$

$$R_2 = \big\{(x,y) \in [s_1, s_2] \times [0, q_1(x)] \cup [s_2, s_3] \times [0, q_2(x)] \cup [s_3, s_4] \times [q_3(x), q_2(x)]\big\},$$

$$R_3 = \big\{(x,y) \in [s_3, s_4] \times [0, q_3(x)] \cup [s_4, 1/2] \times [0, q_2(x)]\big\},$$

$$R_4 = \big\{(x,y) \in [s_1, s_2] \times [0, q_1(x)] \cup [s_4, s_5] \times [q_3(x), \ell_{am}(x)] \cup [s_5, 1/2] \times [q_3(x), q_{12}(x)]\big\},$$

$$R_5 = \big\{(x,y) \in [s_4, 1/2] \times [q_2(x), q_3(x)]\big\}, \quad R_6 = \big\{(x,y) \in [s_5, 1/2] \times [q_{12}(x), \ell_{am}(x)]\big\}.$$

By symmetry, $P_{2N}(N_{PE}^r) = 6\,P\big(\{X_2, X_3\} \subset N_{PE}^r(X_1),\ X_1 \in T_s\big)$.

For $r \in [1, 4/3)$,

$$P\big(\{X_2, X_3\} \subset N_{PE}^r(X_1),\ X_1 \in T_s\big) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{781}{116640}\, r^4,$$

where $A(N_{PE}^r(x_1)) = \frac{\sqrt{3}}{12}\, r^2 \left(y + \sqrt{3}\,x\right)^2$. Hence for $r \in [1, 4/3)$, $P_{2N}(N_{PE}^r) = \frac{781\, r^4}{19440}$. Note that the same results also hold for $r \in [4/3, 3/2)$.

Next, by symmetry, $P_{2G}(N_{PE}^r) = 6\,P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in T_s\big)$ and

$$P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in T_s\big) = \sum_{j=1}^{6} P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_j\big).$$

For $x_1 \in R_1$,

$$P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_1\big) = \int_0^{s_1} \int_0^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r)\right)^2}{A(T(\mathcal{Y}))^3} dy dx$$
$$+ \int_{s_1}^{s_2} \int_{q_1(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r)\right)^2}{A(T(\mathcal{Y}))^3} dy dx = \frac{(211\, r^4 - 1716\, r^3 + 5751\, r^2 - 6696\, r + 2511)(2\, r - 3)^2}{10935\, r^4},$$

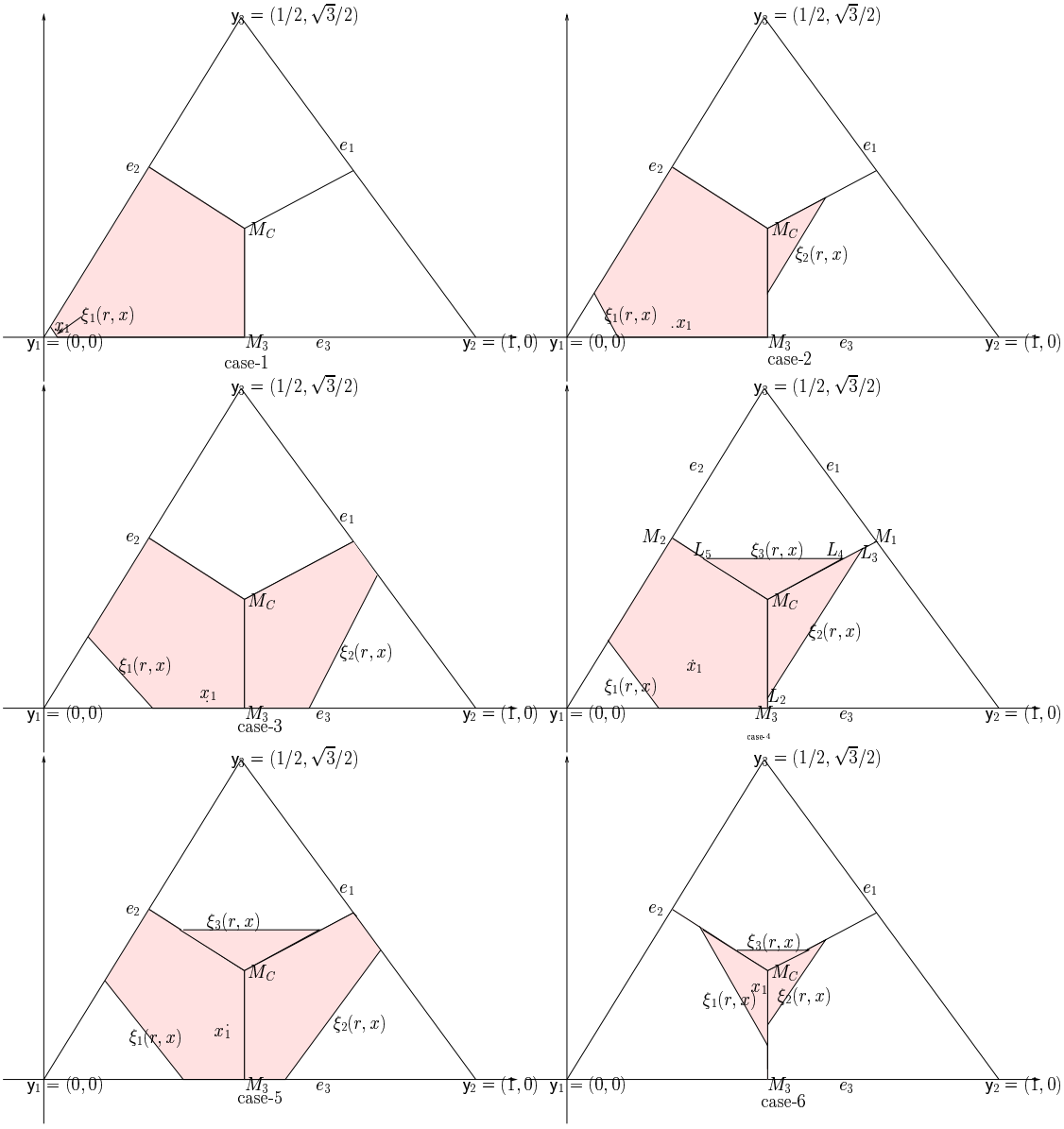


Figure B.2.1: The prototypes of the six cases of $\Gamma_1\left(x_1, N_{PE}^r\right)$ for $x_1 \in T_s$ for $r \in [1, 4/3)$.

where $A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right) = \frac{\sqrt{3}\left(\left(\sqrt{3}\,x+y\right)^2\right)-r^2}{12\,r^2}$.

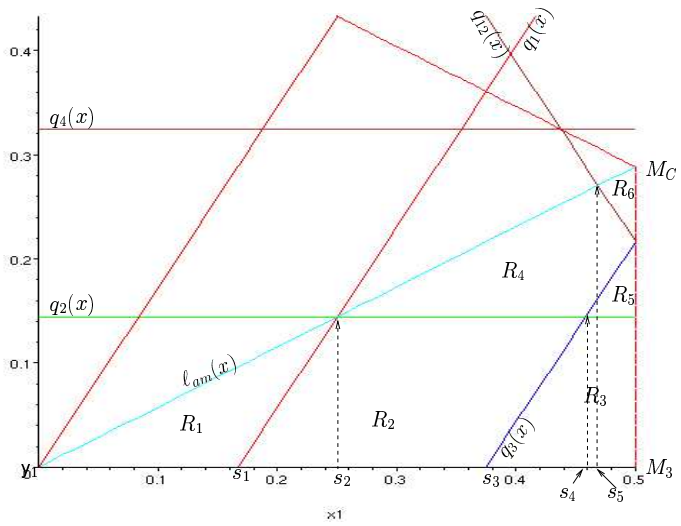


Figure B.2.2: The regions corresponding to the prototypes of the six cases for $r \in [1, 4/3)$ with $r = 1.25$.

For $x_1 \in R_2$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_2\right) = \int_{s_1}^{s_2} \int_0^{\ell_{am}(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_2}^{s_3} \int_0^{q_2(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_3}^{s_4} \int_{q_3(x)}^{q_2(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$= -\frac{(2\,r-3)(440\,r^4 - 4091\,r^3 + 13476\,r^2 - 16506\,r + 6696)}{9720\,r^3}.$$

where $A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right) = \frac{\sqrt{3}\left(-4\,\sqrt{3}\,r\,y - 12\,r + 12\,r\,x + 5\,r^2 + 2\,y^2 + 6\,\sqrt{3}\,y - 8\,x\,\sqrt{3}\,y + 9 - 18\,x + 6\,x^2\right)}{12\,r^2}$.

For $x_1 \in R_3$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_3\right)$$
$$= \int_{s_3}^{s_4} \int_0^{q_3(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_4}^{1/2} \int_0^{q_2(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$= -\frac{(2\,r-3)(21056\,r^5 - 7845\,r^4 + 231300\,r^3 - 943650\,r^2 + 1127520\,r - 428652)}{262440\,r^4}.$$

where $A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right) = -\frac{\sqrt{3}\left(2\,y^2 + 2\,\sqrt{3}\,y + 3 - 6\,x + 6\,x^2 - 2\,r^2\right)}{12\,r^2}$.

For $x_1 \in R_4$,

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_4) = \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{PE}^r))^2}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_4}^{s_5} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{PE}^r))^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_5}^{1/2} \int_{q_3(x)}^{q_{12}(x)} \frac{A(\Gamma_1(x_1, N_{PE}^r))^2}{A(T(\mathcal{Y}))^3} dydx$$
$$= -\frac{12873091}{699840} r^2 + \frac{81239}{648} r + \frac{14714}{27} r^{-1} - \frac{4238}{9} r^{-2} + \frac{656}{3} r^{-3} - \frac{128}{3} r^{-4} - \frac{77123}{216}.$$

where $A(\Gamma_1(x_1, N_{PE}^r)) = \frac{\sqrt{3}\left(9r^2+18-24r+4\sqrt{3}ry-18x+6x^2+14y^2+12rx-8x\sqrt{3}y-6\sqrt{3}y\right)}{12r^2}$.

For $x_1 \in R_5$,

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_5) = \int_{s_4}^{1/2} \int_{q_2(x)}^{q_3(x)} \frac{A(\Gamma_1(x_1, N_{PE}^r))^2}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{(89305r^4 - 364080r^3 + 598320r^2 - 468288r + 145152)(-6+5r)^2}{262440r^4}.$$

where $A(\Gamma_1(x_1, N_{PE}^r)) = \frac{\sqrt{3}\left(9r^2+18-24r+4\sqrt{3}ry-18x+6x^2+14y^2+12rx-8x\sqrt{3}y-6\sqrt{3}y\right)}{12r^2}$.

For $x_1 \in R_6$,

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_6) = \int_{s_5}^{1/2} \int_{q_{12}x}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{PE}^r))^2}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{(1081r^4 - 4672r^3 + 7624r^2 - 5568r + 1536)(-4+3r)^2}{960r^4}.$$

where $A(\Gamma_1(x_1, N_{PE}^r)) = -\frac{\sqrt{3}\left(\sqrt{3}y-2r^2-3+3x+4r-3x^2-3y^2\right)}{2r^2}$.

So,

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r)) = 6\left(\frac{25687}{349920} r^2 - \frac{133}{972} r + \frac{14}{81} r^{-1} - \frac{1}{9} r^{-2} + \frac{1}{90} r^{-4} - \frac{1}{324}\right)$$
$$= \frac{25687r^6 - 47880r^5 - 1080r^4 + 60480r^3 - 38880r^2 + 3888}{58320r^4}.$$

Furthermore, by symmetry, $P_M(N_{PE}^r) = 6\,P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in T_s)$, and

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in T_s)$$
$$= \sum_{j=1}^{6} P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_j).$$

For $x_1 \in R_1$,

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_1) = \int_0^{s_1} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_1}^{s_2} \int_{q_1(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx = -\frac{1}{21870}(143\, r^2 - 744\, r + 558)(2\, r - 3)^4.$$

For $x_1 \in R_2$,

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_2) = \int_{s_1}^{s_2} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_2}^{s_3} \int_0^{q_2(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx + \int_{s_3}^{s_4} \int_{q_3(x)}^{q_2(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{1}{349920}\, r\,(2\, r - 3)(23014\, r^4 - 187311\, r^3 + 517896\, r^2 - 594216\, r + 241056).$$

For $x_1 \in R_3$,

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_3)$$
$$= \int_{s_3}^{s_4} \int_0^{q_3(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx + \int_{s_4}^{1/2} \int_0^{q_2(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{1}{1049760}(2\, r - 3)(874\, r^5 - 297327\, r^4 + 1858392\, r^3 - 4298832\, r^2 + 4280202\, r - 1546209).$$

For $x_1 \in R_4$,

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_4) = \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_4}^{s_5} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx + \int_{s_5}^{1/2} \int_{q_3(x)}^{q_{12}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$= -\frac{1}{466560}\, r\,(1762560\, r - 497664 - 2661120\, r^2 + 201395\, r^5 - 1017720\, r^4 + 2212560\, r^3).$$

For $x_1 \in R_5$,

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_5) = \int_{s_4}^{1/2} \int_{q_2(x)}^{q_3(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{1}{262440}(1570\, r^4 - 1380\, r^3 - 11205\, r^2 + 29700\, r - 19116)(-6 + 5\, r)^2.$$

For $x_1 \in R_6$,

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_6) = \int_{s_5}^{1/2} \int_{q_{12}x}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dy dx$$
$$= \frac{1}{51840}(1485\, r^4 - 2064\, r^3 + 16\, r^2 - 128\, r + 768)(-4 + 3\, r)^2.$$

Thus

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1(X_1, N_{PE}^r))$$
$$= 6\left(\frac{3007}{699840} r^6 - \frac{8}{405} r^5 + \frac{5}{648} r^4 + \frac{1}{9} r^3 - \frac{133}{648} r^2 + \frac{56}{405} r - \frac{143}{4320}\right)$$
$$= \frac{3007}{116640} r^6 - \frac{16}{135} r^5 + \frac{5}{108} r^4 + \frac{2}{3} r^3 - \frac{133}{108} r^2 + \frac{112}{135} r - \frac{143}{720}.$$

Hence

$$\mathbf{E}[h_{12}(N_{PE}^r)\, h_{13}(N_{PE}^r)] = \Big[3007\, r^{10} - 13824\, r^9 + 7743\, r^8 + 77760\, r^7 - 117953\, r^6 + 48888\, r^5$$
$$- 24246\, r^4 + 60480\, r^3 - 38880\, r^2 + 3888\Big] \Big/ \Big[58320\, r^4\Big].$$

Thus

$$\nu(N_{PE}^r) = \Big[3007\, r^{10} - 13824\, r^9 + 898\, r^8 + 77760\, r^7 - 117953\, r^6 + 48888\, r^5 - 24246\, r^4 + 60480\, r^3$$
$$- 38880\, r^2 + 3888\Big] \Big/ \Big[58320\, r^4\Big].$$

For $r \in [4/3, 3/2)$, there are six cases regarding $\Gamma_1(x_1, N_{PE}^r)$ and one case for $N_{PE}^r(x_1)$. Prototypes of the five of the cases for $\Gamma_1(x_1, N_{PE}^r)$ are as in case$-j$ for $j \in \{1, \ldots, 5\}$ in Figure B.2.1 and the new case, case-7, is depicted in Figure B.2.3. Each case j corresponds to the region R_j in Figure B.2.4 where $s_1 = 1 - 2\, r/3,\ s_2 = 3/2 - r,\ s_3 = 1 - r/2,\ s_4 = 3/2 - 5\, r/6,\ s_5 = 3/2 - 3\, r/4$. The explicit forms of R_j for $j \in \{1, 2, 3\}$ are same as before, for $j \in \{4, 5, 7\}$ are

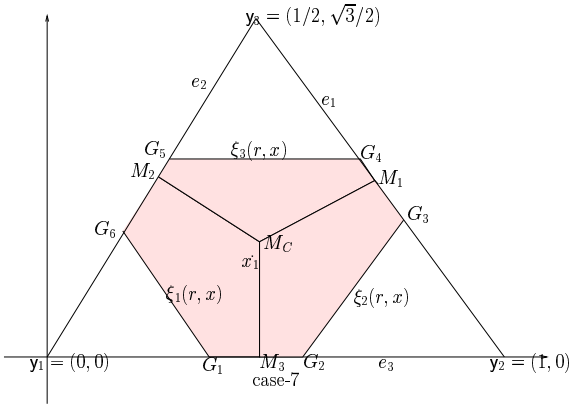


Figure B.2.3: The prototype of the new case for $\Gamma_1\left(x_1, N_{PE}^r\right)$ for $x_1 \in T_s$ for $r \in [4/3, 3/2)$.

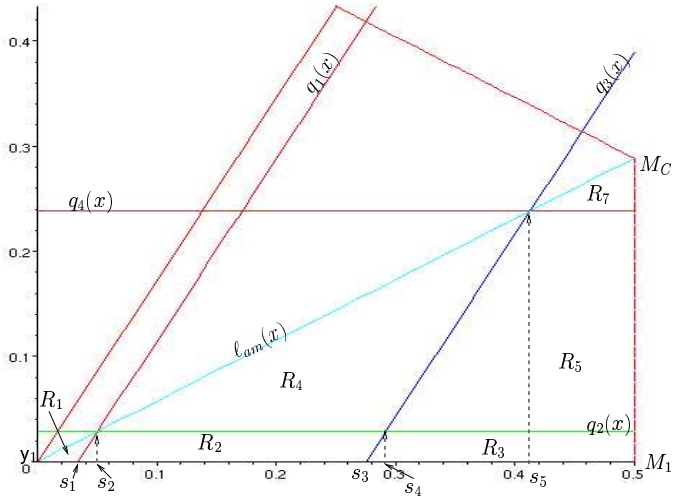


Figure B.2.4: The regions corresponding to the six cases for $r \in [4/3, 3/2)$

given by

$$R_4 = \left\{(x,y) \in [s_2, s_4] \times [q_2(x), \ell_{am}(x)] \cup [s_4, s_6] \times [q_3(x), \ell_{am}(x)]\right\}$$

$$R_5 = \left\{(x,y) \in [s_4, s_6] \times [q_2(x), q_3(x)] \cup [s_6, 1/2] \times [q_2(x), q_4(x)]\right\}$$

$$R_7 = \left\{(x,y) \in [s_6, 1/2] \times [q_4(x), \ell_{am}(x)]\right\}$$

where $\ell_{am}(x) = x/\sqrt{3}$, $q_1(x) = (2\,r - 3)/\sqrt{3} + \sqrt{3}\,x$, $q_2(x) = \sqrt{3}\,(1/2 - r/3)$, $q_3(x) = \sqrt{3}\,(x - 1 + r/2)$, and $q_4(x) = \sqrt{3}\,(1/2 - r/4)$.

Then $P_{2N}(N_{PE}^r) = \frac{781\,r^4}{19440}$. We use the same limits of integration in $\mu(N_{PE}^r)$ calculations with the integrand $A(N_{PE}^r(x_1))^2/A(T(\mathcal{Y}))^3$.

Next, by symmetry, $P_{2G}(N_{PE}^r) = 6\,P\big(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in T_s\big)$, and, let $S_I := \{1,2,3,4,5,7\}$, then

$$P\big(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in T_s\big) = \sum_{j \in S_I} P\left(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_j\right).$$

For $x_1 \in R_j$, $j \in \{1,2,3\}$ we get the same result as before.

For $x_1 \in R_4$,

$$P(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_4) = \int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_4}^{s_6} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx = \frac{9637\,r^4 - 89640\,r^3 + 288360\,r^2 - 362880\,r + 155520}{349920\,r^2}.$$

where $A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right) = \frac{\sqrt{3}\left(9\,r^2+18-24\,r+4\,\sqrt{3}\,r\,y-18\,x+6\,x^2+14\,y^2+12\,r\,x-8\,x\,\sqrt{3}\,y-6\,\sqrt{3}\,y\right)}{12\,r^2}$.

For $x_1 \in R_5$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_5\right)$$
$$= \int_{s_4}^{s_6} \int_{q_2(x)}^{q_3(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_6}^{1/2} \int_{q_2(x)}^{q_4(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{87251\,r^5 + 13219200\,r - 11214720\,r^2 - 5225472 + 3377160\,r^3 - 261288\,r^4}{2099520\,r^3}.$$

where $A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)$ is same as before.

For $x_1 \in R_7$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_7\right) =$$
$$\int_{s_6}^{1/2} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx = \frac{(57\,r^4 + 96\,r^3 - 72\,r^2 - 576\,r + 512)(-4+3\,r)^2}{2880\,r^4}.$$

where $A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right) = -\frac{\sqrt{3}\left(6\,y^2-2\,\sqrt{3}\,y+6-6\,x+6\,x^2-3\,r^2\right)}{12\,r^2}$.

So,

$$\begin{aligned} P_{2G}(N_{PE}^r) &= 6\left(\frac{-47880\,r^5 - 38880\,r^2 + 25687\,r^6 - 1080\,r^4 + 60480\,r^3 + 3888}{349920\,r^4}\right) \\ &= \frac{-47880\,r^5 - 38880\,r^2 + 25687\,r^6 - 1080\,r^4 + 60480\,r^3 + 3888}{58320\,r^4}. \end{aligned}$$

Furthermore,

$$P_M(N_{PE}^r) = \sum_{j \in S_I} P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_j).$$

For $x_1 \in R_j$, $j \in \{1,2,3\}$ we get the same result as before.

For $x_1 \in R_4$,

$$P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_4\right) =$$
$$\int_{s_2}^{s_4} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx + \int_{s_4}^{s_6} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx$$
$$= -\frac{1}{466560}\, r^2 (207360 + 404640\, r^2 - 483840\, r - 142920\, r^3 + 17687\, r^4).$$

For $x_1 \in R_5$,

$$P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_5\right)$$
$$= \int_{s_4}^{s_6} \int_{q_2(x)}^{q_3(x)} \frac{A(N_{PE}^r(x_1))\, A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx + \int_{s_6}^{1/2} \int_{q_2(x)}^{q_4(x)} \frac{A(N_{PE}^r(x_1))\, A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx$$
$$= -\frac{r\,(399064320\, r - 150792192 + 171990000\, r^3 - 391461120\, r^2 - 31140648\, r^4 + 1230359\, r^5}{67184640}.$$

For $x_1 \in R_7$,

$$P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_7\right) = \int_{s_6}^{1/2} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\, A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{1}{829440}\,(2727\, r^4 - 3648\, r^3 - 52736\, r^2 + 166656\, r - 121600)(-4 + 3\, r)^2.$$

Then,

$$P_M(N_{PE}^r) = 6\left(\frac{5467}{2799360}\, r^6 - \frac{35}{2592}\, r^5 + \frac{37}{1296}\, r^4 - \frac{13}{648}\, r^2 + \frac{83}{12960}\right) = \frac{5467}{466560}\, r^6 - \frac{35}{432}\, r^5 + \frac{37}{216}\, r^4 - \frac{13}{108}\, r^2 + \frac{83}{216}.$$

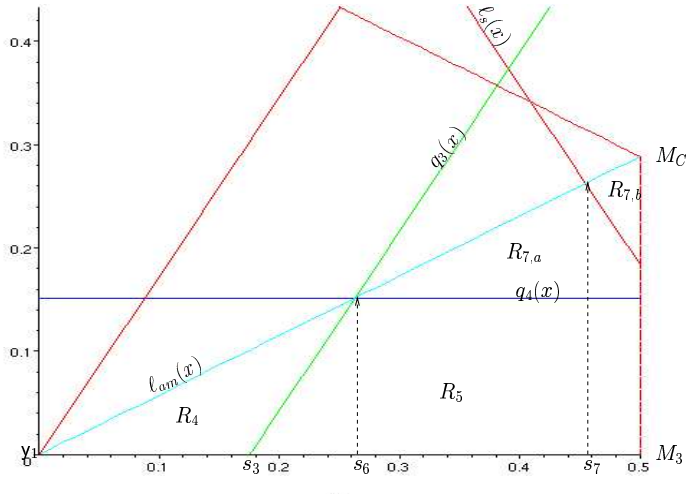


Figure B.2.5: The regions corresponding to the three cases for $r \in [3/2, 2)$ with $r = 1.65$

So,

$$\mathbf{E}\left[h_{12}(N_{PE}^r)\, h_{13}(N_{PE}^r)\right] = \Big[5467\, r^{10} - 37800\, r^9 + 89292\, r^8 + 46588\, r^6 - 191520\, r^5 + 13608\, r^4 + 241920\, r^3 - 155520\, r^2 + 15552\Big] \Big/ \Big[233280\, r^4\Big].$$

Thus, for $r \in [4/3, 3/2)$

$$\nu(N_{PE}^r) = \Big[5467\, r^{10} - 37800\, r^9 + 61912\, r^8 + 46588\, r^6 - 191520\, r^5 + 13608\, r^4 + 241920\, r^3 - 155520\, r^2 + 15552\Big] \Big/ \Big[233280\, r^4\Big].$$

For $r \in [3/2, 2)$, there are three cases regarding $\Gamma_1\left(x_1, N_{PE}^r\right)$ and two cases for $N_{PE}^r(x_1)$. The prototypes of these three cases as in cases 4,5, and 7 of Figures B.2.1 and B.2.3. Each case j, corresponds to the region R_j in Figure B.2.5 where $q_j(x)$ are same as before for $j \in \{3, 4\}$, and s_j, $j \in \{3, 4, 6\}$ are same as before and $s_7 = 3/(4\, r)$. Observe that for $x_1 \in R_4 \cup R_5 \cup R_{7a}$, $N_{PE}^r(x_1) = T_r(x_1) \subsetneq T(\mathcal{Y})$, and for $x_1 \in R_{7b}$, $N_{PE}^r(x_1) = T(\mathcal{Y})$. So there are four regions to consider in order to calculate the covariance.

Then, for $x_1 = (x, y) \in R_j$, $\Gamma_1\left(x_1, N_{PE}^r\right)$ are same as before for $j \in \{4, 5, 7\}$. The explicit

forms of R_j, $j \in \{4, 5, 7a, 7b\}$ are given below (the explicit form of R_7 is same as before):

$$R_4 = \left\{(x,y) \in [0,s_3] \times [0, \ell_{am}(x)] \cup [s_3, s_6] \times [q_3(x), \ell_{am}(x)]\right\}$$

$$R_5 = \left\{(x,y) \in [s_3, s_6] \times [0, q_3(x)] \cup [s_6, 1/2] \times [0, q_4(x)]\right\}$$

$$R_{7,a} = \left\{(x,y) \in [s_6, s_7] \times [q_4(x), \ell_{am}(x)] \cup [s_7, 1/2] \times [q_4(x), \ell_s(r,x)]\right\}$$

$$R_{7,b} = \left\{(x,y) \in [s_7, 1/2] \times [\ell_s(r,x), \ell_{am}(x)]\right\}$$

Now,

$$P(\{X_2, X_3\} \subset N_{PE}^r(X_1), X_1 \in T_s) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dydx =$$
$$\int_0^{s_7} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_7}^{1/2} \int_0^{\ell_s(x)} \frac{A(N_{PE}^r(x_1))^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_7}^{1/2} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T(\mathcal{Y}))} dydx$$
$$= \frac{480 - r^6 - 768\,r + 320\,r^2}{480\,r^2}.$$

Hence $P_{2N}(N_{PE}^r) = -\frac{-480 + r^6 + 768\,r - 320\,r^2}{80\,r^2}$.

Next, by symmetry, $P_{2G}(N_{PE}^r) = 6\,P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in T_s\big)$, and, let $S_I := \{4, 5, 7\}$, then

$$P\big(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in T_s\big) = \sum_{j \in S_I} P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_j).$$

For $x_1 \in R_4$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_4\right) = \int_0^{s_3} \int_0^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r)\right)^2}{A(T(\mathcal{Y}))^3} dydx +$$
$$\int_{s_3}^{s_6} \int_{q_3(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r)\right)^2}{A(T(\mathcal{Y}))^3} dydx = \frac{(237\,r^4 - 956\,r^3 + 1728\,r^2 - 1584\,r + 592)(-2+r)^2}{480\,r^4}.$$

For $x_1 \in R_5$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r),\ X_1 \in R_5\right)$$
$$= \int_{s_3}^{s_6} \int_0^{q_3(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r)\right)^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_6}^{1/2} \int_0^{q_4(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$= \frac{(r-2)(1909\,r^5 - 6142\,r^4 + 10036\,r^3 - 14808\,r^2 + 15024\,r - 6048)}{2880\,r^4}.$$

For $x_1 \in R_7$ the result is same as before. So

$$\begin{aligned} P_{2G}(N_{PE}^r) &= 6\left(\frac{7320\,r^4 - 984\,r^5 + 13\,r^6 - 20480\,r^3 + 27840\,r^2 - 18816\,r + 5152}{1440\,r^4}\right) \\ &= \frac{7320\,r^4 - 984\,r^5 + 13\,r^6 - 20480\,r^3 + 27840\,r^2 - 18816\,r + 5152}{240\,r^4}. \end{aligned}$$

Furthermore,

$$P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in T_s\right) = \int_0^{1/2}\int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx.$$

For $x_1 \in R_4$,

$$\begin{aligned} P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_4) &= \int_0^{s_3}\int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx \\ + \int_{s_3}^{s_6}\int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx &= \frac{1}{1920}\left(99\,r^2 - 16\,r - 84\right)(-2+r)^4. \end{aligned}$$

For $x_1 \in R_5$,

$$\begin{aligned} &P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_5\right) \\ &= \int_{s_3}^{s_6}\int_0^{q_3(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx + \int_{s_6}^{1/2}\int_0^{q_4(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx \\ &= -\frac{1}{92160}(-2+r)(7535\,r^5 - 35210\,r^4 + 9500\,r^3 + 181560\,r^2 - 308880\,r + 147168). \end{aligned}$$

For $x_1 \in R_{7a}$,

$$\begin{aligned} P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_{7a}\right) &= \int_{s_6}^{s_7}\int_{q_4(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx \\ + \int_{s_7}^{1/2}\int_{q_4(x)}^{\ell_s(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx &= \frac{303}{10240}\,r^6 - \frac{91}{768}\,r^5 - \frac{53}{128}\,r^4 + \frac{235}{72}\,r^3 - \frac{173}{24}\,r^2 \\ + \frac{101}{15}\,r + \frac{2}{3}\,r^{-1} - \frac{3}{4}\,r^{-2} - \frac{16}{9}\,r^{-3} + 4\,r^{-4} - \frac{18}{5}\,r^{-5} + \frac{3}{2}\,r^{-6} - \frac{34}{15}. \end{aligned}$$

For $x_1 \in R_{7b}$,

$$\begin{aligned} &P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_{7b}\right) = \\ &\int_{s_7}^{1/2}\int_{\ell_s(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\,A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)}{A(T(\mathcal{Y}))^3} dydx = \frac{(2\,r^4 - 4\,r^2 + 4\,r - 3)(2\,r-3)^2}{12\,r^6}. \end{aligned}$$

Hence

$$P_{2G}(N_{PE}^r) = 6\left(-\left[7\,r^{12}-72\,r^{11}+240\,r^{10}-1440\,r^8+3456\,r^7-10296\,r^6+15360\,r^5+6720\,r^4-40960\,r^3\right.\right.$$
$$\left.\left.+46080\,r^2-27648\,r+8640\right]\Big/\left[11520\,r^6\right]\right) = -\left[7\,r^{12}-72\,r^{11}+240\,r^{10}-1440\,r^8+3456\,r^7\right.$$
$$\left.-10296\,r^6+15360\,r^5+6720\,r^4-40960\,r^3+46080\,r^2-27648\,r+8640\right]\Big/\left[1920\,r^6\right].$$

Thus

$$\mathbf{E}\left[h_{12}(N_{PE}^r)\,h_{13}(N_{PE}^r)\right] = -\left[7\,r^{12}-72\,r^{11}+252\,r^{10}-1492\,r^8+7392\,r^7-43416\,r^6+106496\,r^5\right.$$
$$\left.-110400\,r^4+34304\,r^3+25472\,r^2-27648\,r+8640\right]\Big/\left[960\,r^6\right].$$

Therefore, for $r \in [3/2, 2)$

$$\nu(N_{PE}^r) = -\left[7\,r^{12}-72\,r^{11}+312\,r^{10}-5332\,r^8+15072\,r^7+13704\,r^6-139264\,r^5+273600\,r^4\right.$$
$$\left.-242176\,r^3+103232\,r^2-27648\,r+8640\right]\Big/\left[960\,r^6\right].$$

For $r \in [2, \infty)$, there is only one case regarding $\Gamma_1\left(x_1, N_{PE}^r\right)$, namely R_7, and two cases regarding $N_{PE}^r(x_1)$. Furthermore, s_7, is same as before and $s_8 = 1/r$. Observe that for $x_1 \in R_{7a}$, $N_{PE}^r(x_1) = T_r(x_1) \subsetneq T(\mathcal{Y})$, and for $x_1 \in R_{7b}$, $N_{PE}^r(x_1) = T(\mathcal{Y})$. So there are two regions to consider to calculate the covariance.

For $x_1 = (x, y) \in R_7$, $\Gamma_1\left(x_1, N_{PE}^r\right)$ is same as before. The explicit form of R_7, is same as T_s. For R_{7a} and R_{7b}, see below:

$$R_{7,a} = \left\{(x,y) \in [0, s_7] \times [0, \ell_{am}(x)] \cup [s_7, s_8] \times [0, \ell_s(r,x)]\right\}$$
$$R_{7,b} = \left\{(x,y) \in [s_7, s_8] \times [\ell_s(r,x), \ell_{am}(x)] \cup [s_8, 1/2] \times [0, \ell_{am}(r,x)]\right\}$$

Now,

$$P(\{X_2, X_3\} \subset N_{PE}^r(X_1), X_1 \in T_s) = \int_0^{s_7}\int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))^2}{A(T(\mathcal{Y}))^3}dydx + \int_{s_7}^{s_8}\int_0^{\ell_s(x)} \frac{A(N_{PE}^r(x_1))^2}{A(T(\mathcal{Y}))^3}dydx$$
$$+\int_{s_7}^{s_8}\int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T(\mathcal{Y}))}dydx + \int_{s_8}^{1/2}\int_0^{\ell_{am}(x)} \frac{1}{A(T(\mathcal{Y}))}dydx = -\frac{1}{3}\,r^{-2} + \frac{1}{6}.$$

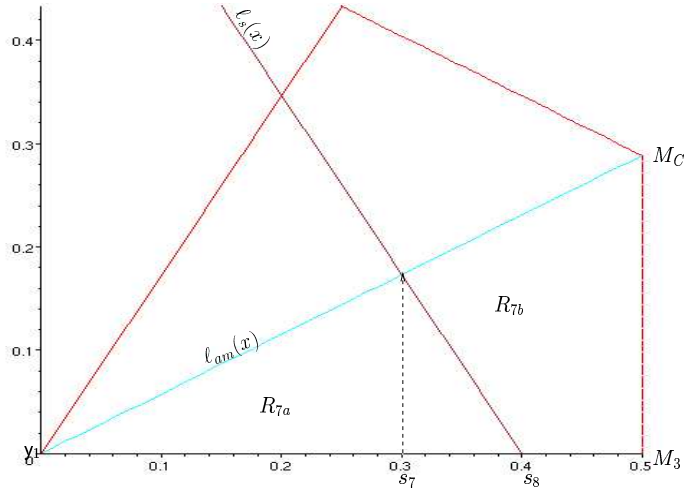


Figure B.2.6: The regions corresponding to the two cases for $N_{PE}^r(x_1)$ for $r \in [2, \infty)$ with $r = 2.5$

Hence $P_{2N}(N_{PE}^r) = 1 - 2\,r^{-2}$.

Next,

$$P\left(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in T_s\right) = P(\{X_2, X_3\} \subset \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_7)$$
$$= \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A\left(\Gamma_1\left(x_1, N_{PE}^r\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx = \frac{34 - 45\,r^2 + 15\,r^4}{90\,r^4}.$$

So $P_{2G}(N_{PE}^r) = \frac{34-45\,r^2+15\,r^4}{15\,r^4}$.

Furthermore,

$$P(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right), X_1 \in T_s) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\,A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$= \int_0^{s_7} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1))\,A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx + \int_{s_7}^{s_8} \int_0^{\ell_s(x)} \frac{A(N_{PE}^r(x_1))\,A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_7}^{s_8} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^2} dydx + \int_{s_8}^{1/2} \int_0^{\ell_{am}(x)} \frac{A(\Gamma_1(x_1, N_{PE}^r))}{A(T(\mathcal{Y}))^2} dydx = \frac{25-48\,r+90\,r^2-90\,r^4+30\,r^6}{180\,r^6}.$$

So $P_M(N_{PE}^r) = \frac{25-48\,r+90\,r^2-90\,r^4+30\,r^6}{30\,r^6}$.

Hence, $\mathbf{E}\left[h_{12}(N_{PE}^r)\,h_{13}(N_{PE}^r)\right] = \frac{60\,r^6-165\,r^4+124\,r^2-48\,r+25}{15\,r^6}$. Thus,

$$\nu(N_{PE}^r) = \frac{15\,r^4 - 11\,r^2 - 48\,r + 25}{15\,r^6},\ \text{for } r \in [2, \infty).$$

For $r = \infty$, it is trivial to see that $\nu(N_{PE}^r) = 0$.

B.3 Derivation of $\mu(N_{PE}^r, \varepsilon)$

We demostrate the derivation of $\mu_S(N_{PE}^r, \varepsilon)$ for segregation with $\varepsilon \in \left[0, \sqrt{3}/8\right)$ and among the intervals of r that do not vanish as $\varepsilon \to 0$. So the resultant expressions can be used in PAE analysis.

First, observe that, by symmetry,

$$\mu_S(N_{PE}^r, \varepsilon) = P\big(X_2 \in N_{PE}^r(X_1, \varepsilon)\big) = 6\, P\big(X_2 \in N_{PE}^r(X_1, \varepsilon), X_1 \in T_s \setminus T(\mathsf{y}_1, \varepsilon)\big).$$

Let $q(\mathsf{y}_j, x)$ be the line parallel to e_j and crossing $T(\mathcal{Y})$ such that $d(\mathsf{y}_j, q(\mathsf{y}_j, x)) = \varepsilon$ for $j \in \{1, 2, 3\}$. Furthermore, let $T_\varepsilon := T(\mathcal{Y}) \setminus \cup_{j=1}^3 T(\mathsf{y}_j, \varepsilon)$. Then $q(\mathsf{y}_1, x) = 2\,\varepsilon - \sqrt{3}\,x$, $q(\mathsf{y}_2, x) = \sqrt{3}\,x - \sqrt{3} + 2\,\varepsilon$, and $q(\mathsf{y}_3, x) = \sqrt{3}/2 - \varepsilon$. Now, let

$$Q_1 = q(\mathsf{y}_1, x) \cap e_3 = \left(2\,\varepsilon/\sqrt{3}, 0\right), \qquad Q_2 = q(\mathsf{y}_2, x) \cap e_3 = \left(1 - 2\,\varepsilon/\sqrt{3}, 0\right),$$
$$Q_3 = q(\mathsf{y}_2, x) \cap e_1 = \left(1 - \varepsilon/\sqrt{3}, \varepsilon\right), \qquad Q_4 = q(\mathsf{y}_3, x) \cap e_1 = \left(1/2 + \varepsilon/\sqrt{3}, \sqrt{3}/2 - \varepsilon\right),$$
$$Q_5 = q(\mathsf{y}_3, x) \cap e_2 = \left(1/2 - \varepsilon/\sqrt{3}, \sqrt{3}/2 - \varepsilon\right), \quad Q_6 = q(\mathsf{y}_1, x) \cap e_2 = \left(\varepsilon/\sqrt{3}, \varepsilon\right).$$

See Figure B.3.1. Then $T(\mathsf{y}_1, \varepsilon) = T(\mathsf{y}_1, Q_1, Q_6)$, $T(\mathsf{y}_2, \varepsilon) = T(Q_2, \mathsf{y}_2, Q_3)$, and $T(\mathsf{y}_3, \varepsilon) = T(Q_4, Q_5, \mathsf{y}_3)$, and for $\varepsilon \in [0, \sqrt{3}/4)$, T_ε is the hexagon with vertices, Q_j, $j \in \{1, \ldots, 6\}$.

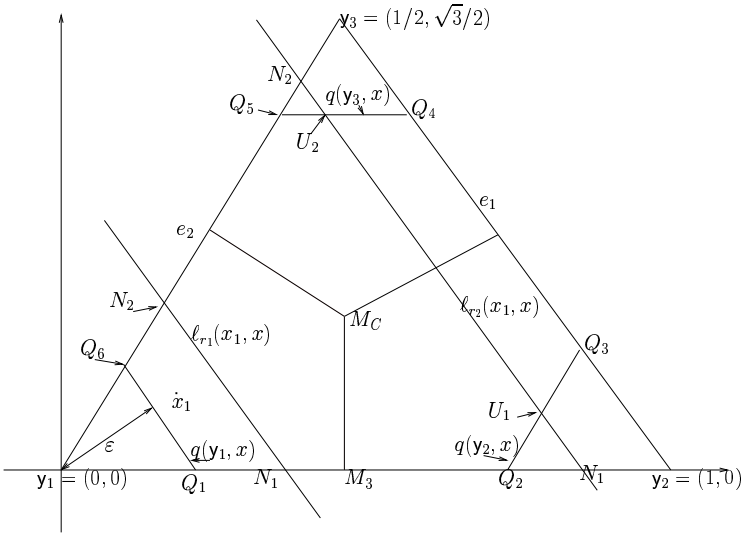


Figure B.3.1: The support under H_ε^S for $\varepsilon \in (0, \sqrt{3}/4)$ and the two types $\ell_r(x_1, x)$ for $r_1 < r_2$.

Now, let $q_2(x)$ be the line such that $r\, d(\mathsf{y}_1, q_2(x)) = d(\mathsf{y}_1, \ell(Q_2)) = d(\mathsf{y}_1, e_1) - \varepsilon$, $q_3(x)$ be

the line such that $r\,d(\mathsf{y}_1, q_3(x)) = d(\mathsf{y}_1, e_1)$. Then $q_2(x) = -\sqrt{3}\,x + \left(\sqrt{3} - 2\,\varepsilon\right)/r$ and $q_3(x)$ is the same as $\ell_s(x)$ before. Let the x coordinate of $q(\mathsf{y}_1, x) \cap \ell_{am}(x)$ be s_1, $q_2(x) \cap \ell_{am}(x)$ be s_3, and $\ell_s(x) \cap \ell_{am}(x)$ be s_5 and $Q_1 = (s_2, 0)$, $q_2(x) \cap e_3 = (s_4, 0)$, and $\ell_s(x) \cap e_3 = (s_6, 0)$. So $s_1 := \sqrt{3}\,\varepsilon/2$, $s_2 = 2\,\varepsilon/\sqrt{3}$, $s_3 = \left(3 - 2\,\varepsilon\,\sqrt{3}\right)/(4\,r)$, $s_4 = \left(3 - 2\,\varepsilon\,\sqrt{3}\right)/(3\,r)$, $s_5 = 3/(4\,r)$, and $s_6 = 1/r$.

See Figure B.3.2 for an $r \in [2, \sqrt{3}/(2\,\varepsilon))$. Furthermore, for $x_1 = (x, y) \in R_{CM}(\mathsf{y}_1)$, let

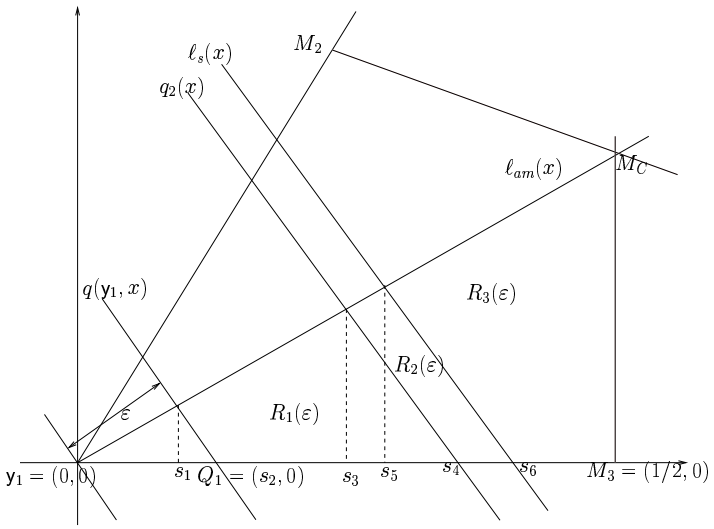


Figure B.3.2: The partition of T_s for different types of $N_{PE}^r(\cdot, \varepsilon)$ under H_ε^S with $r \in [2, \sqrt{3}/(2\,\varepsilon))$.

$$U_1 := q(\mathsf{y}_2, x) \cap \ell_r(x_1, x) = \left(\left(\sqrt{3}\,y/6 + x/2\right)\,r + 1/2 - \varepsilon/\sqrt{3},\, \left(y/2 + \sqrt{3}\,x/2\right)\,r - \sqrt{3}/2 + \varepsilon\right), \text{ and}$$
$$U_2 := q(\mathsf{y}_3, x) \cap \ell_r(x_1, x) = \left(\left(y/\sqrt{3} + x\right)\,r + \varepsilon/\sqrt{3} - 1/2,\, \sqrt{3}/2 - \varepsilon\right).$$

Let $\mathscr{P}(a_1, a_2, \ldots, a_n)$ denote the polygon with vertices $a_1, a_2, \ldots, a_n$. If x_1 is below $q_2(x)$, then $N_{PE}^r(x_1, \varepsilon) = A(N_{PE}^r(x_1)) \setminus T(\mathsf{y}_1, \varepsilon) = \mathscr{P}(Q_1, N_1, N_2, Q_6)$; if x_1 is between $q_2(x)$ and $\ell_s(x)$, then $N_{PE}^r(x_1, \varepsilon) = \mathscr{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)$; and if x_1 is above $\ell_s(x)$, then $N_{PE}^r(x_1, \varepsilon) = T_\varepsilon$.

For $r \in \left[1, 3/2 - \sqrt{3}\,\varepsilon\right)$, since ε small enough that $q_2(x) \cap T_s = \emptyset$, then $N(x, \varepsilon) \subsetneq T_\varepsilon$ for all $x \in T_s \setminus T(\mathsf{y}_1, \varepsilon)$. Then

$$P(X_2 \in N_{PE}^r(X_1, \varepsilon), X_1 \in T_s \setminus T(\mathsf{y}_1, \varepsilon)) = \int_{s_1}^{s_2} \int_{q(\mathsf{y}_1, x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1, \varepsilon))}{A(T_\varepsilon)^2} dy dx$$
$$+ \int_{s_2}^{1/2} \int_{0}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1, \varepsilon))}{A(T_\varepsilon)^2} dy dx = -\frac{(576\,r^2 - 1152)\varepsilon^4 + 288\,\varepsilon^2 - 37\,r^2}{1296\,(2\,\varepsilon - 1)^2 (2\,\varepsilon + 1)^2}.$$

where $A(N_{PE}^r(x_1,\varepsilon)) = A(\mathscr{P}(Q_1,N_1,N_2,Q_6)) = \left(\frac{\sqrt{3}}{12}\,y^2 + \frac{1}{2}\,x\,y + \frac{\sqrt{3}}{4}x^2\right)\, r^2 - \frac{\sqrt{3}}{3}\,\varepsilon^2$ and $A(T_\varepsilon) = \sqrt{3}/4 - \sqrt{3}\,\varepsilon^2$ and $\ell_{am}\,(x) = x/\sqrt{3}$ is the equation of the line segment $\overline{\mathrm{y}_1 M}_C$.

Hence for $r \in [1, 3/2)$,

$$\mu(N_{PE}^r) = -\frac{(576\,r^2 - 1152)\varepsilon^4 + 288\,\varepsilon^2 - 37\,r^2}{216\,(2\,\varepsilon - 1)^2(2\,\varepsilon + 1)^2}.$$

For $r \in \left[3/2, 2 - 4\,\varepsilon/\sqrt{3}\right)$, $\ell_s(x)$ crosses through $\overline{M_3 M}_C$. Since ε small enough so that $q_2(x)$ does the same. So x_1 below $q_2(x)$ is equivalent to $x_1 \in R_1(\varepsilon)$ where

$$R_1(\varepsilon) = \left\{(x,y) \in [s_3, s_5] \times [q_2(x), \ell_{am}\,(x)] \cup [s_2, s_3] \times [q_2(x), \ell_{am}\,(x)] \cup [s_3, 1/2] \times [0, q_2(x)]\right\}.$$

Then

$$P(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_1(\varepsilon)) = \int_0^{1/2} \int_0^{\ell_{am}\,(x)} \frac{A(N_{PE}^r(x_1,\varepsilon))}{A(T_\varepsilon)^2} dydx$$
$$= \int_{s_1}^{s_2} \int_{q(\mathrm{y}_1,x)}^{\ell_{am}\,(x)} \frac{A(\mathscr{P}(Q_1,N_1,N_2,Q_6))}{A(T_\varepsilon)^2} dydx + \int_{s_2}^{s_3} \int_0^{\ell_{am}\,(x)} \frac{A(\mathscr{P}(Q_1,N_1,N_2,Q_6))}{A(T_\varepsilon)^2} dydx +$$
$$\int_{s_3}^{1/2} \int_0^{q_2(x)} \frac{A(\mathscr{P}(Q_1,N_1,N_2,Q_6))}{A(T_\varepsilon)^2} dydx = \Big[(384\,r^2 + 576 - 192\,r^4)\varepsilon^4 + 512\,\varepsilon^3\sqrt{3}\,r + (288\,r^2$$
$$- 1728)\varepsilon^2 + \left(-576\,\sqrt{3}\,r + 864\,\sqrt{3}\right)\varepsilon - 9\,r^4 - 324 + 288\,r\Big] \Big/ \Big[432\,((2\,\varepsilon - 1)^2(2\,\varepsilon + 1)^2 r^2)\Big]$$

where $A(\mathscr{P}(Q_1,N_1,N_2,Q_6))$ is same as before.

Next, x_1 between $q_2(x)$ and $\ell_s(x)$ is equivalent to $x_1 \in R_2(\varepsilon)$ where

$$R_2(\varepsilon) = \left\{(x,y) \in [s_9, s_2] \times [q_1(x), \ell_{am}\,(x)] \cup [s_5, 1/2] \times [q_2(x), \ell_s(x)]\right\}.$$

Then

$$P(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_2(\varepsilon)) = \int_{s_3}^{s_5} \int_{q_2(x)}^{\ell_{am}\,(x)} \frac{A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))}{A(T_\varepsilon)^2} dydx$$
$$+ \int_{s_5}^{1/2} \int_0^{q_2(x)} \frac{A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))}{A(T_\varepsilon)^2} dydx =$$
$$- \frac{2\,\sqrt{3}\,\varepsilon\,(10\,\varepsilon^3\sqrt{3} + (32\,r - 24)\varepsilon^2 + \left(-27\,\sqrt{3} + 12\,\sqrt{3}\,r\right)\,\varepsilon - 18\,r + 27)}{27\,(4\,\varepsilon^2 - 1)^2 r^2}$$

where

$$A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6)) = -\sqrt{3}\,\varepsilon^2 + \left(2 - 2\,r\,y/\sqrt{3} - 2\,r\,x\right)\,\varepsilon + r\,y + \sqrt{3}\,r\,x - \sqrt{3}/2 - \sqrt{3}\,r^2\,y^2/12 - r^2\,x\,y/2 - \sqrt{3}\,r^2\,x^2/4.$$

Furthermore, x_1 above $\ell_s(x)$ is equivalent to $x_1 \in R_3(\varepsilon)$ where $R_3(\varepsilon) = \{(x,y) \in [s_5, 1/2] \times [\ell_{am}(x), \ell_s(x)]\}$.

Then $P(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_3(\varepsilon)) = \int_{s_5}^{1/2} \int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T_\varepsilon)} dy dx = -\frac{(2\,r-3)^2}{6\,(2\,\varepsilon-1)(2\,\varepsilon+1)\,r^2}$.

Hence for $r \in [3/2, 2)$,

$$\begin{aligned}\mu_S(N_{PE}^r,\varepsilon) &= 6\left(-\left[(-384\,r^2+384+192\,r^4)\varepsilon^4 + (-768\,\sqrt{3}+512\,\sqrt{3}\,r)\varepsilon^3 + (1728-2304\,r \right.\right.\\ &\quad \left.\left. +864\,r^2)\varepsilon^2 - 288\,r^2 - 324 + 9\,r^4 + 576\,r)\right] \Big/ \left[432\,((2\,\varepsilon-1)^2(2\,\varepsilon+1)^2\,r^2)\right]\right)\\ &= -\left[(-384\,r^2+384+192\,r^4)\varepsilon^4 + (-768\,\sqrt{3}+512\,\sqrt{3}\,r)\varepsilon^3 + (1728-2304\,r \right.\\ &\quad \left. +864\,r^2)\varepsilon^2 - 288\,r^2 - 324 + 9\,r^4 + 576\,r\right] \Big/ \left[72\,((2\,\varepsilon+1)^2(2\,\varepsilon-1)^2 r^2)\right].\end{aligned}$$

For $r \in [2,\infty)$, $\ell_s(x)$ crosses through $\overline{\mathsf{y}_1 M_3}$, so the same types of $N_{PE}^r(x_1,\varepsilon)$ occur as above. The explicit forms of $R_j(\varepsilon)$ for $j \in \{1,2,3\}$ change and are given by

$$R_1(\varepsilon) = \{(x,y) \in [s_1,s_2] \times [q(\mathsf{y}_1,x), \ell_{am}(x)] \cup [s_2,s_3] \times [0, \ell_{am}(x)] \cup [s_3,s_4] \times [0, q_2(x)(x)]\}$$

$$R_2(\varepsilon) = \{(x,y) \in [s_3,s_5] \times [q_2(x), \ell_{am}(x)] \cup [s_5,s_4] \times [q_2(x),, \ell_s(x)] \cup [s_4,s_6] \times [0, \ell_s(x)]\}$$

$$R_3(\varepsilon) = \{(x,y) \in [s_5,s_6] \times [\ell_s(x), \ell_{am}(x)] \cup [s_6, 1/2] \times [0, \ell_{am}(x)]\}.$$

Then

$$\begin{aligned}P(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_1(\varepsilon)) &= \int_{s_1}^{s_2} \int_{q(\mathsf{y}_1,x)}^{\ell_{am}(x)} \frac{A(\mathscr{P}(Q_1,N_1,N_2,Q_6))}{A(T_\varepsilon)^2} dy dx\\ &\quad + \int_{s_2}^{s_3} \int_0^{\ell_{am}(x)} \frac{A(\mathscr{P}(Q_1,N_1,N_2,Q_6))}{A(T_\varepsilon)^2} dy dx + \int_{s_3}^{s_4} \int_0^{q_2(x)} \frac{A(\mathscr{P}(Q_1,N_1,N_2,Q_6))}{A(T_\varepsilon)^2} dy dx\\ &= -\frac{-9 + (-32\,r^2+16\,r^4+16)\varepsilon^4 - 48\,\varepsilon^2 + 24\,\varepsilon\,\sqrt{3}}{36\,(4\,\varepsilon^2-1)^2 r^2}.\end{aligned}$$

where $A(\mathscr{P}(Q_1,N_1,N_2,Q_6))$ is same as before.

Furthermore,

$$P\left(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_2(\varepsilon)\right) = \int_{s_3}^{s_5}\int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))}{A(T_\varepsilon)^2}dydx$$
$$+\int_{s_5}^{s_4}\int_{q_2(x)}^{\ell_s(x)} \frac{A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))}{A(T_\varepsilon)^2}dydx + \int_{s_4}^{s_6}\int_{0}^{\ell_s(x)} \frac{A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))}{A(T_\varepsilon)^2}dydx$$
$$= \frac{2\sqrt{3}\,\varepsilon\,(-27\,\varepsilon\sqrt{3}-24\,\varepsilon^2+10\,\varepsilon^3\sqrt{3}+27)}{81\,(4\,\varepsilon^2-1)^2 r^2}.$$

where $A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))$ is same as before.

Next,

$$P\left(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_3(\varepsilon)\right) =$$
$$\int_{s_5}^{s_6}\int_{\ell_s(x)}^{\ell_{am}(x)} \frac{1}{A(T_\varepsilon)}dydx + \int_{s_6}^{1/2}\int_{0}^{\ell_{am}(x)} \frac{1}{A(T_\varepsilon)}dydx = \frac{3-r^2}{6\,(4\,\varepsilon^2-1)\,r^2}.$$

Hence for $r \in [2, \sqrt{3}/(2\,\varepsilon)-1)$,

$$\mu_S(N_{PE}^r,\varepsilon) = 6\left(-\frac{(48\,r^4-32-96\,r^2)\varepsilon^4+64\,\varepsilon^3\sqrt{3}+(72\,r^2-144)\varepsilon^2+27-18\,r^2}{108\,(4\,\varepsilon^2-1)^2 r^2}\right)$$
$$= -\frac{(48\,r^4-32-96\,r^2)\varepsilon^4+64\,\varepsilon^3\sqrt{3}+(72\,r^2-144)\varepsilon^2+27-18\,r^2}{18\,(4\,\varepsilon^2-1)^2 r^2}.$$

For $r=\infty$, it is trivial to see that $\mu(N_{PE}^r)=1$. In fact, for fixed $\varepsilon>0$, $\mu(N_{PE}^r)=1$ for $r \geq \sqrt{3}/(2\,\varepsilon)$.

B.3.1 Derivation of $\mu_S(N_{PE}^r,\varepsilon)$ and $\nu_S(N_{PE}^r,\varepsilon)$ for Segregation with $\varepsilon=\sqrt{3}/8$

For the segregation alternative with $\varepsilon=\sqrt{3}/8$, $\mu_S\left(N_{PE}^r,\varepsilon=\sqrt{3}/8\right)=1$ for $r\geq 4$, so we find $\mu_S\left(N_{PE}^r,\varepsilon=\sqrt{3}/8\right)$ for $r\in[1,4)$. In particular, for the mean we partition $[1,4)$ into five intervals, $[1,9/8)$, $[9/8,3/2)$, $[3/2,2)$, $[2,3)$, and $[3,4)$, and for the covariance into twelve intervals, $[1,12/11)$, $[12/11,9/8)$, $[9/8,\sqrt{6}/11)$, $[\sqrt{6}/11,21/16)$, $[21/16,4/3)$, $[4/3,3/2)$, and $[3/2,\sqrt{3})$, $[\sqrt{3},7/4)$, $[7/4,2)$, $[2,3)$, $[3,7/2)$, $[7/2,4)$. We pick the sample intervals $[3/2,2)$ and $[7/4,2)$ to demonstrate the calculations of the mean and the (asymptotic) variance, respectively. Observe that, by symmetry,

$$\mu_S(N_{PE}^r,\varepsilon) = P(X_2 \in N_{PE}^r(X_1,\varepsilon)) = 6\,P(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_\varepsilon\,(\mathrm{y}_1)).$$

Then $q(\mathsf{y}_1, x) = \sqrt{3}\,(1/4 - x)$, $q(\mathsf{y}_2, x) = \sqrt{3}\,(4\,x - 3)/4$, and $q(\mathsf{y}_3, x) = 3\,\sqrt{3}/8$. See Section B.3 for the definition of $q(\mathsf{y}_j, x)$. Hence, $Q_1 = q(\mathsf{y}_1, x) \cap e_3 = (1/4, 0)$, $Q_2 = q(\mathsf{y}_2, x) \cap e_3 = (3/4, 0)$, $Q_3 = q(\mathsf{y}_2, x) \cap e_1 = (7/8, 3/8)$, $Q_4 = q(\mathsf{y}_3, x) \cap e_1 = \left(5/8, 3\,\sqrt{3}/8\right)$, $Q_5 = q(\mathsf{y}_3, x) \cap e_2 = \left(3/8, 3\,\sqrt{3}/8\right)$, $Q_6 = q(\mathsf{y}_1, x) \cap e_2 = (1/8, 3/8)$. Then $T(\mathsf{y}_1, \varepsilon) = T(\mathsf{y}_1, Q_1, Q_6)$, $T(\mathsf{y}_2, \varepsilon) = T(Q_2, y_2, Q_3)$, and $T(\mathsf{y}_3, \varepsilon) = T(Q_4, Q_5, \mathsf{y}_3)$, and for $\varepsilon = \sqrt{3}/8$, T_ε is the hexagon with vertices, Q_j, $j \in \{1, \ldots, 6\}$.

Furthermore, $q_2(x) = -\sqrt{3}\,(x - 3/(4\,r))$ and $q_3(x) = \sqrt{3}\,(1/r - x)$ (i.e. the same as $\ell_s(x)$ before). See Section B.3 for the definition of $q_j(x)$. Let the x coordinate of $q(\mathsf{y}_1, x) \cap \ell_{am}(x)$ be s_1, of $q_2(x) \cap \ell_{am}(x)$ be s_4, and of $\ell_s(x) \cap \ell_{am}(x)$ be s_{10}; $Q_1 = (s_3, 0)$, $q_2(x) \cap e_3 = (s_6, 0)$, and $\ell_s(x) \cap e_3 = (s_{12}, 0)$. So $s_1 = 3/16$, $s_3 = 1/4$, $s_4 := 9/(16\,r)$, $s_6 = 3/(4\,r) = s_{10}$, and $s_{12} = 1/r$. See Figure B.3.3.

For $x_1 = (x, y) \in R_{CM}(\mathsf{y}_1)$, let

$$U_1 := q(\mathsf{y}_2, x) \cap \ell_r(x_1, x) = \left(\sqrt{3}\left(4\,r\,y + 4\,\sqrt{3}\,r\,x + 3\,\sqrt{3}\right)/24, r\,y/2 + \sqrt{3}\,r\,x/2 - 3\,\sqrt{3}/8\right), \text{ and}$$
$$U_2 := q(\mathsf{y}_3, x) \cap \ell_r(x_1, x) = \left(\sqrt{3}\,\left(8\,r\,y + 8\,\sqrt{3}\,r\,x - 3\,\sqrt{3}\right)/24, 3\,\sqrt{3}/8\right).$$

If x_1 is below $q_2(x)$, then $N^r_{PE}(x_1, \varepsilon) = N^r_{PE}(x_1) \setminus T(\mathsf{y}_1, \sqrt{3}/8) = \mathscr{P}(Q_1, N_1, N_2, Q_6)$; if x_1 is between $q_2(x)$ and $\ell_s(x)$, then $N^r_{PE}(x_1, \varepsilon) = \mathscr{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)$; and if x_1 is above $\ell_s(x)$, then $N^r_{PE}(x_1, \varepsilon) = T_\varepsilon$.

For $r \in [3/2, 2)$, $q_3(x)$ crosses through $\overline{M_3 M_C}$ and $q_2(x)$ crosses $\overline{\mathsf{y}_1 M_3}$. So x_1 below $q_2(x)$ is equivalent to $x_1 \in R_1\left(\sqrt{3}/8\right)$ where

$$R_1\left(\sqrt{3}/8\right) = \left\{(x, y) \in [s_1, s_3] \times [q(\mathsf{y}_1, x), \ell_{am}(x)] \cup [s_3, s_4] \times [0, \ell_{am}(x)] \cup [s_4, s_6] \times [0, q_2(x)]\right\}.$$

Then

$$P\left(X_2 \in N^r_{PE}(X_1, \varepsilon), X_1 \in R_1\left(\sqrt{3}/8\right)\right) = \int_{T_s \setminus T(\mathsf{y}_1, \sqrt{3}/8)} \frac{A(N^r_{PE}(x_1, \varepsilon))}{A(T_\varepsilon)^2} dy dx =$$
$$\int_{s_1}^{s_3} \int_{q(\mathsf{y}_1, x)}^{\ell_{am}(x)} \frac{A(\mathscr{P}(Q_1, N_1, N_2, Q_6))}{A(T_\varepsilon)^2} dy dx + \int_{s_3}^{s_4} \int_0^{\ell_{am}(x)} \frac{A(\mathscr{P}(Q_1, N_1, N_2, Q_6))}{A(T_\varepsilon)^2} dy dx$$
$$+ \int_{s_4}^{s_6} \int_0^{q_2(x)} \frac{A(\mathscr{P}(Q_1, N_1, N_2, Q_6))}{A(T_\varepsilon)^2} dy dx = -\frac{r^4 - 2\,r^2 - 63}{676\,r^2}.$$

where $A(\mathscr{P}(Q_1, N_1, N_2, Q_6)) = \frac{\sqrt{3}}{576}\left(12\,r\,x + 4\,\sqrt{3}\,r\,y + 3\right)\left(12\,r\,x + 4\,\sqrt{3}\,r\,y - 3\right)$ is same as be-

fore.

Next, x_1 between $q_2(x)$ and $\ell_s(x)$ is equivalent to $x_1 \in R_2\left(\sqrt{3}/8\right)$ where

$$R_2\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_4, s_6] \times [q_2(x), \ell_{am}(x)] \cup [s_6, 1/2] \times [0, \ell_s(x)]\right\}.$$

Then

$$P\left(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_2\left(\sqrt{3}/8\right)\right) = \int_{s_4}^{s_6}\int_{q_2(x)}^{\ell_{am}(x)} \frac{A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))}{A(T_\varepsilon)^2} dydx$$
$$+ \int_{s_6}^{1/2}\int_{0}^{\ell_s(x)} \frac{A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6))}{A(T_\varepsilon)^2} dydx = \frac{64\,r^4 - 768\,r^3 + 1824\,r^2 - 128\,r - 1467}{2028\,r^2}.$$

where $A(\mathscr{P}(Q_1,Q_2,U_1,U_2,Q_5,Q_6)) = \left(-\frac{\sqrt{3}}{12}y^2 - \frac{1}{2}\,x\,y - \frac{\sqrt{3}}{4}x^2\right) r^2 + \left(\frac{3}{4}\,y + \frac{3\sqrt{3}}{4}\,x\right) r - \frac{19}{64}\sqrt{3}$.

Furthermore, x_1 above $q_3(x)$ is $x_1 \in R_3\left(\sqrt{3}/8\right)$ where $R_3\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_{10}, 1/2] \times [\ell_s(x), \ell_{am}(x)]\right\}$.

Then

$$P\left(X_2 \in N_{PE}^r(X_1,\varepsilon), X_1 \in R_3\left(\sqrt{3}/8\right)\right) = \int_{s_{10}}^{1/2}\int_{q_3(x)}^{\ell_{am}(x)} \frac{1}{A(T_\varepsilon)} dydx = \frac{8\,(4\,r^2 - 12\,r + 9)}{39\,r^2}.$$

Hence for $r \in [3/2, 2)$,

$$\mu_S\left(N_{PE}^r, \sqrt{3}/8\right) =$$
$$6\left(\frac{61\,r^4 - 768\,r^3 + 3494\,r^2 - 5120\,r + 2466}{2028\,r^2}\right) = \frac{61\,r^4 - 768\,r^3 + 3494\,r^2 - 5120\,r + 2466}{338\,r^2}.$$

For $r \geq 4$, it is trivial to see that $\mu(N_{PE}^r) = 1$.

To find the covariance, we need to find the possible types of $\Gamma_1(x_1, N_{PE}^r, \varepsilon)$ and $N_{PE}^r(x_1,\varepsilon)$ for $r \in [1,4)$. The intersection points of $\xi_j(r,x)$ with $\partial(T(\mathcal{Y}))$ and $\partial(R(\mathsf{y}_j))$ for $j \in \{1,2,3\}$, i.e. $G_1 - G_6$ and $L_1 - L_6$ are same as before. Recall also M_C, M_1, M_2, M_3 and $\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3$. Then $\Gamma_1(x_1, N_{PE}^r, \varepsilon)$ is a polygon whose vertices are a subset of the above points.

There are six cases regarding $\Gamma_1(x_1, N_{PE}^r, \varepsilon)$ and one case for $N_{PE}^r(x_1,\varepsilon)$. Each case j, corresponds to the region $R_j\left(\sqrt{3}/8\right)$ in Figure B.3.3 where $q(\mathsf{y}_1, x)$, $q_2(x)$, $q_3(x)$, s_j for $j \in \{1,3,4,6,10\}$ are same as before and $q_4(x) = -\sqrt{3}\,(4\,x - r)/4$, $q_5(x) = \sqrt{3}\,(2 - r)/4$ and $s_2 = (r-1)/4$, $s_5 = (r^2 - 2\,r + 3)/(4\,r)$, $s_7 = 3\,r/16$, $s_8 = (r-1)/2$. (see Figure B.3.3).

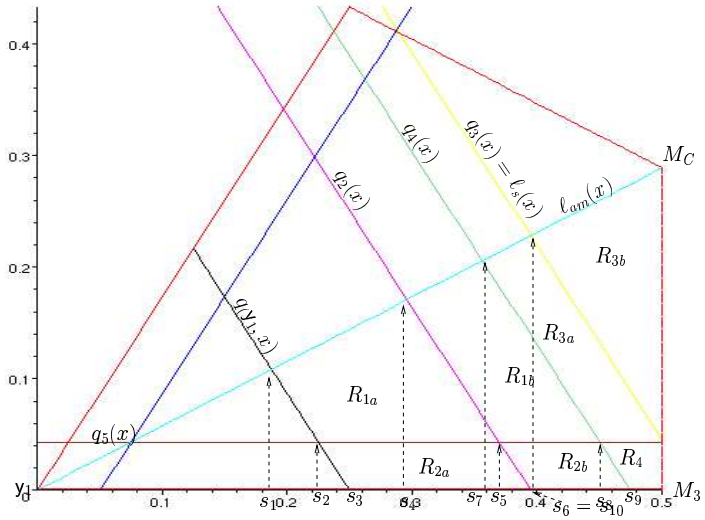


Figure B.3.3: The regions corresponding to the seven cases for $r \in [1, 4)$ with $r = 1.9$

Then, for

$x_1 = (x, y) \in R_1\left(\sqrt{3}/8\right), \Gamma_1\left(x_1, N_{PE}^r\right) = \mathscr{P}(Q_1, G_2, G_3, G_4, G_5, Q_6)$

$x_1 \in R_2\left(\sqrt{3}/8\right), \Gamma_1\left(x_1, N_{PE}^r\right) = \mathscr{P}(Q_1, G_2, G_3, M_2, L_4, L_5, M_3, Q_6)$

$x_1 \in R_3\left(\sqrt{3}/8\right), \Gamma_1\left(x_1, N_{PE}^r\right) = \mathscr{P}(G_1, G_2, G_3, G_4, G_5, G_6)$

$x_1 \in R_4\left(\sqrt{3}/8\right), \Gamma_1\left(x_1, N_{PE}^r\right) = \mathscr{P}(G_1, G_2, G_3, G_4.G_5, G_6)$

The explicit forms of $R_j\left(\sqrt{3}/8\right)$, $j = 1, \ldots, 4$ are as follows:

$$R_1\left(\sqrt{3}/8\right) = \left\{(x, y) \in [s_1, s_2] \times [q(\mathsf{y}_1, x), \ell_{am}(x)] \cup [s_2, s_7] \times [q_5(x), \ell_{am}(x)] \cup [s_7, s_8] \times [q_5(x), q_4(x)]\right\}$$

$$R_2\left(\sqrt{3}/8\right) = \left\{(x, y) \in [s_2, s_3] \times [q(\mathsf{y}_1, x), q_5(x)] \cup [s_3, s_8] \times [0, q_5(x)] \cup [s_8, s_9] \times [0, q_4(x)]\right\}$$

$$R_3\left(\sqrt{3}/8\right) = \left\{(x, y) \in [s_7, s_8] \times [q_3(x), \ell_{am}(x)] \cup [s_8, 1/2] \times [q_5(x), \ell_{am}(x)]\right\}$$

$$R_4\left(\sqrt{3}/8\right) = \left\{(x, y) \in [s_8, s_9] \times [q_4(x), q_5(x)] \cup [s_9, 1/2] \times [0, q_5(x)]\right\}.$$

Let $P_{2N}(N_{PE}^r, \varepsilon) := P\left(\{X_2, X_3\} \subset N_{PE}^r(X_1, \varepsilon)\right)$, $P_{2G}(N_{PE}^r, \varepsilon) := P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{PE}^r, \varepsilon)\right)$, and $P_M(N_{PE}^r, \varepsilon) := P\left(X_2 \in N_{PE}^r(X_1, \varepsilon),\ X_3 \in \Gamma_1(X_1, N_{PE}^r, \varepsilon)\right)$, where $\Gamma_1(X_1,$ $N_{PE}^r, \varepsilon) := \Gamma_1(X_1, N_{PE}^r, M_C) \cap T_\varepsilon$.

Now, by symmetry, $P_{2N}\left(N_{PE}^r, \sqrt{3}/8\right) = 6\, P\left(\{X_2, X_3\} \subset N_{PE}^r\left(X_1, \varepsilon\right),\ X_1 \in T(\mathsf{y}_1, \sqrt{3}/8)\right)$.

For $r \in [7/4, 2)$,

$$P\left(\{X_2, X_3\} \subset N_{PE}^r(X_1, \varepsilon),\ X_1 \in T(\mathsf{y}_1, \sqrt{3}/8)\right) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1, \varepsilon))^2}{A(T_\varepsilon)^3} dy dx$$
$$= -\frac{261\,r^6 - 4608\,r^5 + 29105\,r^4 - 72960\,r^3 + 32575\,r^2 + 78848\,r - 67620}{65910\,r^2}.$$

where $N_{PE}^r(x_1, \varepsilon) = \mathscr{P}(Q_1, N_1, N_2, Q_6)$ for $x_1 \in R_{1a}\left(\sqrt{3}/8\right) \cup R_{2a}\left(\sqrt{3}/8\right)$, $N_{PE}^r(x_1, \varepsilon) = \mathscr{P}(Q_1, Q_2, U_1, U_2, Q_5, Q_6)$ for $x_1 \in R_{1b}\left(\sqrt{3}/8\right) \cup R_{2b}\left(\sqrt{3}/8\right) \cup R_{3a}\left(\sqrt{3}/8\right) \cup R_4$, and $N_{PE}^r(x_1, \varepsilon) = T_\varepsilon$ for $x_1 \in R_{3b}\left(\sqrt{3}/8\right)$, all of whose areas are given above. Hence for $r \in [7/4, 2)$,

$$P(\{X_2, X_3\} \subset N_{PE}^r(X_1, \varepsilon)) = -\frac{261\,r^6 - 4608\,r^5 + 29105\,r^4 - 72960\,r^3 + 32575\,r^2 + 78848\,r - 67620}{10985\,r^2}.$$

Next, by symmetry, $P_{2G}\left(N_{PE}^r, \sqrt{3}/8\right) = 6\,P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, N_{PE}^r, \varepsilon),\ X_1 \in T_s \setminus T(\mathsf{y}_1, \sqrt{3}/8)\right)$, and

$$P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, N_{PE}^r, \varepsilon),\ X_1 \in T_s \setminus T(\mathsf{y}_1, \sqrt{3}/8)\right) = \sum_{j=1}^{4} P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, N_{PE}^r, \varepsilon),\ X_1 \in R_j\left(\sqrt{3}/8\right)\right).$$

For $x_1 \in R_1\left(\sqrt{3}/8\right)$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, N_{PE}^r, \varepsilon),\ X_1 \in R_1\left(\sqrt{3}/8\right)\right) = \int_{s_1}^{s_2} \int_{q(\mathsf{y}_1, x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dy dx$$
$$+ \int_{s_2}^{s_7} \int_{q_5(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dy dx + \int_{s_7}^{s_8} \int_{q_5(x)}^{q_4(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dy dx$$
$$= \frac{18894\,r^6 - 12248\,r^5 - 131375\,r^4 + 45360\,r^3 + 584030\,r^2 - 841816\,r + 337155}{65910\,r^4}.$$

where $A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right) = \frac{\sqrt{3}\left(45\,r^2 + 32\,y\,\sqrt{3}\,x + 96\,x - 48\,x^2 - 80\,y^2 + 32\,\sqrt{3}\,y - 96\right)}{192\,r^2}$.

For $x_1 \in R_2\left(\sqrt{3}/8\right)$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1^r(X_1, N_{PE}^r, \varepsilon),\ X_1 \in R_2\left(\sqrt{3}/8\right)\right) = \int_{s_2}^{s_3} \int_{q(\mathsf{y}_1, x)}^{q_5(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dy dx$$
$$+ \int_{s_3}^{s_8} \int_{0}^{q_5(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dy dx + \int_{s_8}^{s_9} \int_{0}^{q_4(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dy dx$$
$$= -\frac{4\left(4865\,r^6 - 24063\,r^5 + 49460\,r^4 - 73210\,r^3 + 98045\,r^2 - 84107\,r + 29010\right)}{32955\,r^4}.$$

where $A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right) = \frac{\sqrt{3}\left(96+32\,y\,\sqrt{3}\,x+128\,\sqrt{3}\,r\,y-224\,\sqrt{3}\,y-192\,r+96\,x+93\,r^2-48\,x^2+176\,y^2\right)}{192\,r^2}$.

For $x_1 \in R_3\left(\sqrt{3}/8\right)$,

$$
\begin{aligned}
&P\left(\{X_2, X_3\} \subset \Gamma_1^r\left(X_1, N_{PE}^r, \varepsilon\right),\ X_1 \in R_3\left(\sqrt{3}/8\right)\right) \\
&= \int_{s_7}^{s_8}\int_{q_4(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dydx + \int_{s_8}^{1/2}\int_{q_5(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dydx \\
&= -\frac{46293\,r^6 - 100944\,r^5 - 254880\,r^4 + 506880\,r^3 + 829440\,r^2 - 2064384\,r + 1048576}{98865\,r^4}.
\end{aligned}
$$

where $A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right) = \frac{\sqrt{3}\left(3\,r^2+6\,x-6\,x^2-6\,y^2+2\,\sqrt{3}\,y-6\right)}{12\,r^2}$.

For $x_1 \in R_4\left(\sqrt{3}/8\right)$,

$$
\begin{aligned}
&P\left(\{X_2, X_3\} \subset \Gamma_1^r\left(X_1, N_{PE}^r, \varepsilon\right),\ X_1 \in R_4\left(\sqrt{3}/8\right)\right) \\
&= \int_{s_2}^{s_4}\int_{q_2(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dydx + \int_{s_4}^{s_5}\int_{q_3(x)}^{\ell_{am}(x)} \frac{A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)^2}{A(T_\varepsilon)^3} dydx \\
&= \frac{8\left(3577\,r^6 - 20548\,r^5 + 45620\,r^4 - 61760\,r^3 + 79040\,r^2 - 77824\,r + 32256\right)}{32955\,r^4}.
\end{aligned}
$$

where $A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right) = \frac{\sqrt{3}\left(-6\,r+3\,x-3\,x^2+5\,y^2+3\,r^2+3+4\,\sqrt{3}\,r\,y-7\,\sqrt{3}\,y\right)}{6\,r^2}$.

So,

$$
\begin{aligned}
&P_{2G}\left(N_{PE}^r, \sqrt{3}/8\right) = \frac{6\left(19032\,r^6 - 243648\,r^5 + 1118355\,r^4 - 2085120\,r^3 + 1534050\,r^2 - 113664\,r - 233639\right)}{197730\,r^4} \\
&= \frac{19032\,r^6 - 243648\,r^5 + 1118355\,r^4 - 2085120\,r^3 + 1534050\,r^2 - 113664\,r - 233639}{32955\,r^4}.
\end{aligned}
$$

Furthermore, by symmetry,

$$
P_M\left(N_{PE}^r, \sqrt{3}/8\right) = 6\,P\left(X_2 \in N_{PE}^r\left(X_1, \varepsilon\right),\ X_3 \in \Gamma_1^r\left(X_1, N_{PE}^r, \varepsilon\right),\ X_1 \in T_s \setminus T(\mathrm{y}_1, \sqrt{3}/8)\right),
$$

and

$$
\begin{aligned}
&P\left(X_2 \in N_{PE}^r\left(X_1, \varepsilon\right),\ X_3 \in \Gamma_1^r\left(X_1, N_{PE}^r, \varepsilon\right),\ X_1 \in T_s \setminus T(\mathrm{y}_1, \sqrt{3}/8)\right) \\
&\qquad = \sum_j P\left(X_2 \in N_{PE}^r(X_1),\ X_3 \in \Gamma_1\left(X_1, N_{PE}^r\right),\ X_1 \in R_j\left(\sqrt{3}/8\right)\right)
\end{aligned}
$$

where $j \in \{1a,\, 1b,\, 2a,\, 2b, 3a,\, 3b, 4\}$. The explicit forms of these regions are given by

$$R_{1a}\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_1, s_2] \times [q(\mathrm{y}_1, x), \ell_{am}(x)] \cup [s_2, s_4] \times [q_5(x), \ell_{am}(x)] \cup [s_4, s_5] \times [q_5(x), q_2(x)]\right\},$$

$$R_{1b}\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_4, s_7] \times [q_2(x), \ell_{am}(x)] \cup [s_7, s_5] \times [q_2(x), q_4(x)] \cup [s_5, s_8] \times [q_5(x), q_4(x)]\right\},$$

$$R_{2a}\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_2, s_3] \times [q(\mathrm{y}_1, x), q_5(x)] \cup [s_3, s_5] \times [0, q_5(x)] \cup [s_5, s_6] \times [0, q_2(x)]\right\},$$

$$R_{2b}\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_5, s_6] \times [q_2(x), q_5(x)] \cup [s_6, s_8] \times [0, q_5(x)] \cup [s_8, s_9] \times [0, q_4(x)]\right\},$$

$$R_{3a}\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_7, s_{10}] \times [q_4(x), \ell_{am}(x)] \cup [s_{10}, s_8] \times [q_4(x), q_3(x)] \cup [s_8, 1/2] \times [q_5(x), q_4(x)]\right\},$$

$$R_{3b}\left(\sqrt{3}/8\right) = \left\{(x,y) \in [s_{10}, 1/2] \times [q_3(x), \ell_{am}(x)]\right\}.$$

$R_4\left(\sqrt{3}/8\right)$ is the same as before.

For $x_1 \in R_{1a}\left(\sqrt{3}/8\right)$,

$$P\left(X_2 \in N_{PE}^r(X_1, \varepsilon),\ X_3 \in \Gamma_1^r(X_1, N_{PE}^r, \varepsilon),\ X_1 \in R_{1a}\left(\sqrt{3}/8\right)\right) =$$

$$\int_{s_1}^{s_2} \int_{q(\mathrm{y}_1, x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1,\varepsilon))\, A(\Gamma_1(x_1, N_{PE}^r, \varepsilon))}{A(T_\varepsilon)^3} dydx + \int_{s_2}^{s_4} \int_{q_5(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1,\varepsilon))\, A(\Gamma_1(x_1, N_{PE}^r, \varepsilon))}{A(T_\varepsilon)^3} dydx$$

$$+ \int_{s_4}^{s_5} \int_{q_5(x)}^{q_2(x)} \frac{A(N_{PE}^r(x_1,\varepsilon))\, A(\Gamma_1(x_1, N_{PE}^r, \varepsilon))}{A(T_\varepsilon)^3} dydx$$

$$= -\Big[2960\, r^9 - 3045\, r^8 - 21504\, r^7 - 28554\, r^6 + 101040\, r^5 + 205785\, r^4 - 550080\, r^3 + 391392\, r^2$$

$$- 114048\, r + 12150\Big] \Big/ \Big[395460\, r^6\Big].$$

For $x_1 \in R_{1b}\left(\sqrt{3}/8\right)$,

$$P\left(X_2 \in N_{PE}^r(X_1, \varepsilon),\ X_3 \in \Gamma_1^r(X_1, N_{PE}^r, \varepsilon),\ X_1 \in R_{1b}\left(\sqrt{3}/8\right)\right) =$$

$$\int_{s_4}^{s_7} \int_{q_2(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r(x_1,\varepsilon))\, A(\Gamma_1(x_1, N_{PE}^r, \varepsilon))}{A(T_\varepsilon)^3} dydx + \int_{s_7}^{s_5} \int_{q_2(x)}^{q_4(x)} \frac{A(N_{PE}^r(x_1,\varepsilon))\, A(\Gamma_1(x_1, N_{PE}^r, \varepsilon))}{A(T_\varepsilon)^3} dydx$$

$$+ \int_{s_5}^{s_8} \int_{q_5(x)}^{q_4(x)} \frac{A(N_{PE}^r(x_1,\varepsilon))\, A(\Gamma_1(x_1, N_{PE}^r, \varepsilon))}{A(T_\varepsilon)^3} dydx$$

$$= -\Big[(r^2-3)(3111\, r^{10} + 96\, r^9 - 67659\, r^8 + 26528\, r^7 + 341223\, r^6 - 352896\, r^5 - 177291\, r^4 + 229632\, r^3$$

$$+ 23922\, r^2 - 22464\, r + 3078)\Big] \Big/ \Big[395460\, r^6\Big].$$

For $x_1 \in R_{2a}\left(\sqrt{3}/8\right)$,

$$P\left(X_2 \in N_{PE}^r\left(X_1,\varepsilon\right),\ X_3 \in \Gamma_1^r\left(X_1,N_{PE}^r,\varepsilon\right),\ X_1 \in R_{2a}\left(\sqrt{3}/8\right)\right) =$$
$$\int_{s_2}^{s_3}\int_{q(y_1,x)}^{q_5(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx + \int_{s_3}^{s_5}\int_0^{q_5(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx$$
$$+\int_{s_5}^{s_6}\int_0^{q_2(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx$$
$$= \frac{2\left(r-2\right)(r-3)(530\,r^6+1225\,r^5+1697\,r^4-3399\,r^3-3027\,r^2+5274\,r-1188)}{98865\,r^5}.$$

For $x_1 \in R_{2b}\left(\sqrt{3}/8\right)$,

$$P\left(X_2 \in N_{PE}^r\left(X_1,\varepsilon\right),\ X_3 \in \Gamma_1^r\left(X_1,N_{PE}^r,\varepsilon\right),\ X_1 \in R_{2b}\left(\sqrt{3}/8\right)\right) =$$
$$\int_{s_5}^{s_6}\int_{q_2(x)}^{q_5(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx + \int_{s_6}^{s_8}\int_0^{q_5(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx$$
$$+\int_{s_8}^{s_9}\int_0^{q_4(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx$$
$$= \frac{2\left(r-2\right)(r^2-3)(467\,r^8-50\,r^7-7789\,r^6+1290\,r^5+16083\,r^4-4110\,r^3-7311\,r^2-810\,r+702)}{98865\,r^5}.$$

For $x_1 \in R_{3a}\left(\sqrt{3}/8\right)$,

$$P\left(X_2 \in N_{PE}^r\left(X_1,\varepsilon\right),\ X_3 \in \Gamma_1^r\left(X_1,N_{PE}^r,\varepsilon\right),\ X_1 \in R_{3a}\left(\sqrt{3}/8\right)\right) =$$
$$\int_{s_7}^{s_{10}}\int_{q_4(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx + \int_{s_{10}}^{s_8}\int_{q_4(x)}^{q_3(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx$$
$$+\int_{s_8}^{1/2}\int_{q_5(x)}^{q_3(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx$$
$$= \Big[3934\,r^{12}-11040\,r^{11}-2352\,r^{10}-283680\,r^9+1239855\,r^8-751008\,r^7-3225344\,r^6$$
$$+6125568\,r^5-3847680\,r^4+81920\,r^3+1843200\,r^2-2045952\,r+905472\Big]\Big/\Big[395460\,r^6\Big].$$

For $x_1 \in R_{3b}\left(\sqrt{3}/8\right)$,

$$P\left(X_2 \in N_{PE}^r\left(X_1,\varepsilon\right),\ X_3 \in \Gamma_1^r\left(X_1,N_{PE}^r,\varepsilon\right),\ X_1 \in R_{3b}\left(\sqrt{3}/8\right)\right) =$$
$$\int_{s_{10}}^{1/2}\int_{q_3(x)}^{\ell_{am}(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1,N_{PE}^r,\varepsilon)\right)}{A(T_\varepsilon)^3}dydx = \frac{64\left(2\,r^4-4\,r^2+4\,r-3\right)(2\,r-3)^2}{507\,r^6}.$$

For $x_1 \in R_4\left(\sqrt{3}/8\right)$,

$$
\begin{aligned}
&P\left(X_2 \in N_{PE}^r\left(X_1,\varepsilon\right),\ X_3 \in \Gamma_1^r\left(X_1, N_{PE}^r, \varepsilon\right),\ X_1 \in R_4\left(\sqrt{3}/8\right)\right) \\
&= \int_{s_8}^{s_9}\int_{q_4(x)}^{q_5(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)}{A(T_\varepsilon)^3} dydx + \int_{s_9}^{1/2}\int_{0}^{q_5(x)} \frac{A(N_{PE}^r\left(x_1,\varepsilon\right))\,A\left(\Gamma_1(x_1, N_{PE}^r, \varepsilon)\right)}{A(T_\varepsilon)^3} dydx \\
&= -\frac{8\left(48\,r^6 - 22\,r^5 - 73\,r^4 - 3114\,r^3 + 4391\,r^2 + 1574\,r - 2850\right)(r-2)^2}{32955\,r^2}.
\end{aligned}
$$

Then

$$
\begin{aligned}
&P_M(N_{PE}^r,\varepsilon) = 6\left(-\left[49\,r^{12} - 1536\,r^{11} + 17952\,r^{10} - 129280\,r^9 + 609420\,r^8 - 1728768\,r^7 + 2757670\,r^6\right.\right. \\
&\left.\left. - 3013632\,r^5 + 3418140\,r^4 - 3829760\,r^3 + 3026880\,r^2 - 1571616\,r + 445284\right]\Big/\left[395460\,r^6\right]\right) \\
&= -\left[49\,r^{12} - 1536\,r^{11} + 17952\,r^{10} - 129280\,r^9 + 609420\,r^8 - 1728768\,r^7 + 2757670\,r^6\right. \\
&\left. - 3013632\,r^5 + 3418140\,r^4 - 3829760\,r^3 + 3026880\,r^2 - 1571616\,r + 445284\right]\Big/\left[65910\,r^6\right].
\end{aligned}
$$

So

$$
\begin{aligned}
&\mathbf{E}^{S}_{\sqrt{3}/8}[h_{12}(N_{PE}^r)\,h_{13}(N_{PE}^r)] = -\left[49\,r^{12} - 1536\,r^{11} + 18735\,r^{10} - 143104\,r^9 + 677703\,r^8 - 1704000\,r^7 + \right. \\
&\left. 1737040\,r^6 - 691968\,r^5 + 1681230\,r^4 - 3716096\,r^3 + 3260519\,r^2 - 1571616\,r + 445284\right]\Big/\left[3295\,r^6\right].
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\nu_S\left(r, \sqrt{3}/8\right) = \\
&-\left[637\,r^{12} - 19968\,r^{11} + 299370\,r^{10} - 3265792\,r^9 + 24051519\,r^8 - 112023360\,r^7 + 328179640\,r^6\right. \\
&\left. -602490624\,r^5 + 673558110\,r^4 - 427086848\,r^3 + 133604087\,r^2 - 20431008\,r + 5788692\right]\Big/\left[428415\,r^6\right].
\end{aligned}
$$

Derivation of $\mu_S(N_{PE}^r,\varepsilon)$ and $\nu_S(N_{PE}^r,\varepsilon)$ for segregation with $\varepsilon = \sqrt{3}/4$ and with $\varepsilon = 2\sqrt{3}/7$ are similar.

B.4 The Mean $\mu(N_{PE}^r,\varepsilon)$ Under Segregation and Association Alternatives

Derivation of $\mu(N_{PE}^r,\varepsilon)$ involves detailed geometric calculations and partitioning of the space of (r,ε,x_1) for $r \in [1,\infty)$, $\varepsilon \in \left[0,\sqrt{3}/3\right)$, and $x_1 \in T_S = T(\mathrm{y}_1, M_3, M_C)$.

B.4.1 $\mu_S(N^r_{PE}, \varepsilon)$ Under Segregation Alternatives

Under segregation, we compute $\mu_S(N^r_{PE}, \varepsilon)$ explicitly.

For $\varepsilon \in \left[0, \sqrt{3}/8\right)$,

$$\mu_S(N^r_{PE}, \varepsilon) = \sum_{j=1}^{7} \varpi_{1,j}(r, \varepsilon)\, \mathbf{I}(r \in \mathcal{I}_j),$$

where

$$\varpi_{1,1}(r, \varepsilon) = -\frac{576\, r^2 \varepsilon^4 - 1152\, \varepsilon^4 - 37\, r^2 + 288\, \varepsilon^2}{216\, (2\,\varepsilon + 1)^2 (2\,\varepsilon - 1)^2},$$

$$\begin{aligned}\varpi_{1,2}(r, \varepsilon) = -\Big[&576\, r^4\, \varepsilon^4 - 1152\, r^2 \varepsilon^4 + 91\, r^4 + 512\, \sqrt{3}\, r^3 \varepsilon + 2592\, r^2 \varepsilon^2 + 1536\, \sqrt{3}\, r\, \varepsilon^3 + 1152\, \varepsilon^4 \\ &- 768\, r^3 - 2304\, \sqrt{3}\, r^2 \varepsilon - 6912\, r\, \varepsilon^2 - 2304\, \sqrt{3}\, \varepsilon^3 + 1728\, r^2 + 3456\, \sqrt{3}\, r\, \varepsilon + 5184\, \varepsilon^2 \\ &- 1728\, r - 1728\, \sqrt{3}\, \varepsilon + 648\Big] \Big/ \Big[216\, r^2 (2\,\varepsilon + 1)^2 (2\,\varepsilon - 1)^2\Big],\end{aligned}$$

$$\begin{aligned}\varpi_{1,3}(r, \varepsilon) = -\Big[&192\, r^4\, \varepsilon^4 - 384\, r^2 \varepsilon^4 + 9\, r^4 + 864\, r^2 \varepsilon^2 + 512\, \sqrt{3}\, r\, \varepsilon^3 + 384\, \varepsilon^4 - 2304\, r\, \varepsilon^2 - 768\, \sqrt{3}\, \varepsilon^3 \\ &- 288\, r^2 + 1728\, \varepsilon^2 + 576\, r - 324\Big] \Big/ \Big[72\, r^2 (2\,\varepsilon + 1)^2 (2\,\varepsilon - 1)^2\Big],\end{aligned}$$

$$\begin{aligned}\varpi_{1,4}(r, \varepsilon) = -\Big[&192\, r^4\, \varepsilon^4 - 384\, r^2 \varepsilon^4 - 9\, r^4 - 96\, \sqrt{3}\, r^3 \varepsilon + 288\, r^2 \varepsilon^2 - 128\, \varepsilon^4 + 144\, r^3 + 576\, \sqrt{3}\, r^2 \varepsilon + 256 \\ &\sqrt{3}\, \varepsilon^3 - 720\, r^2 - 1152\, \sqrt{3}\, r\, \varepsilon - 576\, \varepsilon^2 + 1152\, r + 768\, \sqrt{3}\, \varepsilon - 612\Big] \Big/ \Big[72\, r^2 (2\,\varepsilon + 1)^2 (2\,\varepsilon - 1)^2\Big],\end{aligned}$$

$$\varpi_{1,5}(r, \varepsilon) = -\frac{48\, r^4 \varepsilon^4 - 96\, r^2 \varepsilon^4 + 72\, r^2 \varepsilon^2 - 32\, \varepsilon^4 + 64\, \sqrt{3}\varepsilon^3 - 18\, r^2 - 144\, \varepsilon^2 + 27}{18\, r^2 (2\,\varepsilon + 1)^2 (2\,\varepsilon - 1)^2},$$

$$\varpi_{1,6}(r, \varepsilon) = \frac{48\, r^4 \varepsilon^4 + 256\, r^3 \varepsilon^4 - 128\, \sqrt{3} r^3 \varepsilon^3 + 288\, r^2 \varepsilon^4 - 192\, \sqrt{3} r^2 \varepsilon^3 + 72\, r^2 \varepsilon^2 + 18\, r^2 + 48\, \sqrt{3}\, \varepsilon - 45}{18\, (2\,\varepsilon + 1)^2 (2\,\varepsilon - 1)^2 r^2},$$

$$\varpi_{1,7}(r, \varepsilon) = 1,$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 3/2 - \sqrt{3}\,\varepsilon\right)$, $\mathcal{I}_2 = \left[3/2 - \sqrt{3}\,\varepsilon, 3/2\right)$, $\mathcal{I}_3 = \left[3/2, 2 - 4\,\varepsilon/\sqrt{3}\right)$, $\mathcal{I}_4 = \left[2 - 4\,\varepsilon/\sqrt{3}, 2\right)$, $\mathcal{I}_5 = \left[2, \sqrt{3}/(2\,\varepsilon) - 1\right)$, $\mathcal{I}_6 = \left[\sqrt{3}/(2\,\varepsilon) - 1, \sqrt{3}/(2\,\varepsilon)\right)$, and $\mathcal{I}_7 = \left[\sqrt{3}/(2\,\varepsilon), \infty\right)$.

For $\varepsilon \in \left[\sqrt{3}/8, \sqrt{3}/6\right)$,

$$\mu_S(N^r_{PE}, \varepsilon) = \sum_{j=1}^{7} \varpi_{2,j}(r, \varepsilon)\, \mathbf{I}(r \in \mathcal{I}_j),$$

where $\varpi_{2,j}(r,\varepsilon) = \varpi_{1,j}(r,\varepsilon)$ for $j \in \{1,2,4,5,6\}$, and for $j \in \{3,7\}$,

$$\varpi_{2,3}(r,\varepsilon) = -\Big[576\,r^4\varepsilon^4 - 1152\,r^2\varepsilon^4 + 37\,r^4 + 224\,\sqrt{3}r^3\,\varepsilon + 864\,r^2\varepsilon^2 - 384\,\varepsilon^4 - 336\,r^3 - 576$$
$$\sqrt{3}r^2\,\varepsilon + 768\,\sqrt{3}\varepsilon^3 + 432\,r^2 - 1728\,\varepsilon^2 + 576\,\sqrt{3}\,\varepsilon - 216\Big] \Big/ \Big[216\,r^2(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2\Big],$$
$$\varpi_{2,7}(r,\varepsilon) = 1,$$

with the corresponding intervals $\mathcal{I}_1 = \Big[1, 3/2 - \sqrt{3}\,\varepsilon\Big)$, $\mathcal{I}_2 = \Big[3/2 - \sqrt{3}\,\varepsilon, 2 - 4\,\varepsilon/\sqrt{3}\Big)$, $\mathcal{I}_3 = \Big[2 - 4\,\varepsilon/\sqrt{3}, 3/2\Big)$, $\mathcal{I}_4 = [3/2, 2)$, $\mathcal{I}_5 = \Big[2, \sqrt{3}/(2\,\varepsilon) - 1\Big)$, $\mathcal{I}_6 = \Big[\sqrt{3}/(2\,\varepsilon) - 1, \sqrt{3}/(2\,\varepsilon)\Big)$, and $\mathcal{I}_5 = \Big[\sqrt{3}/(2\,\varepsilon), \infty\Big)$.

For $\varepsilon \in \Big[\sqrt{3}/6, \sqrt{3}/4\Big)$,

$$\mu_S(N_{PE}^r, \varepsilon) = \sum_{j=1}^{6} \varpi_{3,j}(r,\varepsilon)\,\mathbf{I}(r \in \mathcal{I}_j),$$

where $\varpi_{3,1}(r,\varepsilon) = \varpi_{1,2}(r,\varepsilon)$ and

$$\varpi_{3,2}(r,\varepsilon) = -\Big[576\,r^4\,\varepsilon^4 - 1152\,r^2\varepsilon^4 + 37\,r^4 + 224\,\sqrt{3}\,r^3\varepsilon + 864\,r^2\varepsilon^2 - 384\,\varepsilon^4 - 336\,r^3 - 576\,\sqrt{3}\,r^2\varepsilon$$
$$+ 768\,\sqrt{3}\,\varepsilon^3 + 432\,r^2 - 1728\,\varepsilon^2 + 576\,\sqrt{3}\,\varepsilon - 216\Big] \Big/ \Big[216\,r^2(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2\Big],$$
$$\varpi_{3,3}(r,\varepsilon) = \Big[576\,r^2\varepsilon^4 + 3072\,r\,\varepsilon^4 - 1536\,\sqrt{3}\,r\,\varepsilon^3 + 3456\,\varepsilon^4 - 2304\,\sqrt{3}\,\varepsilon^3 - 37\,r^2 - 224\,\sqrt{3}\,r\,\varepsilon$$
$$+ 864\,\varepsilon^2 + 336\,r + 576\,\sqrt{3}\,\varepsilon - 432\Big] \Big/ \Big[216\,(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2\Big],$$
$$\varpi_{3,4}(r,\varepsilon) = \Big[192\,r^4\varepsilon^4 + 1024\,r^3\varepsilon^4 - 512\,\sqrt{3}\,r^3\varepsilon^3 + 1152\,r^2\varepsilon^4 - 768\,\sqrt{3}\,r^2\varepsilon^3 + 9\,r^4 + 96\,\sqrt{3}\,r^3\varepsilon +$$
$$288\,r^2\varepsilon^2 - 144\,r^3 - 576\,\sqrt{3}\,r^2\varepsilon + 720\,r^2 + 1152\,\sqrt{3}\,r\,\varepsilon - 1152\,r - 576\,\sqrt{3}\,\varepsilon + 540\Big] \Big/ \Big[72\,r^2$$
$$(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2\Big],$$
$$\varpi_{3,5}(r,\varepsilon) = \frac{48\,r^4\,\varepsilon^4 + 256\,r^3\varepsilon^4 - 128\,\sqrt{3}\,r^3\varepsilon^3 + 288\,r^2\varepsilon^4 - 192\,\sqrt{3}\,r^2\varepsilon^3 + 72\,r^2\varepsilon^2 + 18\,r^2 + 48\,\sqrt{3}\,\varepsilon - 45}{18\,r^2(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2},$$
$$\varpi_{3,6}(r,\varepsilon) = 1,$$

with the corresponding intervals $\mathcal{I}_1 = \Big[1, 2 - 4\,\varepsilon/\sqrt{3}\Big)$, $\mathcal{I}_2 = \Big[2 - 4\,\varepsilon/\sqrt{3}, \sqrt{3}/(2\,\varepsilon) - 1\Big)$, $\mathcal{I}_3 = \Big[\sqrt{3}/(2\,\varepsilon) - 1, 3/2\Big)$, $\mathcal{I}_4 = [3/2, 2)$, $\mathcal{I}_5 = \Big[2, \sqrt{3}/(2\,\varepsilon)\Big)$, and $\mathcal{I}_5 = \Big[\sqrt{3}/(2\,\varepsilon), \infty\Big)$.

For $\varepsilon \in \Big[\sqrt{3}/4, \sqrt{3}/3\Big)$,

$$\mu_S(N_{PE}^r, \varepsilon) = \sum_{j=1}^{3} \varpi_{4,j}(r,\varepsilon)\,\mathbf{I}(r \in \mathcal{I}_j),$$

where

$$\varpi_{4,1}(r,\varepsilon) = -\frac{9\,r^2\varepsilon^2 + 2\sqrt{3}\,r^2\varepsilon + 48\,r\,\varepsilon^2 + r^2 - 16\sqrt{3}\,r\,\varepsilon - 90\,\varepsilon^2 - 12\,r + 36\sqrt{3}\,\varepsilon}{18\,\left(3\,\varepsilon - \sqrt{3}\right)^2},$$

$$\begin{aligned}\varpi_{4,2}(r,\varepsilon) = -\Big[&9\,r^4\varepsilon^4 - 4\sqrt{3}\,r^4\varepsilon^3 + 48\,r^3\varepsilon^4 - 48\sqrt{3}\,r^3\varepsilon^3 - 90\,r^2\varepsilon^4 + 36\,r^3\varepsilon^2 + 96\sqrt{3}\,r^2\varepsilon^3 \\ &- 126\,r^2\varepsilon^2 - 32\sqrt{3}\,r\,\varepsilon^3 - 48\,\varepsilon^4 + 36\sqrt{3}\,r^2\varepsilon + 144\,r\,\varepsilon^2 + 96\sqrt{3}\,\varepsilon^3 - 18\,r^2 - 72\sqrt{3}\,r\,\varepsilon \\ &- 216\,\varepsilon^2 + 36\,r + 72\sqrt{3}\,\varepsilon - 27\Big] \Big/ \Big[2\,\left(3\,\varepsilon - \sqrt{3}\right)^4 r^2\Big],\end{aligned}$$

$$\varpi_{4,3}(r,\varepsilon) = 1,$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 3 - 2\,\varepsilon/\sqrt{3}\right)$, $\mathcal{I}_2 = \left[3 - 2\,\varepsilon/\sqrt{3}, \sqrt{3}/\varepsilon - 2\right)$, and $\mathcal{I}_3 = \left[\sqrt{3}/\varepsilon - 2, \infty\right)$.

B.4.2 $\mu_A\left(N^r_{PE}, \varepsilon\right)$ Under Association Alternatives

Under association, we compute $\mu_A\left(N^r_{PE}, \varepsilon\right)$ explicitly.

For $\varepsilon \in \left[0, \left(7\sqrt{3} - 3\sqrt{15}\right)/12 \approx .042\right)$,

$$\mu_A\left(N^r_{PE}, \varepsilon\right) = \sum_{j=1}^{6} \varpi_{1,j}(r,\varepsilon)\,\mathbf{I}(r \in \mathcal{I}_j),$$

where

$$\begin{aligned}\varpi_{1,1}(r,\varepsilon) = -\Big[&3456\,\varepsilon^4 r^4 + 9216\,\varepsilon^4 r^3 - 3072\sqrt{3}\varepsilon^3 r^4 - 17280\,\varepsilon^4 r^2 - 3072\sqrt{3}\varepsilon^3 r^3 + 2304\,\varepsilon^2 r^4 \\ &+ 4608\sqrt{3}\varepsilon^3 r^2 - 2304\,\varepsilon^2 r^3 + 6336\,\varepsilon^4 + 6144\sqrt{3}\varepsilon^3\,r + 6912\,\varepsilon^2 r^2 + 512\sqrt{3}\,\varepsilon\,r^3 \\ &- 101\,r^4 - 6144\sqrt{3}\varepsilon^3 - 11520\,\varepsilon^2\,r - 1536\sqrt{3}\,\varepsilon\,r^2 + 256\,r^3 + 5760\,\varepsilon^2 + 1536\sqrt{3}\,\varepsilon\,r \\ &- 384\,r^2 - 512\sqrt{3}\,\varepsilon + 256\,r - 64\Big] \Big/ \Big[24\,\left(6\,\varepsilon + \sqrt{3}\right)^2 \left(6\,\varepsilon - \sqrt{3}\right)^2 r^2\Big],\end{aligned}$$

$$\begin{aligned}\varpi_{1,2}(r,\varepsilon) = -\Big[&1728\,\varepsilon^4 r^4 - 1536\sqrt{3}\varepsilon^3 r^4 - 31104\,\varepsilon^4 r^2 + 1152\,\varepsilon^2 r^4 + 15552\,\varepsilon^4 + 10368\,\varepsilon^2 r^2 - 37\,r^4 \\ &- 20736\,\varepsilon^2\,r + 10368\,\varepsilon^2\Big] \Big/ \Big[24\,\left(6\,\varepsilon + \sqrt{3}\right)^2 \left(6\,\varepsilon - \sqrt{3}\right)^2 r^2\Big],\end{aligned}$$

$$\varpi_{1,3}(r,\varepsilon) = \Big[-2592\,\varepsilon^4 r^4 - 2304\,\sqrt{3}\varepsilon^3 r^4 - 46656\,\varepsilon^4 r^2 + 1728\,\varepsilon^2 r^4 + 10656\,\varepsilon^4 - 9216\,\sqrt{3}\varepsilon^3\, r + 9072\,\varepsilon^2 r^2$$
$$- 432\,\sqrt{3}\,\varepsilon\, r^3 - 15\, r^4 + 12288\,\sqrt{3}\varepsilon^3 - 13824\,\varepsilon^2\, r + 1728\,\sqrt{3}\,\varepsilon\, r^2 - 216\, r^3 + 4032\,\varepsilon^2$$
$$- 2304\,\sqrt{3}\,\varepsilon\, r + 432\, r^2 + 1024\,\sqrt{3}\,\varepsilon - 384\, r + 128\Big] \Big/ \Big[36\,\left(6\,\varepsilon + \sqrt{3}\right)^2 \left(6\,\varepsilon - \sqrt{3}\right)^2 r^2\Big],$$
$$\varpi_{1,4}(r,\varepsilon) = -\frac{1728\,\varepsilon^4 r^4 - 1536\,\sqrt{3}\varepsilon^3 r^4 - 31104\,\varepsilon^4 r^2 + 1152\,\varepsilon^2 r^4 - 5184\,\varepsilon^4 + 2592\,\varepsilon^2 r^2 - 37\, r^4 - 3456\,\varepsilon^2}{24\,\left(6\,\varepsilon + \sqrt{3}\right)^2 \left(6\,\varepsilon - \sqrt{3}\right)^2 r^2},$$
$$\varpi_{1,5}(r,\varepsilon) = \frac{9}{8}\,\frac{1152\,\varepsilon^4 r^2 + 192\,\varepsilon^4 - 192\,\varepsilon^2 r^2 - r^4 + 128\,\varepsilon^2 + 32\, r^2 - 64\, r + 36}{\left(6\,\varepsilon + \sqrt{3}\right)^2 \left(6\,\varepsilon - \sqrt{3}\right)^2 r^2},$$
$$\varpi_{1,6}(r,\varepsilon) = -\frac{9}{8}\,\frac{(r+6)(r-2)^3}{\left(6\,\varepsilon + \sqrt{3}\right)^2 \left(6\,\varepsilon - \sqrt{3}\right)^2 r^2},$$

with the corresponding intervals $\mathcal{I}_1 = \Big[1, \frac{1+2\,\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}\Big)$, $\mathcal{I}_2 = \Big[\frac{1+2\,\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}, \frac{4\,\left(1-\sqrt{3}\,\varepsilon\right)}{3}\Big)$, $\mathcal{I}_3 = \Big[\frac{4\,\left(1-\sqrt{3}\,\varepsilon\right)}{3}$, $\frac{4\,\left(1+2\,\sqrt{3}\,\varepsilon\right)}{3}\Big)$, $\mathcal{I}_4 = \Big[\frac{4\,\left(1+2\,\sqrt{3}\,\varepsilon\right)}{3}, \frac{3}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}\Big)$, $\mathcal{I}_5 = \Big[\frac{3}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}, 2\Big)$, and $\mathcal{I}_6 = [2,\infty)$.

For $\varepsilon \in \Big[\left(7\,\sqrt{3} - 3\,\sqrt{15}\right)/12, \sqrt{3}/12\Big)$,

$$\mu_A\big(N_{PE}^r, \varepsilon\big) = \sum_{j=1}^{6} \varpi_{2,j}(r,\varepsilon)\, \mathbf{I}(r \in \mathcal{I}_j),$$

where $\varpi_{2,j}(r,\varepsilon) = \varpi_{1,j}(r,\varepsilon)$ for $j \in \{1,3,4,5,6\}$ and

$$\varpi_{2,2}(r,\varepsilon) = \Big[-3456\,\varepsilon^2 r^4 + 111\, r^4 - 5184\,\varepsilon^4 r^4 + 4608\,\sqrt{3}\,\varepsilon^3 r^4 - 336\,\sqrt{3}\,\varepsilon\, r^3 - 168\, r^3 - 13824\,\varepsilon^4 r^3$$
$$+ 4608\,\sqrt{3}\,\varepsilon^3 r^3 + 3456\,\varepsilon^2 r^3 + 144\, r^2 - 6912\,\sqrt{3}\,\varepsilon^3 r^2 - 3888\,\varepsilon^2 r^2 + 576\,\sqrt{3}\,\varepsilon\, r^2 + 25920\,\varepsilon^4$$
$$r^2 + 3168\,\varepsilon^4 + 2880\,\varepsilon^2 - 256\,\sqrt{3}\,\varepsilon - 32 - 3072\,\sqrt{3}\,\varepsilon^3\Big] \Big/ \Big[36\,\left(\sqrt{3} + 6\,\varepsilon\right)^2 \left(-6\,\varepsilon + \sqrt{3}\right)^2 r^2\Big]$$

with the corresponding intervals $\mathcal{I}_1 = \Big[1, \frac{4\,\left(1-\sqrt{3}\,\varepsilon\right)}{3}\Big)$, $\mathcal{I}_2 = \Big[\frac{4\,\left(1-\sqrt{3}\,\varepsilon\right)}{3}, \frac{1+2\,\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}\Big)$, $\mathcal{I}_3 = \Big[\frac{1+2\,\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}$, $\frac{4\,\left(1+2\,\sqrt{3}\,\varepsilon\right)}{3}\Big)$, $\mathcal{I}_4 = \Big[\frac{4\,\left(1+2\,\sqrt{3}\,\varepsilon\right)}{3}, \frac{3}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}\Big)$, $\mathcal{I}_5 = \Big[\frac{3}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}, 2\Big)$, and $\mathcal{I}_6 = [2,\infty)$.

For $\varepsilon \in \Big[\sqrt{3}/12, \sqrt{3}/3\Big)$,

$$\mu_A\big(N_{PE}^r, \varepsilon\big) = \sum_{j=1}^{3} \varpi_{3,j}(r,\varepsilon)\, \mathbf{I}(r \in \mathcal{I}_j),$$

where

$$\varpi_{3,1}(r,\varepsilon) = \frac{2\, r^2 - 1}{6\, r^2},$$

$$\varpi_{3,2}(r,\varepsilon) = \Big[432\,\varepsilon^4 r^4 + 1152\,\varepsilon^4 r^3 - 576\,\sqrt{3}\varepsilon^3 r^4 + 1296\,\varepsilon^4 r^2 - 960\,\sqrt{3}\varepsilon^3 r^3 + 864\,\varepsilon^2 r^4 - 864\,\sqrt{3}\varepsilon^3 r^2$$
$$+ 576\,\varepsilon^2 r^3 - 192\,\sqrt{3}\,\varepsilon\, r^4 - 360\,\varepsilon^4 + 648\,\varepsilon^2 r^2 + 64\,\sqrt{3}\,\varepsilon\, r^3 + 48\, r^4 + 192\,\sqrt{3}\varepsilon^3 - 144\,\sqrt{3}\,\varepsilon\, r^2$$
$$- 64\, r^3 - 504\,\varepsilon^2 + 72\, r^2 + 88\,\sqrt{3}\,\varepsilon - 25\Big] \Big/ \Big[16\,\left(3\,\varepsilon - \sqrt{3}\right)^4 r^2\Big],$$
$$\varpi_{3,3}(r,\varepsilon) = -\frac{-54\,\varepsilon^2 r^2 + 36\,\sqrt{3}\,\varepsilon\, r^2 + 15\,\varepsilon^2 - 18\, r^2 + 2\,\sqrt{3}\,\varepsilon + 20}{6\,\left(-3\,\varepsilon + \sqrt{3}\right)^2 r^2},$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, \frac{1+2\,\sqrt{3}\,\varepsilon}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}\right)$, $\mathcal{I}_3 = \left[\frac{1+2\,\sqrt{3}\,\varepsilon}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}, \frac{3}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}\right)$, and $\mathcal{I}_5 = \left[\frac{3}{2\,\left(1-\sqrt{3}\,\varepsilon\right)}, \infty\right)$.

APPENDIX C

Derivations of Means and Variances in Section 5.5

Note that $\mu(N_{CS}^{\tau})$ is provided in Lemma 5.5.2 for general $M \in T(\mathcal{Y})^o$. In this chapter we assume $M = M_C$.

C.1 Derivation of $\nu(N_{CS}^{\tau})$

Let $P_{2N}(N_{CS}^{\tau}) := P(\{X_2, X_3\} \subset N_{CS}^{\tau}(X_1, M_C))$, $P_{2G}(N_{CS}^{\tau}) := P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C))$ and $P_M(N_{CS}^{\tau}) := P(X_2 \in N_{CS}^{\tau}(X_1, M_C), X_3 \in \Gamma_1(X_1, N_{CS}^{\tau}, M_C)$. Then

$$\nu(N_{CS}^{\tau}) = \mathbf{Cov}\,[h_{12}(N_{CS}^{\tau}, M_C), h_{13}(N_{CS}^{\tau}, M_C)] = \left(P_{2N}(N_{CS}^{\tau}) + 2\,P_M(N_{CS}^{\tau}) + P_{2G}(N_{CS}^{\tau})\right) - [2\,\mu(N_{CS}^{\tau})]^2.$$

To find the covariance, we need to find the possible types of $\Gamma_1(x_1, N_{CS}^{\tau})$ and $N_{CS}^{\tau}(x_1)$ for $\tau \in (0, 1]$. There are four cases regarding $\Gamma_1(x_1, N_{CS}^{\tau})$ and one case for $N_{CS}^{\tau}(x_1)$. See Figure 3.4.6 for the prototypes of these four cases of $\Gamma_1(x_1, N_{CS}^{\tau})$ for $(x_1, y_1) \in T(\mathcal{Y})$, the explicit forms of $\zeta_j(\tau, x)$ are given by

$$\zeta_1(\tau, x) = \frac{\left(\sqrt{3}\,y_1 + 3\,x_1 - 3\,x\right)}{\sqrt{3}\,(1 + 2\,\tau)}, \qquad \zeta_2(\tau, x) = -\frac{\left(-\sqrt{3}\,y_1 + 3\,x_1 - 3\,x\right)}{\sqrt{3}\,(1 + 2\,\tau)},$$

$$\zeta_3(\tau, x) = \frac{\left(3\,x_1 + 3\,\tau - 3\,\tau\,x - 3\,x - \sqrt{3}\,y_1\right)}{\sqrt{3}\,(-1 + \tau)}, \qquad \zeta_4(\tau, x) = -\frac{-\sqrt{3}\,\tau + \sqrt{3}\,\tau\,x - 2\,y_1}{2 + \tau},$$

$$\zeta_5(\tau, x) = \frac{\sqrt{3}\,\tau\,x + 2\,y_1}{2 + \tau}, \qquad \zeta_6(\tau, x) = \frac{\left(-3\,x - 3\,\tau\,x + 3\,x_1 + \sqrt{3}\,y_1\right)}{\sqrt{3}\,(1 - \tau)},$$

$$\zeta_7(\tau, x) = \frac{y_1}{1 - \tau}.$$

Each case j, corresponds to the region R_j in Figure C.1.1, where $\ell_{am}(x) = x/\sqrt{3}$, $q_1(x) = \frac{1-\tau}{2\sqrt{3}}$, $q_2(x) = \frac{(x-1)(\tau-1)}{\sqrt{3}\,(1+\tau)}$, $q_3(x) = \frac{(1-\tau)x}{\sqrt{3}\,(1+\tau)}$, and $s_1 = (1 - \tau)/2$.

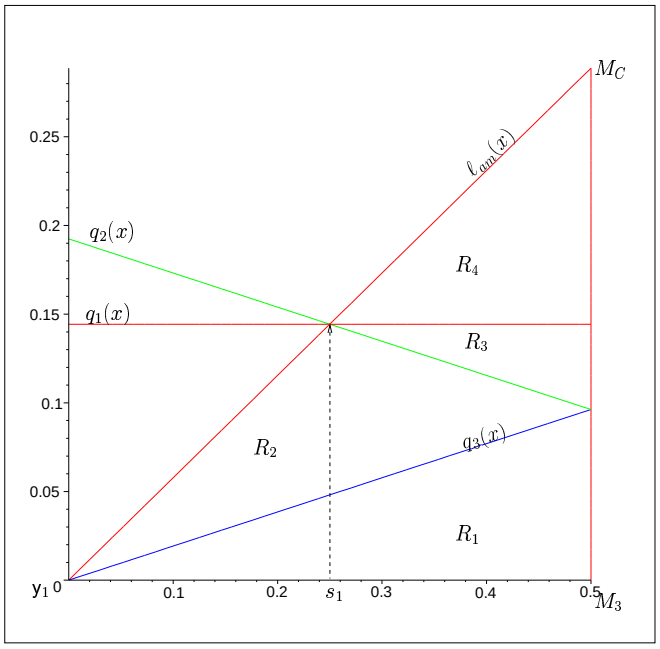


Figure C.1.1: The regions corresponding to the prototypes of the four cases with $\tau = 1/2$.

The explicit forms of R_j, $j \in \{1, \ldots, 4\}$ are given by

$$R_1 = \{(x,y) \in [0,1/2] \times [0, q_3(x)]\},\ R_3 = \{(x,y) \in [s_1,1/2] \times [q_2(x), q_1(x)]\},$$
$$R_2 = \{(x,y) \in [0,s_1] \times [q_3(x), \ell_{am}(x)] \cup [s_1,1/2] \times [q_3(x), q_2(x)]\},$$
$$R_4 = \{(x,y) \in [s_1,1/2] \times [q_1(x), \ell_{am}(x)]\}.$$

By symmetry, $P_{2N}\left(N_{CS}^{\tau}\right) = 6\,P\left(\{X_2, X_3\} \subset N_{CS}^{\tau}(X_1, M_C),\ X_1 \in T_s\right)$, and

$$P(\{X_2, X_3\} \subset N_{CS}^{\tau}(X_1, M_C),\ X_1 \in T_s) = \int_0^{1/2} \int_0^{\ell_{am}(x)} \frac{A(N_{CS}^{\tau}(x_1))^2}{A(T(\mathcal{Y}))^3} dydx = \tau^4/90,$$

where $A(N_{CS}^{\tau}(x_1)) = 3\sqrt{3}\,\tau^2\,y^2$. Hence, $P_{2N}\left(N_{CS}^{\tau}\right) = \tau^4/15$.

Next, by symmetry, $P_{2G}\left(N_{CS}^{\tau}\right) = 6\,P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in T_s\right)$, and

$$P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in T_s\right) = \sum_{j=1}^{4} P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in R_j\right).$$

For $x_1 \in R_1$,

$$P(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in R_1) = \int_0^{1/2} \int_0^{q_3(x)} \frac{A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx = \frac{\tau^4 (1-\tau)}{90\,(1+2\,\tau)^2\,(1+\tau)^5},$$

where $A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right) = \frac{3\sqrt{3}\,\tau^2 y^2}{(\tau-1)^2\,(2\,\tau+1)}$.

For $x_1 \in R_2$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in R_2\right) = \int_0^{s_1} \int_{q_3(x)}^{\ell a m(x)} \frac{A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_1}^{1/2} \int_{q_3(x)}^{q_2(x)} \frac{A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx = \frac{\tau^5\left(4\,\tau^6 + 6\,\tau^5 - 12\,\tau^4 - 21\,\tau^3 + 14\,\tau^2 + 40\,\tau + 20\right)(1-\tau)}{45\,(2\,\tau+1)^2\,(\tau+2)^2\,(\tau+1)^5}.$$

where $A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right) = \frac{3\sqrt{3}\left(x^2\tau + 2\sqrt{3}\,x\,y\,\tau - y^2\tau - x^2 + 2\sqrt{3}\,x\,y - 3\,y^2\right)\tau}{4\,(1-\tau)(2\,\tau+1)(\tau+2)}$.

For $x_1 \in R_3$,

$$\begin{aligned} P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in R_3\right) &= \int_{s_1}^{1/2} \int_{q_2(x)}^{q_1(x)} \frac{A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx \\ &= \frac{\tau^6 (1-\tau)(6\,\tau^6 - 35\,\tau^4 + 130\,\tau^2 + 160\,\tau + 60)}{90\,(2\,\tau+1)^2\,(\tau+2)^2\,(\tau+1)^5}. \end{aligned}$$

where

$$A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right) = -\frac{3\sqrt{3}\,\tau\left(2\,x^2\tau^2 + 2\,y^2\tau^2 - 4\,x^2\tau - 2\,x\,\tau^2 + 4\,y^2\tau + 2\sqrt{3}\,y\,\tau^2 + 2\,x^2 + 4\,x\,\tau + 6\,y^2 + \tau^2 - 2\,x - 2\sqrt{3}\,y - 2\,\tau + 1\right)}{4\,(2\,\tau+1)(\tau-1)^2\,(\tau+2)}.$$

For $x_1 \in R_4$,

$$P\left(\{X_2, X_3\} \subset \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in R_4\right) = \int_{s_1}^{1/2} \int_{q_1(x)}^{\ell a m(x)} \frac{A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx$$
$$+ \int_{s_4}^{s_5} \int_{q_3(x)}^{\ell a m(x)} \frac{A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx + \int_{s_5}^{1/2} \int_{q_3(x)}^{q_{12}(x)} \frac{A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right)^2}{A(T(\mathcal{Y}))^3} dydx = \frac{\tau^6\left(\tau^2 - 5\,\tau + 10\right)}{15\,(2\,\tau+1)^2\,(\tau+2)^2}.$$

where $A\left(\Gamma_1\left(x_1, N_{CS}^{\tau}\right)\right) = -\frac{\sqrt{3}\,\tau\left(3\,x^2 + 3\,y^2 - 3\,x - \sqrt{3}\,y - \tau + 1\right)}{2\,(2\,\tau+1)(\tau+2)}$.

So $P_{2G}\left(N_{CS}^{\tau}\right) = 6\left(-\frac{(\tau^2-7\,\tau-2)\tau^4}{90\,(\tau+1)(2\,\tau+1)(\tau+2)}\right) = -\frac{(\tau^2-7\,\tau-2)\tau^4}{15\,(\tau+1)(2\,\tau+1)(\tau+2)}$.

Furthermore, by symmetry, $P_M\left(N_{CS}^{\tau}\right) = 6\,P\left(X_2 \in N_{CS}^{\tau}(X_1, M_C),\ X_3 \in \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in T_s\right)$, and

$$P\left(X_2 \in N_{CS}^{\tau}(X_1, M_C),\ X_3 \in \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in T_s\right) =$$
$$\sum_{j=1}^{4} P\left(X_2 \in N_{CS}^{\tau}(X_1, M_C),\ X_3 \in \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in R_j\right).$$

where $P(X_2 \in N_{CS}^{\tau}(X_1, M_C),\ X_3 \in \Gamma_1(X_1, N_{CS}^{\tau}, M_C),\ X_1 \in R_j)$ can be calculated with the

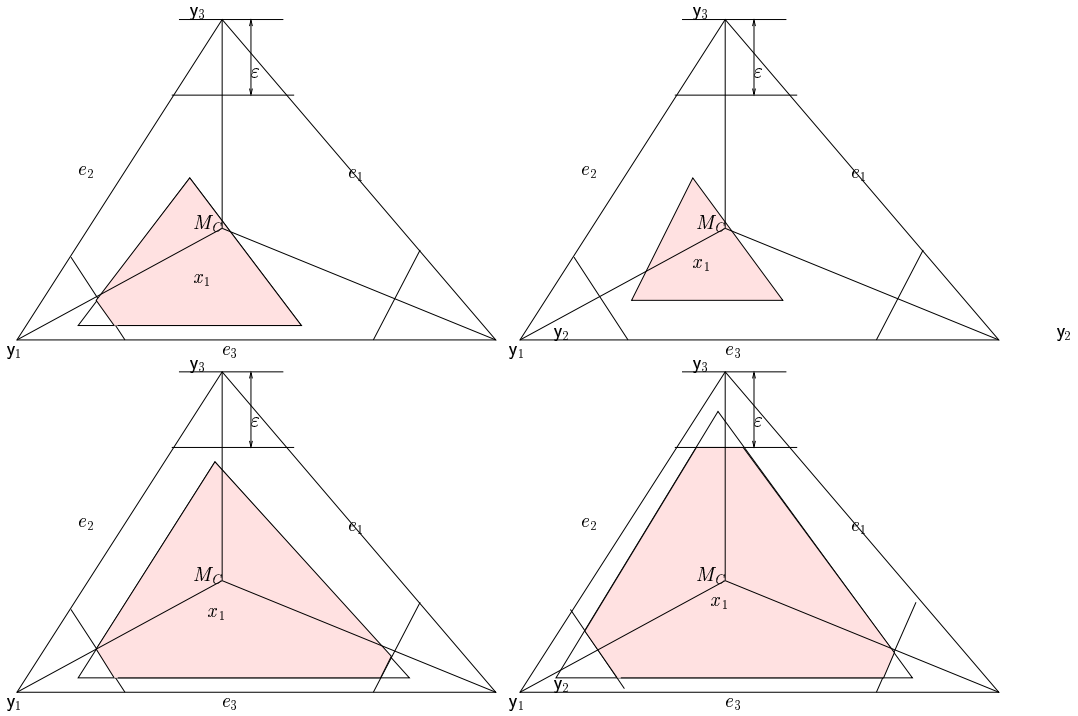


Figure C.2.1: The prototypes of the four cases of $N_{CS}^{\tau}(x)$ for four distinct $x \in R_{CM}(e_3)$ (shaded regions).

same region of integration with integrand being replaced by $\frac{A(N_{CS}^{\tau}(x_1))\, A(\Gamma_1(x_1, N_{CS}^{\tau}))}{A(T(\mathcal{Y}))^3}$.

Then $P_M\big(N_{CS}^{\tau}\big) = 6\left(\frac{(2\,\tau^4-3\,\tau^3-4\,\tau^2+10\,\tau+4)\tau^4}{180\,(2\,\tau+1)(\tau+2)}\right) = \frac{(2\,\tau^4-3\,\tau^3-4\,\tau^2+10\,\tau+4)\tau^4}{30\,(2\,\tau+1)(\tau+2)}$.

Hence

$$\mathbf{E}\left[h_{12}(N_{CS}^{\tau}, M_C)\, h_{13}(N_{CS}^{\tau}, M_C)\right] = \frac{\tau^4\,(2\,\tau^5 - \tau^4 - 5\,\tau^3 + 12\,\tau^2 + 28\,\tau + 8)}{15\,(\tau+1)(2\,\tau+1)(\tau+2)}.$$

Therefore,

$$\nu\big(N_{CS}^{\tau}\big) = \frac{\tau^4\,(6\,\tau^5 - 3\,\tau^4 - 25\,\tau^3 + \tau^2 + 49\,\tau + 14)}{45\,(\tau+1)(2\,\tau+1)(\tau+2)}.$$

For $\tau = 0$, it is trivial to see that $\nu\big(N_{CS}^{\tau}\big) = 0$.

C.2 Derivation of $\mu\big(N_{CS}^{\tau}, \varepsilon\big)$

We pick the interval $\varepsilon \in \left[0, \sqrt{3}/5\right)$ for a demonstrative example in derivation of $\mu_S\big(N_{CS}^{\tau}, \varepsilon\big)$. For $\tau \in \left[1-\sqrt{3}\,\varepsilon, 1\right)$, there are four cases to consider for the form of $N_{CS}^{\tau}(x_1, \varepsilon) := N_{CS}^{\tau}(x_1, M_C) \cap T_{\varepsilon}$). See Figure C.2.1 for the prototypes of these four cases of $N_{CS}^{\tau}(x_1, \varepsilon)$.

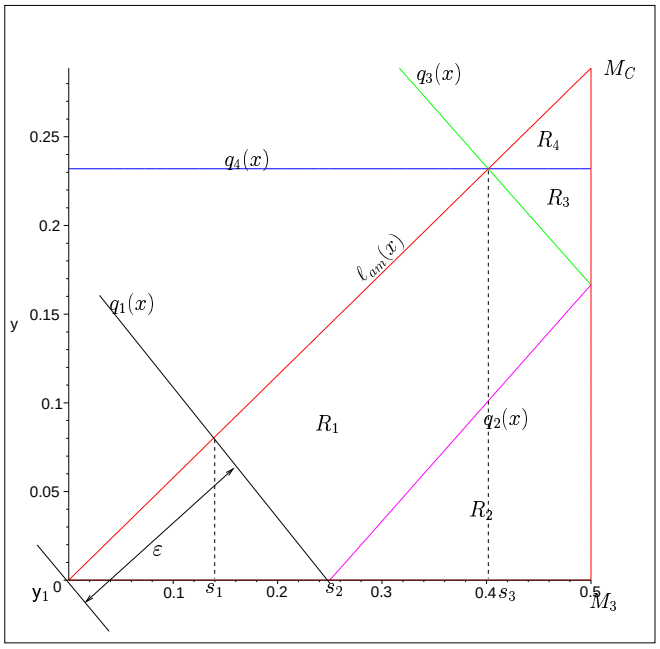


Figure C.2.2: The regions corresponding to the prototypes of the four cases shown in Figure C.2.1.

Each case j, corresponds to the region R_j in Figure C.2.2, where $\ell_{am}(x) = x/\sqrt{3}$, $q_1(x) = -\sqrt{3}\,x + 2\,\varepsilon$, $q_2(x) = \frac{-\sqrt{3}\,x+2\,\varepsilon}{1-4\,\tau}$, $q_3(x) = \frac{\sqrt{3}\,x+2\,\varepsilon-\sqrt{3}}{1-4\,\tau}$, $q_4(x) = \frac{\sqrt{3}-2\,\varepsilon}{2\,(1+2\,\tau)}$, and $s_1 = \sqrt{3}\,\varepsilon/2$, $s_2 = 2\,\sqrt{3}\,\varepsilon/3$, $s_3 = \frac{\sqrt{3}\,(\sqrt{3}-2\,\varepsilon)}{2\,(1+2\,\tau)}$.

The explicit forms of R_j for $j \in \{1, \ldots, 4\}$ are given by

$$R_1 = \{(x,y) \in [s_1, s_2] \times [q_1(x), \ell_{am}(x)] \cup [s_2, s_3] \times [q_2(x), \ell_{am}(x)] \cup [s_3, 1/2] \times [q_2(x), q_3(x)]\},$$

$$R_2 = \{(x,y) \in [s_2, 1/2] \times [0, q_2(x)]\}, \quad R_3 = \{(x,y) \in [s_3, 1/2] \times [q_3(x), q_4(x)(x)]\},$$

$$R_4 = \{(x,y) \in [s_3, 1/2] \times [q_4(x), \ell_{am}(x)]\}.$$

By symmetry, $P(X_2 \in N_{CS}^{\tau}(X_1, \varepsilon)) = 6\,P\,(X_2 \in N_{CS}^{\tau}(X_1, \varepsilon),\ X_1 \in T_s)\,.$

For $x_1 \in R_1$,

$$P(X_2 \in N_{CS}^{\tau}(X_1,\varepsilon),\ X_1 \in R_1) = \int_{s_1}^{s_2}\int_{q_1(x)}^{\ell am(x)} \frac{A(N_{CS}^{\tau}(x_1,\varepsilon))^2}{A(T(\mathcal{Y}))^2}dydx +$$
$$\int_{s_2}^{s_3}\int_{q_2(x)}^{\ell am(x)} \frac{A(N_{CS}^{\tau}(x_1,\varepsilon))^2}{A(T(\mathcal{Y}))^2}dydx + \int_{s_3}^{1/2}\int_{q_2(x)}^{q_3(x)} \frac{A(N_{CS}^{\tau}(x_1,\varepsilon))^2}{A(T(\mathcal{Y}))^2}dydx =$$
$$-\frac{1}{27}\Big[-36 - 1701\,\varepsilon^4\tau^2 + 3182\,\varepsilon^4\tau^3 + 14492\,\varepsilon^4\tau^4 - 1692\,\tau^5 + 414\,\tau^4 + 5832\,\varepsilon^2\tau - 13932\,\varepsilon^2\tau^2$$
$$- 4064\,\varepsilon^4\tau^6 - 128\,\varepsilon^4\tau^7 + 2560\,\varepsilon^4\tau^8 + 968\,\varepsilon^4\tau^5 - 10368\,\tau^5\varepsilon^2 + 31104\,\tau^4\varepsilon^2 + 14256\,\varepsilon^2\tau^3 + 432\,\sqrt{3}\,\varepsilon^3$$
$$- 1467\,\tau^2 + 2268\,\varepsilon^4\tau - 324\,\varepsilon^4 + 2385\,\tau^3 - 20736\,\sqrt{3}\,\tau^4\varepsilon^3 + 4428\,\sqrt{3}\,\tau^2\varepsilon - 5760\,\sqrt{3}\,\tau^3\varepsilon - 1440\,\sqrt{3}\tau\,\varepsilon$$
$$+ 5184\,\sqrt{3}\,\tau^5\varepsilon - 5472\,\sqrt{3}\,\tau^4\varepsilon - 3456\,\tau\,\sqrt{3}\,\varepsilon^3 + 5616\,\sqrt{3}\,\tau^2\,\varepsilon^3 - 5184\,\sqrt{3}\,\tau^3\,\varepsilon^3 - 648\,\varepsilon^2$$
$$+ 396\,\tau + 144\,\sqrt{3}\,\varepsilon\Big]\Big/\Big[(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2(4\,\tau-1)^3(1+2\,\tau)^3\Big],$$

where $A\left(N_{CS}^{\tau}(x_1,\varepsilon)\right) = -\sqrt{3}\left(-2\,\varepsilon + y + 2\,\tau\,y + \sqrt{3}\,x\right)\left(-2\,\varepsilon + \sqrt{3}\,x + y - 10\,\tau\,y\right)/12.$

For $x_1 \in R_2$,

$$\begin{aligned} P\left(X_2 \in N_{CS}^{\tau}(X_1,\varepsilon),\ X_1 \in R_2\right) &= \int_{s_1}^{1/2}\int_{0}^{q_2(x)} \frac{A\left(N_{CS}^{\tau}(x_1,\varepsilon)\right)^2}{A(T(\mathcal{Y}))^2}dydx \\ &= -\frac{\tau^2\left(-9 - 256\,\varepsilon^4 + 48\,\sqrt{3}\,\varepsilon - 288\,\varepsilon^2 + 256\,\sqrt{3}\,\varepsilon^3\right)}{12\,(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2(4\,\tau-1)^3}. \end{aligned}$$

where $A\left(N_{CS}^{\tau}(x_1,\varepsilon)\right) = 3\,\sqrt{3}\,\tau^2\,y^2.$

For $x_1 \in R_3$,

$$P\left(X_2 \in N_{CS}^{\tau}(X_1,\varepsilon),\ X_1 \in R_3\right) = \int_{s_3}^{1/2}\int_{q_3(x)}^{q_4(x)} \frac{A\left(N_{CS}^{\tau}(X_1,\varepsilon)\right)^2}{A(T(\mathcal{Y}))^2}dydx = \frac{2}{3}\Big[-4 - 522\,\varepsilon^4\tau^2 + 1728\,\varepsilon^4\tau^4$$
$$- 396\,\tau^5 - 57\,\tau^4 + 720\,\varepsilon^2\tau - 2259\,\varepsilon^2\tau^2 + 242\,\tau^6 - 4032\,\tau^5\varepsilon^2 + 576\,\tau^6\varepsilon^2 + 3600\,\tau^4\varepsilon^2 + 2196\,\varepsilon^2\tau^3$$
$$+ 48\,\sqrt{3}\,\varepsilon^3 - 207\,\tau^2 + 288\,\varepsilon^4\tau - 36\,\varepsilon^4 + 374\,\tau^3 - 2592\,\sqrt{3}\,\tau^4\varepsilon^3 + 658\,\sqrt{3}\,\tau^2\varepsilon - 922\,\sqrt{3}\,\tau^3\varepsilon - 176\,\sqrt{3}\tau\,\varepsilon$$
$$+ 1376\,\sqrt{3}\,\tau^5\varepsilon - 472\,\sqrt{3}\,\tau^4\varepsilon - 432\,\tau\,\sqrt{3}\,\varepsilon^3 + 1080\,\sqrt{3}\,\tau^2\,\varepsilon^3 - 552\,\sqrt{3}\,\tau^3\,\varepsilon^3 - 72\,\varepsilon^2 - 480\,\sqrt{3}\,\tau^6\varepsilon$$
$$+ 1152\,\sqrt{3}\,\tau^5\varepsilon^3 + 48\,\tau + 16\,\sqrt{3}\,\varepsilon\Big]\Big/\Big[(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2(4\,\tau-1)^3(1+2\,\tau)^4\Big].$$

where

$$A\left(N_{CS}^{\tau}(x_1,\varepsilon)\right) =$$
$$\sqrt{3}\Big(8\,\varepsilon\,y - 32\,\varepsilon\,\tau\,y - 2\,y^2 + 16\,\tau\,y^2 - 2\,\sqrt{3}\,y + 8\,\sqrt{3}\,\tau\,y - 3 + 6\,x - 8\,\varepsilon^2 + 4\,\sqrt{3}\,\varepsilon - 6\,x^2 + 4\,\tau^2y^2\Big)/12.$$

For $x_1 \in R_4$,

$$P\left(X_2 \in N_{CS}^{\tau}(X_1,\varepsilon),\ X_1 \in R_4\right) = \int_{s_3}^{1/2} \int_{q_4(x)}^{\ell_{am}(x)} \frac{A\left(N_{CS}^{\tau}(x_1,\varepsilon)\right)^2}{A(T(\mathcal{Y}))^2} dy dx = -\frac{2}{27}\Big[18 - 90\,\varepsilon^4\tau^2 - 28\,\tau^5$$
$$- 7\,\tau^4 - 360\,\varepsilon^2\tau - 675\,\varepsilon^2\tau^2 + 2\,\tau^6 + 144\,\tau^4\varepsilon^2 + 324\,\varepsilon^2\tau^3 - 216\,\sqrt{3}\,\varepsilon^3 + \tau^2 + 144\,\varepsilon^4\tau + 162\,\varepsilon^4 + 70\,\tau^3$$
$$+ 106\,\sqrt{3}\,\tau^2\varepsilon - 186\,\sqrt{3}\,\tau^3\varepsilon + 152\,\sqrt{3}\tau\,\varepsilon + 32\,\sqrt{3}\,\tau^5\varepsilon - 32\,\sqrt{3}\,\tau^4\varepsilon + 24\,\tau\,\sqrt{3}\,\varepsilon^3 + 384\,\sqrt{3}\,\tau^2\,\varepsilon^3$$
$$+ 24\,\sqrt{3}\,\tau^3\,\varepsilon^3 + 324\,\varepsilon^2 - 56\,\tau - 72\,\sqrt{3}\,\varepsilon\Big] \Big/ \Big[(2\,\varepsilon - 1)^2(2\,\varepsilon + 1)^2(1 + 2\,\tau)^4\Big].$$

where

$$A\left(N_{CS}^{\tau}(x_1,\varepsilon)\right) = \sqrt{3}\Big(-6\,\tau^2 y^2 + 4\,\sqrt{3}\,\varepsilon - 6\,\varepsilon^2 - 3 + \sqrt{3}\,y + 3\,x - 24\,\varepsilon\,\tau\,y - 3\,y^2 + 8\,\sqrt{3}\,\tau\,y - 3\,x^2\Big)/6.$$

So

$$P\left(X_2 \in N_{CS}^{\tau}(X_1,\varepsilon)\right) = -\frac{1}{18}\Big[\tau\,(4\,\tau^3 + 80\,\varepsilon^4\tau^3 - 60\,\tau^2 + 64\,\sqrt{3}\,\tau^2\varepsilon + 16\,\varepsilon^4\tau^2 + 45\,\tau + 288\,\varepsilon^2\tau$$
$$- 108\,\varepsilon^4\tau - 128\,\sqrt{3}\tau\,\varepsilon + 192\,\sqrt{3}\,\varepsilon^3 + 64\,\sqrt{3}\,\varepsilon - 144\,\varepsilon^4 - 16 - 288\,\varepsilon^2)\Big] \Big/ \Big[(1 + 2\,\tau)^2(2\,\varepsilon - 1)^2(2\,\varepsilon + 1)^2\Big].$$

C.3 The Mean $\mu\left(N_{CS}^{\tau},\varepsilon\right)$ Under Segregation and Association Alternatives

Derivation of $\mu\left(N_{CS}^{\tau},\varepsilon\right)$ involves detailed geometric calculations and partitioning of the space of (τ,ε,x_1) for $\tau \in [0,1]$, $\varepsilon \in \left[0,\sqrt{3}/3\right)$, and $x_1 \in T_s = T(\mathsf{y}_1, M_3, M_C)$. A demonstrative calculation is given in Appendix Section C.2 for $\varepsilon \in \left[0,\sqrt{3}/5\right)$ and $\tau \in [1-\sqrt{3}\,\varepsilon, 1]$.

C.3.1 $\mu_S\left(N_{CS}^{\tau},\varepsilon\right)$ Under Segregation Alternatives

Under segregation, we compute $\mu_S\left(N_{CS}^{\tau},\varepsilon\right)$ explicitly.

For $\varepsilon \in \left[0,\sqrt{3}/5\right)$,

$$\mu_S\left(N_{CS}^{\tau},\varepsilon\right) = \sum_{j=1}^{2} \varpi_{1,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j),$$

where

$$\varpi_{1,1}(\tau,\varepsilon) = \frac{(20\,\varepsilon^4\tau - 36\,\varepsilon^4 - 3\,\tau + 3)\tau^2}{18\,(1-\tau)(2\,\varepsilon+1)^2(2\,\varepsilon-1)^2},$$
$$\varpi_{1,2}(\tau,\varepsilon) = -\Big[\Big(80\,\varepsilon^4\tau^3 + 16\,\varepsilon^4\tau^2 - 108\,\varepsilon^4\tau - 144\,\varepsilon^4 + 192\,\sqrt{3}\varepsilon^3 + 288\,\varepsilon^2\tau + 64\,\sqrt{3}\,\varepsilon\,\tau^2 + 4\,\tau^3$$
$$- 288\,\varepsilon^2 - 128\,\sqrt{3}\,\varepsilon\,\tau - 60\,\tau^2 + 64\,\sqrt{3}\,\varepsilon + 45\,\tau - 16\Big)\tau\Big] \Big/ \left[18\,(2\,\tau+1)^2(2\,\varepsilon-1)^2(2\,\varepsilon+1)^2\right],$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 1-\sqrt{3}\,\varepsilon\right)$ and $\mathcal{I}_2 = \left[1-\sqrt{3}\,\varepsilon, 1\right)$.

For $\varepsilon \in \left[\sqrt{3}/5, \sqrt{3}/4\right)$,

$$\mu_S\left(N_{CS}^{\tau}, \varepsilon\right) = \sum_{j=1}^{3} \varpi_{2,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j)$$

where $\varpi_{2,j}(\tau,\varepsilon) = \varpi_{1,j}(\tau,\varepsilon)$ for $j \in \{1,2\}$, and for $j = 3$,

$$\begin{aligned}\varpi_{2,3}(\tau,\varepsilon) = -\Big[\Big(&-96\,\varepsilon^4\tau^4 - 1584\,\varepsilon^4\tau^3 - 3688\,\varepsilon^4\tau^2 + 512\,\sqrt{3}\varepsilon^3\tau^3 - 3636\,\varepsilon^4\tau + 2304\,\sqrt{3}\varepsilon^3\tau^2 \\ &- 1152\,\varepsilon^4 + 3840\,\sqrt{3}\varepsilon^3\tau - 576\,\varepsilon^2\tau^2 + 128\,\sqrt{3}\,\varepsilon\,\tau^3 + 8\,\tau^4 + 1536\,\sqrt{3}\varepsilon^3 - 4320\,\varepsilon^2\tau \\ &- 320\,\sqrt{3}\,\varepsilon\,\tau^2 - 124\,\tau^3 - 2304\,\varepsilon^2 + 640\,\sqrt{3}\,\varepsilon\,\tau + 150\,\tau^2 + 512\,\sqrt{3}\,\varepsilon - 77\,\tau - 128\Big)\tau\Big] \Big/ \\ &\Big[18\,(2\,\tau+1)^2(2\,\varepsilon-1)^2(2\,\varepsilon+1)^2(2\,\tau-1)\Big],\end{aligned}$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 1-\sqrt{3}\,\varepsilon\right)$, $\mathcal{I}_2 = \left[1-\sqrt{3}\,\varepsilon, \sqrt{3}/(2\,\varepsilon) - 3/2\right)$, and $\mathcal{I}_3 = \left[\sqrt{3}/(2\,\varepsilon) - 3/2, 1\right)$.

For $\varepsilon \in \left[\sqrt{3}/4, 2\sqrt{3}/7\right)$,

$$\mu_S\left(N_{CS}^{\tau}, \varepsilon\right) = \sum_{j=1}^{5} \varpi_{3,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j),$$

where

$$\begin{aligned}\varpi_{3,1}(\tau,\varepsilon) = \Big[&(984\,\varepsilon^4\tau^3 - 3452\,\varepsilon^4\tau^2 - 1024\,\sqrt{3}\varepsilon^3\tau^3 + 4992\,\varepsilon^4\tau + 3584\,\sqrt{3}\varepsilon^3\tau^2 + 1152\,\varepsilon^2\tau^3 - 2268\,\varepsilon^4 \\ &- 5120\,\sqrt{3}\varepsilon^3\tau - 4032\,\varepsilon^2\tau^2 - 192\,\sqrt{3}\,\varepsilon\,\tau^3 + 2304\,\sqrt{3}\varepsilon^3 + 5760\,\varepsilon^2\tau + 672\,\sqrt{3}\,\varepsilon\,\tau^2 + 42\,\tau^3 \\ &- 2592\,\varepsilon^2 - 960\,\sqrt{3}\,\varepsilon\,\tau - 141\,\tau^2 + 432\,\sqrt{3}\,\varepsilon + 192\,\tau - 84)\tau^2\Big] \Big/ \Big[32\,(\tau-1)^2\left(3\,\varepsilon - \sqrt{3}\right)^4(2\,\tau-1)\Big], \\ \varpi_{3,2}(\tau,\varepsilon) = \Big[&(3936\,\varepsilon^4\tau^6 - 9872\,\varepsilon^4\tau^5 - 4096\,\sqrt{3}\varepsilon^3\tau^6 + 7144\,\varepsilon^4\tau^4 + 10240\,\sqrt{3}\varepsilon^3\tau^5 + 4608\,\varepsilon^2\tau^6 + 7444\,\varepsilon^4\tau^3 \\ &- 7168\,\sqrt{3}\varepsilon^3\tau^4 - 11520\,\varepsilon^2\tau^5 - 768\,\sqrt{3}\,\varepsilon\,\tau^6 - 4368\,\varepsilon^4\tau^2 - 8064\,\sqrt{3}\varepsilon^3\tau^3 + 7488\,\varepsilon^2\tau^4 + 1792\,\sqrt{3}\,\varepsilon\,\tau^5 \\ &+ 136\,\tau^6 - 1836\,\varepsilon^4\tau + 5056\,\sqrt{3}\varepsilon^3\tau^2 + 10656\,\varepsilon^2\tau^3 - 768\,\sqrt{3}\,\varepsilon\,\tau^4 - 220\,\tau^5 - 144\,\varepsilon^4 + 1536\,\sqrt{3}\varepsilon^3\tau \\ &- 7200\,\varepsilon^2\tau^2 - 2464\,\sqrt{3}\,\varepsilon\,\tau^3 - 154\,\tau^4 + 192\,\sqrt{3}\varepsilon^3 - 1152\,\varepsilon^2\tau + 1664\,\sqrt{3}\,\varepsilon\,\tau^2 + 771\,\tau^3 - 288\,\varepsilon^2 \\ &+ 48\,\sqrt{3}\,\varepsilon\,\tau - 464\,\tau^2 + 64\,\sqrt{3}\,\varepsilon + 28\,\tau - 16)\tau\Big] \Big/ \Big[32\,\left(3\,\varepsilon - \sqrt{3}\right)^4(2\,\tau-1)(2\,\tau+1)^2(\tau-1)^2\Big],\end{aligned}$$

$$\varpi_{3,3}(\tau,\varepsilon) = -\Big[(-2096\,\varepsilon^4\tau^5 + 3376\,\varepsilon^4\tau^4 + 2048\,\sqrt{3}\varepsilon^3\tau^5 - 2204\,\varepsilon^4\tau^3 - 3840\,\sqrt{3}\varepsilon^3\tau^4 - 2304\,\varepsilon^2\tau^5 - 3864\,\varepsilon^4\tau^2$$
$$+ 2304\,\sqrt{3}\varepsilon^3\tau^3 + 4608\,\varepsilon^2\tau^4 + 384\,\sqrt{3}\,\varepsilon\,\tau^5 + 1332\,\varepsilon^4\tau + 4864\,\sqrt{3}\varepsilon^3\tau^2 - 2016\,\varepsilon^2\tau^3 - 704\,\sqrt{3}\,\varepsilon\,\tau^4$$
$$- 68\,\tau^5 + 1152\,\varepsilon^4 - 1536\,\sqrt{3}\varepsilon^3\tau - 6912\,\varepsilon^2\tau^2 + 32\,\sqrt{3}\,\varepsilon\,\tau^3 + 76\,\tau^4 - 1536\,\sqrt{3}\varepsilon^3 + 1728\,\varepsilon^2\tau$$
$$+ 1440\,\sqrt{3}\,\varepsilon\,\tau^2 + 115\,\tau^3 + 2304\,\varepsilon^2 - 208\,\sqrt{3}\,\varepsilon\,\tau - 328\,\tau^2 - 512\,\sqrt{3}\,\varepsilon - 4\,\tau + 128)\tau\Big] \Big/ \Big[32\,(\tau-1)^2$$
$$\left(-3\,\varepsilon + \sqrt{3}\right)^4 (2\,\tau+1)^2\Big],$$
$$\varpi_{3,4}(\tau,\varepsilon) = -\Big[(-2064\,\varepsilon^4\tau^6 + 1696\,\varepsilon^4\tau^5 + 2048\,\sqrt{3}\varepsilon^3\tau^6 + 3292\,\varepsilon^4\tau^4 - 1920\,\sqrt{3}\varepsilon^3\tau^5 - 2304\,\varepsilon^2\tau^6 - 788\,\varepsilon^4\tau^3$$
$$- 2816\,\sqrt{3}\varepsilon^3\tau^4 + 2304\,\varepsilon^2\tau^5 + 384\,\sqrt{3}\,\varepsilon\,\tau^6 + 3948\,\varepsilon^4\tau^2 + 2528\,\sqrt{3}\varepsilon^3\tau^3 + 3168\,\varepsilon^2\tau^4 - 320\,\sqrt{3}\,\varepsilon\,\tau^5$$
$$- 68\,\tau^6 + 5940\,\varepsilon^4\tau - 3872\,\sqrt{3}\varepsilon^3\tau^2 - 4896\,\varepsilon^2\tau^3 - 672\,\sqrt{3}\,\varepsilon\,\tau^4 + 8\,\tau^5 + 1800\,\varepsilon^4 - 7392\,\sqrt{3}\varepsilon^3\tau$$
$$+ 3600\,\varepsilon^2\tau^2 + 1088\,\sqrt{3}\,\varepsilon\,\tau^3 + 191\,\tau^4 - 2400\,\sqrt{3}\varepsilon^3 + 10080\,\varepsilon^2\tau - 304\,\sqrt{3}\,\varepsilon\,\tau^2 - 213\,\tau^3 + 3600\,\varepsilon^2$$
$$- 1968\,\sqrt{3}\,\varepsilon\,\tau - 44\,\tau^2 - 800\,\sqrt{3}\,\varepsilon + 412\,\tau + 200)\tau\Big] \Big/ \Big[32\,(2\,\tau+1)^2\left(-3\,\varepsilon + \sqrt{3}\right)^4 (\tau+1)(\tau-1)^2\Big],$$
$$\varpi_{3,5}(\tau,\varepsilon) = \Big[(1032\,\varepsilon^4\tau^5 + 3280\,\varepsilon^4\tau^4 - 1024\,\sqrt{3}\varepsilon^3\tau^5 + 2186\,\varepsilon^4\tau^3 - 3136\,\sqrt{3}\varepsilon^3\tau^4 + 1152\,\varepsilon^2\tau^5 - 1806\,\varepsilon^4\tau^2$$
$$- 1920\,\sqrt{3}\varepsilon^3\tau^3 + 3456\,\varepsilon^2\tau^4 - 192\,\sqrt{3}\,\varepsilon\,\tau^5 - 2376\,\varepsilon^4\tau + 2384\,\sqrt{3}\varepsilon^3\tau^2 + 2448\,\varepsilon^2\tau^3 - 576\,\sqrt{3}\,\varepsilon\,\tau^4$$
$$+ 36\,\tau^5 - 648\,\varepsilon^4 + 3024\,\sqrt{3}\varepsilon^3\tau - 2736\,\varepsilon^2\tau^2 - 624\,\sqrt{3}\,\varepsilon\,\tau^3 + 108\,\tau^4 + 864\,\sqrt{3}\varepsilon^3 - 4104\,\varepsilon^2\tau$$
$$+ 264\,\sqrt{3}\,\varepsilon\,\tau^2 + 207\,\tau^3 - 1296\,\varepsilon^2 + 768\,\sqrt{3}\,\varepsilon\,\tau + 54\,\tau^2 + 288\,\sqrt{3}\,\varepsilon - 144\,\tau - 72)\tau\Big] \Big/ \Big[16\,(2\,\tau+1)^2$$
$$\left(3\,\varepsilon - \sqrt{3}\right)^4 (\tau+1)(\tau+2)\Big],$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 1-\sqrt{3}\,\varepsilon\right)$, $\mathcal{I}_2 = \left[1-\sqrt{3}\,\varepsilon, \sqrt{3}/(2\,\varepsilon) - 3/2\right)$, $\mathcal{I}_3 = \left[\sqrt{3}/(2\,\varepsilon) - 3/2, \sqrt{3}/\varepsilon - 3\right)$, $\mathcal{I}_4 = \left[\sqrt{3}/\varepsilon - 3, 4\ (1-\sqrt{3}\,\varepsilon)\right)$, and $\mathcal{I}_5 = \left[4\ (1-\sqrt{3}\,\varepsilon), 1\right)$.

For $\varepsilon \in \left[2\sqrt{3}/7, \sqrt{3}/3\right)$,

$$\mu_S\left(N_{CS}^{\tau}, \varepsilon\right) = \sum_{j=1}^{2} \varpi_{4,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j)$$

where $\varpi_{4,j}(\tau,\varepsilon) = \varpi_{3,j}(\tau,\varepsilon)$ for $j \in \{1,\ldots,5\}$, and for $j \in \{6,7,8\}$

$$\varpi_{4,6}(\tau,\varepsilon) = \Big[-254\,\varepsilon^4\tau^4 - 1998\,\varepsilon^4\tau^3 + 256\,\sqrt{3}\varepsilon^3\tau^4 - 4752\,\varepsilon^4\tau^2 + 2160\,\sqrt{3}\varepsilon^3\tau^3 - 288\,\varepsilon^2\tau^4 - 4320\,\varepsilon^4\tau$$
$$+ 5328\,\sqrt{3}\varepsilon^3\tau^2 - 2592\,\varepsilon^2\tau^3 + 48\,\sqrt{3}\,\varepsilon\,\tau^4 - 1296\,\varepsilon^4 + 5184\,\sqrt{3}\varepsilon^3\tau - 6552\,\varepsilon^2\tau^2 + 456\,\sqrt{3}\,\varepsilon\,\tau^3$$
$$- 9\,\tau^4 + 1728\,\sqrt{3}\varepsilon^3 - 6912\,\varepsilon^2\tau + 1152\,\sqrt{3}\,\varepsilon\,\tau^2 - 90\,\tau^3 - 2592\,\varepsilon^2 + 1344\,\sqrt{3}\,\varepsilon\,\tau - 216\,\tau^2$$
$$+ 576\,\sqrt{3}\,\varepsilon - 288\,\tau - 144\Big] \Big/ \Big[16\,(\tau+2)\left(-3\,\varepsilon + \sqrt{3}\right)^4 (\tau+1)\Big],$$

$$\varpi_{4,7}(\tau,\varepsilon) = \Big[-256\,\varepsilon^4\tau^4 - 1536\,\varepsilon^4\tau^3 + 256\,\sqrt{3}\varepsilon^3\tau^4 - 2160\,\varepsilon^4\tau^2$$
$$+ 1664\,\sqrt{3}\varepsilon^3\tau^3 - 288\,\varepsilon^2\tau^4 - 2160\,\varepsilon^4\tau + 2304\,\sqrt{3}\varepsilon^3\tau^2 - 2016\,\varepsilon^2\tau^3 + 48\,\sqrt{3}\,\varepsilon\,\tau^4 - 1296\,\varepsilon^4$$
$$+ 2592\,\sqrt{3}\varepsilon^3\tau - 2664\,\varepsilon^2\tau^2 + 360\,\sqrt{3}\,\varepsilon\,\tau^3 - 9\,\tau^4 + 1728\,\sqrt{3}\varepsilon^3 - 3456\,\varepsilon^2\tau + 432\,\sqrt{3}\,\varepsilon\,\tau^2 - 72\,\tau^3$$
$$- 2592\,\varepsilon^2 + 672\,\sqrt{3}\,\varepsilon\,\tau - 72\,\tau^2 + 576\,\sqrt{3}\,\varepsilon - 144\,\tau - 144\Big] \Big/ \Big[16\,\left(-3\,\varepsilon + \sqrt{3}\right)^4 (\tau+1)\tau\Big],$$

$$\varpi_{4,8}(\tau,\varepsilon) = 1,$$

with the corresponding intervals $\mathcal{I}_1 = \left[1, 1-\sqrt{3}\,\varepsilon\right)$, $\mathcal{I}_2 = \left[1-\sqrt{3}\,\varepsilon, \sqrt{3}/(2\,\varepsilon) - 3/2\right)$, $\mathcal{I}_3 = \left[\sqrt{3}/(2\,\varepsilon) - 3/2, \sqrt{3}/\varepsilon - 3\right)$, $\mathcal{I}_4 = \left[\sqrt{3}/\varepsilon - 3, 4\ (1-\sqrt{3}\,\varepsilon)\right)$, $\mathcal{I}_5 = \left[4\ (1-\sqrt{3}\,\varepsilon)\,, \frac{\sqrt{3}\,(1-\sqrt{3}\,\varepsilon)}{4\ \varepsilon - \sqrt{3}}\right)$, $\mathcal{I}_6 = \left[\frac{\sqrt{3}\,(1-\sqrt{3}\,\varepsilon)}{4\ \varepsilon-\sqrt{3}}, 2\,(\sqrt{3}/\varepsilon - 3)\right)$, $\mathcal{I}_7 = \left[2\ (\sqrt{3}/\varepsilon - 3)\,, \frac{2\,\sqrt{3}\,(1-3\,\varepsilon)}{4\ \varepsilon-\sqrt{3}}\right)$, and $\mathcal{I}_8 = \left[\frac{2\,\sqrt{3}\,(1-3\,\varepsilon)}{4\ \varepsilon-\sqrt{3}}, 1\right)$.

C.3.2 $\mu_A\big(N^\tau_{CS},\varepsilon\big)$ Under Association Alternatives

Under association, we compute $\mu_A\big(N^\tau_{CS},\varepsilon\big)$ explicitly.

For $\varepsilon \in \left[0, \sqrt{3}/21 \approx .0825\right)$,

$$\mu_A\big(N^\tau_{CS},\varepsilon\big) = \sum_{j=1}^{7} \varpi_{1,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j),$$

where

$$\varpi_{1,1}(\tau,\varepsilon) = \Big[(-63936\,\tau^6\varepsilon^4 + 20736\,\sqrt{3}\tau^6\varepsilon^3 - 145728\,\tau^5\varepsilon^4 - 6912\,\tau^6\varepsilon^2 + 46848\,\sqrt{3}\tau^5\varepsilon^3 + 181872\,\tau^4\varepsilon^4$$
$$+ 192\,\sqrt{3}\tau^6\,\varepsilon - 14976\,\tau^5\varepsilon^2 - 60480\,\sqrt{3}\tau^4\varepsilon^3 + 346896\,\tau^3\varepsilon^4 + 36\,\tau^6 + 256\,\sqrt{3}\tau^5\,\varepsilon + 22464\,\tau^4\varepsilon^2$$
$$- 107712\,\sqrt{3}\tau^3\varepsilon^3 - 296640\,\tau^2\varepsilon^4 + 128\,\tau^5 - 1200\,\sqrt{3}\tau^4\,\varepsilon + 28512\,\tau^3\varepsilon^2 + 93696\,\sqrt{3}\tau^2\varepsilon^3$$
$$- 228528\,\tau\,\varepsilon^4 + 63\,\tau^4 + 1056\,\sqrt{3}\tau^3\,\varepsilon - 27360\,\tau^2\varepsilon^2 + 74304\,\sqrt{3}\tau\,\varepsilon^3 + 81648\,\varepsilon^4 - 726\,\tau^3$$
$$- 208\,\sqrt{3}\tau^2\,\varepsilon - 25056\,\tau\,\varepsilon^2 - 25920\,\sqrt{3}\varepsilon^3 + 445\,\tau^2 + 768\,\sqrt{3}\tau\,\varepsilon + 7776\,\varepsilon^2 + 108\,\tau - 108)\tau^2\Big] \Big/$$
$$\Big[18\,(2\,\tau+1)(2\,\tau-1)\left(-6\,\varepsilon + \sqrt{3}\right)^2\left(6\,\varepsilon + \sqrt{3}\right)^2 (\tau+2)^2(\tau-1)^2\Big],$$

$$\varpi_{1,2}(\tau,\varepsilon) = \Big[(-62784\,\tau^7\varepsilon^4 + 19200\,\sqrt{3}\tau^7\varepsilon^3 - 199872\,\tau^6\varepsilon^4 - 4608\,\tau^7\varepsilon^2 + 56832\,\sqrt{3}\tau^6\varepsilon^3 + 34992\,\tau^5\varepsilon^4$$
$$- 320\,\sqrt{3}\tau^7\,\varepsilon - 9216\,\tau^6\varepsilon^2 - 28224\,\sqrt{3}\tau^5\varepsilon^3 + 476640\,\tau^4\varepsilon^4 + 164\,\tau^7 - 1600\,\sqrt{3}\tau^6\,\varepsilon + 22464\,\tau^5\varepsilon^2$$
$$- 151104\,\sqrt{3}\tau^4\varepsilon^3 + 82368\,\tau^3\varepsilon^4 + 484\,\tau^6 - 1872\,\sqrt{3}\tau^5\,\varepsilon + 31392\,\tau^4\varepsilon^2 + 6528\,\sqrt{3}\tau^3\varepsilon^3 - 405216\,\tau^2\varepsilon^4$$
$$- 15\,\tau^5 + 3872\,\sqrt{3}\tau^4\,\varepsilon - 31392\,\tau^3\varepsilon^2 + 153792\,\sqrt{3}\tau^2\varepsilon^3 - 252720\,\tau\,\varepsilon^4 - 1214\,\tau^4 + 3280\,\sqrt{3}\tau^3\,\varepsilon$$
$$- 47520\,\tau^2\varepsilon^2 + 67392\,\sqrt{3}\tau\,\varepsilon^3 - 46656\,\varepsilon^4 - 13\,\tau^3 - 768\,\sqrt{3}\tau^2\,\varepsilon - 7776\,\tau\,\varepsilon^2 + 324\,\tau^2 + 108\,\tau)\tau\Big] \Big/$$
$$\Big[18\,(\tau-1)^2(2\,\tau+1)^2(\tau+2)^2\left(-6\,\varepsilon + \sqrt{3}\right)^2\left(6\,\varepsilon + \sqrt{3}\right)^2\Big],$$

$$\varpi_{1,3}(\tau,\varepsilon) = \Big[(-62784\ \tau^6\varepsilon^4 + 19200\sqrt{3}\tau^6\varepsilon^3 - 74304\ \tau^5\varepsilon^4 - 4608\,\tau^6\varepsilon^2 + 18432\sqrt{3}\tau^5\varepsilon^3 + 183600\,\tau^4\varepsilon^4$$
$$- 320\sqrt{3}\tau^6\,\varepsilon - 65088\sqrt{3}\tau^4\varepsilon^3 + 179424\ \tau^3\varepsilon^4 + 164\ \tau^6 - 960\sqrt{3}\tau^5\,\varepsilon + 22464\ \tau^4\varepsilon^2 - 67584\ \sqrt{3}\tau^3\varepsilon^3$$
$$+ 3456\,\tau^2\varepsilon^4 + 156\,\tau^5 + 48\sqrt{3}\tau^4\,\varepsilon + 21456\,\tau^3\varepsilon^2 + 1728\sqrt{3}\tau^2\varepsilon^3 + 7776\,\tau\,\varepsilon^4 - 327\,\tau^4 - 112\sqrt{3}\tau^3\,\varepsilon$$
$$- 4320\,\tau^2\varepsilon^2 + 10368\sqrt{3}\tau\,\varepsilon^3 + 11664\ \varepsilon^4 - 74\ \tau^3 - 384\ \sqrt{3}\tau^2\,\varepsilon - 3888\,\tau\,\varepsilon^2 + 135\,\tau^2 + 54\ \tau)\tau\Big]\Big/$$
$$\Big[18\,(\tau-1)^2(2\,\tau+1)^2(\tau+2)\left(-6\,\varepsilon+\sqrt{3}\right)^2\left(6\,\varepsilon+\sqrt{3}\right)^2\Big],$$
$$\varpi_{1,4}(\tau,\varepsilon) = \Big[(-63072\,\tau^4\varepsilon^4 + 19584\ \sqrt{3}\tau^4\varepsilon^3 - 267552\,\tau^3\varepsilon^4 - 5184\ \tau^4\varepsilon^2 + 81408\sqrt{3}\tau^3\varepsilon^3 - 389304\ \tau^2\varepsilon^4$$
$$- 192\sqrt{3}\tau^4\,\varepsilon - 20160\,\tau^3\varepsilon^2 + 118176\sqrt{3}\tau^2\varepsilon^3 - 233712\,\tau\,\varepsilon^4 + 132\,\tau^4 - 896\sqrt{3}\tau^3\,\varepsilon - 29520\,\tau^2\varepsilon^2$$
$$+ 71712\sqrt{3}\tau\,\varepsilon^3 - 48600\,\varepsilon^4 + 488\,\tau^3 - 1072\sqrt{3}\tau^2\,\varepsilon - 18576\,\tau\,\varepsilon^2 + 15552\sqrt{3}\varepsilon^3 + 601\,\tau^2 - 384\ \sqrt{3}\tau\,\varepsilon$$
$$- 3888\,\varepsilon^2 + 297\,\tau + 54)\tau^2\Big]\Big/\Big[18\,(2\,\tau+1)^2(\tau+2)(\tau+1)\left(-6\,\varepsilon+\sqrt{3}\right)^2\left(6\,\varepsilon+\sqrt{3}\right)^2\Big],$$
$$\varpi_{1,5}(\tau,\varepsilon) = \Big[-49968\,\tau^5\varepsilon^4 + 15936\sqrt{3}\tau^5\varepsilon^3 - 219384\ \tau^4\varepsilon^4 - 4896\,\tau^5\varepsilon^2 + 64992\sqrt{3}\tau^4\varepsilon^3 - 349920\,\tau^3\varepsilon^4$$
$$+ 32\sqrt{3}\tau^5\,\varepsilon - 18000\,\tau^4\varepsilon^2 + 90720\sqrt{3}\tau^3\varepsilon^3 - 270216\,\tau^2\varepsilon^4 + 58\,\tau^5 + 16\sqrt{3}\tau^4\,\varepsilon - 22032\,\tau^3\varepsilon^2$$
$$+ 51840\sqrt{3}\tau^2\varepsilon^3 - 112752\,\tau\,\varepsilon^4 + 191\,\tau^4 - 9072\,\tau^2\varepsilon^2 + 10368\sqrt{3}\tau\,\varepsilon^3 - 23328\,\varepsilon^4 + 189\,\tau^3$$
$$+ 54\ \tau^2\Big]\Big/\Big[18\,(\tau+2)\left(6\,\varepsilon+\sqrt{3}\right)^2\left(-6\,\varepsilon+\sqrt{3}\right)^2(2\,\tau+1)(\tau+1)\Big],$$
$$\varpi_{1,6}(\tau,\varepsilon) = \Big[-50040\,\tau^6\varepsilon^4 + 16032\ \sqrt{3}\tau^6\varepsilon^3 - 221220\,\tau^5\varepsilon^4 - 5040\,\tau^6\varepsilon^2 + 66864\ \sqrt{3}\tau^5\varepsilon^3 - 368100\,\tau^4\varepsilon^4$$
$$+ 64\ \sqrt{3}\tau^6\,\varepsilon - 19944\,\tau^5\varepsilon^2 + 103728\sqrt{3}\tau^4\varepsilon^3 - 356616\,\tau^3\varepsilon^4 + 50\,\tau^6 + 256\sqrt{3}\tau^5\,\varepsilon - 29880\,\tau^4\varepsilon^2$$
$$+ 88992\sqrt{3}\tau^3\varepsilon^3 - 307152\,\tau^2\varepsilon^4 + 179\,\tau^5 + 304\ \sqrt{3}\tau^4\,\varepsilon - 17712\,\tau^3\varepsilon^2 + 46656\sqrt{3}\tau^2\varepsilon^3$$
$$- 194400\,\tau\,\varepsilon^4 + 185\,\tau^4 + 96\sqrt{3}\tau^3\,\varepsilon - 2592\,\tau^2\varepsilon^2 + 10368\sqrt{3}\tau\,\varepsilon^3 - 46656\,\varepsilon^4$$
$$+ 54\ \tau^3\Big]\Big/\Big[18\,(\tau+1)\left(6\,\varepsilon+\sqrt{3}\right)^2\left(-6\,\varepsilon+\sqrt{3}\right)^2(2\,\tau+1)(\tau+2)\tau\Big],$$
$$\varpi_{1,7}(\tau,\varepsilon) = \Big[3\,(-1512\,\tau^4\varepsilon^4 + 480\sqrt{3}\tau^4\varepsilon^3 - 3780\,\tau^3\varepsilon^4 - 144\,\tau^4\varepsilon^2 + 1200\sqrt{3}\tau^3\varepsilon^3 + 216\,\tau^2\varepsilon^4 - 360\,\tau^3\varepsilon^2$$
$$+ 480\sqrt{3}\tau^2\varepsilon^3 + 4752\,\tau\,\varepsilon^4 + 2\,\tau^4 - 288\,\tau^2\varepsilon^2 + 1728\,\varepsilon^4 + 5\,\tau^3 + 2\,\tau^2)\Big]\Big/\Big[2\ \left(6\,\varepsilon+\sqrt{3}\right)^2$$
$$\left(-6\,\varepsilon+\sqrt{3}\right)^2(2\,\tau+1)(\tau+2)\Big],$$

with the corresponding intervals $\mathcal{I}_1 = \left[0, \frac{3\sqrt{3}\,\varepsilon}{2\left(1-\sqrt{3}\,\varepsilon\right)}\right)$, $\mathcal{I}_2 = \left[\frac{3\sqrt{3}\,\varepsilon}{2\left(1-\sqrt{3}\,\varepsilon\right)}, \frac{2\sqrt{3}\,\varepsilon}{1-2\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_3 = \left[\frac{2\sqrt{3}\,\varepsilon}{1-2\sqrt{3}\,\varepsilon}, \frac{3\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_4 = \left[\frac{3\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}, \frac{3\sqrt{3}\,\varepsilon}{1-4\ \sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_5 = \left[\frac{3\sqrt{3}\,\varepsilon}{1-4\ \sqrt{3}\,\varepsilon}, \frac{6\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_6 = \left[\frac{6\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}, \frac{6\sqrt{3}\,\varepsilon}{1-4\ \sqrt{3}\,\varepsilon}\right)$, and $\mathcal{I}_7 = \left[\frac{6\sqrt{3}\,\varepsilon}{1-4\ \sqrt{3}\,\varepsilon}, 1\right)$.

For $\varepsilon \in \left[\sqrt{3}/21, \sqrt{3}/12\right)$,

$$\mu_A\big(N_{CS}^{\tau}, \varepsilon\big) = \sum_{j=1}^{4} \varpi_{2,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j),$$

where $\varpi_{2,j}(\tau,\varepsilon) = \varpi_{1,j}(\tau,\varepsilon)$ for $j \in \{1,2,3,4\}$ with the corresponding intervals $\mathcal{I}_1 = \left[0, \frac{3\sqrt{3}\,\varepsilon}{2\left(1-\sqrt{3}\,\varepsilon\right)}\right)$, $\mathcal{I}_2 = \left[\frac{3\sqrt{3}\,\varepsilon}{2\left(1-\sqrt{3}\,\varepsilon\right)}, \frac{2\sqrt{3}\,\varepsilon}{1-2\sqrt{3}\,\varepsilon}\right)$, $\mathcal{I}_3 = \left[\frac{2\sqrt{3}\,\varepsilon}{1-2\sqrt{3}\,\varepsilon}, \frac{3\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}\right)$, and $\mathcal{I}_4 = \left[\frac{3\sqrt{3}\,\varepsilon}{1-\sqrt{3}\,\varepsilon}, 1\right)$.

For $\varepsilon \in \left[\sqrt{3}/12, 2\sqrt{3}/15\right)$,

$$\mu_A\left(N_{CS}^{\tau}, \varepsilon\right) = \sum_{j=1}^{2} \varpi_{3,j}(\tau,\varepsilon)\,\mathbf{I}(\tau \in \mathcal{I}_j)$$

where

$$\varpi_{3,1}(\tau,\varepsilon) = \frac{(\tau^2+5\,\tau+9)\tau^2}{18\,(\tau+2)^2},$$

$$\begin{aligned}\varpi_{3,2}(\tau,\varepsilon) = -[&(-216\,\tau^6\varepsilon^4 + 288\,\sqrt{3}\tau^6\varepsilon^3 - 1836\,\tau^5\varepsilon^4 - 432\,\tau^6\varepsilon^2 + 2160\,\sqrt{3}\tau^5\varepsilon^3 - 6786\,\tau^4\varepsilon^4 + 96\,\sqrt{3}\tau^6\,\varepsilon \\ &- 2808\,\tau^5\varepsilon^2 + 6600\,\sqrt{3}\tau^4\varepsilon^3 - 13401\,\tau^3\varepsilon^4 - 24\;\tau^6 + 528\,\sqrt{3}\tau^5\,\varepsilon - 6876\,\tau^4\varepsilon^2 + 9588\,\sqrt{3}\tau^3\varepsilon^3 \\ &- 16074\;\tau^2\varepsilon^4 - 108\,\tau^5 + 1000\,\sqrt{3}\tau^4\,\varepsilon - 6498\,\tau^3\varepsilon^2 + 7500\,\sqrt{3}\tau^2\varepsilon^3 - 10611\,\tau\,\varepsilon^4 - 154\;\tau^4 \\ &+ 484\;\sqrt{3}\tau^3\,\varepsilon - 2178\,\tau^2\varepsilon^2 + 2484\;\sqrt{3}\tau\,\varepsilon^3 - 2916\,\varepsilon^4 - 25\,\tau^3 - 92\,\sqrt{3}\tau^2\,\varepsilon + 162\,\tau\,\varepsilon^2 + 23\,\tau^2 \\ &- 36\,\sqrt{3}\tau\,\varepsilon + 9\,\tau)\tau]/[2\,(2\,\tau-1)(2\,\tau+1)^2(\tau+2)^2\left(-3\,\varepsilon+\sqrt{3}\right)^4],\end{aligned}$$

with the corresponding intervals $\mathcal{I}_1 = \left[0, \frac{3\sqrt{3}\,\varepsilon}{2\left(1-\sqrt{3}\,\varepsilon\right)}\right)$, and $\mathcal{I}_2 = \left[\frac{3\sqrt{3}\,\varepsilon}{2\left(1-\sqrt{3}\,\varepsilon\right)}, 1\right)$.

For $\varepsilon \in \left[2\sqrt{3}/15, \sqrt{3}/3\right)$, $\mu_A\left(N_{CS}^{\tau}, \varepsilon\right) = \varpi_{3,1}(\tau,\varepsilon)\,\mathbf{I}(\tau \in [0,1])$.

Bibliography

[1] P. J. Bickel and A. K. Doksum. *Mathematical Statistics, Basic Ideas and Selected Topics.* Prentice Hall, New Jersey, 1977.

[2] H. Callaert and P. Janssen. The Berry-Esseen theorem for U-statistics. *Annals of Statistics*, 6:417–421, 1978.

[3] E. Ceyhan. The distribution of the domination number of class cover catch digraphs for non-uniform one-dimensional data. Technical Report 646, Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682, 2004.

[4] G. Chartrand and L. Lesniak. *Graphs & Digraphs.* Chapman & Hill, 1996.

[5] V. Chvatal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4:233–235, 1979.

[6] D. A. Coomes, M. Rees, and Turnbull L. Identifying association and association in fully mapped spatial data. *Ecology*, 80:554–565, 1999.

[7] N. A. C. Cressie. *Statistics for Spatial Data.* Wiley, New York, 1991.

[8] J. DeVinney. *The Class Cover Problem and its Applications in Pattern Recognition.* PhD thesis, Johns Hopkins University, Baltimore, MD, 2003.

[9] J. DeVinney and C. E. Priebe. Class cover catch digraphs. *Submitted for publication. (Available as Technical Report No. 633, Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682.)*, 2002.

[10] J. DeVinney, C. E. Priebe, D. J. Marchette, and D. Socolinsky. Random walks and catch digraphs in classification. `http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/%DeVinneyJason.paper.pdf`, 2002. Computing Science and Statistics, Vol. 34.

[11] L. Devroye, L. Gyorfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer Verlag, New York, 1996.

[12] P. Dixon. Testing spatial segregation using a nearest-neighbor contingency table. *Ecology*, 75(7):1940–1948, 1994.

[13] C. V. Eeden. The relation between Pitman's asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics. *The Annals of Mathematical Statistics*, 34(4):1442–1451, 1963.

[14] R. S. Garfinkel and G. L. Nemhauser. *Integer Programming.* John Wiley & Sons, New York, 1972.

[15] Z. Guan. A Berry-Esseen bound for k-sample symmetric statistics. *Northeast. Math. J.*, 10:411–420, 1994.

[16] J. L. Jr. Hodges and E. L. Lehmann. The efficiency of some nonparametric competitors of the t-test. *The Annals of Mathematical Statistics*, 27(2):324–335, 1956.

[17] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19:293–325, 1948.

[18] S. Janson, T. Łuczak, and A. Rucinński. *Random Graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York, 2000.

[19] J. W. Jaromczyk and G. T. Toussaint. Relative neighborhood graphs and their relatives. *Proceedings of IEEE*, 80:1502–1517, 1992.

[20] R. Karp. *Reducibility Among Combinatorial Problems.* Plenum Press, 1972.

[21] A. F. Karr. *Probability.* Springer-Verlag, New York, 1992.

[22] M. Kendall and A. Stuart. *The advanced theory of statistics, volume 2., 4th edition.* Griffin, London, 1979.

[23] C. Kimberling. Encyclopedia of triangle centers. `http://faculty.evansville.edu/ck6/encyclopedia/ETC.html`.

[24] S. N. Lahiri. On consistency of estimators based on spatial data under infill asymptotics. *Sankhya: The Indian Journal of Statistics, Series A*, 58(3):403–417, 1996.

[25] A. J. Lee. *U-Statistics, Theory and Practice.* Marcel Dekker, Inc., 1990.

[26] C. Lee. Domination in digraphs. *Journal of Korean Mathematical Society*, 4:843–853, 1998.

[27] E. L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks.* Prentice-Hall, Upper Saddle River, NJ, 1988.

[28] E. L. Lehmann. *Elements of Large Sample Theory.* Springer, 1999.

[29] D. J. Marchette and C. E. Priebe. Characterizing the scale dimension of a high dimensional classification problem. *Pattern Recognition*, 36(1):45–60, 2003.

[30] A. Okabe, B. Boots, and Sugihara K. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* Wiley, 1992.

[31] A. K. Parekh. Analysis of a greedy heuristic for finding small dominating sets in graphs. *Information Processing Letters*, 39:237–240, 1991.

[32] M. S. Paterson and F. F. Yao. On nearest neighbor graphs. In *Proceedings of* 19^{th} *Int. Coll. Automata, Languages and Programming, Springer LNCS*, volume 623, pages 416–426, 1992.

[33] E. C. Pielou. Segregation and symmetry in two-species populations as studied by nearest-neighbor relationships. *Journal of Ecology*, 49(2):255–269, 1961.

[34] E. J. G. Pitman. *Lecture notes on non-parametric statistics.* Columbia University, 1949.

[35] C. E. Priebe, J. G. DeVinney, and D. J. Marchette. On the distribution of the domination number of random class catch cover digraphs. *Statistics and Probability Letters*, 55:239–246, 2001.

[36] C. E. Priebe, D. J. Marchette, J. DeVinney, and D. Socolinsky. Classification using class cover catch digraphs. *Journal of Classification*, 20(1):3–23, 2003.

[37] C. E. Priebe, J. L. Solka, D. J. Marchette, and B. T. Clark. Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays. *Computational Statistics and Data Analysis on Visualization*, 43-4:621–632, 2003.

[38] R. H. Randles and D. A. Wolfe. *Introduction to the Theory of Nonparametric Statistics.* Wiley, New York, 1979.

[39] M. Sen, S. Das, A.B. Roy, and D.B. West. Interval digraphs: An analogue of interval graphs. *J. of Graph Theory*, 13:189–202, 1989.

[40] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1980.

[41] G. T. Toussaint. The relative neighborhood graphs of a finite planar set. *Pattern Recognition*, 12:261–268, 1980.

[42] Z. Tuza. Inequalities for minimal covering sets in sets in set systems of given rank. *Discrete Applied Mathematics*, 51:187–195, 1994.

[43] E. weisstein. Triangle centers, Eric Weisstein's world of mathematics. `http://mathworld.wolfram.com/TriangleCenter.html`.

[44] D. B. West. *Introduction to Graph Theory, 2nd ed.* Prentice Hall, NJ., 2001.

[45] J. Xie. *Generalizing the Mann-Whitney-Wilcoxon Statistic*. PhD thesis, Johns Hopkins University, Baltimore, MD, 1999.

Vita

Elvan Ceyhan was born on September 20, 1974 in Yozgat, Turkey. He had been living in Turkey for 24 years before he came to U.S. He attended primary school in Yozgat, secondary school in Salihli, lycee in Izmir, and college in Istanbul, Turkey. He studied mathematics at Koç University, Istanbul where he earned a B.S. in 1997. Upon graduation, he entered the Ph.D. program in mathematics at Oklahoma State University at Stillwater, OK and switched to statistics in 1998, where he received a M.S. degree in 2000. In June of 2000, he came to the Department of Mathematical Sciences at The Johns Hopkins University to begin work towards a Ph.D. in mathematical sciences under the direction of his advisor, Carey E. Priebe. Elvan began work first in statistical depth and later switched to proximity maps. In October 2004 he completed a thesis on the proximity catch digraphs.

In the future, Elvan plans on pursuing an academic career, as a researcher pursuing problems in the area of mathematical sciences, in particular statistics, and applying his expertise in this field.

Baltimore, 2004