Statistics and Probability Letters 79 (2009) 223–233

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A CLT for a one-dimensional class cover problem

Pengfei Xiang [a], John C. Wierman [b,*]

[a] Wachovia Bank, USA

[b] Department of Applied Mathematics and Statistics, 302 Whitehead Hall, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

ARTICLE INFO

Article history:
Received 4 January 2008
Received in revised form 28 July 2008
Accepted 31 July 2008
Available online 9 August 2008

ABSTRACT

The central limit theorem is proved for the domination number of random class cover catch digraphs generated by uniform data in one dimension. The class cover problem is motivated by applications in statistical pattern classification.

1. Introduction

The probabilistic behavior of random Class Cover Catch Digraphs (CCCDs) is of considerable recent interest, due to their use in pattern classification methods involving proximity regions. Priebe et al. (2001) found the exact distribution of the domination number of CCCDs for the uniform distribution in one dimension. Under the same conditions, DeVinney and Wierman (2002) proved the SLLN for the domination number. Wierman and Xiang (2004) further extended the SLLN to a class of continuous distributions in one dimension. Xiang (2006) calculated the variance of the domination number for the uniform distribution in one dimension.

Solutions to the CCCD problem can be used to build classifiers. More details and examples of the application of CCCDs to classification are presented in Priebe et al. (2003), Marchette and Priebe (2003), Ceyhan and Priebe (2005), Eveland et al. (2005), Ceyhan and Priebe (2006), DeVinney and Priebe (2006), Ceyhan et al. (2006), and Ceyhan et al. (2007).

1.1. *The class cover problem*

The class cover problem (CCP) is motivated by its applications in pattern classification. The study of the CCP was initiated by Cannon and Cowen (2000), and has been actively pursued recently, because its solution can be directly used to generate classifiers competitive with traditional methods.

For a formal description of the CCP, consider a dissimilarity function $d : \Omega \times \Omega \rightarrow \mathcal{R}$ such that $d(\alpha, \beta) = d(\beta, \alpha) \geq d(\alpha, \alpha) = 0$ for $\forall \alpha, \beta \in \Omega$. We suppose $\mathcal{X} = \{X_i : i = 1, \ldots, n\}$ and $\mathcal{Y} = \{Y_j : j = 1, \ldots, m\}$ are two sets of i.i.d. random variables with class-conditional distribution functions F_X and F_Y, respectively. We assume that each X_i is independent of each Y_j, and all $X_i \in \mathcal{X}$ and all $Y_j \in \mathcal{Y}$ are distinct with probability one. For each X_i, we define its covering ball by $B(X_i) = \{\omega \in \Omega : d(\omega, X_i) < \min_j d(Y_j, X_i)\}$. A *class cover of* $\mathcal{X}$ is a subset of covering balls whose union contains all $X_i \in \mathcal{X}$. Obviously, the set consisting of all covering balls is a class cover. However, we want to choose a class cover to represent class $\mathcal{X}$ that is as small as possible, to make the classifier less complex while keeping most of the relevant information. Therefore, the CCP we consider here is to find a minimum-cardinality class cover.

Furthermore, we can convert the CCP to the graph theory problem of finding dominating sets. The *class cover catch digraph* (CCCD) induced by a CCP is the digraph $D = (V, A)$ with the vertex set $V = \{X_i : i = 1, \ldots, n\}$ and the arc set A such that there is an arc (X_i, X_j) if and only if $X_j \in B(X_i)$. It is easy to see that the CCP is actually equivalent to finding a minimum-cardinality dominating set of the corresponding CCCD.

* Corresponding author. Tel.: +1 410 516 7211; fax: +1 410 516 7459.
E-mail address: wierman@jhu.edu (J.C. Wierman).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2008.07.045

1.2. Previous results

The domination number of a CCCD is the cardinality of the CCCD's minimum dominating set. In the CCCD setting, we denote the domination number by $\Gamma_{n,m}(\mathcal{X}, \mathcal{Y})$, or simply by $\Gamma_{n,m}$.

DeVinney and Wierman (2002) proved the Strong Law of Large Numbers (SLLN) for the special case $\Omega = R$, $F_X = F_Y = U[0, 1]$:

Theorem 1.1. *If $\Omega = \mathbf{R}$, $F_X = F_Y = U[0, 1]$, and $m = \lfloor rn \rfloor$, $r \in (0, \infty)$, then*

$$\lim_{n\to+\infty} \frac{\Gamma_{n,m}}{n} = g(r) \equiv \frac{r(12r + 13)}{3(r + 1)(4r + 3)} \quad \text{a.s.}$$

Wierman and Xiang (2008) generalized this result by proving the SLLN for general distributions in one dimension:

Theorem 1.2. *If $\Omega = \mathbf{R}$, the density functions f_X, f_Y are continuous and bounded on $[a, b]$, and $m/n \to r$, $r \in (0, \infty)$, then*

$$\lim_{n\to\infty} \frac{\Gamma_{n,m}(f_X, f_Y)}{n} = \int_a^b g\left(r \cdot \frac{f_Y(u)}{f_X(u)}\right) \cdot f_X(u)\mathrm{d}u \quad \text{a.s.},$$

where $g(r)$ is the same as in Theorem 1.1.

As a first step in proving the Central Limit Theorem (CLT) for the domination number $\Gamma_{n,m}$, Xiang (2006) calculated the limiting variance for $\Gamma_{n,m}$. The calculation is very technical and lengthy, but the final result can be simply stated as follows, with an outline of the calculation given in Appendix A:

Theorem 1.3. *If $\Omega = \mathbf{R}$, $F_X = F_Y = U[0, 1]$, and $m/n \to r$, $r \in (0, \infty)$, then*

$$\frac{\mathrm{Var}(\Gamma_{n,m})}{m} \to v(r) = \frac{1536r^5 + 6848r^4 + 11536r^3 + 8836r^2 + 2793r + 180}{9(r + 1)^3(4r + 3)^4}.$$

1.3. A central limit theorem

In this paper, we prove the CLT for the domination number generated by uniformly distributed data:

Theorem 1.4. *If $\Omega = \mathbf{R}$, $F_X = F_Y = U[0, 1]$, and $m/n \to r$, $r \in (0, \infty)$, then*

$$\frac{1}{m^{1/2}}\left(\Gamma_{n,m} - E[\Gamma_{n,m}]\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\sigma^2 = \lim_{m\to\infty} \frac{\mathrm{Var}[\Gamma n,m]}{m}$ (the exact limiting value is given in Theorem 1.3).

In our proof of the CLT, we extensively use the concept of *negative association*, so in the next sub-section we introduce some of its basic properties and consequences.

1.4. Negative association

The concept of negatively associated (NA) random variables was introduced and carefully studied by Joag-Dev and Proschan (1983). Since then, limit theorems for this type of random variables have been well established. (See Newman (1984) and Taylor et al. (2002).)

Definition 1.1. Random variables X_i, $i = 1, \ldots, k$, are said to be negatively associated (NA) if for every pair of disjoint subsets I, J of $\{1, \ldots, k\}$, and any increasing functions f_I and f_J the following covariance exists,

$$\mathrm{Cov}\left\{f_I(X_i, i \in I), f_J(X_j, j \in J)\right\} \le 0,$$

"NA" may also refer to the random vector $(X_1, \ldots, X_k)$.

The following proposition is obvious from the definition.

Proposition 1.1. *Increasing functions defined on disjoint subsets of a set of NA random variables are NA.*

Furthermore, several types of random vectors were proven to be NA by Joag-Dev and Proschan (1983), particularly:

Proposition 1.2. *A multinomial random vector is NA.*

Proposition 1.3. *Let $X_1, \dots, X_k$ be k independent random variables with PF_2 (log-concave) densities. Then, given $\sum_{i=1}^{k} X_i$, the random vector $(X_1, \dots, X_k)$ is NA.*

In some situations, a dependence condition called negative dependence, which is weaker than negative association, is used.

Definition 1.2. Random variables $X_i, i = 1, \dots, k$, are said to be negatively dependent (ND) if for any real numbers $x_i, i = 1, \dots, k$,

$$P(X_i > x_i, i = 1, \dots, k) \le \prod_{i=1}^{k} P(X_i > x_i),$$

and

$$P(X_i \le x_i, i = 1, \dots, k) \le \prod_{i=1}^{k} P(X_i \le x_i).$$

Note. By choosing the functions in Definition 1.1 to be indicator functions of the events $\{X_i > x_i\}$ and the events $\{X_i \le x_i\}$, it is readily apparent that negative association implies negative dependence.

Taylor et al. (2002) proved that the SLLN holds for row-wise ND random variable arrays. We apply one part of their theorem:

Theorem 1.5. *Let $\{X_{k,m} : 1 \le k \le m, m \ge 1\}$ be row-wise ND random variable arrays such that $E[X_{k,m}] = 0$ for each k and m. If $|X_{k,m}| \le M$ for all k and m for some constant $M < \infty$, then*

$$\frac{1}{m^{1/p}} \sum_{k=1}^{m} X_{k,m} \xrightarrow{a.s.} 0, \quad 0 < p < 2.$$

Newman (1984, Theorem 11) established the CLT for ND sequences. First, a distributional limit theorem for row-wise ND random variable arrays was proved:

Theorem 1.6. *Suppose $X_{k,m}$ and $Y_{k,m}$, $1 \le k \le m, m \ge 1$, are triangular arrays such that for each m and k, random variable $X_{k,m}$ is equidistributed with $Y_{k,m}$. Assume for each m, the random variables $X_{k,m}, k = 1, \dots, m$ are ND, but $Y_{k,m}, k = 1, \dots, m$ are independent. If in addition,*

$$\lim_{m\to\infty} \frac{\sum_{1\le i<j\le m} \mathrm{Cov}(X_{i,m}, X_{j,m})}{m} = 0,$$

then $\frac{1}{m^{1/2}} \sum_{k=0}^{m} X_{k,m}$ converges in distribution if and only if $\frac{1}{m^{1/2}} \sum_{k=0}^{m} Y_{k,m}$ converges in distribution to the same limit distribution.

By applying the classical CLT for bounded i.i.d. random variable arrays to the $\{Y_{k,m}\}$ in the theorem above, we conclude that:

Theorem 1.7. *Let $\{X_{k,m} : 1 \le k \le m, m \ge 1\}$ be identically distributed row-wise ND random variable arrays such that $E[X_{k,m}] = 0$ for each k and m. If $|X_{k,m}| \le M$, and*

$$\lim_{m\to\infty} \frac{\sum_{1\le k<l\le m} \mathrm{Cov}(X_{k,m}, X_{l,m})}{m} = 0,$$

then

$$\frac{1}{m^{1/2}} \sum_{k=1}^{m} X_{k,m} \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where $\sigma^2 = \lim_{m\to\infty} \frac{\mathrm{Var}[\sum_{k=1}^{m} X_{k,m}]}{m}$.

Remark. Since negative association implies negative dependence, Theorems 1.5–1.7 will all hold if the $\{X_{k,m} : 1 \le k \le m, m \ge 1\}$ are row-wise NA.

In Appendix B, we show that the $\{N_{j,m} : 0 \le j \le m, m \ge 1\}$ in the CCCD problem are row-wise NA.

2. Proof of the CLT

2.1. Basic idea

For $j = 1, 2, 3, \ldots, m$, let $Y_{(j)}$ denote the jth order statistic of $Y_1, \ldots, Y_m$, and define $Y_{(0)} = 0$ and $Y_{(m+1)} = 1$. Define $\alpha_{j,m}$ as the minimum number of covering balls required to cover the $N_{j,m}$ X-points contained in $[Y_{(j)}, Y_{(j+1)}]$. It should be noted that $\Gamma_{n,m} = \sum_{j=0}^{m} \alpha_{j,m}$; thus, the original CCP is decomposed into $m + 1$ sub-CCPs of finding the domination number $\alpha_{j,m}$ in the interval $[Y_{(j)}, Y_{(j+1)}]$. Since $\alpha_{j,m}, j = 0, \ldots, m$, are dependent, the classical CLT theorem cannot be directly applied. However, given $N_{j,m}, j = 0, \ldots, m$, the domination numbers $\alpha_{j,m}$, $0 = 1, \ldots, m$, are conditionally independent. Therefore, we can calculate the conditional characteristic function f_m for $\sum_{j=0}^{m} \alpha_{j,m}$ on the σ-field generated by $N_{j,m}, j = 1, \ldots, m$. Using the Taylor expansion, a lengthy calculation shows that f_m can be expressed in terms of $E[\alpha_{j,m} \mid N_{j,m}]$ and $\mathrm{Var}[\alpha_{j,m} \mid N_{j,m}]$. We prove that these two sequences of random variables are both negatively associated. Thus, applying limit theorems for row-wise NA arrays, we conclude that f_m converges to a constant almost surely. By the dominated convergence theorem, the unconditional characteristic function $E[f_m]$ goes to the same constant, hence the result follows by the convergence theorem for characteristic functions.

2.2. Detailed proof

Denote $F_m = \sigma(N_{0,m}, \ldots, N_{m,m})$, the σ-field generated by $N_{j,m}, j = 0, \ldots, m$. Let

$$Z_{j,m} = \frac{1}{m^{1/2}} \left(\alpha_{j,m} - E[\alpha_{j,m}]\right)$$

and

$$f_m(t) = E\left[\mathrm{e}^{\mathrm{i}t \sum_{j=0}^{m} Z_{j,m}} \mid F_m \right].$$

By Lemma A.3, the $Z_{j,m}, j = 0, \ldots, m$ are conditionally independent given F_m, so

$$f_m(t) = \prod_{j=0}^{m} E\left[\mathrm{e}^{\mathrm{i}tZ_{j,m}} \mid F_m\right].$$

Again by Lemma A.3, we know that $Z_{j,m}$ only depends on $N_{j,m}$ given F_m, so

$$f_m(t) = \prod_{j=0}^{m} E\left[\mathrm{e}^{\mathrm{i}tZ_{j,m}} \mid N_{j,m}\right].$$

Using the Taylor expansion

$$\mathrm{e}^{\mathrm{i}z} = 1 + \mathrm{i}z - \frac{1}{2}z^2 + A(z), \quad \text{where } |A(z)| \le \frac{|z|^3}{6},$$

the conditional characteristic function of $Z_{j,m}$ can be written as

$$E\left[\mathrm{e}^{\mathrm{i}tZ_{j,m}} \mid F_m\right] = 1 + \mathrm{i}tE\left[Z_{j,m} \mid N_{j,m}\right] - \frac{t^2}{2}E\left[Z_{j,m}^2 \mid N_{j,m}\right] + r_{j,m}^{(1)},$$

where

$$|r_{j,m}^{(1)}| = E\left[|A(tZ_{j,m})| \mid N_{j,m}\right] \le E\left[\frac{|tZ_{j,m}|^3}{6} \middle| N_{j,m} \right].$$

Therefore, by substituting the formula for $E\left[\mathrm{e}^{\mathrm{i}tZ_{j,m}} \mid F_m\right]$ into the expression for $f_m(t)$, we obtain

$$\begin{aligned} \log f_m(t) &= \sum_{j=0}^{m} \log E\left[\mathrm{e}^{\mathrm{i}tZ_{j,m}} \mid F_{j,m}\right] \\ &= \sum_{j=0}^{m} \log\left(1 + \mathrm{i}tE[Z_{j,m} \mid N_{j,m}] - \frac{t^2}{2}E\left[Z_{j,m}^2 \mid N_{j,m}\right] + r_{j,m}^{(1)}\right). \end{aligned}$$

Again, by the Taylor expansion

$$\log(1 + \delta) = \delta - \frac{\delta^2}{2} + r(\delta), \quad \text{where } |r(\delta)| \le \frac{|\delta|^3}{24} \text{ for } |\delta| < 1, \tag{1}$$

$\log f_m(t)$ can be further expanded as

$$\log f_m(t) = \sum_{j=0}^{m} \left(\mathrm{it}E\left[Z_{j,m} \mid N_{j,m}\right] - \frac{t^2}{2} E\left[Z_{j,m}^2 \mid N_{j,m}\right] + r_{j,m}^{(1)} - \frac{1}{2}\left(\mathrm{it}E[Z_{j,m} \mid N_{j,m}] - \frac{t^2}{2}E\left[Z_{j,m}^2 \mid N_{j,m}\right] + r_{j,m}^{(1)}\right)^2 + r_{j,m}^{(2)} \right), \tag{2}$$

where $\left|r_{j,m}^{(2)}\right| \le \frac{1}{24}\left|\mathrm{it}E\left[Z_{j,m} \mid N_{j,m}\right] - \frac{t^2}{2}E\left[Z_{j,m}^2 \mid N_{j,m}\right] + r_{j,m}^{(1)}\right|^3$.

Recall that $\left|r_{j,m}^{(1)}\right|$ is bounded by $E\left[\frac{|tZ_{j,m}|^3}{6} \mid N_{j,m}\right]$. Since $|\alpha_{j,m}|$ is bounded by 2 as shown in Priebe et al. (2001), $|Z_{j,m}| = \left|\frac{1}{m^{1/2}}\left(\alpha_{j,m} - E[\alpha_{j,m}]\right)\right| \le \frac{4}{\sqrt{m}}$. Thus

$$\left|r_{j,m}^{(1)}\right| \le \frac{|t|^3 4^3}{6m^{3/2}} \equiv C_1 \frac{1}{m^{3/2}} \tag{3}$$

where $C_1 = C_1(t) \equiv \frac{32}{3}|t|^3$.

We now proceed to consider the quadratic term in Eq. (2). By the same reasoning used to derive the bound on $\left|r_{j,m}^{(1)}\right|$, we conclude that

$$\left|E[Z_{j,m} \mid N_{j,m}]\right| \le \frac{4}{m^{1/2}}, \tag{4}$$

$$\left|E\left[Z_{j,m}^2 \mid N_{j,m}\right]\right| \le \frac{16}{m}. \tag{5}$$

Using Inequalities (3)–(5), we may write the quadratic term in Eq. (2) as

$$\frac{t^2}{2}E[Z_{j,m} \mid N_{j,m}]^2 + r_{j,m}^{(3)}, \tag{6}$$

where

$$\left|r_{j,m}^{(3)}\right| \le C_3 \frac{1}{m^{3/2}}. \tag{7}$$

It remains to check $\left|r_{j,m}^{(2)}\right|$. Based on the bound for $\left|r_{j,m}^{(2)}\right|$ given in Formula (2), and Inequalities (3)–(5), when m is sufficiently large,

$$\left|r_{j,m}^{(2)}\right| \le \frac{1}{24}\left|\mathrm{it}E[Z_{j,m} \mid N_{j,m}] - \frac{t^2}{2}E\left[Z_{j,m}^2 \mid N_{j,m}\right] + r_{j,m}^{(1)}\right|^3 \le C_2 \frac{1}{m^{3/2}}. \tag{8}$$

Now we put all the pieces together. By substituting Formula (6) into Eq. (2), we have

$$\begin{aligned} \log f_m(t) &= \sum_{j=0}^{m}\left(\mathrm{it}E[Z_{j,m} \mid N_{j,m}] - \frac{t^2}{2}E\left[Z_{j,m}^2 \mid N_{j,m}\right] + \frac{t^2}{2}E[Z_{j,m} \mid N_{j,m}]^2 + r_{j,m}^{(1)} + r_{j,m}^{(2)} + r_{j,m}^{(3)}\right) \\ &= \mathrm{it}\sum_{j=0}^{m}E[Z_{j,m} \mid N_{j,m}] - \frac{t^2}{2}\sum_{j=0}^{m}\mathrm{Var}[Z_{j,m} \mid N_{j,m}] + \sum_{j=0}^{m}\left(r_{j,m}^{(1)} + r_{j,m}^{(2)} + r_{j,m}^{(3)}\right), \end{aligned} \tag{9}$$

so

$$f_m(t) = \mathrm{e}^{\mathrm{it}\sum_{j=0}^{m}E[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{-\frac{t^2}{2}\sum_{j=0}^{m}\mathrm{Var}[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{\sum_{j=0}^{m}\left(r_{j,m}^{(1)}+r_{j,m}^{(2)}+r_{j,m}^{(3)}\right)},$$

and taking expectations yields that the characteristic function $E[f_m(t)]$ equals

$$E\left[\mathrm{e}^{\mathrm{it}\sum_{j=0}^{m}E[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{-\frac{t^2}{2}\sum_{j=0}^{m}\mathrm{Var}[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{\sum_{j=0}^{m}\left(r_{j,m}^{(1)}+r_{j,m}^{(2)}+r_{j,m}^{(3)}\right)}\right]. \tag{10}$$

Note that by Lemma B.2, the random variable array $\{N_{j,m}\}$ is NA. Because $E[\alpha_{j,m} \mid N_{j,m}] - E[\alpha_{j,m}]$ is an increasing function of $N_{j,m}$, by Proposition 1.1, the random variable array $\left\{E[\alpha_{j,m} \mid N_{j,m}] - E[\alpha_{j,m}]\right\}$ is also NA. Hence, by the CLT for row-wise NA arrays (Theorem 1.7), we get

$$\sum_{j=0}^{m}E[Z_{j,m} \mid N_{j,m}] = \frac{1}{m^{1/2}}\sum_{j=0}^{m}\left(E[\alpha_{j,m} \mid N_{j,m}] - E[\alpha_{j,m}]\right) \xrightarrow{\mathscr{L}} N(0, {\sigma_1}^2),$$

where

$$\sigma_1{}^2 = \lim_{m\to\infty} \left\{\mathrm{Var}\left(E[\alpha_{1,m} \mid N_{1,m}]\right) + (m-1)\mathrm{Cov}(\alpha_{1,m}, \alpha_{2,m})\right\}.$$

Therefore, by the convergence theorem of characteristic functions, we obtain

$$E\left[\mathrm{e}^{\mathrm{i}t\sum_{j=0}^{m} E[Z_{j,m}|N_{j,m}]}\right] \longrightarrow \mathrm{e}^{-\frac{t^2}{2}\sigma_1{}^2}. \tag{11}$$

Similarly, we can prove that the random variable array $\left\{\mathrm{Var}[\alpha_{j,m} \mid N_{j,m}]\right\}$ is also NA, hence by the SLLN theorem for row-wise NA arrays (Theorem 1.5), we have

$$\sum_{j=0}^{m} \mathrm{Var}[Z_{j,m} \mid N_{j,m}] = \frac{\sum_{j=0}^{m} \mathrm{Var}[\alpha_{j,m} \mid N_{j,m}]}{m} \xrightarrow{a.s} \sigma_2{}^2,$$

where $\sigma_2{}^2 \equiv \lim_{m\to\infty} E\left[\mathrm{Var}[\alpha_{1,m} \mid N_{j,m}]\right]$. Thus,

$$\mathrm{e}^{-\frac{t^2}{2}\sum_{j=0}^{m}\mathrm{Var}[Z_{j,m}|N_{j,m}]} \xrightarrow{a.s} \mathrm{e}^{-\frac{t^2}{2}\sigma_2{}^2}.$$

From the bounds we obtained in (3), (7) and (8),

$$\mathrm{e}^{\sum_{j=0}^{m}\left(r_{j,m}^{(1)}+r_{j,m}^{(2)}+r_{j,m}^{(3)}\right)} \to \mathrm{e}^0 = 1.$$

The two convergence results immediately above produce the following:

$$\mathrm{e}^{-\frac{t^2}{2}\sum_{j=0}^{m}\mathrm{Var}[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{\sum_{j=0}^{m}\left(r_{j,m}^{(1)}+r_{j,m}^{(2)}+r_{j,m}^{(3)}\right)} \xrightarrow{a.s} \mathrm{e}^{-\frac{t^2}{2}\sigma_2{}^2}. \tag{12}$$

Therefore, by Eq. (10),

$$\begin{aligned}
&\left|E[f_m(t)] - E\left[\mathrm{e}^{\mathrm{i}t\sum_{j=0}^{m} E[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{-\frac{t^2}{2}\sigma_2{}^2}\right]\right| \\
&\quad \le E\left[\left|\mathrm{e}^{\mathrm{i}t\sum_{j=0}^{m} E[Z_{j,m}|N_{j,m}]}\right| \cdot \left|\mathrm{e}^{-\frac{t^2}{2}\sum_{j=0}^{m}\mathrm{Var}[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{\sum_{j=0}^{m}\left(r_{j,m}^{(1)}+r_{j,m}^{(2)}+r_{j,m}^{(3)}\right)} - \mathrm{e}^{-\frac{t^2}{2}\sigma_2{}^2}\right|\right] \\
&\quad = E\left[\left|\mathrm{e}^{-\frac{t^2}{2}\sum_{j=0}^{m}\mathrm{Var}[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{\sum_{j=0}^{m}\left(r_{j,m}^{(1)}+r_{j,m}^{(2)}+r_{j,m}^{(3)}\right)} - \mathrm{e}^{-\frac{t^2}{2}\sigma_2{}^2}\right|\right] \longrightarrow 0
\end{aligned}$$

by the dominated convergence theorem.

Combining this with Formula (11), we obtain

$$\lim_{m\to\infty} E[f_m(t)] = \lim_{m\to\infty} \mathrm{e}^{\mathrm{i}t\sum_{j=0}^{m} E[Z_{j,m}|N_{j,m}]} \cdot \mathrm{e}^{-\frac{t^2}{2}\sigma_2{}^2} = \mathrm{e}^{-\frac{t^2}{2}\sigma_1{}^2} \cdot \mathrm{e}^{-\frac{t^2}{2}\sigma_2{}^2} = \mathrm{e}^{-\frac{t^2}{2}\sigma^2},$$

where

$$\begin{aligned}
\sigma^2 &\equiv \sigma_1{}^2 + \sigma_2{}^2 \\
&= \lim_{m\to\infty}\left(\mathrm{Var}\left[E[\alpha_{1,m} \mid N_{1,m}]\right] + (m-1)\mathrm{Cov}(\alpha_{1,m}, \alpha_{2,m})\right) + \lim_{m\to\infty} E\left[\mathrm{Var}[\alpha_{1,m} \mid N_{j,m}]\right] \\
&= \lim_{m\to\infty} \frac{\mathrm{Var}[\Gamma_{n,m}]}{m}.
\end{aligned}$$

Recalling the definitions of $f_m(t)$ and $Z_{j,m}$ given at the beginning of the proof, we finally obtain

$$E\left[\mathrm{e}^{\mathrm{i}t\frac{1}{m^{1/2}}\left(\Gamma_{n,m}-E[\Gamma_{n,m}]\right)}\right] \to \mathrm{e}^{-\frac{t^2}{2}\sigma^2}.$$

Thus, the result follows by the convergence theorem for characteristic functions.

3. Further research

In this paper, we have established the CLT for the domination number of CCCDs in the case of the uniform distribution in one dimension. Further research directions consist of extending the CLT to more general distributions in one dimension, and finally obtaining a similar result in higher dimensions. As many applications of CCCDs arise in higher dimensions, proving the CLT in this situation would significantly benefit evaluation of CCCD-classifiers.

Acknowledgments

The authors gratefully acknowledge financial support for this research from Johns Hopkins University through the Acheson J. Duncan Fund for the Advancement of Research in Statistics. They also thank Carey Priebe for encouragement and helpful discussions regarding research on class cover catch digraphs and related statistical classification methods.

Appendix A. Calculation of the variance of the domination number

In this appendix, we outline the calculation of the exact and limiting variance of $\Gamma_{n,m}$. Recall that $Y_{(0)} \equiv 1$, $Y_{(m+1)} \equiv 1$, $Y_{(j)}$ denotes the jth order statistic of $Y_1, \ldots, Y_m$, and the random variable $\alpha_{j,m}$ is the minimum number of covering balls needed to cover the $N_{j,m}$ X-points located between $Y_{(j)}$ and $Y_{(j+1)}$. The random variables $\alpha_{j,m}, j = 0, m$, are referred to as the external components, and $\alpha_{j,m}, j = 1, \ldots, m-1$ as the internal components.

A.1. Expressions in terms of expectations

First, compute the conditional moments of $\alpha_{j,m}$, using the conditional distribution formulas given in the following theorem by Priebe et al. (2001):

Theorem A.1. *If $F_X = F_Y = U[0, 1]$, then*

- *for $j \in \{0, 1, \ldots, m\}$, if $N_{j,m} = 0$, then $\alpha_{j,m} = 0$;*
- *for $j \in \{0, m\}$, if $N_{j,m} > 0$, then $\alpha_{j,m} = 1$;*
- *for $j \in \{1, 2, \ldots, m-1\}$, if $N_{j,m} = n_{j,m} > 0$, then*

$$\begin{aligned} P(\alpha_{j,m} = 1 \mid N_{j,m} = n_{j,m}) &= 1 - P(\alpha_{j,m} = 2 \mid N_{j,m} = n_{j,m}) \\ &= \frac{5}{9} + \frac{4}{9}\frac{1}{4^{n_{j,m}-1}}. \end{aligned}$$

Based on the above formulas, straightforward calculations yield

$$E(\alpha_{j,m} \mid N_{j,m}) = \begin{cases} 0 & N_{j,m} = 0, j = 0, \ldots, m \\ 1 & N_{j,m} > 0, j = 0, m \\ \dfrac{13}{9} - \dfrac{16}{9}\dfrac{1}{4^{N_{j,m}}} & N_{j,m} > 0, j = 1, \ldots, m-1, \end{cases} \tag{A.1}$$

and

$$\mathrm{Var}(\alpha_{j,m} \mid N_{j,m}) = \begin{cases} 0 & N_{j,m} = 0, j = 0, \ldots, m \\ 0 & N_{j,m} > 0, j = 0, m \\ \dfrac{20}{81} - \dfrac{16}{81}\dfrac{1}{4^{N_{j,m}}} - \dfrac{256}{81}\dfrac{1}{4^{2N_{j,m}}} & N_{j,m} > 0, j = 1, \ldots, m-1. \end{cases} \tag{A.2}$$

To determine the marginal and joint distributions of $\{N_{j,m}\}$, we use the following two lemmas:

Lemma A.1. *Given $L_{j,m} \equiv Y_{(j+1)} - Y_{(j)} = l_{j,m}, j = 0, 1, \ldots, m$, the random vector $\{N_{j,m}\}$ is multinomially distributed with parameters $\{n, l_{j,m} : \sum_{j=0}^{m} l_{j,m} = 1\}$.*

Lemma A.2. *The density function of $L_{j,m}$ is*

$$f(l_{j,m}) = m(1 - l_{j,m})^{m-1},$$

and the joint density function of $L_{j_1,m}$ and $L_{j_2,m}$ is

$$f(l_{j_1,m}, l_{j_2,m}) = m(m-1)(1 - l_{j_1,m} - l_{j_2,m})^{m-2}.$$

Two consequences of these two lemmas are used repeatedly in our calculations:

$$P(N_{j,m}=0)=\frac{m}{m+n}$$
$$P(N_{j_1,m}=0,N_{j_2,m}=0)=\frac{m(m-1)}{(m+n)(m+n-1)}.$$

Using the identity $\mathrm{Var}(\alpha_{j,m})=E[\mathrm{Var}(\alpha_{j,m}\mid N_{j,m})]+\mathrm{Var}[E(\alpha_{j,m})\mid N_{j,m}]$, we can calculate $\mathrm{Var}(\alpha_{j,m})$ from Eqs. (A.1) and (A.2):

$$\mathrm{Var}(\alpha_{j,m})=\begin{cases}\dfrac{mn}{(m+n)^2} & j=0,m\\ \dfrac{20}{81}\dfrac{n}{m+n}+\dfrac{169}{81}\dfrac{mn}{(m+n)^2}-\left(\dfrac{16}{81}+\dfrac{416}{81}\dfrac{m}{m+n}\right)\mu_{n,m}-\dfrac{256}{81}\mu_{n,m}{}^2 & j=1,\ldots,m-1,\end{cases}\tag{A.3}$$

where $\mu_{n,m}\equiv E\left(\frac{1}{4^{N_{j,m}}}I_{\{N_{j,m}>0\}}\right)$.

Similarly, we can calculate the covariance $\mathrm{Cov}(\alpha_{j_1,m},\alpha_{j_2,m})$ between any two components. An additional fact used in the calculation is the following lemma:

Lemma A.3. *For distinct j_1, j_2, given $N_{j_1,m}$ and $N_{j_2,m}$, the corresponding components $\alpha_{j_1,m}$ and $\alpha_{j_2,m}$ are conditionally independent. In addition, $\alpha_{j_1,m}$ is only dependent on $N_{j_1,m}$.*

Using Lemma A.3, we immediately obtain

$$\begin{aligned}\mathrm{Cov}(\alpha_{j_1,m},\alpha_{j_2,m})&=E(\alpha_{j_1,m}\alpha_{j_2,m})-E(\alpha_{j_1,m})E(\alpha_{j_2,m})\\&=E\left[E(\alpha_{j_1,m}\mid N_{j_1,m},N_{j_2,m})E(\alpha_{j_2,m}\mid N_{j_1,m},N_{j_2,m})\right]-E\left[E(\alpha_{j_1,m}\mid N_{j_1,m})\right]E\left[E(\alpha_{j_2,m}\mid N_{j_2,m})\right]\\&=E\left[E(\alpha_{j_1,m}\mid N_{j_1,m})E(\alpha_{j_2,m}\mid N_{j_2,m})\right]-E\left[E(\alpha_{j_1,m}\mid N_{j_1,m})\right]E[E(\alpha_{j_2,m}\mid N_{j_2,m})\right].\end{aligned}$$

The rest of the calculation of $\mathrm{Cov}(\alpha_{j_1,m},\alpha_{j_2,m})$ is similar to that of $\mathrm{Var}(\alpha_{j,m})$, obtaining the formulas

$$\mathrm{Cov}(\alpha_{j_1,m},\alpha_{j_2,m})=\begin{cases}-\dfrac{mn}{(m+n)^2(m+n-1)} & j_1=0,j_2=m\\ -\dfrac{13}{9}\dfrac{mn}{(m+n)^2(m+n-1)}-\dfrac{16}{9}\delta_{n,m} & j_1=0,j_2\neq m\\ -\dfrac{169}{81}\dfrac{mn}{(m+n)^2(m+n-1)}-\dfrac{416}{81}\delta_{n,m}+\dfrac{256}{81}(\nu_{n,m}-\mu_{n,m}{}^2) & j_1,j_2=1,\ldots,m-1,\end{cases}\tag{A.4}$$

where

$$\delta_{n,m}\equiv E\left(\frac{1}{4^{N_{j_2,m}}}I_{\{N_{j_1,m}>0,N_{j_2,m}>0\}}\right)-\frac{n}{m+n}E\left(\frac{1}{4^{N_{j_2,m}}}I_{\{N_{j_2,m}>0\}}\right),$$
$$\nu_{n,m}\equiv E\left(\frac{1}{4^{N_{j_1,m}+N_{j_2,m}}}I_{\{N_{j_1,m}>0,N_{j_2,m}>0\}}\right).$$

A.2. Expressions in terms of series

In the previous sub-section, we converted the variance and covariances of the components into expressions determined by $\mu_{m,n}$, $\delta_{n,m}$ and $\nu_{n,m}$. Our next step is to compute series expressions for $\mu_{m,n}$, $\delta_{n,m}$ and $\nu_{n,m}$. For $\mu_{n,m}$, we have

$$\begin{aligned}\mu_{n,m}&=E\left[\frac{1}{4^{N_{j,m}}}\left(I_{\{N_{j,m}\geq 0\}}-I_{\{N_{j,m}=0\}}\right)\right]\\&=E\left(\frac{1}{4^{N_{j,m}}}\right)-P(N_{j,m}=0)=E\left(\frac{1}{4^{N_{j,m}}}\right)-\frac{m}{m+n}.\end{aligned}$$

From Lemma A.1, we have

$$E\left(\frac{1}{4^{N_{j,m}}}\mid L_{j,m}=l_{j,m}\right)=\sum_{q=0}^{n}\frac{1}{4^q}\binom{n}{q}l_{j,m}^q(1-l_{j,m})^{n-q}.$$

Use the distribution of $L_{j,m}$ from Lemma A.2 to write $\mu_{n,m}$ as an integral:

$$\mu_{n,m}=\int_0^1\sum_{0\leq q\leq n}\frac{1}{4^q}\binom{n}{q}l_{j,m}{}^q(1-l_{j,m})^{n-q}m(1-l_{j,m})^{m-1}\mathrm{d}l_{j,m}-\frac{m}{m+n}.$$

Interchange integration and summation, and integrate, to obtain

$$\mu_{n,m} = \frac{m}{m+n}\left(\sum_{q=0}^{n}\frac{1}{4^q}\frac{n!(m+n-q-1)!}{(n-q)!(m+n-1)!} - 1\right). \tag{A.5}$$

Similarly,

$$\delta_{n,m} = \sum_{q=0}^{n}\frac{1}{4^q}\frac{n!(m+n-q-2)!}{(n-q)!(m+n)!}\frac{mn-m^2q}{m+n} - \frac{mn}{(m+n)^2(m+n-1)},$$

and

$$\begin{aligned}\nu_{n,m} - {\mu_{n,m}}^2 = \frac{mn}{(m+n)^2(m+n-1)}&\left\{\frac{m(m+n-1)}{n}\left\{\sum_{q=0}^{n}\frac{1}{4^q}\frac{n!(m+n-q-2)!}{(n-q)!(m+n-2)!}\right.\right.\\ &\left.\cdot\left(q+1+\frac{-2(m+n)q}{m^2+mn-m+n}\right) - \left(\sum_{q=0}^{n}\frac{1}{4^q}\frac{n!(m+n-q-1)!}{(n-q)!(m+n-1)!}\right)^2\right\}\\ &\left.+\left\{\sum_{q=0}^{n}\frac{1}{4^q}\frac{n!(m+n-q-2)!}{(n-q)!(m+n-2)!}\cdot\left(q+1+\frac{-2(m+n)q}{m^2+mn-m+n}\right)-1\right\}\right\}.\end{aligned} \tag{A.6}$$

A.3. Asymptotic results

In the calculation of the exact limiting values μ_r, δ_r, ν_r, we rely heavily on the following version of the dominated convergence theorem (DCT).

Theorem A.2. *If $D_n(q) \xrightarrow{n\to\infty} D(q)$, and $|D_n(q)| \le D^*(q)$ where $\sum_{q=0}^{\infty} D^*(q) < \infty$, then $\sum_{q=0}^{\infty} D_n(q) \xrightarrow{n\to\infty} \sum_{q=0}^{\infty} D(q)$.*

For $\mu_{n,m}$, let $D_n(q) = \begin{cases}(\frac{1}{4})^q\frac{n!(m+n-q-1)!}{(n-q)!(m+n-1)!} & q \le n\\ 0 & q > n\end{cases}$, then $\mu_{n,m}$ can be written as $\frac{m}{m+n}\left(\sum_{q=0}^{\infty} D_n(q) - 1\right)$. If $m/n \to r$, then it can be easily checked that $D_n(q) \xrightarrow{n\to\infty} \left(\frac{1}{4(r+1)}\right)^q$, and $|D_n(q)| \le (\frac{1}{4})^q$, where $\sum_{q=0}^{\infty}(\frac{1}{4})^q = \frac{4}{3} < \infty$. Therefore, by Theorem A.2, as $m/n \to r$, the limiting value of $\mu_{n,m}$ is

$$\frac{r}{r+1}\left(\sum_{q=0}^{\infty}\left(\frac{1}{4(r+1)}\right)^q - 1\right) = \frac{r}{r+1}\left(\frac{1}{1-\frac{1}{4(r+1)}} - 1\right),$$

thus,

$$\mu_{n,m} = \frac{r}{(r+1)(4r+3)} + o(1). \tag{A.7}$$

By the same technique, but a more complicated calculation, we obtain

$$\delta_{n,m} = \frac{r(-4r^2+3)}{(r+1)^3(4r+3)^2}\frac{1}{n} + o\left(\frac{1}{n}\right), \tag{A.8}$$

$$\nu_{n,m} - {\mu_{n,m}}^2 = -\frac{r(4r^2-3)^2}{(r+1)^3(4r+3)^4}\frac{1}{n} + oo\left(\frac{1}{n}\right). \tag{A.9}$$

Recalling that $\Gamma_{n,m} = \sum_{j=0}^{m}\alpha_{j,m}$, we have

$$\begin{aligned}\mathrm{Var}(\Gamma_{n,m}) = {} & 2\mathrm{Var}(\alpha_{0,m}) + (m-1)\mathrm{Var}(\alpha_{1,m}) + \mathrm{Cov}(\alpha_{0,m}, \alpha_{m,m}) + 2(m-1)\mathrm{Cov}(\alpha_{0,m}, \alpha_{1,m})\\ & + m(m-1)\mathrm{Cov}(\alpha_{1,m}, \alpha_{2,m}).\end{aligned}$$

By plugging Formula (A.7)–(A.9) into Eqs. (A.3) and (A.4), and substituting the generated expressions for $\mathrm{Var}(\alpha_{j,m})$ and $\mathrm{Cov}(\alpha_{j_1,m}, \alpha_{j_2,m})$ into the above equation, we can finally get the desired result stated in Theorem 1.3.

Appendix B. Proof of negative association of $\{N_{j,m} : 0 \le j \le m, m \ge 1\}$

To prove the random vector $(N_{0,m}, \ldots, N_{m,m})$ is NA for any $m \ge 1$, we need the following lemma:

Lemma B.1. *If $F_X = F_Y = U[0, 1]$, then $(L_{0,m}, \ldots, L_{m,m})$ is NA.*

Proof. Recall that $L_{j,m} = Y_{(j+1)} - Y_{(j)}$ and suppose that $Z_0, \ldots, Z_m$ are i.i.d. random variables with an exponential distribution, where $\{Z_0, \ldots, Z_m\}$ are independent of $\{L_{0,m}, \ldots, L_{m,m}\}$. Since the exponential distribution is log-concave, from Proposition 1.3 we know that given $\sum_{j=0}^{m} Z_i$, the random vector $(Z_0, \ldots, Z_m)$ is NA. Hence by Definition 1.1, we know that for any pair of disjoint subsets I, J of $\{0, \ldots, m\}$ and any increasing functions f_I and f_J such that the following covariance exists,

$$\operatorname{Cov}\left\{ f_I(Z_i, i \in I), f_J(Z_j, j \in J) \;\middle|\; \sum_{k=0}^{m} Z_k = a \right\} \le 0, \quad \text{where } a > 0.$$

Since $f_I(\frac{Z_i}{a}, i \in I)$ and $f_J(\frac{Z_j}{a}, j \in J)$ are still increasing functions of $Z_i, i \in I$ and $Z_j, j \in J$, respectively, we have

$$\operatorname{Cov}\left\{ f_I\left(\frac{Z_i}{a}, i \in I\right), f_J\left(\frac{Z_j}{a}, j \in J\right) \;\middle|\; \sum_{k=0}^{m} Z_k = a \right\} \le 0,$$

i.e.,

$$\operatorname{Cov}\left\{ f_I\left(\frac{Z_i}{\sum_{k=0}^{m} Z_k}, i \in I\right), f_J\left(\frac{Z_j}{\sum_{k=0}^{m} Z_k}, j \in J\right) \;\middle|\; \sum_{k=0}^{m} Z_k = a \right\} \le 0.$$

Note that given $\sum_{k=0}^{m} Z_k = a$, the distribution of $\left(\frac{Z_0}{\sum_{i=0}^{m} Z_i}, \frac{Z_1}{\sum_{i=0}^{m} Z_i}, \ldots, \frac{Z_m}{\sum_{i=0}^{m} Z_i}\right)$ is independent of a, so the above inequality yields

$$\operatorname{Cov}\left\{ f_I\left(\frac{Z_i}{\sum_{k=0}^{m} Z_k}, i \in I\right), f_J\left(\frac{Z_j}{\sum_{k=0}^{m} Z_k}, j \in J\right) \right\} \le 0.$$

Therefore, the random vector

$$\left(\frac{Z_0}{\sum_{i=0}^{m} Z_i}, \frac{Z_1}{\sum_{i=0}^{m} Z_i}, \ldots, \frac{Z_m}{\sum_{i=0}^{m} Z_i}\right)$$

is NA. However, $\left(\frac{Z_0}{\sum_{i=0}^{m} Z_i}, \frac{Z_1}{\sum_{i=0}^{m} Z_i}, \ldots, \frac{Z_m}{\sum_{i=0}^{m} Z_i}\right)$ and $(L_{0,m}, \ldots, L_{m,m})$ have the same distribution, so $(L_{0,m}, \ldots, L_{m,m})$ is also NA. □

Lemma B.2. *If $F_X = F_Y = U[0, 1]$, then the random vector $(N_{0,m}, \ldots, N_{m,m})$ is NA.*

Proof. It is easy to show that given $L_{j,m} = l_{j,m}, j = 1, \ldots, m$, the random vector $(N_{0,m}, \ldots, N_{m,m})$ is multinomially distributed, hence it is NA (Proposition 1.2). From the definition of negative association, we know that for any disjoint subset I, J of $\{0, \ldots, m\}$ and increasing functions f_I, f_J, the following inequality holds:

$$\begin{aligned} &E\left[f_I(N_{i,m}, i \in I) \cdot f_J(N_{j,m}, j \in J) \mid L_{k,m}, k = 0, \ldots, m\right] \\ &\quad \le E\left[f_I(N_{i,m}, i \in I) \mid L_{k,m}, k = 0, \ldots, m\right] \cdot E\left[f_J(N_{j,m}, j \in J) \mid L_{k,m}, k = 0, \ldots, m\right]. \end{aligned}$$

Note that given $L_{k,m} = l_{k,m}, k = 0, \ldots, m$, the joint distribution of $\{N_{i,m}, i \in I\}$ only depends on $L_{i,m}, i \in I$, thus $E\left[f_I(N_{i,m}, i \in I) \mid L_{k,m}, k = 0, \ldots, m\right] = E\left[f_I(N_{i,m}, i \in I) \mid L_{i,m}, i \in I\right]$; similarly, $E\left[f_J(N_{j,m}, j \in J) \mid L_{k,m}, k = 0, \ldots, m\right] = E\left[f_J(N_{j,m}, j \in J) \mid L_{j,m}, j \in J\right]$. Therefore,

$$E\left[f_I(N_{i,m}, i \in I) \cdot f_J(N_{j,m}, j \in J) \mid L_{k,m}, k = 0, \ldots, m\right] = E\left[f_I(N_{i,m}, i \in I) \mid L_{i,m}, i \in I\right] \cdot E\left[f_J(N_{j,m}, j \in J) \mid L_{j,m}, j \in J\right].$$

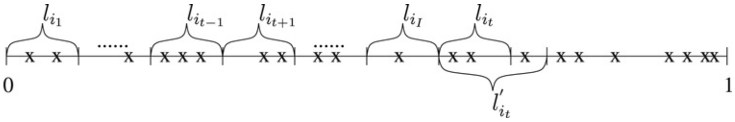

Fig. B.1. A coupling argument with respect to Inequality (B.1).

Taking expectation on both sides of the above inequality yields

$$E\left[E\left[f_I(N_{i,m}, i \in I) \cdot f_J(N_{j,m}, j \in J) \mid L_{k,m}, k = 0, \ldots, m\right]\right]$$
$$\leq E\left[E\left[f_I(N_{i,m}, i \in I) \mid L_{i,m}, i \in I\right] \cdot E\left[f_J(N_{j,m}, j \in J) \mid L_{j,m}, j \in J\right]\right].$$

Since Lemma B.1 showed that $(L_{0,m}, \ldots, L_{m,m})$ is NA, and $E\left[f_I(N_{i,m}, i \in I) \mid L_{i,m}, i \in I\right]$ and $E\left[f_J(N_{j,m}, j \in J) \mid L_{j,m}, j \in J\right]$ are actually increasing functions of $\{L_{i,m}, i \in I\}$ and $\{L_{j,m}, j \in J\}$, respectively (see Remark B), applying the definition of NA random vectors yields

$$E\left[E\left[f_I(N_{i,m}, i \in I) \mid L_{i,m}, i \in I\right] \cdot E\left[f_J(N_{j,m}, j \in J) \mid L_{j,m}, j \in J\right]\right]$$
$$\leq E\left[E\left[f_I(N_{i,m}, i \in I) \mid L_{i,m}, i \in I\right]\right] \cdot E\left[E\left[f_J(N_{j,m}, j \in J) \mid L_{j,m}, j \in J\right]\right].$$

Connecting the two inequalities above produces

$$E\left[E\left[f_I(N_{i,m}, i \in I) \cdot f_J(N_{j,m}, j \in J) \mid L_{k,m}, k = 0, \ldots, m\right]\right]$$
$$\leq E\left[E\left[f_I(N_{i,m}, i \in I) \mid L_{i,m}, i \in I\right]\right] \cdot E\left[E\left[f_J(N_{j,m}, j \in J) \mid L_{j,m}, j \in J\right]\right],$$

thus

$$E\left[f_I(N_{i,m}, i \in I) \cdot f_J(N_{j,m}, j \in J)\right] \leq E\left[f_I(N_{i,m}, i \in I)\right] \cdot E\left[f_J(N_{j,m}, j \in J)\right].$$

Remark B. It suffices to show that for any subset $I = \{i_1, \ldots, i_I\}$ of $\{0, \ldots, m\}$, if $l_{i_t,m} < l'_{i_t,m}$, $t \in I$, then

$$E[f_I(N_{i,m}, i \in I) \mid L_{i,m} = l_{i,m}, i \in I] \leq E[f_I(N_{i,m}, i \in I) \mid L_{i,m} = l_{i,m} \text{ for } i \in I - \{i_t\}, L_{i_t,m} = l'_{i_t,m}]. \quad \text{(B.1)}$$

As illustrated in Fig. B.1, suppose n X-points are independently uniformly distributed in [0, 1], and denote $N_{i,m}$, $i \in I$ as the number of X-points falling in the interval with length $L_{i,m} = l_{i,m}$. If the length $l_{i_t,m}$ of the most right interval increases to $l'_{i_t,m}$, then $N_{i_t,m}$ will not decrease (possibly increase). This means that when $\left\{L_{i,m} = l_{i,m} \text{ for } i \in I - \{i_t\}, L_{i_t,m} = l'_{i_t,m}\right\}$, the random variable $N_{j,m}$ is stochastically larger than the original one when $L_{i,m} = l_{i,m}$, $i \in I$. By considering the fact that f_I is an increasing function of $N_{i,m}$, $i \in I$, it follows that Inequality (B.1) indeed holds. □

References

Cannon, A., Cowen, L., 2000. Approximation algorithms for the class cover problem. In: 6th International Symposium on Artificial Intelligence and Mathematics.
Ceyhan, E., Priebe, C.E., 2005. The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association. Statist. Probab. Lett. 73, 37–50.
Ceyhan, E., Priebe, C.E., 2006. On the distribution of the domination number of a new family of parametrized random digraphs. Model Assisted Statist. Appl. 1, 231–255.
Ceyhan, E., Priebe, C.E., Marchette, D.J., 2007. A new family of graphs for testing spatial segregation. Canadian J. Stat. 35, 27–50.
Ceyhan, E., Priebe, C.E., Wierman, J.C., 2006. Relative density of the random r-factor proximity catch digraph for testing spatial patterns of segregation and association. Comp. Stat. Data Anal. 50, 1925–1964.
DeVinney, J.G., Priebe, C.E., 2006. A new family of proximity graphs: Class cover catch digraphs. Discrete Appl. Math. 154, 1975–1982.
DeVinney, J., Wierman, J.C., 2002. A SLLN for a one-dimensional class cover problem. Statist. Probab. Lett. 59, 425–435.
Eveland, C.K., Socolinsky, D.A., Priebe, C.E., Marchette, D.J., 2005. A hierarchical methodology for one-class problems with skewed priors. J. Classification 22, 17–48.
Joag-Dev, K., Proschan, F., 1983. Negative association of random variables, with applications. Ann. Stat. 11, 286–295.
Marchette, D.J., Priebe, C.E., 2003. Characterizing the scale dimension of a high-dimensional classification problem. Patt. Recog. 36, 45–60.
Newman, C.M., 1984. Asymptotic independence and limit theorems for positively and negatively dependent random variables. In: Inequalities in Statistics and Probability. In: IMS Lecture Notes, vol. 5. pp. 127–140.
Priebe, C.E., DeVinney, J., Marchette, D.J., 2001. On the distribution of the domination number for random class cover catch digraphs. Statist. Probab. Lett. 55, 239–246.
Priebe, C.E., Marchette, D.J., DeVinney, J., Socolinsky, D., 2003. Classification using class cover catch digraphs. J. Classification 20, 3–23.
Taylor, R.L., Patterson, R.F., Bozorgnia, A., 2002. A strong law of large numbers for arrays of row-wise negatively dependent random variables. Stochastic Anal. Appl. 20, 643–656.
Wierman, J.C., Xiang, P., 2008. A general SLLN for the one-dimensional class cover problem. Statist. Probab. Lett. 78, 1110–1118.
Xiang, P., 2006. Limit theory for the domination number of random class cover catch digraphs. Ph.D. Dissertation. Johns Hopkins University, Department of Applied Mathematics and Statistics.