The Distribution of the Domination Number of Class Cover Catch Digraphs for Non-uniform One-dimensional Data

Elvan Ceyhan*

October 10, 2008

Abstract

For two or more classes of points in \( \mathbb{R}^d \) with \( d \geq 1 \), the class cover catch digraphs (CCCDs) can be constructed using the relative positions of the points from one class with respect to the points from the other class. The CCCDs were introduced by Priebe et al. [2001] who investigated the case of two classes, \( \mathcal{X} \) and \( \mathcal{Y} \). They calculated the exact (finite sample) distribution of the domination number of the CCCDs based on \( \mathcal{X} \) points relative to \( \mathcal{Y} \) points both of which were uniformly distributed on a bounded interval. We investigate the distribution of the domination number of the CCCDs based on data from non-uniform \( \mathcal{X} \) points on an interval with end points from \( \mathcal{Y} \). Then we extend these calculations for multiple \( \mathcal{Y} \) points on bounded intervals.

Keywords: Class Cover Catch Digraph; Domination number; Non-uniform Distribution; Proximity Map; Random digraph

*Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD, 21218.
1 Introduction

In 2001, a new classification method was developed which was based on the relative positions of the data points from various classes [Priebe et al. (2001)] introduced the class cover catch digraphs (CCCDs) in \( \mathbb{R} \) and gave the exact distribution of the domination number of the CCCDs for two classes, \( \mathcal{X} \) and \( \mathcal{Y} \), with uniform distribution on a bounded interval in \( \mathbb{R} \). DeVinney and Wierman (2003) proved a SLLN result for the one-dimensional class cover problem. DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a), and Priebe et al. (2003b) extended the CCCDs to higher dimensions and demonstrated that CCCDs are a competitive alternative to the existing methods in classification. The classification method based on CCCDs involves data reduction (condensing) by using approximate — rather than exact — minimum dominating proximity regions are defined for one point, \( \Gamma_{B} \subseteq \mathcal{X} \). Involving data reduction (condensing) by using approximate — rather than exact — minimum dominating regions are defined for one point, \( \Gamma_{B} \subseteq \mathcal{X} \). However, for finding a dominating set of CCCDs on the real line, a simple linear time algorithm is available [Priebe et al. (2001)]. But unfortunately, the exact and the asymptotic distributions of the domination number of the CCCDs are not analytically tractable in multiple dimensions.

To address the latter issue of intractability of the distribution of the domination number in multiple dimensions, Ceyhan and Priebe (2003, 2005) introduced the central similarity proximity maps and \( r \)-factor proportional-edge proximity maps and the associated random proximity catch digraphs. Proximity catch digraphs are a generalization of the CCCDs. The asymptotic distribution of the domination number of the latter is calculated and then used in testing spatial patterns between two or more classes. See Ceyhan and Priebe (2005) for more detail.

In this article, we generalize the original result of [Priebe et al. (2001)] to the case of non-uniform \( \mathcal{X} \) points with support being the interval with end points from \( \mathcal{Y} \), and then to multiple \( \mathcal{Y} \) points in a bounded interval \( (c, d) \subset \mathbb{R} \) with \( c < d \). These generalizations will also serve as the bases for extension of the results for the uniform and non-uniform data in higher dimensions.

2 Data-random Class Cover Catch Digraphs

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 sets of \( \Omega \)-valued random variables from classes \( \mathcal{X} \) and \( \mathcal{Y} \), respectively, with joint probability distribution \( F_{X,Y} \). Let \( d(\cdot, \cdot) : \Omega \times \Omega \rightarrow [0, \infty) \) be any distance function. The class cover problem for a target class, say \( \mathcal{X} \), refers to finding a collection of neighborhoods, \( N_i \) around \( X_i \) such that (i) \( \mathcal{X}_n \subseteq \{\cup_i N_i\} \) and (ii) \( \mathcal{Y}_m \cap \{\cup_i N_i\} = \emptyset \). A collection of neighborhoods satisfying both conditions is called a class cover. A cover satisfying condition (i) is a proper cover of class \( \mathcal{X} \) while a cover satisfying condition (ii) is a pure cover relative to class \( \mathcal{Y} \). This article is on the minimum cardinality class covers; that is, class covers satisfying both (i) and (ii) with the smallest number of neighborhoods. See Priebe et al. (2001).

Consider the map \( N : \Omega \rightarrow 2^\Omega \) where \( 2^\Omega \) represents the power set of \( \Omega \). Then given \( \mathcal{Y}_m \subseteq \Omega \), the proximity map \( N_\mathcal{Y} : \Omega \rightarrow 2^\Omega \) associates with each point \( x \in \Omega \) a proximity region \( N_\mathcal{Y}(x) \subseteq \Omega \). For \( B \subseteq \Omega \), the \( \Gamma_1 \)-region is the image of the map \( \Gamma_1(\cdot, N_\mathcal{Y}) : 2^\Omega \rightarrow 2^\Omega \) that associates the region \( \Gamma_1(B, N_\mathcal{Y}) := \{z \in \Omega : B \subseteq N_\mathcal{Y}(z)\} \) with the set \( B \). For a point \( x \in \Omega \), we denote \( \Gamma_1(\{x\}, N_\mathcal{Y}) \) as \( \Gamma_1(x, N_\mathcal{Y}) \). Notice that while the proximity regions are defined for one point, \( \Gamma_1 \)-regions are defined for sets of points.

The data-random CCCD has the vertex set \( \mathcal{V} = \mathcal{X}_n \) and arc set \( \mathcal{A} \) defined by \( (X_i, X_j) \in \mathcal{A} \iff X_j \in N_\mathcal{Y}(X_i) \). In particular, we use \( N_\mathcal{Y}(X_i) = B(X_i, r_i) \), the open ball around \( X_i \) with radius \( r_i := \min_{Y \in \mathcal{Y}_m} d(X_i, Y) \), as the proximity map as in Priebe et al. (2001). We call such a digraph a \( \mathcal{D}_{n,m} \).
digraph. A $\mathcal{D}_{n,m}$-digraph is a pseudo digraph according some authors if loops are allowed (see, e.g., Chartrand and Lesniak (1996)).

A data-random CCCD for $\Omega = \mathbb{R}^d$ and $N_i = B(X_i, r_i)$ is referred to as $\mathcal{C}_{n,m}$-graph in Priebe et al (2001). We change the notation to emphasize the fact that $\mathcal{D}_{n,m}$ is a digraph. Furthermore, Ceyhan and Priebe (2003) call the proximity map $N_i = B(X_i, r_i)$ a spherical proximity map.

The $\mathcal{D}_{n,m}$-digraphs are closely related to the proximity graphs of Jaromczyk and Toussaint (1992) and might be considered as a special case of covering sets of Tuza (1994) and intersection digraphs of Sen et al. (1989). Our data-random proximity digraph is a vertex-random proximity digraph and not a standard one (see, e.g., Janson et al. (2000)). The randomness of a $\mathcal{D}_{n,m}$-digraph lies in the fact that the vertices are random with the joint distribution $F_{X,Y}$, but arcs $(X_i, X_j)$ are deterministic functions of the random variable $X_j$ and the random set $N_i$.

3 Domination Number of Random $\mathcal{D}_{n,m}$-digraphs

In a digraph $D = (\mathcal{V}, \mathcal{A})$ of order $|\mathcal{V}| = n$, a vertex $v$ dominates itself and all vertices of the form $\{u : (v, u) \in \mathcal{A}\}$. 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, denoted $\gamma(D)$, is defined as $\gamma(D) := |S^*_D|$, where $|\cdot|$ is the set cardinality functional (West (2001)). If a minimum dominating set consists of only one vertex, we call that vertex a dominating vertex. The vertex set $\mathcal{V}$ itself is always a dominating set, so $\gamma(D) \leq n$.

Let $\mathcal{F}(\mathbb{R}^d) := \{F_{X,Y} \text{ on } \mathbb{R}^d \text{ with } P(X = Y) = 0\}$. As in Priebe et al. (2001), in this article, we consider $\mathcal{D}_{n,m}$-digraphs for which $\mathcal{X}_n$ and $\mathcal{Y}_m$ are random samples from $F_X$ and $F_Y$, respectively, and the joint distribution of $X, Y$ is $F_{X,Y} \in \mathcal{F}(\mathbb{R}^d)$. We call such digraphs as $\mathcal{F}(\mathbb{R}^d)$-random $\mathcal{D}_{n,m}$-digraphs and focus on the random variable $\gamma(D)$. To make the dependence on sample sizes explicit, we use $\gamma(D_{n,m})$ instead of $\gamma(D)$. It is trivial to see that $1 \leq \gamma(D_{n,m}) \leq n$, and $\gamma(D_{n,m}) < n$ for nontrivial digraphs.

4 The Distribution of the Domination Number of $\mathcal{F}(\mathbb{R})$-random $\mathcal{D}_{n,m}$-digraphs

In $\mathbb{R}$, the data-random CCCD is a special case of interval catch digraphs (see, e.g., Sen et al. (1989) and Prisner (1994)). Let $\mathcal{X}_n$ and $\mathcal{Y}_m$ be two samples from $\mathcal{F}(\mathbb{R})$ and $Y_{(j)}$ be the $j^{th}$ order statistic of $\mathcal{Y}_m$ for $j = 1, 2, \ldots, m$. Then $Y_{(j)}$ partition $\mathbb{R}$ into $(m + 1)$ intervals. Let $-\infty := Y_{(0)} < Y_{(1)} < \ldots < Y_{(m)} < Y_{(m+1)} := \infty$, and $I_j := (Y_{(j-1)}, Y_{(j)})$, $\mathcal{X}^j := \mathcal{X}_n \cap I_j$, and $\mathcal{Y}^j := \{Y_{(j-1)}, Y_{(j)}\}$ for $j = 1, 2, \ldots, (m + 1)$. This yields a disconnected digraph with subdigraphs $\mathcal{D}^j$ for $j = 1, 2, \ldots, (m + 1)$, each of which might be null or itself disconnected. Let $\gamma(\mathcal{D}^j)$ denote the the cardinality of the minimum dominating set for the component of the random $\mathcal{D}_{n,m}$-digraph induced by the pair $\mathcal{X}^j$ and $\mathcal{Y}^j$, $n_j := |\mathcal{X}^j|$, and $F_j$ be the density $F_X$ restricted to $I_j$. Then $\gamma(D_{n,m}) = \sum_{j=1}^{m+1} \gamma(\mathcal{D}^j)$. We study the simpler random variable $\gamma(\mathcal{D}^j)$ first. The following lemma follows trivially (see Priebe et al. (2001)).

Lemma 4.1. For $j \in \{1, (m + 1)\}$, $\gamma(\mathcal{D}^j) = \mathbf{1}(n_j > 0)$ where $\mathbf{1}(\cdot)$ is the indicator function.
For $j = 2, \ldots, m$ and $n_j > 0$, we prove that $\gamma(D^j) \in \{1, 2\}$ with the distribution dependent probabilities $1 - p_{n_j}(F_j), p_{n_j}(F_j)$, respectively, where $p_{n_j}(F_j) = P(\gamma(D^j) = 2)$. A quick investigation shows that $\gamma(D^j) = 2$ iff $\mathcal{X}^j \cap \left(\frac{\max(\mathcal{X}^j) + Y_{j-1}}{2}, \frac{\min(\mathcal{X}^j) + Y_{j}}{2}\right) = \emptyset$; that is, $\mathcal{X}^j \subset B(x, r(x))$ if and only if $x \in \left(\frac{\max(\mathcal{X}^j) + Y_{j-1}}{2}, \frac{\min(\mathcal{X}^j) + Y_{j}}{2}\right)$

where $r(x) = \min(x - Y_{j-1}, Y_j - x)$. Hence $\Gamma_1(\mathcal{X}^j, N_{Y^j}) = \left(\frac{\max(\mathcal{X}^j) + Y_{i-1}}{2}, \frac{\min(\mathcal{X}^j) + Y_{j}}{2}\right) \subset I_j$. By definition, if $\mathcal{X}^j \cap \Gamma_1(\mathcal{X}^j, N_{Y^j}) \neq \emptyset$, then $\gamma(D^j) = 1$; hence the name $\Gamma_1$-region and the notation $\Gamma_1(\cdot, N_{Y^j})$.

**Theorem 4.2.** For $j = 2, \ldots, m$, $\gamma(D^j) \sim 1 + \text{Bernoulli}(p_{n_j}(F_j))$ for $n_j > 0$.

**Proof:** See [3] for the proof.

The probability $P(\gamma(D^j) = 2) = P(\mathcal{X}^j \cap \Gamma_1(\mathcal{X}^j, N_{Y^j}) = \emptyset)$ depends on the conditional distribution $F_{X|Y}$ and the interval $\Gamma_1(\mathcal{X}^j, N_{Y^j})$, which, if known, will make possible the calculation of $p_{n_j}(F_j)$. As an immediate result of Lemma [1.1] and Theorem 1.2 we have the following upper bound for $\gamma(D_{n,m})$.

**Theorem 4.3.** Let $D_{n,m}$ be an $\mathcal{F}(\mathbb{R})$-random $\mathcal{D}_{n,m}$-digraph with $n > 0$, $m > 0$ and $k_1$ and $k_2$ be two natural numbers defined as $k_1 := \sum_{j=2}^{m} I(|\mathcal{X}_j \cap I_j| > 1)$ and $k_2 := \sum_{j=2}^{m} I(|\mathcal{X}_j \cap I_j| = 1) + \sum_{j \in \{1, (m+1)\}} I(|\mathcal{X}_j \cap I_j| \neq \emptyset)$. Then $1 \leq \gamma(D_{n,m}) \leq 2k_1 + k_2 \leq \min(n, 2m)$.

In the special case of fixed $\mathcal{Y}_2 = \{y_1, y_2\}$ and $\mathcal{X}_n$ a random sample from $\mathcal{U}(y_1, y_2)$, the uniform distribution on $(y_1, y_2)$, we have a $\mathcal{D}_{n,2}$-digraph for which $F_{X} = \mathcal{U}(y_1, y_2)$ and $F_{Y}$ is a degenerate distribution. We call such digraphs as $\mathcal{U}(y_1, y_2)$-random $\mathcal{D}_{n,2}$-digraphs and provide an exact result on the distribution of their domination number in the next section.

### 4.1 The Exact Distribution of the Domination Number of $\mathcal{U}(y_1, y_2)$-random $\mathcal{D}_{n,2}$-digraphs

Suppose $\mathcal{Y}_2 = \{y_1, y_2\} \subset \mathbb{R}$ with $-\infty < y_1 < y_2 < \infty$ and $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ a set of iid random variables from $\mathcal{U}(y_1, y_2)$. Any $\mathcal{U}(y_1, y_2)$ random variable can be transformed into a $\mathcal{U}(0, 1)$ random variable by $\phi(x) = (x - y_1)/(y_2 - y_1)$, which maps intervals $(t_1, t_2) \subseteq (y_1, y_2)$ to intervals $(\phi(t_1), \phi(t_2)) \subseteq (0, 1)$. So, without loss of generality, we can assume $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ is a set of iid random variables from the $\mathcal{U}(0, 1)$ distribution. That is, the distribution of $\gamma(D_{n,2})$ does not depend on the support interval $(y_1, y_2)$. Recall that $\gamma(D_{n,2}) = 2$ if $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{Y^2}) = \emptyset$, then $P(\gamma(D_{n,2}) = 2) = 4/9 - (16/9)4^{-n}$. For more detail, see [3]. Hence, for $\mathcal{U}(y_1, y_2)$ data, we have

$$
\gamma(D_{n,2}) = \begin{cases} 
1 & \text{w.p. } 5/9 + (16/9)4^{-n}, \\
2 & \text{w.p. } 4/9 - (16/9)4^{-n},
\end{cases} \quad \text{for all } n \geq 1,
$$

(1)

where w.p. stands for “with probability”. Then the asymptotic distribution of $\gamma(D_{n,2})$ for $\mathcal{U}(y_1, y_2)$ data is given by

$$
\lim_{n \to \infty} \gamma(D_{n,2}) = \begin{cases} 
1 & \text{w.p. } 5/9, \\
2 & \text{w.p. } 4/9.
\end{cases}
$$

(2)

For $m > 2$, [3] computed the exact distribution of $\gamma(D_{n,m})$. However, independence of the distribution of the domination number from the support interval does not hold in general; that is, for $X_i \overset{iid}{\sim} F$ with support $S(F) \subseteq (y_1, y_2)$, the exact and asymptotic distribution of $\gamma(D_{n,2})$ will depend on $F$ and $\mathcal{Y}_2$. 

4
4.2 The Distribution of the Domination Number for $\mathcal{F}(\mathbb{R})$-random $\mathcal{D}_{n,2}$-digraphs

For $\gamma_2 = \{(y_1, y_2) \in \mathbb{R} \mid -\infty < y_1 < y_2 < \infty\}$, a quick investigation shows that the $\Gamma_1$-region is $\Gamma_1(\mathcal{X}_n, N_2) = \left(\frac{y_1 + x_{i+1}}{2}, \frac{y_1 + x_{i+2}}{2}\right)$. Note that $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_2)$ is the set of all dominating vertices, which is empty when $\gamma(D_{n,2}) > 1$. To make the dependence on $F$ explicit and for brevity of notation, we will denote the domination number of the $F((y_1, y_2))$-random $\mathcal{D}_{n,2}$-digraphs as $\gamma_n(F)$.

Let $p_n(F) := P(\gamma_n(F) = 2)$ and $p(F) := \lim_{n \to \infty} P(\gamma_n(F) = 2)$. Then the exact (finite sample) and asymptotic distributions of $\gamma_n(F)$ are $1 + \text{Bernoulli}(p_n(F))$ and $1 + \text{Bernoulli}(p(F))$, respectively. That is, for finite $n$, we have

$$\gamma_n(F) = \begin{cases} 1 & \text{w.p. } 1 - p_n(F) \text{ for all } n \geq 1. \\ 2 & \text{w.p. } p_n(F) \end{cases}$$

(3)

The asymptotic distribution is similar.

With $\mathcal{Y}_2 = \{0, 1\}$, let $F$ be a distribution with support $S(F) \subseteq (0, 1)$ and density $f$ and let $\mathcal{X}_n$ be a set of $n$ iid random variables from $F$. Since $\gamma_n(F) \in \{1, 2\}$, to find the distribution of $\gamma_n(F)$, it suffices to find $P(\gamma_n(F) = 1)$ or $P(\gamma_n(F) = 2)$. For computational convenience, we employ the latter in our calculations.

Then

$$p_n(F) = \int_{S(F) \setminus \Gamma_1(\mathcal{X}_n, N_2)} \left[1 - \frac{F((1 + x_1)/2) - F(x_n/2)}{F(x_n) - F(x_1)}\right]^{n-2} f_1(x_1, x_n) dx_1 dx_n,$$

(4)

where $f_1(x_1, x_n) = n(n-1) \left[(F(x_n) - F(x_1))^{n-2} f(x_1) f(x_n) \mathbf{1}(0 < x_1 < x_n < 1)\right]$ which is the joint probability density function of $X(1), X(n)$.

If the support $S(F) = (0, 1)$, then the region of integration becomes

$$\{(x_1, x_n) \in (0, 1)^2 : (1 + x_1)/2 \leq x_n \leq 1; \ 0 \leq x_1 \leq 1/3 \text{ or } 2x_1 \leq x_n \leq 1; \ 1/3 \leq x_1 \leq 1/2\}.$$

The integrand in Equation (4) simplifies to

$$H(x_1, x_n) := n(n-1) f(x_1) f(x_n) \left[F(x_n) + F(x_n/2) - (F((1 + x_1)/2) + F(x_1))\right]^{n-2}.$$

(5)

Let $\mathcal{X}_n$ be a set of iid random variables from a continuous distribution $F$ with $S(F) \subseteq (0, 1)$. The simplest of such distributions is $U(0, 1)$, the uniform distribution on $(0, 1)$, which yields the simplest exact distribution for $\gamma_n(F)$. If $X \sim F$, then by probability integral transform, $F(X) \sim U(0, 1)$. So for any continuous $F$, we can construct a proximity map depending on $F$ for which the distribution of the domination number for the associated digraph will have the same distribution as that of $\gamma_n(U(0, 1))$.

**Proposition 4.4.** Let $X_i \sim F$ which is an (absolutely) continuous distribution with support $S(F) = (0, 1)$ and $\mathcal{X}_n = \{X_1, \ldots, X_n\}$. Define the proximity map $N_F(x) := F^{-1}(N_{\mathcal{Y}}(F(x))) = F^{-1}(B(F(x), r(F(x))))$ where $r(F(x)) = \min(F(x), 1 - F(x))$. Then the domination number of the digraph based on $N_F$, $\mathcal{X}_n$, and $\mathcal{Y}_2 = \{0, 1\}$, is equal in distribution to $\gamma_n(U(0, 1))$.

**Proof:** Let $U_i := F(X_i)$ for $i = 1, \ldots, n$ and $\mathcal{U}_n = \{U_1, \ldots, U_n\}$. Hence, by probability integral transform, $U_i \sim U(0, 1)$. Let $U_{(k)}$ be the $k$th order statistic of $\mathcal{U}_n$ for $k = 1, \ldots, n$. Furthermore, such an $F$ preserves order; that is, for $x \leq y$, $F(x) \leq F(y)$. So the image of $N_F(x)$ under $F$ is $F(N_F(x)) = N_{\mathcal{Y}}(F(x)) = B(F(x), r(F(x)))$ for (almost) all $x \in (0, 1)$. Then $F(N_F(X_i)) = N_{\mathcal{Y}}(F(X_i)) = N_{\mathcal{Y}}(U_i)$ for
4.2.1 The Exact Distribution of $\gamma_n$ 

$i = 1, \ldots, n$. Since $U_i \overset{iid}{\sim} U(0, 1)$, the distribution of the domination number of the digraph based on $N_Y$, $U_n$ and $\{0, 1\}$ is given in Equation (1). Observe that $X_j \in N_Y(x_i)$ if $X_j \in F^{-1}(B(F(X_i), r(F(X_i))))$ if $F(X_j) \in B(F(X_i), r(F(X_i)))$ if $U_j \in B(U_i, r(U_i))$ for $i, j = 1, \ldots, n$. Hence $P(X_n \in N_Y(x_i)) = P(U_n \in N_Y(U_j))$ for all $i = 1, \ldots, n$. Therefore, $X_n \cap \Gamma_1(x_n, N_Y) = \emptyset$ if $U_n \cap \Gamma_1(U_n, N_Y) = \emptyset$, which implies that the domination number of the digraph based on $N_Y$, $X_n$, and $Y_2 = \{0, 1\}$ is 2 with probability $4/9 - (16/9) 4^{-n}$. Hence the desired result follows.

**Example 4.6.** Then $\gamma_n(1, 0)$ becomes

$$N_F(x) = \begin{cases} (0, \sqrt{2}x) & \text{for } x \in [0, 1/2], \\ (\sqrt{2}x^2 - 1, 1) & \text{for } x \in (1/\sqrt{2}, 1]. \end{cases}$$

There is also a stochastic ordering between $\gamma_n(F)$ and $\gamma_n(U(0, 1))$ provided that $F$ satisfies some conditions which are given in the following proposition.

**Proposition 4.5.** Suppose $X_n = \{X_1, \ldots, X_n\}$ is a random sample from a continuous distribution $F$ with $S(F) \subseteq (0, 1)$ and let $X_{(j)}$ be the $j^{th}$ order statistic of $X_n$ for $j = 1, \ldots, n$. If

$$F(X_{(n)}/2) < F(X_{(n)})/2 \text{ and } F(X_{(1)}) < 2 F((1 + X_{(1)})/2) - 1 \text{ hold a.s.,}$$

then $\gamma_n(F) <^{ST} \gamma_n(U(0, 1))$. If $<$ in expression (5) are replaced with $>$, then $\gamma_n(F) >^{ST} \gamma_n(U(0, 1))$. If $<$ in expression (5) are replaced with $=$, then $\gamma_n(F) \overset{d}{=} \gamma_n(U(0, 1))$ where $\overset{d}{=}$ stands for equality in distribution.

**Proof:** Let $U_i := F(X_i)$ for $i = 1, \ldots, n$ and $U_n = \{U_1, \ldots, U_n\}$. Then, by probability integral transform, $U_i \overset{iid}{\sim} U(0, 1)$. Let $U_{(j)}$ be the $j^{th}$ order statistic of $U_n$ for $j = 1, \ldots, n$. The $\Gamma_1$-region for $U_n$ based on $N_Y$ is $\Gamma_1(U_n, N_Y) = (U_{(n)}/2, (1 + U_{(1)})/2)$; likewise, $\Gamma_1(X_n, N_Y) = (X_{(n)}/2, (1 + X_{(1)})/2)$.

But the conditions in expression (5) imply that $\Gamma_1(U_n, N_Y) \subseteq F(\Gamma_1(X_n, N_Y))$. So $U_n \cap F(\Gamma_1(X_n, N_Y)) = \emptyset$ implies that $U_n \cap \Gamma_1(U_n, N_Y) = \emptyset$ and $U_n \cap F(\Gamma_1(X_n, N_Y)) = \emptyset$ iff $X_n \cap \Gamma_1(X_n, N_Y) = \emptyset$. Hence

$$p_n(F) = P(X_n \cap \Gamma_1(X_n, N_Y) = \emptyset) < P(U_n \cap \Gamma_1(U_n, N_Y) = \emptyset) = p_n(U(0, 1)).$$

Then $\gamma_n(F) <^{ST} \gamma_n(U(0, 1))$ follows. The other cases can be shown similarly.

For more on the comparison of $\gamma_n(F)$ for general $F$ against $\gamma_n(U(0, 1))$, see Section 4.2.2 of the technical report by [Ceyhan 2004].

4.2.1 The Exact Distribution of $\gamma_n(F)$ for $F$ with Piecewise Constant Density

Let $\mathcal{Y}_2 = \{0, 1\}$. We can find the exact distribution of $\gamma_n(F)$ for $F$ whose density is piecewise constant. Note that the simplest of such distributions is the uniform distribution $U(0, 1)$. Below we give some examples for such densities.

**Example 4.6.** Consider the distribution $F$ with density $f(\cdot)$ which is of the form $f(x) = \frac{1}{1-2\delta} I(\delta < x < 1 - \delta)$ with $\delta \in [0, 1/2)$. Then $F(x) = \frac{x-\delta}{1-2\delta} I(\delta < x < 1 - \delta) + I(x \geq 1 - \delta)$. The integrand in Equation (5) becomes

$$H(x_1, x_n) = \frac{n(n-1)}{(1-2\delta)^2} \left( \frac{3(x_n-x_1) - 1}{2(1-2\delta)} \right)^{n-2}.$$
Then for $\delta \in [0, 1/3]$

$$p_n(F) = \int_{\delta}^{1/3} \int_{(1+\delta)/2}^{1-\delta} H(x_1, x_n) \, dx_1 \, dx_n + \int_{1/3}^{(1-\delta)/2} \int_{2x_1}^{1-\delta} H(x_1, x_n) \, dx_1 \, dx_n$$

$$= \left(4/9 - (16/9) \cdot 4^{-n}\right) \left(\frac{1 - 3\delta}{1 - 2\delta}\right)^n, \quad (7)$$

which converges to 0 as $n \to \infty$ at (an exponential) rate $O\left((\frac{1 - 3\delta}{1 - 2\delta})^n\right)$. For $\delta \in [1/3, 1/2]$, it is easy to see that $\gamma_n(F) = 1$ a.s. In fact, for $\delta \in [1/3, 1/2]$ the corresponding digraph is a complete digraph of order $n$, since $X_n \subset N(X_i)$ for each $i = 1, \ldots, n$. Furthermore, if $\delta = 0$, then $F = U(0,1)$ which yields $p_n(F) = 4/9 - (16/9) \cdot 4^{-n}$. □

**Example 4.7.** Consider the distribution $F$ with density $f(\cdot)$ which is of the form

$$f(x) = \frac{1}{1 - 2\delta} I(x \in (0, 1) \setminus (1/2 - \delta, 1/2 + \delta)) \quad \text{with} \quad \delta \in [0, 1/6].$$

Then the cumulative distribution function (cdf) is given by

$$F(x) = F_1(x) I(0 < x < 1/2 - \delta) + F_2(x) I(1/2 - \delta < x < 1/2 + \delta) + F_3(x) I(1/2 + \delta < x < 1) + I(x \geq 1),$$

where

$$F_1(x) = x/(1 - 2\delta), \quad F_2(x) = 1/2, \quad \text{and} \quad F_3(x) = (x - 2\delta)/(1 - 2\delta).$$

There are four cases regarding the relative position of $X_n/2, (1 + X_{(i)})/2$ and $1/2 - \delta, 1/2 + \delta$ that yield $\gamma_n(F) = 2$:

- **case (1)** $(X_n)/2, (1 + X_{(i)})/2 \subseteq (1/2 - \delta, 1/2 + \delta)$;
- **case (2)** $X_n/2 < 1/2 - \delta < (1 + X_{(i)})/2 < 1/2 + \delta$;
- **case (3)** $1/2 - \delta < X_n/2 < 1/2 + \delta < (1 + X_{(i)})/2$;
- **case (4)** $X_n/2 < 1/2 - \delta < 1/2 + \delta < (1 + X_{(i)})/2$.

Let $E_j(n)$ be the event for which **case (j)** holds for $j = 1, 2, 3, 4$, for example,

$$E_1(n) := \{(X_n)/2, (1 + X_{(i)})/2 \subseteq (1/2 - \delta, 1/2 + \delta)\}.$$

Then $p_n(F) = \sum_{j=1}^{4} P(\gamma_n(F) = 2, E_j(n))$. Furthermore, **cases (2)** and **(3)** are symmetric; i.e., $P(\gamma_n(F) = 2, E_2(n)) = P(\gamma_n(F) = 2, E_3(n))$. Then in **case (1)**, we obtain $P(\gamma_n(F) = 2, E_1(n)) = 1 - 2 \left(\frac{1 - 6\delta}{1 - 2\delta}\right)^n + \left(\frac{1 - 6\delta}{1 - 2\delta}\right)^n$. Note that $P(\Gamma_1(X_n, N_Y) \subseteq (1/2 - \delta, 1/2 + \delta)) \to 1$ as $n \to \infty$, hence it suffices to use this case to show that $p_n(F) \to 1$ as $n \to \infty$ at an exponential rate since $P(E_1(n)) \leq p_n(F)$.

In **cases (2)** and **(3)**, we obtain $P(\gamma_n(F) = 2, E_2(n)) = \frac{4}{3} \left(1 - \frac{1}{4}\right) \left(\frac{1 - 6\delta}{1 - 2\delta}\right)^n - \left(\frac{1 - 6\delta}{1 - 2\delta}\right)^n$ and in **case (4)**, $P(\gamma_n(F) = 2, E_4(n)) = \frac{4}{3} \left(1 - 4^{-n+1}\right) \left(\frac{1 - 6\delta}{1 - 2\delta}\right)^n$. See Ceyhan (2004) for the details of the computations.

Combining the results from the cases, for $\delta \in [0, 1/6]$ we have

$$P(\gamma_n(F) = 2) = 1 + \left(\frac{1 - 6\delta}{1 - 2\delta}\right)^n (9/19 + (32/9)4^{-n}) - \left(\frac{1 - 4\delta}{1 - 2\delta}\right)^n (1/3 + (16/3)4^{-n}),$$

which converges to 1 as $n \to \infty$ at rate $O\left((\frac{1 - 4\delta}{1 - 2\delta})^n\right)$.

Notice that if $\delta = 0$, then $F = U(0, 1)$. The exact distribution for $\delta \in (1/6, 1/3]$ can be found in a similar fashion. Furthermore, if $\delta \in [1/3, 1/2]$, then $p_n(F) = 1 - 2\delta^n$. See Ceyhan (2004) also for the details of the computations. □
Example 4.8. Consider the distribution $F$ with density $f(\cdot)$ which is of the form $f(x) = (1 + \delta) I(x \in (0, 1/2)) + (1 - \delta) I(x \in [1/2, 1])$ with $\delta \in [-1, 1]$.

Then
\[
 p_n(F) = \frac{4(1 - \delta^2)}{9 - \delta^2} - \frac{8 \cdot 4^{-n}(1 - \delta^2)}{3} \left( \frac{(1 + \delta)^{n-1}}{3 - \delta} + \frac{(1 - \delta)^{n-1}}{3 + \delta} \right).
\] (9)

See Cevher (2004) for the derivation. Hence $\lim_{n \to \infty} p_n(F) = \frac{4(1 - \delta^2)}{9 - \delta^2} =: p_F(\delta)$, with the rate of convergence $O \left( \left( \frac{1 + \delta}{1 - \delta} \right)^n \right)$. Note that $p_F(\delta) \in [0, 4/9]$ is continuous in $\delta$ and decreases as $|\delta|$ increases. If $\delta = 0$, then $F = U(0, 1)$ and $p_F(\delta = 0) = 4/9$. Note also that $p_F(\delta = \pm 1) = 0$. □

Example 4.9. Consider the distribution $F$ with density $f(\cdot)$ which is of the form
\[
f(x) = (1 + \delta) I(0 < x < 1/4) + (1 - \delta) I(1/4 \leq x < 3/4) + (1 + \delta) I(3/4 \leq x < 1) \text{ with } \delta \in [-1, 1].
\]

The exact value of $p_n(F)$ is available, but it is rather a lengthy expression (see Cevher (2004) for the expression and its derivation). But the limit is as follows: $p_n(F) \to \frac{4(1 + \delta)^2}{(3 + \delta)^2} =: p_F(\delta)$ as $n \to \infty$ with the rate of convergence $O \left( \left( \frac{2 + \delta}{2 - \delta} \right)^n \right)$. So $p_F(\delta)$ is increasing in $\delta$. Notice here that $p_n(F)$ and $p_F(\delta)$ are continuous in $\delta$ and $p_F(\delta) > 0$ for all $\delta \in (-1, 1)$. Moreover, $p_F(\delta = 1) = 1$ and $p_F(\delta = -1) = 0$. □

Note that extra care should be taken if the points of discontinuity in the above examples are different from $\{1/4, 3/4\}$ or $1/2$, since the symmetry in the probability calculations no longer exists in such cases.

4.2.2 The Exact Distribution of $\gamma_n(F)$ for Polynomial $f$ Using Multinomial Expansions

The exact distribution of $\gamma_n(F)$ for (piecewise) polynomial $f(x)$ with at least one piece is of degree 1 or higher can be obtained using the multinomial expansion of the term $(\cdot)^{n-2}$ in Equation (5) with careful bookkeeping. However, the resulting expression for $p_n(F)$ is extremely lengthy and not that informative.

The simplest example is with $f(x) = 2x$ and $F(x) = x^2$. Then $p_n(F) = P(\gamma_n(F) = 2) = \Lambda_1(n) + \Lambda_2(n)$, where $\Lambda_1(n) := \int_0^{1/3} \int_{1 + x_1/2}^{x_1} H(x_1, x_n) dx_n dx_1$, $\Lambda_2(n) := \int_{1/3}^{1/2} \int_{1/2}^{x_1} H(x_1, x_n) dx_n dx_1$, and $H(x_1, x_n) = n(n - 1)x_1 x_n (5x_n^2 - 1 - 2x_1 - 5x_1^2)^{-n-2}$. Then
\[
\Lambda_1(n) = \int_0^{1/3} (8n x_1/5) (1 - x_1/2 - 5x_1^2/4)^{n-1} - (8n x_1/5) (1/16 + x_1/2 - 15x_1^2/16)^{n-1} dx_1.
\]

Using the multinomial expansion of $(\cdot)^{n-1}$ with respect to $x_1$ in the integral above, we have
\[
\Lambda_1(n) = \sum_{Q_2} \binom{n-1}{q_1, q_2, q_3} 8n(-1)^{q_2+q_3} 2^{-q_2-q_3} q_1 3^{-2-q_2-q_3} q_3 15q_1 3^{-2-q_2-q_3} q_3 2^q_2 + (n(-1)^{q_2+q_3} 2^{-q_2-q_3} q_1 3^{-2-q_2-q_3} q_3 15q_1 3^{-2-q_2-q_3} q_3 2^q_2 = \sum_{Q_2} \binom{n-1}{q_1, q_2, q_3} 8n(-1)^{q_2+q_3} 2^{-q_2-q_3} q_1 3^{-2-q_2-q_3} q_3 15q_1 3^{-2-q_2-q_3} q_3 2^q_2 + (n(-1)^{q_2+q_3} 2^{-q_2-q_3} q_1 3^{-2-q_2-q_3} q_3 15q_1 3^{-2-q_2-q_3} q_3 2^q_2
\]
where $Q_2 = \{q_1, q_2, q_3 \in \mathbb{N} : q_1 + q_2 + q_3 = n - 1\}$.

Similarly, the second piece follows as
\[
\Lambda_2(n) = \int_{1/3}^{1/2} (8n x_1/5) (1 - x_1/2 - 5x_1^2/4)^{n-1} - (8n x_1/5) (1/16 + x_1/2 - 15x_1^2/16)^{n-1} dx_1.
\]
Again, using the multinomial expansion of the \((\cdot)^{n-1}\) term above, we get
\[
\Lambda_2(n) = \sum_{Q_3} \left( \frac{n-1}{r_1, r_2, r_3} \right) \left[ 2n \left( 9 (-1)^{r_2+r_1} 5^{r_1} 4^{-2r_1-r_2} + 9 (-1)^{1+r_3+r_2} 15^{r_1} 4^{-2r_1-r_3-r_2} + 4 (-1)^{1+r_2+r_1} 6^{-r_2-r_1} 5^{r_1} + (-1)^{r_1+r_2} 4^{1-r_3} 6^{-r_2-r_1} 12^{-r_3} 5^{r_1} \right) \right] / \left[ 90 + 45 r_2 + 90 r_1 \right]
\]
where \(Q_3 = \{ r_1, r_2, r_3 \in \mathbb{N} : r_1 + r_2 + r_3 = n - 1 \} \). See Ceyhan (2004) for more detail and examples.

For fixed numeric \(n\), one can obtain \(p_n(F)\) for \(F\) (omitted for the sake of brevity) with the above densities by numerical integration of the below expression.
\[
p_n(F) = P(\gamma_n(F) = 2) = \int_0^{1/3} \int_{(1+x_1)/2}^{1} H(x_1, x_n) + \int_{1/3}^{1} \int_{2x_1}^{1} H(x_1, x_n) dx_n dx_1,
\]
where \(H(x_1, x_n)\) is given in Equation [3].

Recall the \(\mathcal{F}(\mathbb{R}^d)\)-random \(\mathcal{D}_{n,m}\)-digraphs. We call the digraph which obtains in the special case of \(\mathcal{Y}_m = \{ y_1, y_2 \}\) and support of \(F_X\) in \((y_1, y_2)\), \(\mathcal{F}(y_1, y_2)\)-random \(\mathcal{D}_{n,2}\)-digraph. Below, we provide asymptotic results pertaining to the distribution of such digraphs.

5 The Asymptotic Distribution of the Domination Number of \(\mathcal{F}(y_1, y_2)\)-random \(\mathcal{D}_{n,2}\)-digraphs

Although the exact distribution of \(\gamma_n(F)\) is not analytically available in a simple closed form for \(F\) whose density is not piecewise constant, the asymptotic distribution of \(\gamma_n(F)\) is available for larger families of distributions. First, we present the asymptotic distribution of \(\gamma_n(F)\) for \(\mathcal{D}_{n,2}\)-digraphs with \(\mathcal{Y}_2 = \{ y_1, y_2 \}\) \(\subset\) \(\mathbb{R}\) with \(y_1 < y_2\) for various \(F\) with support \(S(F) \subseteq (y_1, y_2)\). Then we will extend this to the case with \(\mathcal{Y}_m \subset \mathbb{R}\) for \(m > 2\).

For \(\varepsilon \in (0, (y_1 + y_2)/2)\), consider the family of distributions given by
\[
\mathcal{F}(y_1, y_2, \varepsilon) = \left\{ F : (y_1, y_1 + \varepsilon) \cup (y_2 - \varepsilon, y_2) \cup ((y_1 + y_2)/2 - \varepsilon, (y_1 + y_2)/2 + \varepsilon) \subseteq S(F) \subseteq (y_1, y_2) \right\}.
\]

Let the \(k^{th}\) order right (directed) derivative at \(x\) be defined as \(f^{(k)}(x^+) := \lim_{h \to 0^+} \frac{f^{(k-1)}(x+h) - f^{(k-1)}(x)}{h}\) for all \(k \geq 1\) and the right limit at \(c\) be defined as \(f^{(c^+)} := \lim_{h \to 0^+} f(c + h)\). The left derivatives and limits are defined similarly with +’s being replaced by −’s. Furthermore, let \(\vec{h} = (h_1, h_2)\) and \(\vec{c} = (c_1, c_2)\) and the directionlal limit at \((c_1, c_2)\) \(\in\) \(\mathbb{R}^2\) for \(g(x, y)\) in the first quadrant in \(\mathbb{R}^2\) be \(g(c_1^+, c_2^+) := \lim_{\|h\| \to 0, h_1, h_2 \geq 0} g(\vec{c} + \vec{h})\) and the directional partial derivatives at \((c_1, c_2)\) along paths in the first quadrant be
\[
\frac{\partial^{k+1} g(c_1^+, c_2^+)}{\partial x^{k+1}} := \lim_{\|h\| \to 0, h_1, h_2 \geq 0} \left( \frac{\partial^k g(\vec{c} + \vec{h})}{\partial x^k} \right) \text{ for } k \geq 1.
\]

\textbf{Theorem 5.1.} Let \(\mathcal{Y}_2 = \{ y_1, y_2 \} \subset \mathbb{R}\) with \(-\infty < y_1 < y_2 < \infty\) and \(\mathcal{X}_n = \{ X_1, \ldots, X_n \}\) with \(X_i \sim F \in \mathcal{F}(y_1, y_2, \varepsilon)\). Let \(\mathcal{D}_{n,2}\) be the random \(\mathcal{D}_{n,2}\)-digraph based on \(\mathcal{X}_n\) and \(\mathcal{Y}_2\). Suppose \(k \geq 0\) is the smallest integer for which \(F(\cdot)\) has continuous right derivatives up to order \((k + 1)\) at \(y_1, (y_1 + y_2)/2,\)
\( f^{(k)}(y_1^+) + 2^{-(k+1)} f^{(k)} \left( \left( \frac{y_1 + y_2}{2} \right)^+ \right) \neq 0 \) and \( f^{(j)}(y_1^+) = 0 \) for all \( j = 0, 1, \ldots, k-1 \); and \( \ell \geq 0 \) is the smallest integer for which \( F(\cdot) \) has continuous left derivatives up to order \( (\ell + 1) \) at \( y_2 \), \( (y_1 + y_2)/2 \), \( f^{(\ell)}(y_2^-) + 2^{-(\ell+1)} f^{(\ell)}(\left( \frac{y_1 + y_2}{2} \right)^- \) \neq 0 \) and \( f^{(j)}(y_2^-) = 0 \) for all \( j = 0, 1, \ldots, \ell-1 \). Then \( \gamma_n(F) \sim 1 + \text{Bernoulli}(p_n(F)) \) where \( p_n(F) := P(\gamma_n(F) = 2) \) and for bounded \( f^{(k)}(\cdot) \) and \( f^{(\ell)}(\cdot) \), we have the following limit

\[
\lim_{n \to \infty} p_n(F) = \frac{\int f^{(k)}(y_1^+) f^{(\ell)}(y_2^-)}{\left[ f^{(k)}(y_1^+) + 2^{-(k+1)} f^{(k)} \left( \left( \frac{y_1 + y_2}{2} \right)^+ \right) \right] \left[ f^{(\ell)}(y_2^-) + 2^{-(\ell+1)} f^{(\ell)}(\left( \frac{y_1 + y_2}{2} \right)^- \right].}
\]

Note also that \( p_1(F) = 0 \).

**Proof:** First suppose \((y_1, y_2) = (0,1)\). Recall that \( \Gamma_1(X_n, N_Y) = (X(0), (1 + X(1))/2) \subset (0,1) \) and \( \gamma_n(F) = 2 \) iff \( X_0 \cap \Gamma_1(X_n, Y_0) = \emptyset \). Then for finite \( n \),

\[
p_n(F) = P(\gamma_n(F) = 2) = \int_{S(F) \setminus \Gamma_1(X_n, N_Y)} H(x_1, x_n) \, dx_n \, dx_1,
\]

where \( H(x_1, x_n) \) is as in Equation 5.

Let \( \varepsilon \in (0,1/3) \). Then \( P(X(1) < \varepsilon, X_n > 1 - \varepsilon) \to 1 \) as \( n \to \infty \) with the rate of convergence depending on \( F \). So for sufficiently large \( n \),

\[
p_n(F) \approx \int_0^\varepsilon \int_{1-\varepsilon}^1 n (n-1) f(x_1) f(x_n) \left[ F(x_n) - F(x_1) + F(x_n/2) - F((1 + x_1)/2) \right]^{n-2} \, dx_n \, dx_1.
\]

Let

\[
G(x_1, x_n) = F(x_n) - F(x_1) + F(x_n/2) - F((1 + x_1)/2).
\]

The integral in Equation (10) is critical at \((x_1, x_n) = (0,1)\), since \( G(0,1) = 1 \) and for \((x_1, x_n) \in (0,1)^2\) the integral converges to 0 as \( n \to \infty \). So we make the change of variables \( z_1 = x_1 \) and \( z_n = 1 - x_n \), then \( G(z_1, z_n) \) becomes

\[
G(z_1, z_n) = F(1 - z_n) - F(z_1) + F((1 - z_n)/2) - F((1 + z_1)/2),
\]

and Equation (10) becomes

\[
p_n(F) \approx \int_0^\varepsilon \int_0^\varepsilon n (n-1) f(z_1) f(1 - z_n) [G(z_1, z_n)]^{n-2} \, dz_n \, dz_1.
\]

The new integral is critical at \((z_1, z_n) = (0,0)\). Note that \( \frac{\partial^{s+r} G(z_1, z_n)}{\partial z_1^s \partial z_n^r} = 0 \) for all \( r, s \geq 1 \). Let \( \alpha_i := \frac{\partial^{i+1} G(z_1, z_n)}{\partial z_1^{i+1}} \big|_{(0^+,0^+)} = f^{(i)}(0^+) + 2^{-i+1} f^{(i)} \left( \frac{1}{2} \right) \) and \( \beta_j := \frac{\partial^{j+1} G(z_1, z_n)}{\partial z_n^{j+1}} \big|_{(0^+,0^+)} = f^{(j)}(1^-) + 2^{-j+1} f^{(j)} \left( \frac{1}{2} \right) \).
Then by the hypothesis of the theorem, we have \( \alpha_i = 0 \) and \( f^{(i)} \left( \frac{1}{2} \right) = 0 \) for all \( i = 0, 1, \ldots, (k-1) \); and \( \beta_j = 0 \) and \( f^{(j)} \left( \frac{1}{2} \right) = 0 \) for all \( j = 0, 1, \ldots, (\ell - 1) \). So the Taylor series expansions of \( f(z_1) \) around \( z_1 = 0^+ \) up to order \( k \) and \( f(1 - z_n) \) around \( z_n = 0^+ \) up to order \( \ell \), and \( G(z_1, z_n) \) around \((0^+,0^+)\) up to order \( (k+1) \) and \((\ell+1)\) in \( z_1, z_n \), respectively, so that \((z_1, z_n) \in (0,\varepsilon)^2\), are as follows.

\[
f(z_1) = \frac{1}{k!} f^{(k)}(0^+) z_1^k + O \left( z_1^{k+1} \right); \quad f(1 - z_n) = \frac{(-1)^{\ell}}{\ell!} f^{(\ell)}(1^-) z_n^\ell + O \left( z_n^{\ell+1} \right);
\]

\[
G(z_1, z_n) = G(0^+, 0^+) + \frac{1}{(k+1)!} \left( \frac{\partial^{k+1} G(0^+, 0^+)}{\partial z_1^{k+1}} \right) z_1^{k+1} + \frac{1}{(\ell+1)!} \left( \frac{\partial^{\ell+1} G(0^+, 0^+)}{\partial z_n^{\ell+1}} \right) z_n^{\ell+1} + O \left( z_1^{k+2} \right) + O \left( z_n^{\ell+2} \right).
\]

\[
\approx 1 - \frac{\alpha_k}{(k+1)!} z_1^{k+1} + \frac{(-1)^{\ell+1} \beta_\ell}{(\ell+1)!} z_n^{\ell+1} + O \left( z_1^{k+2} \right) + O \left( z_n^{\ell+2} \right).
\]

10
Then substituting these expansions in Equation (11), we obtain

\[
p_n(F) \approx \int_0^\infty \int_0^\infty n (n-1) \left[ \frac{1}{k!} f^{(k)}(0^+) z_k + O \left( \frac{\ell}{n} \right) \right] \left[ \frac{(-1)^\ell}{\ell!} f^{(\ell)}(1^-) z_{n-1} + O \left( \frac{\ell}{n} \right) \right] \left[ 1 - \frac{\alpha_k}{(k+1)!} z_1 + \frac{(-1)^\ell \beta_\ell}{(\ell+1)!} z_{n-1} + O \left( \frac{\ell}{n} \right) \right] dzn dz_1.
\]

Now we let \( z_1 = w n^{-1/(k+1)} \), \( z_n = v n^{-1/(\ell+1)} \), and \( \nu = \min(k, \ell) \) to obtain

\[
p_n(F) \approx \int_0^\infty n \left[ 1 - \frac{1}{n} \left( \frac{\alpha_k}{(k+1)!} w + \frac{(-1)^\ell \beta_\ell}{(\ell+1)!} v \right) + O \left( \frac{\ell}{(k+1)(\ell+1)} \right) \right] \left[ \frac{(-1)^\ell}{n^{1/(k+1)}} f^{(\ell)}(1^-) w^{k+1} + O \left( \frac{\ell}{(k+1)(\ell+1)} \right) \right] \left[ \frac{1}{n^{1/(\ell+1)}} \right] dv dw = \int_0^\infty n (n-1) \left[ \frac{(-1)^\ell}{n^2 \ell!} f^{(k)}(0^+) f^{(\ell)}(1^-) w^{k+1} + O \left( \frac{\ell}{(k+1)(\ell+1)} \right) \right] \left[ 1 - \frac{1}{n} \left( \frac{\alpha_k}{(k+1)!} w + \frac{(-1)^\ell \beta_\ell}{(\ell+1)!} v \right) + O \left( \frac{\ell}{(k+1)(\ell+1)} \right) \right] dvn dw,
\]

letting \( n \to \infty \),

\[
\approx \int_0^\infty \int_0^\infty \left( \frac{(-1)^\ell}{k! \ell!} f^{(k)}(0^+) f^{(\ell)}(1^-) w^{k+1} v^{\ell+1} \right) \left[ \frac{\alpha_k}{(k+1)!} w^{k+1} + \frac{(-1)^\ell \beta_\ell}{(\ell+1)!} v^{\ell+1} \right] dv dw = \frac{f^{(k)}(0^+) f^{(\ell)}(1^-)}{k! \ell! (k+1)! (\ell+1)!} \frac{\alpha_k \beta_\ell}{(k+1)! (\ell+1)!} = \frac{f^{(k)}(0^+) f^{(\ell)}(1^-)}{k! \ell! (k+1)! (\ell+1)!} \frac{\alpha_k \beta_\ell}{(k+1)! (\ell+1)!} = \frac{f^{(k)}(0^+) f^{(\ell)}(1^-)}{k! \ell! (k+1)! (\ell+1)!} \frac{\alpha_k \beta_\ell}{(k+1)! (\ell+1)!} = \frac{f^{(k)}(0^+) f^{(\ell)}(1^-)}{k! \ell! (k+1)! (\ell+1)!} \frac{\alpha_k \beta_\ell}{(k+1)! (\ell+1)!}
\]

as \( n \to \infty \) at rate \( O(c(f) \cdot n^{-m}) \) where \( c(f) \) is a constant depending on \( f \).

For the general case of \( \mathcal{Y} = \{y_1, y_2\} \), the transformation \( \phi(x) = \frac{x - y_1}{y_2 - y_1} \) maps \( \mathcal{Y} \) to \((0,1)\) and the transformed random variables \( \phi(X_i) \) are distributed with density \( g(x) = (y_2 - y_1) f \left( \frac{x - y_1}{y_2 - y_1} \right) \) on \((0,1)\). Substituting \( f(x) \) by \( g(x) \) in Equation (12), the desired result follows.

Note that

- if \( \min(f^{(k)}(y_1^+), f^{(\ell)}(y_2^-)) = f^{(k)} \left( \frac{y_1 + y_2}{2} \right) \neq 0 \) and \( \min(f^{(k)}(y_1^-), f^{(\ell)}(y_2^+)) = f^{(\ell)} \left( \frac{y_1 + y_2}{2} \right) \neq 0 \) then \( p_n(F) \to 0 \) as \( n \to \infty \), at rate \( O(c(f) \cdot n^{-m}) \) and

- if \( \min(f^{(k)}(y_1^+), f^{(\ell)}(y_2^-)) = f^{(k)} \left( \frac{y_1 + y_2}{2} \right) = 0 \) and \( f^{(k)} \left( \frac{y_1 + y_2}{2} \right) = f^{(\ell)} \left( \frac{y_1 + y_2}{2} \right) = 0 \) then \( p_n(F) \to 1 \) as \( n \to \infty \), at rate \( O(c(f) \cdot n^{-m}) \).

For example, with \( F = \mathcal{U}(y_1, y_2) \), in Theorem 5.1 we have \( k = \ell = 0 \), \( f(y_1^+) = f(y_2^-) = f \left( \frac{y_1 + y_2}{2} \right) = 0 \).
Remark 5.4. For $F$ with density $f(x) = (x + 1/2)I(0 < x < 1)$, we have $k = \ell = 0$, $f(0^+) = 1/2$, $f(1^-) = 3/2$ and $f\left(\frac{1}{2}\right) = f\left(\frac{1}{2}\right) = 1$. Thus $\lim_{n \to \infty} p_n(F) = 3/8 = 0.375$. The numerically computed (by numerical integration) value of $p_n(F)$ with $n = 1000$ is $\hat{p}_{1000}(F) \approx 0.3753$. \qed

Remark 5.3. Let $p_F := \lim_{n \to \infty} p_n(F)$. Then the finite sample mean and variance of $\gamma_n(F)$ are given by $1 + p_n(F)$ and $p_n(F) (1 - p_n(F))$, respectively; and the asymptotic mean and variance of $\gamma_n(F)$ are given by $1 + p_F$ and $p_F (1 - p_F)$, respectively. \qed

Remark 5.4. In Theorem 5.1 we assume that $f^{(k)}(\cdot)$ and $f^{(\ell)}(\cdot)$ are bounded on $(y_1, y_2)$. Suppose either $f^{(k)}(\cdot)$ or $f^{(\ell)}(\cdot)$ or both are not bounded on $(y_1, y_2)$ for $k, \ell \geq 0$, in particular at $y_1, (y_1 + y_2)/2, y_2$, for example, $\lim_{x \to y_1^+} f^{(k)}(x) = \infty$. Then we find $p(F)$ as

$$p(F) = \lim_{\delta \to 0^+} \frac{f^{(k)}(y_1 + \delta) f^{(\ell)}(y_2 - \delta)}{f^{(k)}(y_1 + \delta) + 2^{-(k+1)} f^{(k)}\left(\frac{y_1 + y_2}{2} + \delta\right)} \frac{f^{(\ell)}(y_2 - \delta) + 2^{-(\ell+1)} f^{(\ell)}\left(\frac{y_1 + y_2}{2} - \delta\right)}{f^{(\ell)}(y_2 - \delta) + 2^{-(\ell+1)} f^{(\ell)}(y_2 - \delta)}$$.

\qed

Example 5.5. Consider the distribution with density function $f(x) = \frac{1}{\pi \sqrt{1 - x^2}} I(0 < x < 1)$. Note that $Y_2 = \{0, 1\}$ and $f(x)$ is unbounded at $x \in \{0, 1\}$. See Figure 1 (left) for the plot of $f(x)$. Instead of $f(x)$, we consider $g(x) = \frac{2}{\pi \arcsin(1 - 2x)} I(\delta < x < 1 - \delta)$ with cdf $G(x)$. For $g(x)$, we have $k = \ell = 0$ in Theorem 5.3 and then $\lim_{n \to \infty} p_n(F) = \lim_{\delta \to 0^+} \lim_{n \to \infty} p_n(G) = 1$ using Remark 5.4. The numerically computed value of $p_{1000}(F)$ is $\hat{p}_{1000}(F) \approx 1.000$. \qed

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{density_plot.png}
\caption{Graph of the density in Examples 5.5 (left) and 5.7 (right).}
\end{figure}

Remark 5.6. The rate of convergence in Theorem 5.1 depends on $f$. From the proof of Theorem 5.1, it follows that for sufficiently large $n$,

$$p_n(F) \approx \frac{f^{(k)}(y_1^+) f^{(\ell)}(y_2^+)}{f^{(k)}(y_1^+) + 2^{-(k+1)} f^{(k)}\left(\frac{y_1 + y_2}{2} + \delta\right)} \frac{f^{(\ell)}(y_2^+) + 2^{-(\ell+1)} f^{(\ell)}\left(\frac{y_1 + y_2}{2} - \delta\right)}{f^{(\ell)}(y_2^+) + 2^{-(\ell+1)} f^{(\ell)}(y_2^+)} + \frac{c(f)}{n^m},$$

where

$$c(f) = \frac{s_1 s_3^{\ell+1} \Gamma\left(\ell + 2\right)}{(k + 1) (\ell + 1) s_3^{\ell+1} s_4^{\ell+1}} \frac{s_2 s_3^{\ell+1} \Gamma\left(\ell + 2\right)}{s_3^{\ell+1} s_4^{\ell+1}} + \frac{0.4}{0.9},$$

and

$$\hat{p}_{1000}(F) \approx 0.3753.$$
So consider the distribution of the piecewise constant functions in Section 4.2.1, Theorem 5.1 applies. See Section 6.1 in Ceyhan (2004).

\[ \lim_{n \to \infty} p_n(F) = \frac{4 - a^2}{9 - a^2} =: p_F(a). \]

Note that \( p_F(a) \in [0, 4/9] \) is continuous in \( a \) and decreases as \( |a| \) increases. If \( a = 0 \), then \( F = U(0,1) \), and \( p_F(a = 0) = 4/9 \). Moreover, \( p_F(a = \pm 2) = 0 \); that is, for \( a = \pm 2 \), the asymptotic distribution of \( \gamma_n(F) \) is degenerate. □

Example 5.9. Consider the distribution of the piecewise constant function \( F(x) = a x + b \) for \( x \in (0, 1) \) with \( |a| \leq 2, b = 1 - a/2 \).

So \( k = \ell = 0 \) and \( f(0^+) = b, f(1^-) = a + b \) and \( f\left(\frac{1}{2}\right) = f\left(\frac{-1}{2}\right) = a/2 + b \). Then by Theorem 5.1 we have

\[ \lim_{n \to \infty} p_n(F) = \frac{4 - a^2}{9 - a^2} =: p_F(a). \]

Example 5.10. Consider the normal distribution \( \mathcal{N}(\mu, \sigma^2) \) restricted to the interval \((0, 1)\) with \( \mu \in \mathbb{R} \) and \( \sigma > 0 \). Then the corresponding density function is given by

\[ f(x, \mu, \sigma) = \kappa \left( \frac{1}{\sqrt{2\pi} \sigma} \right) \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right) \mathbf{1}(0 < x < 1), \]
where $\kappa = \Phi \left( \frac{1-\mu}{\sigma} \right) - \Phi \left( \frac{-\mu}{\sigma} \right)^{-1}$ with $\Phi(\cdot)$ being the cdf of the standard normal distribution $\mathcal{N}(0,1)$. Note that $k = \ell = 0$, then by Theorem 5.1

$$\lim_{n \to \infty} p_n(F) = \frac{4}{(2 + \exp \left( \frac{4-4\mu}{\sqrt{8}\sigma} \right)) (2 + \exp \left( \frac{4-4\mu}{\sqrt{8}\sigma} \right))} =: p_F(\mu, \sigma).$$

Observe that $p_F(\mu, \sigma) \in [0, 4/9]$ is continuous in $\mu$ and $\sigma$ and increases as $\sigma$ increases for fixed $\mu$. Furthermore, for fixed $\mu$, $\lim_{\sigma \to \infty} p_F(\mu, \sigma) = 4/9$ and $\lim_{\sigma \to 0} p_F(\mu, \sigma) = 0$. For fixed $\sigma > 0$, $\lim_{\mu \to \pm \infty} p_F(\mu, \sigma) = 0$, $p_F(\mu, \sigma)$ decreases as $|\mu - 1/2|$ increases, and $p_F(\mu, \sigma)$ is maximized at $\mu = 1/2$. □

**Example 5.11.** Consider the distribution $F$ with density $f(\cdot)$ which is of the form

$$f(x) = 2^q(q+1) \left[ x^q I(0 < x < 1/2) + (x - 1/2)^q I(1/2 \leq x < 1) \right] \quad \text{with } q \in [0, \infty].$$

See Figure 2 (left) with $q = 2$. Since $f(0^+) = f \left( \frac{1^+}{2} \right) = 0$, we can apply Theorem 5.1 with $k = q$ and $l = 0$. Then $f^{(q)}(0^+) = (q+1)!2^q$, $f(1^-) = (q+1)$, $f \left( \frac{1^-}{2} \right) = (q+1)$, and $f^{(q)} \left( \frac{1^-}{2} \right) = (q+1)!2^q$. By Theorem 5.1 we have

$$\lim_{n \to \infty} p_n(F) = \frac{2^q+2}{3(1 + 2q+1)} =: p_F(q).$$

Note that $p_F(q) \in [4/9, 2/3]$ is a continuous increasing function of $q$. If $q = 0$, then $F = \mathcal{U}(0,1)$. □

**Example 5.12.** Consider the distribution $F$ with density $f(\cdot)$ which is of the form

$$f(x) = (\delta + 12(1-\delta)x^2) I(0 < x < 1/2) + (\delta + 12(1-\delta)(x - 1/2)^2) I(1/2 \leq x < 1) \quad \text{with } \delta \in [0, 1].$$

See Figure 2 with $\delta = 0$ (left) and $\delta = 2/3$ (right). Since $f(0^+) = \delta$, $f(1^-) = (3-2\delta)$, $f \left( \frac{1^-}{2} \right) = (3-2\delta)$, and $f \left( \frac{1^+}{2} \right) = \delta$, for $\delta \in (0,1]$ we have $k = \ell = 0$ and so by Theorem 5.1

$$\lim_{n \to \infty} p_n(F) = 4/9$$

for $\delta \in (0,1]$. Note that if $\delta = 1$, then $F = \mathcal{U}(0,1)$. For $\delta = 0$, we can apply Theorem 5.1 with $k = 2$ and $l = 0$. Hence we get $p_F(\delta = 0) = 16/27$. Observe that in this example, $\gamma_n(F)$ has two distinct non-degenerate distributions at different values of $\delta$. □

![Figure 2: Left plot is for the density in Example 5.11 with $q = 2$ or for the density in Example 5.12 with $\delta = 0$. Right plot is for the density in Example 5.12 with $\delta = 2/3.](image)
Remark 5.13. If, in Theorem 5.1, we have \( f^{(k)}(0^+) = f^{(k)} \left( \frac{1}{2} \right) \) and \( f^{(\ell)}(1^-) = f^{(\ell)} \left( \frac{1}{2} \right) \), then
\[
\lim_{n \to \infty} p_n(F) = \left( \frac{1}{1 + 2^{-(k+1)}} \right) \left( \frac{1}{1 + 2^{-(\ell+1)}} \right).
\]
In particular, if \( k = \ell = 0 \), then \( \lim_{n \to \infty} p_n(F) = 4/9 \) (i.e., \( \gamma_n(F) \) and \( \gamma_n(U(0, 1)) \) have the same asymptotic distributions). \( \square \)

Example 5.14. Beta\((\nu_1, \nu_2)\) with \( \nu_1, \nu_2 \geq 1 \). The density function is
\[
f(x, \nu_1, \nu_2) = \frac{x^{\nu_1-1}(1-x)^{\nu_2-1}}{\beta(\nu_1, \nu_2)} \mathbf{I}(0 < x < 1)
\]
where \( \beta(\nu_1, \nu_2) = \frac{\Gamma(\nu_1) \Gamma(\nu_2)}{\Gamma(\nu_1 + \nu_2)} \).

Then \( \lim_{n \to \infty} p_n(\text{Beta}(\nu_1, \nu_2)) = 0 \) at rate \( O(n^{-(\nu_1+\nu_2-2)}) \). Let \( p_n(\nu_1, \nu_2) \) denote the \( P(\gamma_n(F) = 2) \) for \( F = \text{Beta}(\nu_1, \nu_2) \). The numerically computed values of \( p_n(\nu_1, \nu_2) \) for \( n = 1000 \) are \( \hat{p}_{1000}(4, 1) = \hat{p}_{1000}(1, 4) \approx 0.000005 \), \( \hat{p}_{1000}(4, 2) = \hat{p}_{1000}(2, 4) < 0.00001 \) and \( \hat{p}_{1000}(2, 2) \approx 0.000001 \). \( \square \)

Here is an example with general support \((y_1, y_2)\).

Example 5.15. Consider the distribution \( F \) with density \( f(\cdot) \) which is of the form \( f(x) = a x + b \) with \( b = \frac{1}{(y_2 - y_1)}(1 - a(y_2^2 - y_1^2)/2) \) and \( |a| \leq \frac{2}{(y_2 - y_1)^2} \). Using Theorem 5.1, we obtain \( p_F = \frac{a^2(y_2-y_1)^{\nu_2-4}}{a^2(y_2-y_1)^{\nu_2-9}} \).

If \( y_1, y_2 = 0, 1 \), then \( b = 1 - a/2 \) and \( p_F(a) = \frac{\sqrt{a}}{\sqrt{2}} \). In both cases, \( p_F(a) \) is maximized for the uniform case; i.e., when \( a = 0 \), then we have \( p_F(a = 0) = 4/9 \). Furthermore, \( \gamma_n(F) \) is degenerate in the limit when \( a = \pm \frac{2}{(y_2 - y_1)^2} \), since \( p_n(F) \to 0 \) as \( n \to \infty \) at rate \( O(n^{-1}) \). \( \square \)

For more detail and examples, see Sections 6.4 and 7.1 in Ceyhan (2004).

6 The Distribution of the Domination Number of \( \mathcal{D}_{n,m} \)-digraphs

In this section, we attempt the more challenging case of \( m > 2 \). For \( c < d \) in \( \mathbb{R} \), define the family of distributions
\[
\mathcal{H}(\mathbb{R}) := \{ F_{X,Y} : (X_i, Y_i) \sim F_{X,Y} \text{ with support } S(F_{X,Y}) = (c, d)^2 \subseteq \mathbb{R}^2, X_i \sim F_X \text{ and } Y_i \overset{iid}{\sim} F_Y \}.
\]

We provide the exact distribution of \( \gamma(D_{n,m}) \) for \( \mathcal{H}(\mathbb{R}) \)-random digraphs in the following theorem. Let \( \lfloor n \rfloor := \{0, 1, \ldots, m-1\} \) and \( \Theta_{a,b}^S := \{ (u_1, \ldots, u_b) : \sum_{i=1}^b u_i = a, u_i \in S, \forall i \} \). Let \( \mathcal{Y}_m = \{ Y_1, Y_2, \ldots, Y_m \} \) whose order statistics are denoted as \( Y_j \) for \( j = 1, 2, \ldots, m \). Note that the order statistics are distinct a.s. provided \( Y \) has a continuous distribution. Let \( \gamma(D^X) \) be the domination number of the digraph induced by \( X^j \) and \( Y^j \) (see Section 4). Given \( Y(j) = y(j) \) for \( j = 1, \ldots, m \), let \( F_j \) be the (conditional) marginal distribution of \( X \) restricted to \( I_j = (y(j-1), y(j)) \) for \( j = 1, \ldots, (m + 1) \), \( \tilde{n} \) be the vector of numbers of \( X \) points \( n_j \) falling into intervals \( I_j \). Let \( f_{\tilde{n}}(\tilde{y}) \) be the joint distribution of the order statistics of \( \mathcal{Y}_m \), i.e.,
\[
f_{\tilde{n}}(\tilde{y}) = \frac{1}{m!} \prod_{j=1}^m f(y_j) I(c < y_1 < \ldots < y_m < d), \quad f_{\tilde{n}}(\tilde{y}) \text{ be the joint distribution of } Y_{(\tilde{n})}. \]

Then we have the following theorem which is a generalization of the main result of Priebe et al. (2001).

Theorem 6.1. Let \( D \) be an \( \mathcal{H}(\mathbb{R}) \)-random \( \mathcal{D}_{n,m} \)-digraph. Then the probability mass function of the domination number of \( D \) is given by
\[
P(\gamma(D_{n,m}) = k) = \int_{\mathcal{H}} \sum_{\tilde{n} \in \Theta_{n,(m+1)^\nu_2}} \sum_{\tilde{k} \in \Theta_{k,(m+1)^\nu_1}} P(\tilde{N} = \tilde{n}) \tilde{\zeta}(k_1, \ldots, k_{m+1}) \prod_{j=2}^m \eta(k_j, n_j) f_{\tilde{n}}(\tilde{y}) \, dy_1 \ldots dy_m,
\]
where $P(\bar{N} = \vec{n})$ is the joint probability of $n_j$ points falling into intervals $I_j$ for $j = 1, 2, \ldots, (m + 1)$, and $k_j \in \{0, 1, 2\}$, and

$$
\zeta(k_j, n_j) = \max (I(n_j = k_j = 0), I(n_j \geq k_j = 1)) \text{ for } j = 1, (m + 1), \text{ and }
\eta(k_j, n_j) = \max (I(n_j = k_j = 0), I(n_j \geq k_j \geq 1)) \cdot p_{n_j}(F_j)^{I(k_j=2)} \cdot (1 - p_{n_j}(F_j))^{I(k_j=1)}
$$

for $j = 2, \ldots, m$, and the region of integration is given by

$$
\mathcal{J} := \{(y_1, y_2, \ldots, y_m) \in (c, d)^2 : c < y_1 < y_2 < \ldots < y_m < d\}.
$$

**Proof:** For $\gamma(D_{n,m}) = \sum_{j=1}^{m+1} \gamma(D^j) = k$, we must have $\gamma(D^j) = k_j$ for $j = 1, \ldots, (m + 1)$ so that $\sum_{j=1}^{m+1} k_j = k$ and $\sum_{j=1}^{m+1} n_j = n$. By definition, $\Theta^{[n+1]}_{n,(m+1)}$ is the collection of such $\vec{n}$ and since $k_j \in \{0, 1, 2\}$ for all $j = 1, \ldots, (m + 1)$, $\Theta^{[3]}_{k,(m+1)}$ is the collection of such $\vec{k}$. We treat the end intervals, $I_1$ and $I_{m+1}$, separately. The indicator functions in the statement of the theorem guarantees that the pairs $n_j, k_j$ are compatible for $j \in \{1, (m + 1)\}$; that is, incompatible pairs such as $(n_j = 0, k_j > 0)$ are eliminated. The $\zeta$ terms equal unity if $(n_j, k_j)$ are compatible. Therefore we have

$$
P(\gamma(D_{n,m}) = k) = \int_{\mathcal{J}} \sum_{\vec{n} \in \Theta^{[n+1]}_{n,(m+1)}} \sum_{\vec{k} \in \Theta^{[3]}_{k,(m+1)}} P(\bar{N} = \vec{n}) \prod_{j=1}^{m+1} \eta(k_j, n_j) f_\mathcal{J}(\vec{y}) \, dy_1 \ldots dy_m
$$

where we have used the conditional pairwise independence of $\gamma(D^j)$. The $\eta$ terms are based on the compatibility of pairs $(n_j, k_j)$ for $j = 1, \ldots, (m + 1)$ and $p_{n_j}(F_j) = P(\gamma(D^j) = 2)$.

For $n, m < \infty$, the expected value of domination number is

$$
E[\gamma(D_{n,m})] = P(X_{(1)} < Y_{(1)}) + P(X_{(m)} > Y_{(m)}) + \sum_{j=2}^{m} \sum_{k=1}^{n} P(N_j = k) E[\gamma(D^j)]
$$

where

$$
P(N_j = k) = \int_{c}^{d} \int_{y_{(j-1)}}^{d} f_{j-1,j}(y_{(j-1)}, y_{(j)}) \left[F_X(y_{(j)}) - F_X(y_{(j-1)})\right]^{k} \left[1 - (F_X(y_{(j)}) - F_X(y_{(j-1)})\right]^{n-k} \, dy_{(j)} \, dy_{(j-1)}
$$

and $E[\gamma(D^j)] = 1 + p_k(F_j)$.

**Corollary 6.2.** For $F_X, Y \in \mathcal{H}(\mathbb{R})$ with support $S(F_X) \cap S(F_Y)$ of positive measure, $\lim_{n \to \infty} E[\gamma(D_{n,n})] = \infty$.

**Proof:** Consider the intersection of the supports $S(F_X) \cap S(F_Y)$ that has positive (Lebesgue) measure. For $S(Y) \setminus S(X)$; i.e., in the intervals $I_j$ falling outside the intersection $S(F_X) \cap S(F_Y)$, the domination
number of the component $D^j$ is $\gamma(D^j) = 0$ w.p. 1 but inside the intersection, $\gamma(D^j) > 0$ w.p. 1 for infinitely many $j$. That is,

$$
E[\gamma(D_{n,m})] = P(X_{(1)} < Y_{(1)}) + P(X_{(n)} > Y_{(n)}) + \sum_{j=2}^{n} \sum_{k=1}^{n} P(N_j = k) E[\gamma_{N_j}(F_j)]
$$

$$
> \sum_{j=2}^{n} \sum_{k=1}^{n} P(N_j = k) E[\gamma_{N_j}(F_j)] = \sum_{j=2}^{n} \sum_{k=1}^{n} P(N_j = k) (1 + p_{N_j}(F_j))
$$

$$
> \sum_{j=2}^{n} \sum_{k=1}^{n} P(N_j = k) > \sum_{j=2}^{n} P(N_j \geq 1)
$$

$$
\approx n \quad \text{(for sufficiently large } n) \nonumber
$$

where $E[\gamma_{N_j}(F_j)] = (1 + p_{N_j}(F_j))$ follows from the fact that $\gamma_{N_j}(F_j) \sim 1 + \text{Bernoulli}(p_{N_j}(F_j))$ from Theorem 6.3. Furthermore, $P(N_j \geq 1) \approx 1$ for sufficiently large $n$. Then the desired result follows. \[\square\]

**Theorem 6.3.** Let $D_{n,m}$ be an $\mathcal{H}(\mathbb{R})$-random $\mathcal{B}_{n,m}$-digraph. Then (i) for fixed $n < \infty$, $\lim_{n \to \infty} \gamma(D_{n,m}) = n$ a.s. (ii) for fixed $m < \infty$, $\lim_{n \to \infty} \gamma(D_{n,m}) \equiv m + 1 + \sum_{j=1}^{m} B_j$, where $B_j \sim \text{Bernoulli}(p_{F_j})$ where $d$ stands for equality in distribution.

**Proof:** Part (i) is trivial. As for part (ii), first note that $N_j \to \infty$ as $n \to \infty$ for all $j$ a.s., hence $\lim_{n \to \infty} \gamma(D^1) = \lim_{n \to \infty} \gamma(D^{m+1}) = 1$ a.s. and $\lim_{n \to \infty} \gamma(D^j) = 1 + \text{Bernoulli}(p_{F_j})$ a.s. for $j = 2, \ldots, m$

where

$$
p_{F_j} = \int_{c}^{d} \int_{y_{(j-1)}}^{d} H^*(y_{(j-1)}, y_{(j)}) f_{j-1,j}(y_{(j-1)}, y_{(j)}) dy_{(j)} dy_{(j-1)}
$$

with $H^*(y_{(j-1)}, y_{(j)}) = \lim_{n \to \infty}(p_{N_j}(F_j))$ which is given in Theorem 6.1 for $F_j$ with density $f_j$ whose support is $(y_{(j-1)}, y_{(j)})$. Then the desired result follows. \[\square\]

So far, $Y_m$ is assumed to be a random sample, so $P(\gamma(D_{n,m}) = k)$ includes the integration with respect to $f_{\hat{\gamma}}(\hat{y})$ which can be lifted by conditioning. Conditional on $Y_m = \{y_{(1)}, \ldots, y_{(m)}\}$, by Theorem 6.1 we have

$$
P(\gamma(D_{n,m}) = k) = \sum_{\pi \in \Theta_{[n+1]}^{[m+1]}} \sum_{\pi \in \Theta_{[3]}^{[m+1]}} P(\hat{N} = \pi) \zeta(k_1, n_1) \zeta(k_{m+1}, n_{m+1}) \prod_{j=2}^{m} \eta(k_j, n_j),
$$

where $\zeta(k_j, n_j)$ and $\eta(k_j, n_j)$ are as in Theorem 6.1 and the expected domination number $E[\gamma(D_{n,m})]$ is as in Equation 6.3 with $P(N_j = k) = [F_X(y_{(j)}) - F_X(y_{(j-1)})]^k [1 - (F_X(y_{(j)}) - F_X(y_{(j-1)}))]^{n-k}$; and $\lim_{n \to \infty} \gamma(D_{n,m}) \equiv m + 1 + \sum_{j=1}^{m} B_j$, where $B_j \sim \text{Bernoulli}(p_{F_j})$ with $p_{F_j} := \lim_{n \to \infty} p_{N_j}(F_j)$.

Let $F_X$ be a distribution with support $S(F_X) \subseteq (0, 1)$ and density $f_X(x)$. Conditional on $Y_m = \{y_{(1)}, \ldots, y_{(m)}\}$, let $F_j$ be the distribution with density $f_j(x) = \frac{1}{(y_{(j)} - y_{(j-1)})} f_X \left( \frac{x - y_{(j-1)}}{y_{(j)} - y_{(j-1)}} \right)$ for $j = 2, \ldots, m$, and $S(F_j(x))$ is non-empty for $j \in \{1, (m+1)\}$. By this construction, the independence of the distribution of $\gamma_n(F_j)$ from $I_j$ obtains; i.e., $\gamma_n(F_j) \equiv \gamma_n(F_X)$ for all $j \in \{1, \ldots, (m+1)\}$. Now consider the family $\mathcal{H}(\mathbb{R})$ defined as

$$
\mathcal{H}(\mathbb{R}) := \{F_X, Y_j : (X_i, Y_i) \sim F_{X,Y}, Y_j \overset{iid}{\sim} U(c, d) \text{ for } (c, d) \subseteq \mathbb{R}, \text{ and } X_i | Y_m \overset{iid}{\sim} F_j\}.
$$

Clearly $\mathcal{H}(\mathbb{R}) \subseteq \mathcal{H}(\mathbb{R})$. 

17
Corollary 6.4. Suppose $F_{X,Y} \in \mathcal{H}_{\mathbb{U}}(\mathbb{R})$. Then

$$P(\gamma(D_{n,m}) = k) = \sum_{\vec{n} \in \Theta_{n,(m+1)}^{n+1}} \sum_{k \in \Theta_{k,(m+1)}^{k+1}} P(\vec{N} = \vec{n}) \zeta(k_1, n_1) \zeta(k_{m+1}, n_{m+1}) \prod_{j=2}^{m} \eta(k_j, n_j)$$

where $\zeta(k_j, n_j)$ and $\eta(k_j, n_j)$ are as in Theorem 6.4.

Note that if in addition, $P_{F_j}(X \in I_j) = P_{\mathcal{U}}(X \in I_j)$ for all $j$, then $P(\vec{N} = \vec{n}) = \binom{n+m}{n}^{-1}$, since each $\vec{n} \in \Theta_{n,(m+1)}^{n+1}$ occurs with probability $\binom{n+m}{n}^{-1}$. Moreover, $F = \mathcal{U}(c, d)$ is a special case of Corollary 6.4. For $n, m < \infty$, we have the explicit form of $p_{n_j}(F_j)$ for $F_j$ with piecewise constant density $f_j$.

Here are some examples which are generalized from piecewise-constant densities so that now the distribution of $\gamma(D^j)$ is independent from the support $(y_{j-1}, y_j)$. Hence Corollary 6.4 applies to these examples.

Example 6.5. Let $u_j := \frac{(y_{j-1}+y_j)}{2}$ and $v_j := y_j - y_{j-1}$.

- If $f(\cdot)$ is of the form
  $$f(x) = \frac{1}{(1 - 2 \delta) v_j} \mathbf{I}(x \in (y_{j-1} + \delta v_j, y_j - \delta v_j)) \quad \text{with } \delta \in [0, 1/3]$$
  then $p_n(F)$ is as in Equation (7).

- If $f(\cdot)$ is of the form
  $$f(x) = \frac{1}{(1 - 2 \delta) v_j} \mathbf{I}(x \in (y_{j-1}, u_j - \delta v_j) \cup [u_j + \delta v_j, y_j)) \quad \text{with } \delta \in [0, 1/3],$$
  then $p_n(F)$ is as in Equation (8).

- If $f(\cdot)$ is of the form
  $$f(x) = \frac{(1 + \delta)}{v_j} \mathbf{I}(x \in (y_{j-1}, u_j)) + \frac{(1 - \delta)}{v_j} \mathbf{I}(x \in [u_j, y_j]),$$
  then $p_n(F)$ is as in Equation (9).

- If $f(\cdot)$ is of the form
  $$f(x) = f_1(x) \mathbf{I}(x \in (y_{j-1}, t_j)) + f_2(x) \mathbf{I}(x \in [t_j, w_j]) + f_3(x) \mathbf{I}(x \in [w_j, y_j])$$
  where $t_j = \frac{y_{j-1} y_{j-1}}{4}, w_j = \frac{3y_{j-1}}{4}, f_1(x) = \frac{(1+\delta)}{v_j}, f_2(x) = \frac{(1-\delta)}{v_j}$ and $f_3(x) = \frac{1+\delta}{v_j}$, then $p_n(F)$ is as in Example 6.9.

Theorem 6.6. Let $D$ be an $\mathcal{H}_{\mathcal{U}}(\mathbb{R})$-random $\mathcal{D}_{n,m}$-digraph with the additional assumption that $P_{F_j}(X \in I_j) = P_{\mathcal{U}}(X \in I_j)$ for all $j$. Then

$$E[\gamma(D_{n,m})] = \frac{2n}{n+m} + \frac{n! m (m-1)!}{(n+m)!} \sum_{i=1}^{n} \frac{(n+m-i-1)!}{(n-i)!} (1 + p_i(F))$$

where $p_i(F) = P(\gamma(D_{i,2}) = 2)$. 

As another generalization direction, we also devise proximity maps depending on \( \gamma \) distribution identical to that of Proposition 5.8. Furthermore, this article will form the foundation of the generalizations and calculations have a regional relationship to determine the inclusion of a point in the proximity region.

\[ \gamma \]

In our calculations, the domination number of a CCCD, \( Y \) associated CCCD, \( \gamma \) distribution the domination number of a CCCD, \( Y \) given \( F \) and \( Y \) numerical integration. However, the asymptotic distribution of \( Y \) provided the exact (finite sample) distribution of the class cover catch digraphs (CCCDs) based on \( X \) and \( Y \) both of which were sets of iid random variables from a uniform distribution on \((-\infty, c, d, \infty)\) and the proximity map \( N_Y(x) \cap N_Y(y) = \emptyset \) for \( x, y \) in distinct intervals. The straightforward extension to multiple dimensions (i.e., \( \mathbb{R}^d \) with \( d > 1 \)) does not have a nice ordering structure; and \( Y \) does not readily partition the support, but we can use the Delaunay tessellation based on \( Y \). Furthermore, in multiple dimensions finding a minimum dominating set is an NP-hard problem; and \( \Gamma \)-regions are not readily available (in fact for \( n_j \geq 3 \), complexity of finding the \( \Gamma \)-regions is an open problem). In addition, in multiple dimensions the components of the digraph restricted to Delaunay cells are not necessarily disconnected from each other, since \( N_Y(x) \cap N_Y(y) \neq \emptyset \) might hold for \( x, y \) in distinct Delaunay cells. These have motivated us to generalize the proximity map \( Y \) in order to avoid the difficulties above. See Ceyhan and Priebe (2003, 2005), where two new families of proximity maps are introduced, and the generalization of CCCD are called proximity catch digraphs. The distribution of the domination number of these proximity maps is still a topic of ongoing research. □

7 Discussion

This article generalizes the main result of [Priebe et al. (2001)] in several directions. [Priebe et al. (2001)] provided the exact (finite sample) distribution of the class cover catch digraphs (CCCDs) based on \( X \) and \( Y \) both of which were sets of iid random variables from a uniform distribution on \((c, d) \subset \mathbb{R} \) with \(-\infty < c < d < \infty \) and the proximity map \( N_Y(x) := B(x, r(x)) \) where \( r(x) := \min_{y \in Y} d(x, y) \). First, given \( Y_2 = \{y_1, y_2\} \subset \mathbb{R} \), we lift the uniformity assumption of \( X \) by assuming it to be from a non-uniform distribution \( F \) with support \( \mathcal{S}(F) \subseteq (y_1, y_2) \). The exact distribution of the domination number of the associated CCCD, \( \gamma_n(F) \), is calculated for \( F \) that has piecewise constant density \( f \) on \((y_1, y_2)\). For more general \( F \), the exact distribution is not analytically available in simple closed form, so we compute it by numerical integration. However, the asymptotic distribution of \( \gamma_n(F) \) is tractable, which is the one of the main results of this article. Unfortunately, the distribution of \( \gamma_n(F) \) depends on \( Y_2 \), hence the distribution of the domination number of a CCCD, \( \gamma(D_{n,m}) \), for \( X \) and \( Y \) with \( m > 2 \), for general \( F \) includes integration with respect to order statistics of \( Y \). We provide the conditions that make \( \gamma(D_{n,m}) \) independent of \( Y \). As another generalization direction, we also devise proximity maps depending on \( F \) that will yield the distribution identical to that of \( \gamma_n(U(y_1, y_2)) \). Our set-up is more general than the one given in [Priebe et al. (2001)]. The definition of the proximity map is generalized to any probability space and is only assumed to have a regional relationship to determine the inclusion of a point in the proximity region.

The exact (finite sample) distribution of \( \gamma_n(F) \) characterizes \( F \) up to a special type of symmetry (see Proposition 5.8). Furthermore, this article will form the foundation of the generalizations and calculations
for uniform and non-uniform cases in multiple dimensions. As in [Ceyhan and Priebe (2005)], we can use the domination number in testing one-dimensional spatial point patterns and our results will help make the power comparisons possible for large families of distributions.

Acknowledgments

I would like to thank the anonymous referees, whose constructive comments and suggestions greatly improved the presentation and flow of this article.

References

Ceyhan, E. (2004). 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, The Johns Hopkins University, Baltimore, MD, 21218.

Ceyhan, E. and Priebe, C. (2003). Central similarity proximity maps in Delaunay tessellations. In Proceedings of the Joint Statistical Meeting, Statistical Computing Section, American Statistical Association.

Ceyhan, E. and Priebe, C. E. (2005). The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association. Statistics and Probability Letters, 73:37–50.

Chartrand, G. and Lesniak, L. (1996). Graphs & Digraphs. Chapman & Hall/CRC Press LLC, Florida.

DeVinney, J., Priebe, C. E., Marchette, D. J., and Socolinsky, D. (2002). Random walks and catch digraphs in classification. http://www.galaxy.gmu.edu/interface/102/12002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf

DeVinney, J. and Wierman, J. C. (2003). A SLLN for a one-dimensional class cover problem. Statistics and Probability Letters, 59(4):425–435.

Janson, S., Luczak, T., and Rucinski, A. (2000). Random Graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York.

Jaromczyk, J. W. and Toussaint, G. T. (1992). Relative neighborhood graphs and their relatives. Proceedings of IEEE, 80:1502–1517.

Marchette, D. J. and Priebe, C. E. (2003). Characterizing the scale dimension of a high dimensional classification problem. Pattern Recognition, 36(1):45–60.

Priebe, C. E., DeVinney, J. G., and Marchette, D. J. (2001). On the distribution of the domination number of random class catch cover digraphs. Statistics and Probability Letters, 55:239–246.

Priebe, C. E., Marchette, D. J., DeVinney, J., and Socolinsky, D. (2003a). Classification using class cover catch digraphs. Journal of Classification, 20(1):3–23.

Priebe, C. E., Solka, J. L., Marchette, D. J., and Clark, B. T. (2003b). 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.

Prisner, E. (1994). Algorithms for interval catch digraphs. Discrete Applied Mathematics, 51:147–157.
Sen, M., Das, S., Roy, A., and West, D. (1989). Interval digraphs: An analogue of interval graphs. Journal of Graph Theory, 13:189–202.

Tuza, Z. (1994). Inequalities for minimal covering sets in sets in set systems of given rank. Discrete Applied Mathematics, 51:187–195.

West, D. B. (2001). Introduction to Graph Theory, 2nd Ed. Prentice Hall, N.J.