Available online at www.sciencedirect.com
ScienceDirect
Discrete Mathematics 308 (2008) 5376–5393
DISCRETE MATHEMATICS
www.elsevier.com/locate/disc

The distribution of the domination number of class cover catch digraphs for non-uniform one-dimensional data

Elvan Ceyhan*

Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218, United States

Received 27 December 2004; received in revised form 3 October 2007; accepted 3 October 2007
Available online 20 February 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 one or all of the other classes. The CCCDs were introduced by Priebe et al. [C.E. Priebe, J.G. DeVinney, D.J. Marchette, On the distribution of the domination number of random class catch cover digraphs. Statistics and Probability Letters 55 (2001) 239–246] who investigated the case of two classes, $\mathcal{X}$ and $\mathcal{Y}$. They calculated the exact (i.e., 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

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. [10] 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 [6] proved a SLLN result for the one-dimensional class cover problem. DeVinney et al. [5], Marchette and Priebe [9], and Priebe et al. [11,12] 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 sets as *prototype sets*, since finding the exact minimum dominating set for CCCDs is an NP-hard problem in general. However for finding a dominating set of CCCDs on the real line, a simple linear time algorithm is available [10]. But unfortunately, the exact and the asymptotic distributions of the domination number of the CCCDs are not analytically tractable in multiple dimensions.

* Tel.: +1 410 516 4058; fax: +1 410 516 7459.
E-mail addresses: elvan@cis.jhu.edu, elceyhan@ku.edu.tr.

0012-365X/$ - see front matter
doi:10.1016/j.disc.2007.10.003

To address the latter issue of intractability of the distribution of the domination number in multiple dimensions, Ceyhan and Priebe [2,3] 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 r-factor proportional-edge proximity catch digraphs is calculated and then used in testing spatial patterns between two or more classes. See [3] for more detail.

In this article, we generalize the original result of Priebe et al. [10] 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 extensions 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 $\mathcal{X}_n = \{X_1, \dots, X_n\}$ and $\mathcal{Y}_m = \{Y_1, \dots, Y_m\}$ be two sets of Ω-valued random variables from classes $\mathcal{X}$ and $\mathcal{Y}$, respectively, with joint probability distribution function $F_{X,Y}$. Let $d(\cdot,\cdot) : \Omega \times \Omega \to [0, \infty)$ 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 [10] for more detail.

Consider the map $N_{\mathcal{Y}} : \Omega \to 2^{\Omega}$: where 2^{Ω} represents the power set of Ω. Then given $\mathcal{Y}_m \subseteq \Omega$, the *proximity map* $N_{\mathcal{Y}}(\cdot) : \Omega \to 2^{\Omega}$ associates with each point $x \in \Omega$ a *proximity region* $N_{\mathcal{Y}}(x) \subseteq \Omega$. For $B \subseteq \Omega$, the Γ_1-region is the image of the map $\Gamma_1(\cdot, N_{\mathcal{Y}}) : 2^{\Omega} \to 2^{\Omega}$ 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, Γ_1-regions can be 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 [10]. We call such a digraph a $\mathscr{D}_{n,m}$-digraph. A $\mathscr{D}_{n,m}$-digraph is a *pseudo-digraph* according some authors if loops are allowed (see, e.g., [4]). A data-random CCCD for $\Omega = \mathbb{R}^d$ and $N_i = B(X_i, r_i)$ is referred to as $\mathscr{C}_{n,m}$-graph in [10]. We change the notation to emphasize the fact that $\mathscr{D}_{n,m}$ is a digraph. Furthermore, Ceyhan and Priebe [2] call the proximity map $N_i = B(X_i, r_i)$ a *spherical proximity map*.

The $\mathscr{D}_{n,m}$-digraphs are closely related to the *proximity graphs* of Jaromczyk and Toussaint [8] and might be considered as a special case of *covering sets* of Tuza [15] and *intersection digraphs* of Sen et al. [14]. Our data-random proximity digraph is a *vertex-random proximity digraph* and not a standard one (see e.g., [7]). The randomness of a $\mathscr{D}_{n,m}$-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_i and the random set N_i.

3. Domination number of random $\mathscr{D}_{n,m}$-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 [16]. 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 $\gamma(D) \le n$.

Let $\mathcal{F}\left(\mathbb{R}^d\right) := \{F_{X,Y} \text{ on } \mathbb{R}^d \text{ with } P(X = Y) = 0\}$. As in [10], in this article, we consider $\mathscr{D}_{n,m}$-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 $F_{X,Y} \in \mathcal{F}\left(\mathbb{R}^d\right)$ where $\mathcal{F}(\mathbb{R}^d) := \{F_{X,Y} \text{ on } \mathbb{R}^d \text{ with } P(X_i = Y_j) = 0 \text{ for all } i, j;\ P(X_i = X_j) = 0 \text{ for } i \ne j \text{ and } P(Y_k = Y_l) = 0 \text{ for } k \ne l\}$. We call such digraphs as $\mathcal{F}\left(\mathbb{R}^d\right)$*-random* $\mathscr{D}_{n,m}$*-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 $\gamma(D_{n,m})$ is not defined for $m = 0$; $\gamma(D_{n,m}) = 0$ for $n = 0$ and $m \ge 1$; $1 \le \gamma(D_{n,m}) \le n$ for $n \ge 1$ and $m \ge 1$; and for nontrivial digraphs $\gamma(D_{n,m}) < n$.

4. The Distribution of the domination number of $\mathcal{F}(\mathbb{R})$-random $\mathscr{D}_{n,m}$-digraphs

In $\mathbb{R}$, the data-random CCCD is a special case of *interval catch digraphs* (see, e.g., [14,13]). Let $\mathcal{X}_n$ and $\mathcal{Y}_m$ be two samples from $\mathcal{F}(\mathbb{R})$ and $Y_{(j)}$ be the jth order statistic of $\mathcal{Y}_m$ for $j = 1, 2, \ldots, m$. Then $Y_{(j)}$ partition $\mathbb{R}$ into $(m+1)$ intervals a.s. Let $-\infty =: Y_{(0)} \le Y_{(1)} \le \cdots \le Y_{(m)} \le Y_{(m+1)} := \infty$. But since $\mathcal{Y}_m$ is from $\mathcal{F}(\mathbb{R})$, $Y_{(i)} < Y_{(j)}$ for $i < j$ a.s. Also let $\mathcal{I}_j := (Y_{(j-1)}, Y_{(j)})$, $\mathcal{X}^j := \mathcal{X}_n \cap \mathcal{I}_j$, and $\mathcal{Y}^j := \{Y_{(j-1)}, Y_{(j)}\}$ for $j = 1, 2, \ldots, (m+1)$. This yields a disconnected digraph with subdigraphs D^j for $j = 1, 2, \ldots, (m+1)$, each of which is induced by the vertices $\mathcal{X}^j$ whose support is in $\mathcal{Y}^j$. Notice that each subdigraph D^j might be null or itself disconnected. Let $\gamma(D^j)$ denote the domination number of the subdigraph D^j, $n_j := |\mathcal{X}^j|$, and F_j be the density F_X restricted to $\mathcal{I}_j$. Then $\gamma(D_{n,m}) = \sum_{j=1}^{m+1} \gamma(D^j)$. We study the simpler random variable $\gamma(D^j)$ first. The following lemma follows trivially (see [10]).

Lemma 4.1. *For $j \in \{1, (m+1)\}$, $\gamma(D^j) = \mathbf{I}(n_j > 0)$ where $\mathbf{I}(\cdot)$ 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, $\mathcal{X}^j \subset B(x, r(x))$ iff $x \in \left(\frac{\max(\mathcal{X}^j)+Y_{(j-1)}}{2}, \frac{\min(\mathcal{X}^j)+Y_{(j)}}{2}\right)$ where $r(x) = \min(x - Y_{(j-1)}, Y_{(j)} - x)$. Hence $\Gamma_1(\mathcal{X}^j, N_{\mathcal{Y}}) = \left(\frac{\max(\mathcal{X}^j)+Y_{(j-1)}}{2}, \frac{\min(\mathcal{X}^j)+Y_{(j)}}{2}\right) \subseteq \mathcal{I}_j$. By definition, if $\mathcal{X}^j \cap \Gamma_1(\mathcal{X}^j, N_{\mathcal{Y}}) \neq \emptyset$, then $\gamma(D^j) = 1$; hence the name Γ_1-*region* and the notation $\Gamma_1(\cdot, N_{\mathcal{Y}})$.

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

Proof. See [10] for the proof. ■

The probability $P(\gamma(D^j) = 2) = P(\mathcal{X}^j \cap \Gamma_1(\mathcal{X}^j, N_{\mathcal{Y}}) = \emptyset)$ depends on the conditional distribution $F_{X|Y}$ and the interval $\Gamma_1(\mathcal{X}^j, N_{\mathcal{Y}})$, which, if known, will make possible the calculation of $p_{n_j}(F_j)$. As an immediate result of Lemma 4.1 and Theorem 4.2, 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 $\mathscr{D}_{n,m}$-digraph with $n > 0$, $m > 0$ and k_1 and k_2 be two natural numbers defined as $k_1 := \sum_{j=2}^{m} \mathbf{I}(|\mathcal{X}_n \cap \mathcal{I}_j| > 1)$ and $k_2 := \sum_{j=2}^{m} \mathbf{I}(|\mathcal{X}_n \cap \mathcal{I}_j| = 1) + \sum_{j \in \{1,(m+1)\}} \mathbf{I}(\mathcal{X}_n \cap \mathcal{I}_j \neq \emptyset)$. Then $1 \le \gamma(D_{n,m}) \le 2\,k_1 + k_2 \le \min(n, 2\,m)$.*

In the special case of fixed $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\}$ and $\mathcal{X}_n$ a random sample from $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$, the uniform distribution on $(\mathsf{y}_1, \mathsf{y}_2)$, we have a $\mathscr{D}_{n,2}$-digraph for which $F_X = \mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$ and F_Y is a degenerate distribution. We call such digraphs as $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$-*random* $\mathscr{D}_{n,2}$-*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 $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$-random $\mathscr{D}_{n,2}$-digraphs

Suppose $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\} \subset \mathbb{R}$ with $-\infty < \mathsf{y}_1 < \mathsf{y}_2 < \infty$ and $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ a set of iid random variables from $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$. Any $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$ random variable can be transformed into a $\mathcal{U}(0, 1)$ random variable by $\phi(x) = (x - \mathsf{y}_1)/(\mathsf{y}_2 - \mathsf{y}_1)$, which maps intervals $(t_1, t_2) \subseteq (\mathsf{y}_1, \mathsf{y}_2)$ to intervals $(\phi(t_1), \phi(t_2)) \subseteq (0, 1)$ and $\phi(X_i) \stackrel{\text{iid}}{\sim} \mathcal{U}(0, 1)$. 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 $(\mathsf{y}_1, \mathsf{y}_2)$. Recall that $\gamma(D_{n,2}) = 2$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) = \emptyset$, then $P(\gamma(D_{n,2}) = 2) = 4/9 - (16/9)\,4^{-n}$. For more detail, see [10]. Hence, for $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$ data, we have

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

where w.p. stands for "with probability". Then the asymptotic distribution of $\gamma(D_{n,2})$ for $\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$ data is given by

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

For $m > 2$, Priebe et al. [10] 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 $X_i \stackrel{iid}{\sim} F$ with support $\mathcal{S}(F) \subseteq (\mathsf{y}_1, \mathsf{y}_2)$, the exact and asymptotic distribution of $\gamma(D_{n,2})$ will depend on F and $\mathcal{Y}_2$.

4.2. The distribution of the domination number for $\mathcal{F}(\mathbb{R})$-random $\mathscr{D}_{n,2}$-digraphs

For $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\} \subset \mathbb{R}$ with $-\infty < \mathsf{y}_1 < \mathsf{y}_2 < \infty$, a quick investigation shows that the Γ_1-region is $\Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) = \left(\frac{\mathsf{y}_1+X_{(n)}}{2}, \frac{\mathsf{y}_2+X_{(1)}}{2}\right)$. Note that $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}})$ 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 $F\,((\mathsf{y}_1, \mathsf{y}_2))$-random $\mathscr{D}_{n,2}$-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 $1 +$ Bernoulli $(p_n(F))$ and $1 +$ Bernoulli $(p(F))$, respectively. That is, for finite n, we have

$$\gamma_n(F) = \begin{cases} 1 & \text{w.p. } 1 - p_n(F) \\ 2 & \text{w.p. } p_n(F) \end{cases} \qquad \text{for all } n \geq 1. \tag{3}$$

The asymptotic distribution is similar.

With $\mathcal{Y}_2 = \{0, 1\}$, let F be a distribution with support $\mathcal{S}(F) \subseteq (0, 1)$ 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

$$p_n(F) = \int_{\mathcal{S}(F)\setminus\Gamma_1(\mathcal{X}_n,N_{\mathcal{Y}})} \left[1 - \frac{F((1+x_1)/2) - F(x_n/2)}{F(x_n) - F(x_1)}\right]^{n-2} f_{1n}(x_1, x_n)\mathrm{d}x_n\mathrm{d}x_1, \tag{4}$$

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

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

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

The integrand in Eq. (4) simplifies to

$$H(x_1, x_n) := n\,(n-1) f(x_1) f(x_n)\,[F(x_n) + F\,(x_n/2) - (F\,((1 + x_1)/2) + F(x_1))]^{n-2}\,. \tag{5}$$

Let $\mathcal{X}_n$ be a set of iid random variables from a continuous distribution F with $\mathcal{S}(F) \subseteq (0, 1)$. The simplest of such distributions is $\mathcal{U}(0, 1)$, 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, $F(X) \sim \mathcal{U}(0, 1)$. 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 $\gamma_n(\mathcal{U}(0, 1))$.

Proposition 4.4. *Let $X_i \stackrel{iid}{\sim} F$ which is an (absolutely) continuous distribution with support $\mathcal{S}(F) = (0, 1)$ 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 $\gamma_n(\mathcal{U}(0, 1))$.*

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, $U_i \stackrel{iid}{\sim} \mathcal{U}(0, 1)$. Let $U_{(k)}$ be the kth 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 $i = 1, \ldots, n$. Since $U_i \stackrel{iid}{\sim} \mathcal{U}(0, 1)$, the distribution of the domination number of the digraph based on $N_{\mathcal{Y}}$, $\mathcal{U}_n$ and $\{0, 1\}$ is given in Eq. (1). Observe that $X_j \in N_F(X_i)$ iff $X_j \in F^{-1}(B(F(X_i), r(F(X_i))))$ iff $F(X_j) \in B(F(X_i), r(F(X_i)))$ iff $U_j \in B(U_i, r(U_i))$ for $i, j = 1, \ldots, n$. Hence $P(\mathcal{X}_n \subset N_F(X_i)) = P(\mathcal{U}_n \subset N_{\mathcal{Y}}(U_i))$ for all $i = 1, \ldots, n$. Therefore, $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_F) = \emptyset$

iff $\mathcal{U}_n \cap \Gamma_1(\mathcal{U}_n, N_{\mathcal{Y}}) = \emptyset$, which implies that the domination number of the digraph based on N_F, $\mathcal{X}_n$, and $\mathcal{Y}_2 = \{0, 1\}$ is 2 with probability $4/9 - (16/9)\,4^{-n}$. Hence the desired result follows. ■

For example for $F(x) = x^2\,\mathbf{I}(0 \le x \le 1) + \mathbf{I}(x > 1)$,

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

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

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

$$F(X_{(n)}/2) < F(X_{(n)})/2 \quad \text{and} \quad F(X_{(1)}) < 2\,F((1 + X_{(1)})/2) - 1 \text{ hold a.s.,} \tag{6}$$

then $\gamma_n(F) <^{ST} \gamma_n(\mathcal{U}(0, 1))$. If $<$'s in expression (6) are replaced with $>$'s, then $\gamma_n(F) >^{ST} \gamma_n(\mathcal{U}(0, 1))$. If $<$'s in expression (6) are replaced with $=$'s, then $\gamma_n(F) \stackrel{d}{=} \gamma_n(\mathcal{U}(0, 1))$ where $\stackrel{d}{=}$ stands for equality in distribution.

Proof. Let $U_i := F(X_i)$ for $i = 1, \ldots, n$ and $\mathcal{U}_n = \{U_1, \ldots, U_n\}$. Then, by probability integral transform, $U_i \stackrel{\text{iid}}{\sim} \mathcal{U}(0, 1)$. Let $U_{(j)}$ be the jth order statistic of $\mathcal{U}_n$ for $j = 1, \ldots, n$. The Γ_1-region for $\mathcal{U}_n$ based on $N_{\mathcal{Y}}$ is $\Gamma_1(\mathcal{U}_n, N_{\mathcal{Y}}) = (U_{(n)}/2, (1 + U_{(1)})/2)$; likewise, $\Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) = (X_{(n)}/2, (1 + X_{(1)})/2)$. But the conditions in expression (6) imply that $\Gamma_1(\mathcal{U}_n, N_{\mathcal{Y}}) \subsetneq F(\Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}))$. So $\mathcal{U}_n \cap F(\Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}})) = \emptyset$ implies that $\mathcal{U}_n \cap \Gamma_1(\mathcal{U}_n, N_{\mathcal{Y}}) = \emptyset$ and $\mathcal{U}_n \cap F(\Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}})) = \emptyset$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) = \emptyset$. Hence

$$p_n(F) = P(\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) = \emptyset) < P(\mathcal{U}_n \cap \Gamma_1(\mathcal{U}_n, N_{\mathcal{Y}}) = \emptyset) = p_n(\mathcal{U}(0, 1)).$$

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

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

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 $\mathcal{U}(0, 1)$. Below we give some examples for such densities.

Example 4.6. Consider the distribution F with density $f(\cdot)$ which is of the form $f(x) = \frac{1}{1-2\,\delta}\,\mathbf{I}\,(\delta < x < 1 - \delta)$ with $\delta \in [0, 1/2)$. Then $F(x) = \frac{x-\delta}{1-2\,\delta}\,\mathbf{I}\,(\delta < x < 1 - \delta) + \mathbf{I}\,(x \ge 1 - \delta)$. The integrand in Eq. (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 $\delta \in [0, 1/3]$

$$\begin{aligned} p_n(F) &= \int_{\delta}^{1/3} \int_{(1+x_1)/2}^{1-\delta} H(x_1, x_n)\,\mathrm{d}x_n \mathrm{d}x_1 + \int_{1/3}^{(1-\delta)/2} \int_{2\,x_1}^{1-\delta} H(x_1, x_n)\,\mathrm{d}x_n \mathrm{d}x_1 \\ &= \left(4/9 - (16/9)\,4^{-n}\right) \left(\frac{1 - 3\,\delta}{1 - 2\,\delta}\right)^n, \end{aligned} \tag{7}$$

which for $\delta \in (0, 1/3]$, converges to 0 as $n \to \infty$ at (an exponential) rate $O((\frac{1-3\,\delta}{1-2\,\delta})^n)$. 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 $\mathcal{X}_n \subset N(X_i)$ for each $i = 1, \ldots, n$. Furthermore, if $\delta = 0$, then $F = \mathcal{U}(0, 1)$ which yields $p_n(F) = 4/9 - (16/9)\,4^{-n}$. □

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

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

Then the cumulative distribution function (cdf) is given by

$$\begin{aligned} F(x) = {} & F_1(x)\,\mathbf{I}\,(0 < x < 1/2-\delta) + F_2\,(x)\,\mathbf{I}\,(1/2-\delta < x < 1/2+\delta) \\ & + F_3\,(x)\,\mathbf{I}\,(1/2+\delta < x < 1) + \mathbf{I}\,(x \geq 1)\,, \end{aligned}$$

where

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

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

case(1) $\left(X_{(n)}/2, \left(1+X_{(1)}\right)/2\right) \subseteq (1/2-\delta, 1/2+\delta)$;
case(2) $X_{(n)}/2 < 1/2-\delta < \left(1+X_{(1)}\right)/2 < 1/2+\delta$;
case(3) $1/2-\delta < X_{(n)}/2 < 1/2+\delta < \left(1+X_{(1)}\right)/2$;
case(4) $X_{(n)}/2 < 1/2-\delta < 1/2+\delta < \left(1+X_{(1)}\right)/2$.

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

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

Then $p_n(F) = \sum_{j=1}^{4} P\left(\gamma_n(F) = 2,\ E_j(n)\right)$. 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 $P(\gamma_n(F) = 2, E_1(n)) = 1 - 2\left(\frac{1-4\,\delta}{1-2\,\delta}\right)^n + \left(\frac{1-6\,\delta}{1-2\,\delta}\right)^n$. Note that $P\left(\Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) \subseteq (1/2-\delta, 1/2+\delta)\right) \to 1$ 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 $P(\gamma_n(F) = 2, E_2(n)) = \frac{2}{3}\left(1 - \frac{4}{4^n}\right)\left(\left(\frac{1-4\,\delta}{1-2\,\delta}\right)^n - \left(\frac{1-6\,\delta}{1-2\,\delta}\right)^n\right)$ and in **case (4)**, $P(\gamma_n(F) = 2, E_4(n)) = \frac{4}{9}\left(1 - 4^{-n+1}\right)\left(\frac{1-6\,\delta}{1-2\,\delta}\right)^n$. See [1] for the details of the computations.

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

$$P\left(\gamma_n(F) = 2\right) = 1 + \left(\frac{1-6\,\delta}{1-2\,\delta}\right)^n \left(1/9 + (32/9)4^{-n}\right) - \left(\frac{1-4\,\delta}{1-2\,\delta}\right)^n \left(2/3 + (16/3)4^{-n}\right), \tag{8}$$

which, for $\delta \in (0, 1/6]$, converges to 1 as $n \to \infty$ at rate $O\left(\left(\frac{1-4\,\delta}{1-2\,\delta}\right)^n\right)$.

Notice that if $\delta = 0$, then $F = \mathcal{U}(0,1)$. 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 [1] also for the details of the computations. □

Example 4.8. Consider the distribution F with density $f(\cdot)$ which is of the form $f(x) = (1+\delta)\,\mathbf{I}\,(x \in (0, 1/2)) + (1-\delta)\,\mathbf{I}\,(x \in [1/2, 1))$ with $\delta \in [-1, 1]$.

Then

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

See [1] 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 $O\left(\left(\frac{1+\delta}{4}\right)^n\right)$. Note that $p_F(\delta) \in [0, 4/9]$ is continuous in δ and decreases as $|\delta|$ increases. If $\delta = 0$, then $F = \mathcal{U}(0,1)$ 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

$$f(x) = (1+\delta)\,\mathbf{I}(0 < x < 1/4) + (1-\delta)\,\mathbf{I}(1/4 \leq x < 3/4) + (1+\delta)\,\mathbf{I}(3/4 \leq x < 1) \text{ with } \delta \in [-1, 1].$$

The exact value of $p_n(F)$ is available, but it is rather a lengthy expression (see [1] 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 $O\left(\left(\frac{5-\delta}{8}\right)^n\right)$. So $p_F(\delta)$ is increasing in δ. Notice here that $p_n(F)$ and $p_F(\delta)$ are continuous in δ 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 Eq. (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 $\Lambda_1(n) := \int_0^{1/3}\int_{(1+x_1)/2}^1 H(x_1, x_n)\mathrm{d}x_n\mathrm{d}x_1$, $\Lambda_2(n) := \int_{1/3}^{1/2}\int_{2x_1}^1 H(x_1, x_n)\mathrm{d}x_n\mathrm{d}x_1$, and $H(x_1, x_n) = n(n-1)x_1 x_n \left(5x_n^2 - 1 - 2x_1 - 5x_1^2\right)^{n-2}$. Then

$$\Lambda_1(n) = \int_0^{1/3} (8n x_1/5)(1 - x_1/2 - 5x_1^2/4)^{n-1} - (8n x_1/5)(1/16 + x_1/2 - 15x_1^2/16)^{n-1}\,\mathrm{d}x_1.$$

Using the multinomial expansion of $(\cdot)^{n-1}$ with respect to x_1 in the integral above, we have

$$\Lambda_1(n) = \sum_{Q_2}\binom{n-1}{q_1, q_2, q_3}\frac{8n(-1)^{q_2+q_1}5^{-1+q_1}2^{-q_2-2q_1}3^{-2-q_2-2q_1}}{2+q_2+2q_1} + \frac{n(-1)^{1+q_1}2^{3-3q_2-4q_3-4q_1}15^{q_1}3^{-2-q_2-2q_1}}{5(2+q_2+2q_1)}$$

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

$$\Lambda_2(n) = \int_{1/3}^{1} (8n x_1/5)(1 - x_1/2 - 5x_1^2/4)^{n-1} - (8n x_1/5)(15/x_1^2/4 - 1/4 - x_1/2)^{n-1}\,\mathrm{d}x_1.$$

Again, using the multinomial expansion of the $(\cdot)^{n-1}$ term above, we get

$$\Lambda_2(n) = \sum_{Q_3}\binom{n-1}{r_1, r_2, r_3}[2n(9(-1)^{r_2+r_1}5^{r_1}4^{-2r_1-r_2} + 9(-1)^{1+r_3+r_2}15^{r_1}4^{-2r_1-r_3-r_2} + 4(-1)^{1+r_2+r_1}6^{-r_2-2r_1}5^{r_1} + (-1)^{r_3+r_2}4^{1-r_3}6^{-r_2}12^{-r_1}5^{r_1})]/[90 + 45r_2 + 90r_1]$$

where $Q_3 = \{r_1, r_2, r_3 \in \mathbb{N} : r_1 + r_2 + r_3 = n-1\}$. See [1] 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.

$$p_n(F) = P(\gamma_n(F) = 2) = \int_0^{1/3}\int_{(1+x_1)/2}^1 H(x_1, x_n) + \int_{1/3}^{1/2}\int_{2x_1}^1 H(x_1, x_n)\,\mathrm{d}x_n\mathrm{d}x_1,$$

where $H(x_1, x_n)$ is given in Eq. (5).

Recall the $\mathcal{F}(\mathbb{R}^d)$-random $\mathscr{D}_{n,m}$-digraphs. We call the digraph which obtains in the special case of $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2\}$ and support of F_X in $(\mathsf{y}_1, \mathsf{y}_2)$, *$\mathcal{F}((\mathsf{y}_1, \mathsf{y}_2))$-random $\mathscr{D}_{n,2}$-digraph*. Below, we provide asymptotic results pertaining to the distribution of such digraphs.

5. The asymptotic distribution of the domination number of $\mathcal{F}((\mathsf{y}_1, \mathsf{y}_2))$-random $\mathscr{D}_{n,2}$-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 $\mathscr{D}_{n,2}$-digraphs with $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\} \subset \mathbb{R}$ with $\mathsf{y}_1 < \mathsf{y}_2$ for various F with support $\mathcal{S}(F) \subseteq (\mathsf{y}_1, \mathsf{y}_2)$. Then we will extend this to the case with $\mathcal{Y}_m \subset \mathbb{R}$ for $m > 2$. For $\varepsilon \in (0, (\mathsf{y}_1+\mathsf{y}_2)/2)$, consider the family of distributions given by

$$\mathcal{F}((\mathsf{y}_1, \mathsf{y}_2), \varepsilon) = \{F : (\mathsf{y}_1, \mathsf{y}_1 + \varepsilon) \cup (\mathsf{y}_2 - \varepsilon, \mathsf{y}_2) \cup ((\mathsf{y}_1 + \mathsf{y}_2)/2 - \varepsilon, (\mathsf{y}_1 + \mathsf{y}_2)/2 + \varepsilon) \subseteq \mathcal{S}(F) \subseteq (\mathsf{y}_1, \mathsf{y}_2)\}.$$

Let the kth 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 directional limit at $(c_1, c_2) \in \mathbb{R}^2$ for $g(x, y)$ in the first quadrant in $\mathbb{R}^2$ be $g(c_1^+, c_2^+) := \lim_{\substack{\|\vec{h}\| \to 0 \\ h_1, h_2 > 0}} g(\vec{c} + \vec{h})$ and the directional partial derivatives at (c_1, c_2) along paths in the first quadrant be

$$\frac{\partial^{k+1} g(c_1^+, c_2^+)}{\partial x^{k+1}} := \lim_{\substack{\|\vec{h}\|\to 0 \\ h_1,h_2>0}} \frac{1}{\|h\|} \left(\frac{\partial^k g(\vec{c} + h)}{\partial x^k} - \frac{\partial^k g(\vec{c})}{\partial x^k} \right) \quad \text{for } k \geq 1.$$

Theorem 5.1. *Let $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\} \subset \mathbb{R}$ with $-\infty < \mathsf{y}_1 < \mathsf{y}_2 < \infty$ and $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ with $X_i \overset{iid}{\sim} F \in \mathcal{F}((\mathsf{y}_1, \mathsf{y}_2), \varepsilon)$. Let $D_{n,2}$ be the random $\mathscr{D}_{n,2}$-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, $(\mathsf{y}_1 + \mathsf{y}_2)/2$, $f^{(k)}(\mathsf{y}_1^+) + 2^{-(k+1)} f^{(k)}\left(\left(\frac{\mathsf{y}_1+\mathsf{y}_2}{2}\right)^+\right) \neq 0$ and $f^{(j)}(\mathsf{y}_1^+) = 0$ 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, $(\mathsf{y}_1 + \mathsf{y}_2)/2$, $f^{(\ell)}(\mathsf{y}_2^-) + 2^{-(\ell+1)} f^{(\ell)}\left(\left(\frac{\mathsf{y}_1+\mathsf{y}_2}{2}\right)^-\right) \neq 0$ and $f^{(j)}(\mathsf{y}_2^-) = 0$ for all $j = 0, 1, \ldots, \ell-1$. Then $\gamma_n(F) \sim 1 + \mathrm{Bernoulli}(p_n(F))$ where $p_n(F) := P(\gamma_n(F) = 2)$ and for bounded $f^{(k)}(\cdot)$ and $f^{(\ell)}(\cdot)$, we have the following limit*

$$\lim_{n\to\infty} p_n(F) = \frac{f^{(k)}(\mathsf{y}_1^+)\, f^{(\ell)}(\mathsf{y}_2^-)}{\left[f^{(k)}(\mathsf{y}_1^+) + 2^{-(k+1)} f^{(k)}\left(\left(\frac{\mathsf{y}_1+\mathsf{y}_2}{2}\right)^+\right)\right] \left[f^{(\ell)}(\mathsf{y}_2^-) + 2^{-(\ell+1)} f^{(\ell)}\left(\left(\frac{\mathsf{y}_1+\mathsf{y}_2}{2}\right)^-\right)\right]}.$$

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

Proof. First suppose $(\mathsf{y}_1, \mathsf{y}_2) = (0, 1)$. Recall that $\Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) = \left(X_{(n)}/2, \left(1 + X_{(1)}\right)/2\right) \subset (0, 1)$ and $\gamma_n(F) = 2$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N_{\mathcal{Y}}) = \emptyset$. Then for finite n,

$$p_n(F) = P(\gamma_n(F) = 2) = \int_{\mathcal{S}(F)\setminus\Gamma_1(\mathcal{X}_n,N_{\mathcal{Y}})} H(x_1, x_n)\, \mathrm{d}x_n \mathrm{d}x_1,$$

where $H(x_1, x_n)$ is as in Eq. (5).

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

$$p_n(F) \approx \int_0^{\varepsilon} \int_{1-\varepsilon}^{1} n(n-1) f(x_1) f(x_n) \left[F(x_n) - F(x_1) + F(x_n/2) - F((1+x_1)/2)\right]^{n-2} \mathrm{d}x_n \mathrm{d}x_1. \tag{10}$$

Let

$$G(x_1, x_n) = F(x_n) - F(x_1) + F(x_n/2) - F((1+x_1)/2).$$

The integral in Eq. (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 $G(x_1, x_n)$ becomes

$$G(z_1, z_n) = F(1 - z_n) - F(z_1) + F((1 - z_n)/2) - F((1 + z_1)/2),$$

and Eq. (10) becomes

$$p_n(F) \approx \int_0^{\varepsilon} \int_0^{\varepsilon} n(n-1) f(z_1) f(1 - z_n) \left[G(z_1, z_n)\right]^{n-2} \mathrm{d}z_n \mathrm{d}z_1. \tag{11}$$

The new integral is critical at $(z_1, z_n) = (0, 0)$. Note that $\frac{\partial^{r+s} G(z_1,z_n)}{\partial z_1^r \, \partial z_n^s} = 0$ for all $r, s \geq 1$. Let $\alpha_i := \left.\frac{\partial^{i+1} G(z_1,z_n)}{\partial z_1^{i+1}}\right|_{(0^+,0^+)} = f^{(i)}(0^+) + 2^{-(i+1)} f^{(i)}\left(\frac{1}{2}^+\right)$ and $\beta_j := \left.\frac{\partial^{j+1} G(z_1,z_n)}{\partial z_n^{j+1}}\right|_{(0^+,0^+)} = f^{(j)}(1^-) + 2^{-(j+1)} f^{(j)}\left(\frac{1}{2}^-\right)$. 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 ℓ, 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.

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

$$\begin{aligned}
G(z_1, z_n) \;&=\; G(0^+, 0^+) + \frac{1}{(k+1)!}\left(\frac{\partial^{k+1} G(0^+, 0^+)}{\partial z_1^{k+1}}\right) z_1^{k+1} \\
&\quad + \frac{1}{(\ell+1)!}\left(\frac{\partial^{\ell+1} G(0^+, 0^+)}{\partial z_n^{\ell+1}}\right) z_n^{\ell+1} + O\left(z_1^{k+2}\right) + O\left(z_n^{\ell+2}\right) \\
&=\; 1 - \frac{\alpha_k}{(k+1)!}\, z_1^{k+1} + \frac{(-1)^{\ell+1}\beta_\ell}{(\ell+1)!}\, z_n^{\ell+1} + O\left(z_1^{k+2}\right) + O\left(z_n^{\ell+2}\right).
\end{aligned}$$

Then substituting these expansions in Eq. (11), we obtain

$$\begin{aligned}
p_n(F) \approx \int_0^\varepsilon \int_0^\varepsilon n(n-1) &\left[\frac{1}{k!} f^{(k)}(0^+)\, z_1^k + O\left(z_1^{k+1}\right)\right]\left[\frac{(-1)^\ell}{\ell!} f^{(\ell)}(1^-)\, z_n^\ell + O\left(z_n^{\ell+1}\right)\right] \\
&\times \left[1 - \frac{\alpha_k}{(k+1)!}\, z_1^{k+1} - \frac{(-1)^\ell \beta_\ell}{(\ell+1)!}\, z_n^{\ell+1} + O\left(z_1^{k+2}\right) + O\left(z_n^{\ell+2}\right)\right]^{n-2} \mathrm{d}z_n \mathrm{d}z_1.
\end{aligned}$$

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

$$\begin{aligned}
p_n(F) &\approx \int_0^{\varepsilon\, n^{1/(k+1)}} \int_0^{\varepsilon\, n^{1/(\ell+1)}} n\,(n-1)\left[\frac{1}{n^{k/(k+1)}\, k!} f^{(k)}(0^+) w^k + O\left(n^{-1}\right)\right] \\
&\quad \times \left[\frac{(-1)^\ell}{n^{\ell/(\ell+1)}\, \ell!} f^{(\ell)}(1^-) v^\ell + O\left(n^{-1}\right)\right] \\
&\quad \times \left[1 - \frac{1}{n}\left(\frac{\alpha_k}{(k+1)!}\, w^{k+1} + \frac{(-1)^\ell \beta_\ell}{(\ell+1)!}\, v^{\ell+1}\right) + O\left(n^{-(\nu+2)/(\nu+1)}\right)\right]^{n-2} \\
&\quad \times \left(\frac{1}{n^{1/(k+1)}}\right)\left(\frac{1}{n^{1/(\ell+1)}}\right) \mathrm{d}v \mathrm{d}w \\
&= \int_0^{\varepsilon\, n^{1/(k+1)}} \int_0^{\varepsilon\, n^{1/(\ell+1)}} n\,(n-1)\left[\frac{(-1)^\ell}{n^2\, k!\, \ell!} f^{(k)}(0^+) f^{(\ell)}(1^-) w^k v^\ell \right. \\
&\quad + O\left(n^{-(2k+3)/(k+1)}\right) + O\left(n^{-(2\ell+3)/(\ell+1)}\right) \\
&\quad \left. + O\left(n^{-2(k+2)(\ell+2)/((k+1)(\ell+1))}\right)\right]\left[1 - \frac{1}{n}\left[\frac{\alpha_k}{(k+1)!}\, w^{k+1} + \frac{(-1)^\ell\, \beta_\ell}{(\ell+1)!}\, v^{\ell+1}\right]\right. \\
&\quad \left. + O\left(n^{-(\nu+2)/(\nu+1)}\right)\right]^{n-2} \mathrm{d}v \mathrm{d}w,
\end{aligned}$$

letting $n \to \infty$,

$$\begin{aligned}
&\approx \int_0^\infty \int_0^\infty \frac{(-1)^\ell}{k!\, \ell!} f^{(k)}(0^+) f^{(\ell)}(1^-) w^k v^\ell \exp\left[-\frac{\alpha_k}{(k+1)!}\, w^{k+1} - \frac{(-1)^\ell\, \beta_\ell}{(\ell+1)!}\, v^{\ell+1}\right] \mathrm{d}v \mathrm{d}w \\
&= \frac{f^{(k)}(0^+)\, f^{(\ell)}(1^-)\, (-1)^\ell\, (k+1)!(\ell+1)!}{k!\, \ell!\, (-1)^\ell\, (k+1)(\ell+1)\, \alpha_k\, \beta_\ell} = \frac{f^{(k)}(0^+)\, f^{(\ell)}(1^-)}{\alpha_k\, \beta_\ell}
\end{aligned}$$

$$= \frac{f^{(k)}(0^+)\, f^{(\ell)}(1^-)}{\left[f^{(k)}(0^+) + 2^{-(k+1)}\, f^{(k)}\left(\frac{1}{2}^+\right)\right]\left[f^{(\ell)}(1^-) + 2^{-(\ell+1)}\, f^{(\ell)}\left(\frac{1}{2}^-\right)\right]}, \tag{12}$$

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 $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2\}$, the transformation $\phi(x) = \frac{x-\mathsf{y}_1}{\mathsf{y}_2-\mathsf{y}_1}$ maps $(\mathsf{y}_1, \mathsf{y}_2)$ to $(0, 1)$ and the transformed random variables $\phi(X_i)$ are distributed with density $g(x) = (\mathsf{y}_2 - \mathsf{y}_1)\, f\left(\frac{x-\mathsf{y}_1}{\mathsf{y}_2-\mathsf{y}_1}\right)$ on $(0, 1)$. Substituting $f(x)$ by $g(x)$ in Eq. (12), the desired result follows. ■

Note that

- if $\min\left(f^{(k)}(\mathsf{y}_1^+), f^{(\ell)}(\mathsf{y}_2^-)\right) = 0$ and $\min\left(f^{(k)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^+\right), f^{(\ell)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^-\right)\right) \neq 0$ then $p_n(F) \to 0$ as $n \to \infty$, at rate $O\left(c(f) \cdot n^{-m}\right)$ and
- if $\min\left(f^{(k)}(\mathsf{y}_1^+), f^{(\ell)}(\mathsf{y}_2^-)\right) \neq 0$ and $f^{(k)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^+\right) = f^{(\ell)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^-\right) = 0$ then $p_n(F) \to 1$ as $n \to \infty$, at rate $O\left(c(f) \cdot n^{-m}\right)$.

For example, with $F = \mathcal{U}(\mathsf{y}_1, \mathsf{y}_2)$, in Theorem 5.1 we have $k = \ell = 0$, $f(\mathsf{y}_1^+) = f(\mathsf{y}_2^-) = f\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^+\right) = f\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^-\right) = 1/(\mathsf{y}_2 - \mathsf{y}_1)$. Then $\lim_{n\to\infty} p_n(F) = 4/9$, which agrees with the result given in Eq. (2).

Example 5.2. For F with density $f(x) = (x + 1/2)\, \mathbf{I}(0 < x < 1)$, 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 $\widehat{p}_{1000}(F) \approx 0.3753$. □

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. □

Remark 5.4. In Theorem 5.1, we assume that $f^{(k)}(\cdot)$ and $f^{(\ell)}(\cdot)$ are bounded on $(\mathsf{y}_1, \mathsf{y}_2)$. Suppose either $f^{(k)}(\cdot)$ or $f^{(\ell)}(\cdot)$ or both are not bounded on $(\mathsf{y}_1, \mathsf{y}_2)$ for $k, l \geq 0$, in particular at $\mathsf{y}_1, (\mathsf{y}_1 + \mathsf{y}_2)/2, \mathsf{y}_2$, for example, $\lim_{x\to \mathsf{y}_1^+} f^{(k)}(x) = \infty$. Then we find $p(F)$ as

$$p(F) = \lim_{\delta\to 0^+} \frac{f^{(k)}(\mathsf{y}_1 + \delta)\, f^{(\ell)}(\mathsf{y}_2 - \delta)}{\left[f^{(k)}(\mathsf{y}_1 + \delta) + 2^{-(k+1)}\, f^{(k)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2} + \delta\right)\right]\left[f^{(\ell)}(\mathsf{y}_2 - \delta) + 2^{-(\ell+1)}\, f^{(\ell)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2} - \delta\right)\right]}. \quad \square$$

Example 5.5. Consider the distribution with density function $f(x) = \frac{1}{\pi\sqrt{x\,(1-x)}}\, \mathbf{I}(0 < x < 1)$. Note that $\mathcal{Y}_2 = \{0, 1\}$ and $f(x)$ is unbounded at $x \in \{0, 1\}$. See Fig. 1 (left) for the plot of $f(x)$. Instead of $f(x)$, we consider $g(x) = \frac{\pi\, f(x)}{2\, \arcsin(1-2\delta)}\, \mathbf{I}(\delta < x < 1 - \delta)$ with cdf $G(x)$. For $g(x)$, we have $k = \ell = 0$ in Theorem 6.3 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 $\widehat{p}_{1000}(F) \approx 1.000$. □

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,

$$p_n(F) \approx \frac{f^{(k)}(\mathsf{y}_1^+)\, f^{(\ell)}(\mathsf{y}_2^-)}{\left[f^{(k)}(\mathsf{y}_1^+) + 2^{-(k+1)}\, f^{(k)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^+\right)\right]\left[f^{(\ell)}(\mathsf{y}_2^-) + 2^{-(\ell+1)}\, f^{(\ell)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)}{2}^-\right)\right]} + \frac{c(f)}{n^m},$$

where

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

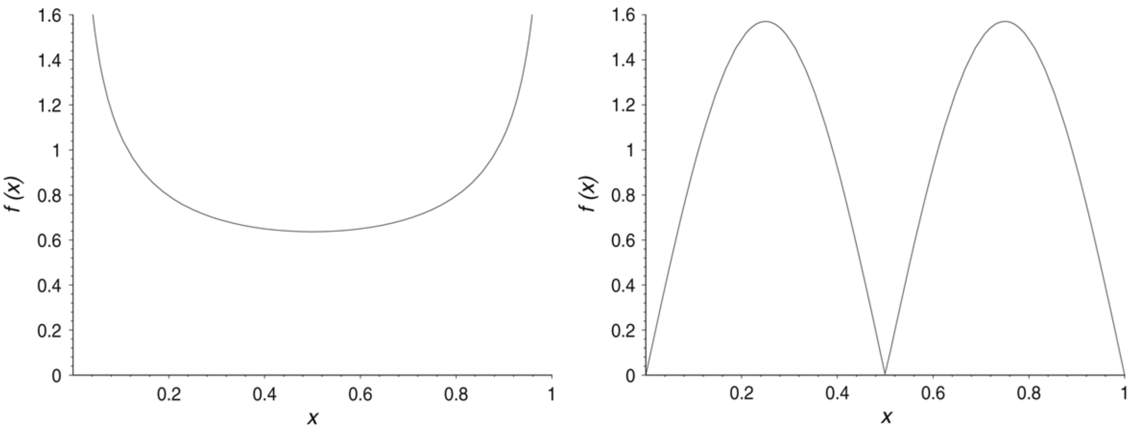

Fig. 1. Graph of the density in Example 5.5 (left) and Example 5.7 (right).

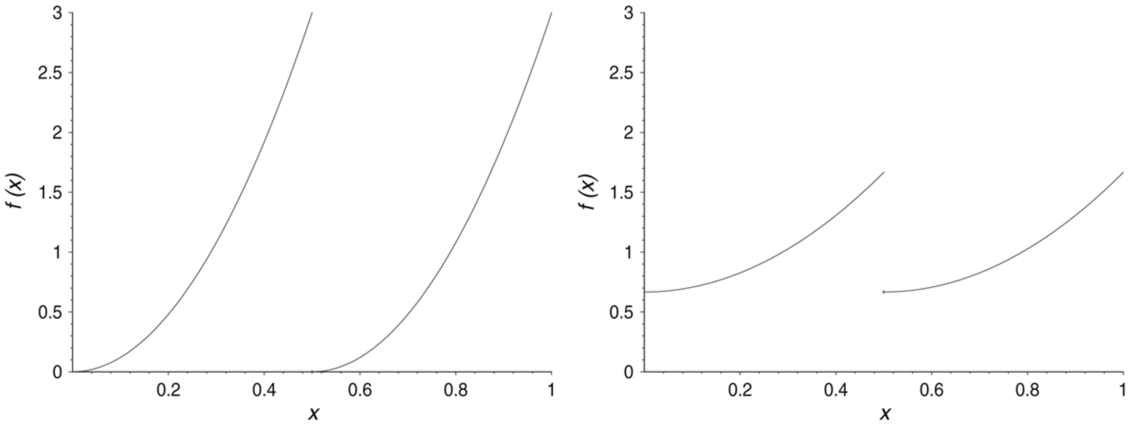

Fig. 2. Left plot is for the density in Example 5.11 with $q = 2$ or for the density in Example 5.12 with $\delta = 0$. Right plot is for the density in Example 5.12 with $\delta = 2/3$.

with $\Gamma(x) = \int_0^\infty \exp(-t)t^{(x-1)}\,dt$ and

$$s_1 = \frac{1}{n^{\frac{k+\ell+1}{\ell+1}}} \frac{(-1)^{\ell+1}}{k!\,(\ell+1)!} f^{(k)}(\mathsf{y}_1^+)\, f^{(\ell+1)}(\mathsf{y}_2^-), \quad s_3 = \frac{1}{(k+1)!}\left(f^{(k)}(\mathsf{y}_1^+) + 2^{-(k+1)} f^{(k)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)^+}{2}\right)\right),$$

$$s_2 = \frac{1}{n^{\frac{k+\ell+1}{k+1}}} \frac{(-1)^{\ell}}{l!\,(k+1)!} f^{(k+1)}(\mathsf{y}_1^+)\, f^{(\ell)}(\mathsf{y}_2^-), \quad s_4 = \frac{(-1)^{\ell+1}}{(\ell+1)!}\left(f^{(\ell)}(\mathsf{y}_2^-) + 2^{-(\ell+1)} f^{(\ell)}\left(\frac{(\mathsf{y}_1+\mathsf{y}_2)^-}{2}\right)\right),$$

provided the derivatives exist. □

Example 5.7. Consider the distribution with absolute sine density $f(x) = \pi/2\,|\sin(2\pi x)|\ \mathbf{I}(0 < x < 1)$. See Fig. 1 (right) for the plot of $f(x)$. Then $\mathcal{Y}_2 = \{0, 1\}$ and since $f(0^+) = f\left(\frac{1}{2}^+\right) = 0$ and $f(1^-) = f\left(\frac{1}{2}^-\right) = 0$ and $f'(0) = f'\left(\frac{1}{2}^+\right) = \pi^2$ and $f'(1^-) = f'\left(\frac{1}{2}^-\right) = -\pi^2$, we apply Theorem 5.1 with $k = \ell = 1$. Then $\lim_{n\to\infty} p_n(F) = 16/25 = 0.64$. The numerically computed value (by numerical integration) of $p_{1000}(F)$ is $\widehat{p}_{1000}(F) \approx 0.6400$. □

The distribution of $\gamma_n(F)$ depends on the distribution of $r(X_i) = \min(d(X_i, \mathsf{y}_1), d(X_i, \mathsf{y}_2))$. Based on this, we have the following symmetry result.

Proposition 5.8. *Let F_1 and F_2 be two distributions with support $\mathcal{S}(F_j) \subseteq (\mathsf{y}_1, \mathsf{y}_2)$ for $j = 1, 2$ such that $F_1(\mathsf{y}_1 + x) = 1 - F_2(\mathsf{y}_2 - x)$ for all $x \in (0, \mathsf{y}_2 - \mathsf{y}_1)$ (hence $f_1(\mathsf{y}_1 + x) = f_2(\mathsf{y}_2 - x)$). Also, let $\mathcal{X}_n^j$ be a set of iid random variables from F_j for $j = 1, 2$. Then the distributions of $\gamma_n(F_j)$ are identical for $j = 1, 2$.*

Proof. By the change of variable $X = \varphi(U) = \mathsf{y}_2 - \mathsf{y}_1 - U$ for $U \in (0, \mathsf{y}_2 - \mathsf{y}_1)$, we get $F_2(\mathsf{y}_1 + u) = 1 - F_1(\mathsf{y}_2 - u)$. Furthermore, $\varphi(u)$ transforms $\Gamma_1(\mathcal{X}_n^1, N_{\mathcal{Y}})$ into $\Gamma_1(\mathcal{X}_n^2, N_{\mathcal{Y}})$ for $\mathcal{X}_n^2$, so $P(\gamma_n(F_j) = 2)$ are same for both $j = 1, 2$. Hence the desired result follows. ■

Below are asymptotic distributions of $\gamma_n(F)$ for various families of distributions. Recall that $p_F = \lim_{n\to\infty} p_n(F) = \lim_{n\to\infty} P(\gamma_n(F) = 2)$. The asymptotic distribution of $\gamma_n(F)$ is $1 + \text{Bernoulli}(p_F)$. For the piecewise constant functions in Section 4.2.1, Theorem 5.1 applies. See Section 6.1 in [1].

Example 5.9. Consider the distribution F with density $f(\cdot)$ which is of the form

$$f(x) = (a\,x + b)\,\mathbf{I}(x \in (0, 1)) \text{ with } |a| \le 2,\ \ b = 1 - a/2.$$

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

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

Note that $p_F(a) \in [0, 4/9]$ is continuous in a and decreases as $|a|$ increases. If $a = 0$, then $F = \mathcal{U}(0, 1)$, 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 5.10. 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

$$f(x, \mu, \sigma) = \kappa \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) \exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right)\ \mathbf{I}(0 < x < 1),$$

where $\kappa = \left[\Phi\left(\frac{1-\mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)\right]^{-1}$ 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

$$\lim_{n\to\infty} p_n(F) = \frac{4}{\left(2 + \exp\left(\frac{4\mu-1}{8\sigma^2}\right)\right)\left(2 + \exp\left(\frac{3-4\mu}{8\sigma^2}\right)\right)} =: p_F(\mu, \sigma).$$

Observe that $p_F(\mu, \sigma) \in [0, 4/9)$ is continuous in μ and σ and increases as σ increases for fixed μ. Furthermore, for fixed μ, $\lim_{\sigma\to\infty} p_F(\mu, \sigma) = 4/9$ 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

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

See Fig. 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

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

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

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

See Fig. 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

$$\lim_{n\to\infty} p_n(F) = 4/9 \text{ for } \delta \in (0, 1].$$

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 δ. □

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 $\gamma_n(\mathcal{U}(0, 1))$ have the same asymptotic distributions). □

Example 5.14. Consider the Beta(ν_1, ν_2) distribution with $\nu_1, \nu_2 \geq 1$. The density function is

$$f(x, \nu_1, \nu_2) = \frac{x^{\nu_1-1}(1 - x)^{\nu_2-1}}{\beta(\nu_1, \nu_2)}\,\mathbf{I}(0 < x < 1) \text{ where } \beta(\nu_1, \nu_2) = \frac{\Gamma(\nu_1)\,\Gamma(\nu_2)}{\Gamma(\nu_1 + \nu_2)}.$$

Then $\lim_{n\to\infty} p_n(\text{Beta}(\nu_1, \nu_2)) = 0$ at rate $O\left(n^{-(\nu_1+\nu_2-2)}\right)$. Let $p_n(\nu_1, \nu_2)$ denote the $P(\gamma_n(F) = 2)$ for $F =$ Beta(ν_1, ν_2). The numerically computed values of $p_n(\nu_1, \nu_2)$ for $n = 1000$ are $\widehat{p}_{1000}(4, 1) = \widehat{p}_{1000}(1, 4) \approx 0.000005$, $\widehat{p}_{1000}(4, 2) = \widehat{p}_{1000}(2, 4) < 0.00001$ and $\widehat{p}_{1000}(2, 2) \approx 0.000001$. □

Here is an example with general support $(\mathsf{y}_1, \mathsf{y}_2)$.

Example 5.15. Consider the distribution F with density $f(\cdot)$ which is of the form $f(x) = (ax + b)\,\mathbf{I}(\mathsf{y}_1 < x < \mathsf{y}_2)$ with $b = \frac{1}{(\mathsf{y}_2-\mathsf{y}_1)}\left(1 - a\,(\mathsf{y}_2^2 - \mathsf{y}_1^2)/2\right)$ and $|a| \leq \frac{2}{(\mathsf{y}_2-\mathsf{y}_1)^2}$. Using Theorem 5.1, we obtain $p_F = \frac{a^2\,(\mathsf{y}_2-\mathsf{y}_1)^4-4}{a^2\,(\mathsf{y}_2-\mathsf{y}_1)^4-9}$. If $(\mathsf{y}_1, \mathsf{y}_2) = (0, 1)$, then $b = 1 - a/2$ and $p_F(a) = \frac{a^2-4}{a^2-9}$. 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 $a = \pm\frac{2}{(\mathsf{y}_2-\mathsf{y}_1)^2}$, since $p_n(F) \to 0$ as $n \to \infty$ at rate $O\left(n^{-1}\right)$. □

For more detail and examples, see Section 6.4 and 7.1 in [1].

6. The distribution of the domination number of $\mathscr{D}_{n,m}$-digraphs

In this section, we attempt the more challenging case of $m > 2$. For $c < d$ in $\mathbb{R}$, define the family of distributions

$$\mathscr{H}(\mathbb{R}) := \left\{F_{X,Y} :\ (X_i, Y_i) \sim F_{X,Y} \text{ with support } \mathcal{S}(F_{X,Y}) = (c, d)^2 \subsetneq \mathbb{R}^2,\ \ X_i \sim F_X \text{ and } Y_i \overset{\text{iid}}{\sim} F_Y\right\}.$$

We provide the exact distribution of $\gamma(D_{n,m})$ for $\mathscr{H}(\mathbb{R})$-random digraphs in the following theorem. Let $[m] := \{0, 1, \ldots, m - 1\}$ and $\Theta^S_{a,b} := \{(u_1, \ldots u_b) :\ \sum_{i=1}^{b} u_i = a :\ u_i \in S,\ \ \forall i\}$. 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 $\gamma(D^j)$ be the domination number of the digraph induced by $\mathcal{X}^j$ and $\mathcal{Y}^j$ (see Section 4). Given $Y_{(j)} = \mathsf{y}_{(j)}$ for $j = 1, \ldots, m$, let F_j be the (conditional) marginal distribution of X restricted to $\mathcal{I}_j = \left(\mathsf{y}_{(j-1)}, \mathsf{y}_{(j)}\right)$ for $j = 1, \ldots, (m + 1)$, $\vec{n}$ be the vector of numbers of $\mathcal{X}$ points n_j falling into intervals $\mathcal{I}_j$. Let $f_{\vec{Y}}(\vec{\mathsf{y}})$ be the joint distribution of the order statistics of $\mathcal{Y}_m$, i.e., $f_{\vec{Y}}(\vec{\mathsf{y}}) = \frac{1}{m!}\prod_{j=1}^{m} f(\mathsf{y}_j)\,\mathbf{I}(c < \mathsf{y}_1 < \cdots < \mathsf{y}_m < d)$, and $f_{j,k}(\mathsf{y}_j, \mathsf{y}_k)$ be the joint distribution of $Y_{(j)}, Y_{(k)}$. Then we have the following theorem which is a generalization of the main result of [10].

Theorem 6.1. *Let D be an $\mathscr{H}(\mathbb{R})$-random $\mathscr{D}_{n,m}$-digraph. Then the probability mass function of the domination number of D is given by*

$$P(\gamma(D_{n,m})=k)=\int_{\mathscr{S}}\sum_{\vec{n}\in\Theta^{[n+1]}_{n,(m+1)}}\sum_{\vec{k}\in\Theta^{[3]}_{k,(m+1)}}P(\vec{N}=\vec{n})\,\zeta(k_1,n_1)\,\zeta(k_{m+1},\,n_{m+1})\times\prod_{j=2}^{m}\eta(k_j,n_j)f_{\vec{Y}}(\vec{\mathsf{y}})\,\mathrm{d}\mathsf{y}_1\cdots\mathrm{d}\mathsf{y}_m$$

where $P(\vec{N}=\vec{n})$ is the joint probability of n_j points falling into intervals $\mathcal{I}_j$ for $j=1,2,\ldots,(m+1)$, $k_j\in\{0,1,2\}$, and

$$\zeta(k_j,n_j)=\max\big(\mathbf{I}(n_j=k_j=0),\mathbf{I}(n_j\geq k_j=1)\big)\,\textit{for}\,j=1,(m+1),\textit{ and}$$

$$\eta(k_j,n_j)=\max\big(\mathbf{I}(n_j=k_j=0),\mathbf{I}(n_j\geq k_j\geq 1)\big)\cdot p_{n_j}(F_j)^{\mathbf{I}(k_j=2)}\,\big(1-p_{n_j}(F_j)\big)^{\mathbf{I}(k_j=1)}$$

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

$$\mathscr{S}:=\{(\mathsf{y}_1,\mathsf{y}_2,\ldots,\mathsf{y}_m)\in(c,d)^2:\,c<\mathsf{y}_1<\mathsf{y}_2<\cdots<\mathsf{y}_m<d\}.$$

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, $\mathcal{I}_1$ and $\mathcal{I}_{m+1}$, 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 ζ terms equal unity if (n_j,k_j) are compatible. Therefore we have

$$\begin{aligned}P(\gamma(D_{n,m})=k)&=\int_{\mathscr{S}}\sum_{\vec{n}\in\Theta^{[n+1]}_{n,(m+1)}}\sum_{\vec{k}\in\Theta^{[3]}_{k,(m+1)}}P(\vec{N}=\vec{n})\prod_{j=1}^{m+1}\eta(k_j,n_j)f_{\vec{Y}}(\vec{\mathsf{y}})\,\mathrm{d}\mathsf{y}_1\cdots\mathrm{d}\mathsf{y}_m\\&=\int_{\mathscr{S}}\sum_{\vec{n}\in\Theta^{[n+1]}_{n,(m+1)}}\sum_{\vec{k}\in\Theta^{[3]}_{k,(m+1)}}P(\vec{N}=\vec{n})\prod_{j\in\{1,(m+1)\}}\eta(k_j,n_j)\prod_{j=2}^{m}\eta(k_j,n_j)f_{\vec{Y}}(\vec{\mathsf{y}})\,\mathrm{d}\mathsf{y}_1\cdots\mathrm{d}\mathsf{y}_m\\&=\int_{\mathscr{S}}\sum_{\vec{n}\in\Theta^{[n+1]}_{n,(m+1)}}\sum_{\vec{k}\in\Theta^{[3]}_{k,(m+1)}}P(\vec{N}=\vec{n})\,\zeta(k_1,n_1)\,\zeta(k_{m+1},\,n_{m+1})\prod_{j=2}^{m}\eta(k_j,n_j)f_{\vec{Y}}(\vec{\mathsf{y}})\,\mathrm{d}\mathsf{y}_1\cdots\mathrm{d}\mathsf{y}_m\end{aligned}$$

where we have used the conditional pairwise independence of $\gamma(D^j)$. The η 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

$$\mathbf{E}[\gamma(D_{n,m})]=P\left(X_{(1)}<Y_{(1)}\right)+P\left(X_{(n)}>Y_{(m)}\right)+\sum_{j=2}^{m}\sum_{k=1}^{n}P(N_j=k)\,\mathbf{E}[\gamma(D^j)]\tag{13}$$

where

$$P(N_j=k)=\int_c^d\int_{\mathsf{y}_{(j-1)}}^{d}f_{j-1,j}\left(\mathsf{y}_{(j-1)},\mathsf{y}_{(j)}\right)\left[F_X\left(\mathsf{y}_{(j)}\right)-F_X\left(\mathsf{y}_{(j-1)}\right)\right]^k\times\left[1-\left(F_X\left(\mathsf{y}_{(j)}\right)-F_X\left(\mathsf{y}_{(j-1)}\right)\right)\right]^{n-k}\,\mathrm{d}\mathsf{y}_{(j)}\mathrm{d}\mathsf{y}_{(j-1)}$$

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

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

Proof. Consider the intersection of the supports $\mathcal{S}(F_X) \cap \mathcal{S}(F_Y)$ that has positive (Lebesgue) measure. For $\mathcal{S}(Y) \setminus \mathcal{S}(X)$; i.e., in the intervals $\mathcal{I}_j$ falling outside the intersection $\mathcal{S}(F_X) \cap \mathcal{S}(F_Y)$, 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,

$$\begin{aligned}\mathbf{E}[\gamma(D_{n,n})] &= P\left(X_{(1)} < Y_{(1)}\right) + P(X_{(n)} > Y_{(n)}) + \sum_{j=2}^{n}\sum_{k=1}^{n} P(N_j = k)\,\mathbf{E}[\gamma_{N_j}(F_j)] \\ &> \sum_{j=2}^{n}\sum_{k=1}^{n} P(N_j = k)\,\mathbf{E}[\gamma_{N_j}(F_j)] = \sum_{j=2}^{n}\sum_{k=1}^{n} P(N_j = k)\,(1 + p_{N_j}(F_j)) \\ &> \sum_{j=2}^{n}\sum_{k=1}^{n} P(N_j = k) > \sum_{j=2}^{n} P(N_j \geq 1) \\ &= O(n) \quad (\text{as } n \to \infty)\end{aligned}$$

where $\mathbf{E}[\gamma_{N_j}(F_j)] = (1 + p_{N_j}(F_j))$ follows from the fact that $\gamma_{N_j}(F_j) \sim 1 + \text{Bernoulli}(p_{N_j}(F_j))$ from Theorem 4.2. Furthermore, $P(N_j \geq 1) \approx 1$ for sufficiently large n. Then the desired result follows. ■

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

$$p_{F_j} = \int_c^d \int_{\mathsf{y}_{(j-1)}}^d H^*\left(\mathsf{y}_{(j-1)}, \mathsf{y}_{(j)}\right) f_{j-1,j}\left(\mathsf{y}_{(j-1)}, \mathsf{y}_{(j)}\right) \mathrm{d}\mathsf{y}_{(j)}\mathrm{d}\mathsf{y}_{(j-1)}$$

with $H^*\left(\mathsf{y}_{(j-1)}, \mathsf{y}_{(j)}\right) = \lim_{n_j\to\infty}(p_{n_j}(F_j))$ which is given in Theorem 5.1 for F_j with density f_j whose support is $\left(\mathsf{y}_{(j-1)}, \mathsf{y}_{(j)}\right)$. Then the desired result follows. ■

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

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

where $\zeta(k_j, n_j)$ and $\eta(k_j, n_j)$ are as in Theorem 6.1; and the expected domination number $\mathbf{E}[\gamma(D_{n,m})]$ is as in Eq. (13) with $P(N_j = k) = \left[F_X\left(\mathsf{y}_{(j)}\right) - F_X\left(\mathsf{y}_{(j-1)}\right)\right]^k \left[1 - \left(F_X\left(\mathsf{y}_{(j)}\right) - F_X\left(\mathsf{y}_{(j-1)}\right)\right)\right]^{n-k}$; and $\lim_{n\to\infty} \gamma(D_{n,m}) \stackrel{d}{=} m + 1 + \sum_{j=1}^{m} B_j$, where $B_j \sim$ Bernoulli(p_{F_j}) with $p_{F_j} := \lim_{n_j\to\infty} p_{n_j}(F_j)$.

Let F_X be a distribution with support $\mathcal{S}(F_X) \subseteq (0, 1)$ and density $f_X(x)$. Conditional on $\mathcal{Y}_m = \left\{\mathsf{y}_{(1)}, \ldots, \mathsf{y}_{(m)}\right\}$, let F_j be the distribution with density $f_j(x) = \frac{1}{(\mathsf{y}_{(j)} - \mathsf{y}_{(j-1)})} f_X\left(\frac{x - \mathsf{y}_{(j-1)}}{\mathsf{y}_{(j)} - \mathsf{y}_{(j-1)}}\right)$ for $j = 2, \ldots, m$, and $\mathcal{S}(F_j(x))$ is nonempty for $j \in \{1, (m+1)\}$. By this construction, the independence of the distribution of $\gamma_n(F_j)$ from $\mathcal{I}_j$ obtains; i.e., $\gamma_n(F_j) \stackrel{d}{=} \gamma_n(F_X)$ for all $j \in \{1, \ldots, (m+1)\}$. Now consider the family $\mathscr{H}_{\mathcal{U}}(\mathbb{R})$ defined as

$$\mathscr{H}_{\mathcal{U}}(\mathbb{R}) := \left\{F_{X,Y} : (X_i, Y_i) \sim F_{X,Y},\ Y_j \stackrel{\text{iid}}{\sim} \mathcal{U}(c, d) \text{ for } (c, d) \subsetneq \mathbb{R},\ \text{and } X_i|\mathcal{Y}_m \stackrel{\text{iid}}{\sim} F_j\right\}.$$

Clearly $\mathscr{H}_{\mathcal{U}}(\mathbb{R}) \subsetneq \mathscr{H}(\mathbb{R})$.

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

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

where $\zeta(k_j, n_j)$ *and* $\eta(k_j, n_j)$ *are as in* Theorem 6.1.

Note that if in addition, $P_{F_j}(X \in \mathcal{I}_j) = P_{\mathcal{U}}(X \in \mathcal{I}_j)$ for all j, then $P(\vec{N} = \vec{n}) = \binom{n+m}{n}^{-1}$, since each $\vec{n} \in \Theta^{[n+1]}_{n,(m+1)}$ 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 $(\mathsf{y}_{(j-1)}, \mathsf{y}_{(j)})$. Hence Corollary 6.4 applies to these examples.

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

- If $f(\cdot)$ is of the form
$$f(x) = \frac{1}{(1-2\,\delta)\,v_j}\,\mathbf{I}\left(x \in \left(\mathsf{y}_{(j-1)} + \delta\, v_j, \mathsf{y}_{(j)} - \delta\, v_j\right)\right) \text{ with } \delta \in [0, 1/3]$$
then $p_n(F)$ is as in Eq. (7).
- If $f(\cdot)$ is of the form
$$f(x) = \frac{1}{(1-2\,\delta)\,v_j}\,\mathbf{I}\left(x \in \left(\mathsf{y}_{(j-1)}, u_j - \delta\, v_j\right) \cup \left[u_j + \delta\, v_j, \mathsf{y}_{(j)}\right)\right) \text{ with } \delta \in [0, 1/3],$$
then $p_n(F)$ is as in Eq. (8).
- If $f(\cdot)$ is of the form
$$f(x) = \frac{(1+\delta)}{v_j}\,\mathbf{I}\left(x \in \left(\mathsf{y}_{(j-1)}, u_j\right)\right) + \frac{(1-\delta)}{v_j}\,\mathbf{I}\left(x \in \left[u_j, \mathsf{y}_{(j)}\right)\right),$$
then $p_n(F)$ is as in Eq. (9).
- If $f(\cdot)$ is of the form
$$f(x) = f_1(x)\,\mathbf{I}\left(x \in \left(\mathsf{y}_{(j-1)}, t_j\right)\right) + f_2(x)\,\mathbf{I}\left(x \in \left[t_j, w_j\right)\right) + f_3(x)\,\mathbf{I}\left(x \in \left[w_j, \mathsf{y}_{(j)}\right)\right)$$
where $t_j = \frac{\mathsf{y}_{(j)}+3\,\mathsf{y}_{(j-1)}}{4}$, $w_j = \frac{3\mathsf{y}_{(j)}+\mathsf{y}_{(j-1)}}{4}$, $f_1(x) = \frac{(1+\delta)}{v_j}$, $f_2(x) = \frac{(1-\delta)}{v_j}$ and $f_3(x) = \frac{(1+\delta)}{v_j}$, then $p_n(F)$ is as in Example 4.9. □

Theorem 6.6. *Let* D *be an* $\mathscr{H}_{\mathcal{U}}(\mathbb{R})$*-random* $\mathscr{D}_{n,m}$*-digraph with the additional assumption that* $P_{F_j}(X \in \mathcal{I}_j) = P_{\mathcal{U}}(X \in \mathcal{I}_j)$ *for all* j*. Then*
$$\mathbf{E}[\gamma(D_{n,m})] = \frac{2\,n}{n+m} + \frac{n!\,m\,(m-1)}{(n+m)!}\sum_{i=1}^{n}\frac{(n+m-i-1)!}{(n-i)!}\,(1 + p_i(F))$$
where $p_i(F) = P(\gamma(D_{i,2}) = 2)$.

Proof. Similar to the Proof of Theorem 4 in [10]. ■

Furthermore, from Corollary 6.2, we have $\mathbf{E}[\gamma(D_{n,n})] \to \infty$ as $n \to \infty$.

Theorem 6.7. *Let* $D_{n,m}$ *be an* $\mathscr{H}_{\mathcal{U}}(\mathbb{R})$*-random* $\mathscr{D}_{n,m}$*-digraph. Then* (i) *for fixed* $n < \infty$, $\lim_{m\to\infty}\gamma(D_{n,m}) = n$ *a.s.* (ii) *for fixed* $m < \infty$, $\lim_{n\to\infty}\gamma(D_{n,m}) \overset{d}{=} m + 1 + B$, *where* $B \sim Binomial(m-1, p_F)$ *where* $p_F = \lim_{n\to\infty} P(\gamma(D_{n,2}) = 2)$.

Proof. Similar to the Proof of Theorem 5 in [10]. ■

Remark 6.8 (*Extension to Multi-dimensional Case*). The existence of ordering of points in $\mathbb{R}$ is crucial in our calculations. The order statistics of $\mathcal{Y}_m$ partition the support (c, d) into disjoint intervals a.s. which can also be viewed as the Delaunay tessellation of $\mathbb{R}$ based on $\mathcal{Y}_m$. This nice structure in $\mathbb{R}$ avails a minimum dominating set and hence the domination number, both in polynomial time. Furthermore, the Γ_1-region is readily available by the order statistics of $\mathcal{X}_n$; also the components of the digraph restricted to intervals $\mathcal{I}_j$ (see Section 4) are not connected to each other, since $N_{\mathcal{Y}}(x) \cap N_{\mathcal{Y}}(y) = \emptyset$ 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 $\mathcal{Y}_m$ does not readily partition the support, but we can use the Delaunay tessellation based on $\mathcal{Y}_m$. Furthermore, in multiple dimensions finding a minimum dominating set is an NP-hard problem; and Γ_1-regions are not readily available (in fact for $n_j > 3$, complexity of finding the Γ_1-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 $N_{\mathcal{Y}}(x) \cap N_{\mathcal{Y}}(y) \neq \emptyset$ might hold for x, y in distinct Delaunay cells. These have motivated us to generalize the proximity map $N_{\mathcal{Y}}$ in order to avoid the difficulties above. See [2,3], 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. [10] in several directions. Priebe et al. [10] provided the exact (finite sample) distribution of the class cover catch digraphs (CCCDs) based on $\mathcal{X}_n$ and $\mathcal{Y}_m$ 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 $N_{\mathcal{Y}}(x) := B(x, r(x))$ where $r(x) := \min_{\mathsf{y} \in \mathcal{Y}_m} d(x, \mathsf{y})$. First, given $\mathcal{Y}_2 = \{\mathsf{y}_1, \mathsf{y}_2\} \subset \mathbb{R}$, we lift the uniformity assumption of $\mathcal{X}_n$ by assuming it to be from a non-uniform distribution F with support $\mathcal{S}(F) \subseteq (\mathsf{y}_1, \mathsf{y}_2)$. 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 $(\mathsf{y}_1, \mathsf{y}_2)$. 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 $\mathcal{Y}_2$, hence the distribution of the domination number of a CCCD, $\gamma(D_{n,m})$, for $\mathcal{X}_n$ and $\mathcal{Y}_m$ with $m > 2$, for general F includes integration with respect to order statistics of $\mathcal{Y}_m$. We provide the conditions that make $\gamma(D_{n,m})$ independent of $\mathcal{Y}_m$. As another generalization direction, we also devise proximity maps depending on F that will yield the distribution identical to that of $\gamma_n(\mathcal{U}(\mathsf{y}_1, \mathsf{y}_2))$. Our set-up is more general than the one given in [10]. 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 [3], 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

[1] E. Ceyhan, The distribution of the domination number of class cover catch digraphs for non-uniform one-dimensional data, Technical Report 646, Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218, 2004.
[2] E. Ceyhan, C. Priebe, Central similarity proximity maps in Delaunay tessellations, in: Proceedings of the Joint Statistical Meeting, Statistical Computing Section, American Statistical Association, 2003.
[3] E. Ceyhan, C.E. Priebe, The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association, Statistics and Probability Letters 73 (2005) 37–50.
[4] G. Chartrand, L. Lesniak, Graphs & Digraphs, Chapman & Hall/CRC Press LLC, Florida, 1996.
[5] J. DeVinney, C.E. Priebe, D.J. Marchette, D. Socolinsky, Random walks and catch digraphs in classification. in: Proceedings of the 34th Symposium on the Interface: Computing Science and Statistics, vol. 34, 2002. http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf.
[6] J. DeVinney, J.C. Wierman, A SLLN for a one-dimensional class cover problem, Statistics and Probability Letters 59 (4) (2003) 425–435.
[7] S. Janson, T. Łuczak, A. Rucinński, Random Graphs, in: Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., New York, 2000.
[8] J.W. Jaromczyk, G.T. Toussaint, Relative neighborhood graphs and their relatives, Proceedings of IEEE 80 (1992) 1502–1517.
[9] D.J. Marchette, C.E. Priebe, Characterizing the scale dimension of a high dimensional classification problem, Pattern Recognition 36 (1) (2003) 45–60.

[10] C.E. Priebe, J.G. DeVinney, D.J. Marchette, On the distribution of the domination number of random class catch cover digraphs, Statistics and Probability Letters 55 (2001) 239–246.
[11] C.E. Priebe, D.J. Marchette, J. DeVinney, D. Socolinsky, Classification using class cover catch digraphs, Journal of Classification 20 (1) (2003) 3–23.
[12] C.E. Priebe, J.L. Solka, D.J. Marchette, B.T. Clark, Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays, Computational Statistics and Data Analysis on Visualization 43 (4) (2003) 621–632.
[13] E. Prisner, Algorithms for interval catch digraphs, Discrete Applied Mathematics 51 (1994) 147–157.
[14] M. Sen, S. Das, A. Roy, D. West, Interval digraphs: An analogue of interval graphs, Journal of Graph Theory 13 (1989) 189–202.
[15] Z. Tuza, Inequalities for minimal covering sets in sets in set systems of given rank, Discrete Applied Mathematics 51 (1994) 187–195.
[16] D.B. West, Introduction to Graph Theory, 2nd ed., Prentice Hall, NJ, 2001.