On the Distribution of the Domination Number of a New Family of Parametrized Random Digraphs

Elvan Ceyhan* & Carey E. Priebe†

February 9, 2022

Abstract

We derive the asymptotic distribution of the domination number of a new family of random digraph called proximity catch digraph (PCD), which has application to statistical testing of spatial point patterns and to pattern recognition. The PCD we use is a parametrized digraph based on two sets of points on the plane, where sample size and locations of the elements of one is held fixed, while the sample size of the other whose elements are randomly distributed over a region of interest goes to infinity. PCDs are constructed based on the relative allocation of the random set of points with respect to the Delaunay triangulation of the other set whose size and locations are fixed. We introduce various auxiliary tools and concepts for the derivation of the asymptotic distribution. We investigate these concepts in one Delaunay triangle on the plane, and then extend them to the multiple triangle case. The methods are illustrated for planar data, but are applicable in higher dimensions also.

Keywords: random graph; domination number; proximity map; Delaunay triangulation; proximity catch digraph

* This research was supported by the Defense Advanced Research Projects Agency as administered by the Air Force Office of Scientific Research under contract DOD F49620-99-1-0213 and by Office of Naval Research Grant N00014-95-1-0777.
* Corresponding author.
E-mail address: elceyhan@ku.edu.tr (E. Ceyhan)

* Department of Mathematics, Koç University, Sarıyer, 34450, Istanbul, Turkey
† Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD, 21218, USA
1 Introduction

The proximity catch digraphs (PCDs) are a special type of proximity graphs which were introduced by Toussaint (1980). A digraph is a directed graph with vertices \( V \) and arcs (directed edges) each of which is from one vertex to another based on a binary relation. Then the pair \((p, q) \in V \times V\) is an ordered pair which stands for an arc (directed edge) from vertex \( p \) to vertex \( q \). For example, the nearest neighbor (di)graph of Paterson and Yao (1992) is a proximity digraph. The nearest neighbor digraph has the vertex set \( V \) and \((p, q)\) an arc if \( q \) is a nearest neighbor of \( p \).

Our PCDs are based on the proximity maps which are defined in a fairly general setting. Let \((\Omega, \mathcal{M})\) be a measurable space. The proximity map \( N(\cdot) \) is defined as \( N : \Omega \to 2^\Omega \), where \( 2^\Omega \) is the power set of \( \Omega \). The proximity region of \( x \in \Omega \), denoted \( N(x) \), is the image of \( x \in \Omega \) under \( N(\cdot) \). The points in \( N(x) \) are thought of as being “closer” to \( x \in \Omega \) than are the points in \( \Omega \setminus N(x) \). Hence the term “proximity” in the name proximity catch digraph. Proximity maps are the building blocks of the proximity graphs of Toussaint (1980); an extensive survey on proximity maps and graphs is available in Jaromczyk and Toussaint (1992).

The proximity catch digraph \( D \) has the vertex set \( V = \{p_1, \ldots, p_n\} \) and the arc set \( A \) is defined by \((p_i, p_j) \in A\) if \( p_j \in N(p_i) \) for \( i \neq j \). Notice that the proximity catch digraph \( D \) depends on the proximity map \( N(\cdot) \) and if \( p_i \in N(p_j) \), then we call \( N(p_i) \) (and hence point \( p_i \)) catches \( p_j \). Hence the term “catch” in the name proximity catch digraph. If arcs of the form \((p_j, p_i)\) (i.e., loops) were allowed, \( D \) would have been called a pseudodigraph according to some authors (see, e.g., Chartrand and Lesniak (1996)).

In a digraph \( D = (V, A) \), a vertex \( v \in V \) dominates itself and all vertices of the form \( \{u : (v, u) \in A\} \). A dominating set \( S_D \) for the digraph \( D \) is a subset of \( V \) such that each vertex \( v \in V \) is dominated by a vertex in \( S_D \). A minimum dominating set \( S_D^* \) is a dominating set of minimum cardinality and the domination number \( \gamma(D) \) is defined as \( \gamma(D) := |S_D^*| \) (see, e.g., Lee (1998)) where \(|\cdot|\) denotes the set cardinality functional. See Chartrand and Lesniak (1996) and West (2001) for more on graphs and digraphs. If a minimum dominating set is of size one, we call it a dominating point.

Note that for \(|V| = n > 0\), \( 1 \leq \gamma(D) \leq n \), since \( V \) itself is always a dominating set.

In recent years, a new classification tool based on the relative allocation of points from various classes has been developed. Priebe et al. (2001) introduced the class cover catch digraphs (CCCDs) and gave the exact and the asymptotic distribution of the domination number of the CCCD based on two sets, \( X_n \) and \( Y_m \), which are of size \( n \) and \( m \), from classes, \( X \) and \( Y \), respectively, and are sets of iid random variables from uniform distribution on a compact interval in \( \mathbb{R} \). Deviney and Priebe (2003), DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a,b) applied the concept in higher dimensions and demonstrated relatively good performance of CCCD in classification. The methods employed involve data reduction (condensing) by using approximate minimum dominating sets as prototype sets (since finding the exact minimum dominating set is an NP-hard problem in general — e.g., for CCCD in multiple dimensions — (see DeVinney (2003)). DeVinney and Wierman (2003) proved a SLLN result for the domination number of CCCDs for one-dimensional data. Although intuitively appealing and easy to extend to higher dimensions, exact and asymptotic distribution of the domination number of the CCCDs are not analytically tractable in \( \mathbb{R}^2 \) or higher dimensions. As alternatives to CCCD, two new families of PCDs are introduced in Ceyhan and Priebe (2003, 2005) and are applied in testing spatial point patterns (see, Ceyhan et al. (2005, 2004)). These new families are both applicable to pattern classification also. They are designed to have better distributional and mathematical properties. For example, the distribution of the relative density (of arcs) is derived for one family in Ceyhan et al. (2005) and for the other family in Ceyhan et al. (2004). In this article, we derive the asymptotic distribution of the domination number of the latter family called \( r \)-factor proportional-edge PCD. During the derivation process, we introduce auxiliary tools, such as, proximity region (which is the most crucial concept in defining the PCD), \( \Gamma_1 \)-region, superset region, closest edge extrema, asymptotically accurate distribution, and so on. We utilize these special regions, extrema, and asymptotic expansion of the distribution of these extrema. The choice of the change of variables in the asymptotic expansion is also dependent on the type of the extrema used and crucial in finding the limits of the improper integrals we encounter. Our methodology is instructive in finding the distribution of the number of similar PCDs in \( \mathbb{R}^2 \) or higher dimensions.

In addition to the mathematical tractability and applicability to testing spatial patterns and classification, this new family of PCDs is more flexible as it allows choosing an optimal parameter for best performance in hypothesis testing or pattern classification.
The domination number of PCDs is first investigated for data in one Delaunay triangle (in $\mathbb{R}^2$) and the analysis is generalized to data in multiple Delaunay triangles. Some trivial proofs are omitted, shorter proofs are given in the main body of the article; while longer proofs are deferred to the Appendix.

2 Proximity Maps and the Associated PCDs

We construct the proximity regions using two data sets $X_n$ and $Y_m$ from two classes $X$ and $Y$, respectively. Given $Y_m \subseteq \Omega$, the proximity map $N_Y(\cdot) : \Omega \to 2^\Omega$ associates a proximity region $N_Y(x) \subseteq \Omega$ with each point $x \in \Omega$. The region $N_Y(x)$ is defined in terms of the distance between $x$ and $Y_m$. More specifically, our $r$-factor proximity maps will be based on the relative position of points from $X_n$ with respect to the Delaunay tessellation of $Y_m$. In this article, a triangle refers to the closed region bounded by its edges. See Figure 1 for an example with $n = 200$ $X$ points iid $\mathcal{U}((0,1) \times (0,1))$, the uniform distribution on the unit square and the Delaunay triangulation is based on $m = 10$ $Y$ which are points also iid $\mathcal{U}((0,1) \times (0,1))$. 

![Figure 1: A realization of 200 $X$ points (crosses) and the Delaunay triangulation based on 10 $Y$ points (circles).](image)

If $X_n = \{X_1, \ldots, X_n\}$ is a set of $\Omega$-valued random variables then $N_Y(X_i)$ are random sets. If $X_i$ are iid then so are the random sets $N_Y(X_i)$. We define the data-random proximity catch digraph $D$ — associated with $N_Y(\cdot)$ — with vertex set $X_n = \{X_1, \ldots, X_n\}$ and arc set $A$ by

$$(X_i, X_j) \in A \iff X_j \in N_Y(X_i).$$

Since this relationship is not symmetric, a digraph is used rather than a graph. The random digraph $D$ depends on the (joint) distribution of $X_i$ and on the map $N_Y(\cdot)$. For $X_n = \{X_1, \ldots, X_n\}$, a set of iid random variables from $F$, the domination number of the associated data-random proximity catch digraph based on the proximity map $N(\cdot)$, denoted $\gamma(X_n, N)$, is the minimum number of point(s) that dominate all points in $X_n$.

The random variable $\gamma(X_n, N)$ depends explicitly on $X_n$ and $N(\cdot)$ and implicitly on $F$. Furthermore, in general, the distribution, hence the expectation $\mathbb{E}[\gamma(X_n, N)]$, depends on $n$, $F$, and $N$; $1 \leq \mathbb{E}[\gamma(X_n, N)] \leq n$. In general, the variance of $\gamma(X_n, N)$ satisfies, $1 \leq \text{Var}[\gamma(X_n, N)] \leq n^2 / 4$. 

For example, the CCCD of Priebe et al. (2001) can be viewed as an example of PCDs and is briefly discussed in the next section. We use many of the properties of CCCD in $\mathbb{R}$ as guidelines in defining PCDs in higher dimensions.
2.1 Spherical Proximity Maps

Let \( Y_m = \{y_1, \ldots, y_m\} \subset \mathbb{R} \). Then the proximity map associated with CCCD is defined as the open ball 
\( N_S(x) := B(x, r(x)) \) for all \( x \in \mathbb{R} \), where 
\( r(x) := \min_{y \in Y_m} d(x, y) \) (see [Priebe et al. (2001)]) with \( d(x, y) \) being the Euclidean distance between \( x \) and \( y \). That is, there is an arc from \( X_i \) to \( X_j \) iff there exists an open ball centered at \( X_i \) which is “pure” (or contains no elements) of \( Y_m \) in its interior, and simultaneously contains (or “catches”) point \( X_j \). We consider the closed ball, \( \overline{B}(x, r(x)) \) for \( N_S(x) \) in this article. Then for \( x \in Y_m \), we have \( N_S(x) = \{x\} \). Notice that a ball is a sphere in higher dimensions, hence the notation \( N_S \). Furthermore, dependence on \( Y_m \) is through \( r(x) \). Note that in \( \mathbb{R} \) this proximity map is based on the intervals \( I_j = (y_{(j-1)} + y_{j,m}) / 2 \) for \( j = 0, \ldots, m + 1 \) with \( y_{0m} = -\infty \) and \( y_{(m+1) + m} = \infty \), where \( y_{j,m} \) is the \( j^{th} \) order statistic in \( Y_m \). This interval partitioning can be viewed as the Delaunay tessellation of \( \mathbb{R} \) based on \( Y_m \). So in higher dimensions, we use the Delaunay triangulation based on \( Y_m \) to partition the support.

A natural extension of the proximity region \( N_S(x) \) to \( \mathbb{R}^d \) with \( d > 1 \) is obtained as 
\( N_S(x) := B(x, r(x)) \) where 
\( r(x) := \min_{y \in Y_m} d(x, y) \) which is called the spherical proximity map. The spherical proximity map \( N_S(x) \) is well-defined for all \( x \in \mathbb{R}^d \) provided that \( Y_m \neq \emptyset \). Extensions to \( \mathbb{R}^2 \) and higher dimensions with the spherical proximity map — with applications in classification — are investigated by [DeVinney and Priebe (2006), DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003a)]1. However, finding the minimum dominating set of CCCD (i.e., the PCD associated with \( N_S(\cdot) \)) is an NP-hard problem and the distribution of the domination number is not analytically tractable for \( d > 1 \). This drawback has motivated us to define new types of proximity maps. [Cevhan and Priebe (2005)] introduced \( r \)-factor proportional-edge PCD, where the distribution of the domination number of \( r \)-factor PCD with \( r = 3/2 \) is used in testing spatial patterns of segregation or association. [Cevhan et al. (2006)] computed the asymptotic distribution of the relative density of the \( r \)-factor PCD and used it for the same purpose. [Cevhan and Priebe (2003)] introduced the central similarity proximity maps and the associated PCDs, and [Cevhan et al. (2003)] computed the asymptotic distribution of the relative density of the parameterized version of the central similarity PCDs and applied the method to testing spatial patterns. An extensive treatment of the PCDs based on Delaunay tessellations is available in [Cevhan (2004)].

The following property (which is referred to as Property (1)) of CCCDs in \( \mathbb{R} \) plays an important role in defining proximity maps in higher dimensions.

**Property (1)** For \( x \in I_j \), \( N_S(x) \) is a proper subset of \( I_j \) for almost all \( x \in I_j \).

In fact, Property (1) holds for all \( x \in I_j \setminus \{y_{(j-1) + m} + y_{j,m}/2\} \) for CCCDs in \( \mathbb{R} \). For \( x \in I_j \), \( N_S(x) = I_j \) iff 
\( x = (y_{(j-1) + m} + y_{j,m}) / 2 \). We define an associated region for such points in the general context. The superset region for any proximity map \( N(\cdot) \) in \( \mathbb{R} \) is defined to be

\[ S(N) := \{x \in \Omega : N(x) = \Omega\} \]

For example, for \( \Omega = I_j \subseteq \mathbb{R} \), \( S(N_S) := \{x \in I_j : N_S(x) = I_j\} = \{y_{(j-1) + m} + y_{j,m}/2\} \) and for \( \Omega = T_j \subseteq \mathbb{R}^d \), \( S(N_S) := \{x \in T_j : N_S(x) = T_j\} \), where \( T_j \) is the \( j^{th} \) Delaunay cell in the Delaunay tessellation. Note that for \( x \in I_j \), \( \lambda(N_S(x)) \leq \lambda(I_j) \) and \( \lambda(N_S(x)) = \lambda(I_j) \) iff \( x \in S(N_S) \) where \( \lambda(\cdot) \) is the Lebesgue measure on \( \mathbb{R} \). So the proximity region of a point in \( S(N_S) \) has the largest Lebesgue measure. Note also that given \( Y_m \), \( S(N_S) \) is not a random set, but \( I(\cdot) \) is a random variable, where \( I(\cdot) \) stands for the indicator function. Property (1) also implies that \( S(N_S) \) has zero \( \mathbb{R} \)-Lebesgue measure.

Furthermore, given a set \( B \) of size \( n \) in \( \{y_{1:m}, y_{m+1:m}\} \setminus Y_m \), the number of disconnected components in the PCD based on \( N_S(\cdot) \) is at least the cardinality of the set \( \{j \in \{1, 2, \ldots, m\} : B \cap I_j \neq \emptyset\} \), which is the set of indices of the intervals that contain some point(s) from \( B \).

Since the distribution of the domination number of spherical PCD (or CCCD) is tractable in \( \mathbb{R} \), but not in \( \mathbb{R}^d \) with \( d > 1 \), we try to mimic its properties in \( \mathbb{R} \) while defining new PCDs in higher dimensions.

3 The \( r \)-Factor Proportional-Edge Proximity Maps

First, we describe the construction of the \( r \)-factor proximity maps and regions, then state some of its basic properties and introduce some auxiliary tools.
3.1 Construction of the Proximity Map

Let $\mathcal{Y}_m = \{y_1, \ldots, y_m\}$ be $m$ points in general position in $\mathbb{R}^d$ and $T_j$ be the $j^{th}$ Delaunay cell for $j = 1, \ldots, J_m$, where $J_m$ is the number of Delaunay cells. Let $\mathcal{X}'_n$ be a set of iid random variables from distribution $F$ in $\mathbb{R}^d$ with support $\mathcal{S}(F) \subseteq C_H(\mathcal{Y}_m)$.

In particular, for illustrative purposes, we focus on $\mathbb{R}^2$ where a Delaunay tessellation is a triangulation, provided that no more than three points in $\mathcal{Y}_m$ are cocircular (i.e., lie in the same circle). Furthermore, for simplicity, let $\mathcal{Y}_3 = \{y_1, y_2, y_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}_3) = T(y_1, y_2, y_3)$ be the triangle with vertices $\mathcal{Y}_3$. Let $\mathcal{X}'_n$ be a set of iid random variables from $F$ with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. If $F = \mathcal{U}(T(\mathcal{Y}_3))$, a composition of translation, rotation, reflection, and scaling will take any given triangle $T(\mathcal{Y}_3)$ to the basic triangle $T_b = T((0, 0), (1, 0), (c_1, c_2))$ with $0 < c_1 \leq 1/2$, $c_2 > 0$, and $(1 - c_1)^2 + c_2^2 \leq 1$, preserving uniformity. That is, if $X \sim \mathcal{U}(T(\mathcal{Y}_3))$ is transformed in the same manner to, say $X'$, then we have $X' \sim \mathcal{U}(T_b)$.

For $r \in [1, \infty]$, define $N_{PE}^r(\cdot, M) := N(\cdot, M; r, \mathcal{Y}_3)$ to be the $r$-factor proportional-edge proximity map with $M$-vertex regions as follows (see also Figure 2 with $M = M_C$ and $r = 2$). For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $\nu(x) \in \mathcal{Y}_3$ be the vertex whose region contains $x$, i.e., $x \in R_\nu(x)$. In this article $M$-vertex regions are constructed by the lines joining any point $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$ to a point on each of the edges of $T(\mathcal{Y}_3)$. Preferably, $M$ is selected to be in the interior of the triangle $T(\mathcal{Y}_3)^o$. For such an $M$, the corresponding vertex regions can be defined using the line segment joining $M$ to $e_j$, which lies on the line joining $y_j$ to $M$; e.g., see Figure 3 (left) for vertex regions based on center of mass $M_C$, and (right) incenter $M_I$. With $M_C$, the lines joining $M$ and $\mathcal{Y}_3$ are the median lines, that cross edges at $M_j$ for $j = 1, 2, 3$. $M$-vertex regions, among many possibilities, can also be defined by the orthogonal projections from $M$ to the edges. See Ceyhan (2004) for a more general definition. The vertex regions in Figure 2 are center of mass vertex regions or $CM$-vertex regions. If $x$ falls on the boundary of two $M$-vertex regions, we assign $\nu(x)$ arbitrarily. Let $\ell(x)$ be the edge of $T(\mathcal{Y}_3)$ opposite of $\nu(x)$. Let $\ell(\nu(x), x)$ be the line parallel to $\ell(x)$ through $x$. Let $d(\nu(x), \ell(\nu(x), x))$ be the Euclidean (perpendicular) distance from $\nu(x)$ to $\ell(\nu(x), x)$. For $r \in [1, \infty]$, let $\ell_r(\nu(x), x)$ be the line parallel to $\ell(x)$ such that

$$d(\nu(x), \ell_r(\nu(x), x)) = r d(\nu(x), \ell(\nu(x), x)) \text{ and } d(\ell(\nu(x), x), \ell_r(\nu(x), x)) < d(\nu(x), \ell_r(\nu(x), x)).$$

Let $T_r(x)$ be the triangle similar to and with the same orientation as $T(\mathcal{Y}_3)$ having $\nu(x)$ as a vertex and $\ell_r(\nu(x), x)$ as the opposite edge. Then the $r$-factor proportional-edge proximity region $N_{PE}^r(x, M)$ is defined to be $T_r(x) \cap T(\mathcal{Y}_3)$. Notice that $\ell(\nu(x), x)$ divides the edges of $T_r(x)$ (other than the one lies on $\ell_r(\nu(x), x)$) proportionally with the factor $r$. Hence the name $r$-factor proportional edge proximity region.

![Figure 2: Construction of $r$-factor proximity region, $N_{PE}^{r=2}(x)$ (shaded region).](image)

Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M)$ for all $x \in T(\mathcal{Y}_3)$. Furthermore, $\lim_{r \to \infty} N_{PE}^r(x, M) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, so we define $N_{PE}^\infty(x, M) = T(\mathcal{Y}_3)$ for all such $x$. For $x \in \mathcal{Y}_3$, we define $N_{PE}^r(x, M) = \{x\}$ for all $r \in [1, \infty]$. 

5
Figure 3: The vertex regions constructed with center of mass $M_C$ (left) and incenter $M_I$ (right) using the line segments on the line joining $M$ to $Y_3$.

Figure 4: A realization of 7 $X$ points generated iid $\mathcal{U}(T(Y_3))$ (left) and the corresponding arcs of $r$-factor proportional edge PCD with $r = 3/2$ and $M = M_C$.

Hence, $r$-factor proportional edge PCD has vertices $X_n$ and arcs $(x_i, x_j)$ iff $x_j \in N_{PE}^r(x_i, M)$. See Figure 1 for a realization of $X_n$ with $n = 7$ and $m = 3$. The number of arcs is 12 and $\gamma_n(r = 2, M_C) = 1$. By construction, note that as $x$ gets closer to $M$ (or equivalently further away from the vertices in vertex regions), $N_{PE}^r(x, M)$ increases in area, hence it is more likely for the outdegree of $x$ to increase. So if more $X$ points are around the center $M$, then it is more likely for $\gamma_n$ to decrease, on the other hand, if more $X$ points are around the vertices $Y_3$, then the regions get smaller, hence it is more likely for the outdegree for such points to be smaller, thereby implying $\gamma_n$ to increase. This probabilistic behaviour is utilized in Ceyhan and Priebe (2005) for testing spatial patterns.

Note also that, $N_{PE}^r(x, M)$ is a homothetic transformation (enlargement) with $r \geq 1$ applied on the region $N_{PE}^{r-1}(x, M)$. Furthermore, this transformation is also an affine similarity transformation.

3.2 Some Basic Properties and Auxiliary Concepts

First, notice that $X_i \overset{\text{iid}}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional probability density function $f$ exists with support $\mathcal{S}(F) \subseteq T(Y_3)$, imply that the special case in the construction of $N_{PE}^r$ — $X$ falls on the boundary of two vertex regions — occurs with probability zero. Note that for such an $F$, $N_{PE}^r(X)$ is a triangle a.s.

The similarity ratio of $N_{PE}^r(x, M)$ to $T(Y_3)$ is given by $\frac{\min\left(\frac{d(v(x), e(x))}{d(v(x), \ell(v(x), x))} \right)}{\min\left(\frac{d(v(x), e(x))}{d(v(x), \ell(v(x), x))} \right)}$, that is, $N_{PE}^r(x, M)$ is similar to $T(Y_3)$ with the above ratio. Property (1) holds depending on the pair $M$ and $r$. That is,
there exists an $r_0$ and a corresponding point $M(r_0) \in T(T_3)^o$ so that $N^o_{PE}(x, M)$ satisfies Property (1) for all $r \leq r_0$, but fails to satisfy it otherwise. Property (1) fails for all $M$ when $r = \infty$. With $CM$-vertex regions, for all $r \in [1, \infty]$, the area $A(N^o_{PE}(x, M))$ is a continuous function of $d(t_x, v(x), v(x))$ which is a continuous function of $d(\ell(v(x), x), v(x))$ which is a continuous function of $x$.

Note that if $x$ is close enough to $M$, we might have $N^o_{PE}(x, M) = T(T_3)$ for $r = \sqrt{2}$ also.

In $T(T_3)$, drawing the lines $q_j(r, x)$ such that $d(y_j, e_j) = r d(y_j, q_j(r, x))$ for $j \in \{1, 2, 3\}$ yields a triangle, denoted $\mathcal{T}_r$, for $r < 3/2$. See Figure 5 for $\mathcal{T}_r$ with $r = \sqrt{2}$.

The functional form of $\mathcal{T}_r$ in the basic triangle $T_b$ is given by

$$
\mathcal{T}_r = T(t_1(r), t_2(r), t_3(r)) = \{(x, y) \in T_b : y \geq \frac{c_2(r-1)}{r}; \ y \leq \frac{c_2(r(x-1)+1)}{r c_1}\} \tag{2}
$$

There is a crucial difference between the triangles $\mathcal{T}_r$ and $T(M_1, M_2, M_3)$. More specifically $T(M_1, M_2, M_3) \subseteq \mathcal{A}_S(r, M)$ for all $M$ and $r \geq 2$, but $(\mathcal{T}_r)^o$ and $\mathcal{A}_S(r, M)$ are disjoint for all $M$ and $r$. So if $M \in (\mathcal{T}_r)^o$, then $\mathcal{A}_S(r, M) = \emptyset$; if $M \in \partial(\mathcal{T}_r)$, then $\mathcal{A}_S(r, M) = \{M\}$; and if $M \notin \mathcal{T}_r$, then $\mathcal{A}_S(r, M)$ has positive area. Thus $N^o_{PE}(r, M)$ fails to satisfy Property (1) if $M \notin \mathcal{T}_r$. See Figure 5 for two examples of superset regions with $M$ that corresponds to circumcenter $M_{CC}$ in this triangle and the vertex regions are constructed using orthogonal projections. For $r = 2$, note that $\mathcal{T}_r = \emptyset$ and the superset region is $T(M_1, M_2, M_3)$ (see Figure 6 (left)), while for $r = \sqrt{2}$, $\mathcal{A}_S(r = \sqrt{2})$ and $\mathcal{A}_S(r = \sqrt{2})^o$ are disjoint (see Figure 6 (right))

The triangle $\mathcal{T}_r$ given in Equation (2) and the superset region $\mathcal{A}_S(r, M)$ play a crucial role in computing the distribution of the domination number of the $r$-factor PCD.

### 3.3 Main Result

Next, we present the main result of this article. Let $\gamma_n(r, M) := \gamma(X_n, N^*_{PE}(r, M)$ be the domination number of the PCD based on $N^*_{PE}$ with $X_n$, a set of iid random variables from $U(T(T_3))$, with $M$-vertex regions.

The domination number $\gamma_n(r, M)$ of the PCD has the following asymptotic distribution. As $n \to \infty$,

$$
\gamma_n(r, M) \sim \begin{cases}
2 + \text{BER}(1-p_r), & \text{for } r \in [1, 3/2] \text{ and } M \in \{t_1(r), t_2(r), t_3(r)\}, \\
1, & \text{for } r > 3/2, \\
3, & \text{for } r \in [1, 3/2] \text{ and } M \in \mathcal{T}_r \setminus \{t_1(r), t_2(r), t_3(r)\},
\end{cases} \tag{3}
$$

where $\text{BER}(p)$ stands for Bernoulli distribution with probability of success $p$, $\mathcal{T}_r$ and $t_j(r)$ are defined in Equation (2), and for $r \in [1, 3/2]$ and $M \in \{t_1(r), t_2(r), t_3(r)\}$,

$$
p_r = \int_0^\infty \int_0^\infty \frac{64 r^2}{9 (r-1)^2} w_1 w_3 \exp \left( \frac{4 r}{3 (r-1)} \left( w_1^2 + w_3^2 + 2 r (r-1) w_1 w_3 \right) \right) dw_3 w_1; \tag{4}
$$

Figure 5: The triangle $\mathcal{T}_r$ with $r = \sqrt{2}$ (the hatched region).
Figure 6: The superset regions (the shaded regions) constructed with circumcenter $M_{CC}$ with $r = \sqrt{2}$ (left) and $r = 2$ (right) with vertex regions constructed with orthogonal projections to the edges.

Figure 7: Plotted is the probability $p_r = \lim_{n \to \infty} P(\gamma_n(r, M) = 2)$ given in Equation (4) as a function of $r$ for $r \in [1, 3/2)$ and $M \in \{t_1(r), t_2(r), t_3(r)\}$.

for example for $r = 3/2$ and $M = M_{CC}$, $p_r \approx .7413$.

In Equation (3), the first line is referred as the non-degenerate case, the second and third lines are referred as degenerate cases with a.s. limits 1 and 3, respectively.

In the following sections, we define a region associated with $\gamma = 1$ case in general. Then we give finite sample and asymptotic upper bounds for $\gamma_n(r, M)$. Then we derive the asymptotic distribution of $\gamma_n(r, M)$.

4 The $\Gamma_1$-Regions for $N_{PE}^r$

First, we define $\Gamma_1$-regions in general, and describe the construction of $\Gamma_1$-region of $N_{PE}^r$ for one point and multiple point data sets, and provide some results concerning $\Gamma_1$-regions.

4.1 Definition of $\Gamma_1$-Regions

Let $(\Omega, M)$ be a measurable space and consider the proximity map $N : \Omega \to 2^\Omega$. For any set $B \subseteq \Omega$, the $\Gamma_1$-region of $B$ associated with $N(\cdot)$, is defined to be the region $\Gamma_1^N(B) := \{z \in \Omega : B \subseteq N(z)\}$. For $x \in \Omega$, we denote $\Gamma_1^N(\{x\})$ as $\Gamma_1^N(x)$.

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

Note that $\gamma(\mathcal{X}_n, N) = 1$ iff $\mathcal{X}_n \cap \Gamma_1^N(\mathcal{X}_n) \neq \emptyset$. Hence the name $\Gamma_1$-region.
It is trivial to see the following.

**Proposition 4.1.** For any proximity map $N$ and set $B \subseteq \Omega$, $\mathcal{R}_S(N) \subseteq \Gamma_1^N(B)$.

**Lemma 4.2.** For any proximity map $N$ and $B \subseteq \Omega$, $\Gamma_1^N(B) = \cap_{x \in B} \Gamma_1^N(x)$.

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

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

For example, in $\mathbb{R}$ with $\mathcal{Y}_2 = \{0, 1\}$, and $\mathcal{X}_n$ a set of iid random variables of size $n > 1$ from $F$ in $(0, 1)$, $\Gamma_1^{N_3}(\mathcal{X}_n) = \left(\frac{X_{n:n}}{2}, \frac{1 + X_{1:n}}{2}\right)$. So the extrema (minimum and maximum) of the set $\mathcal{X}_n$ are sufficient to determine the $\Gamma_1$-region; i.e., $G = \{X_{1:n}, X_{n:n}\}$ for $\mathcal{X}_n$ a set of iid random variables from a continuous distribution on $(0, 1)$. Unfortunately, in the multi-dimensional case, there is no natural ordering that yields natural extrema such as minimum or maximum.

### 4.2 Construction of $\Gamma_1$-Region of a Point for $N_{PE}$

For $N_{PE}(\cdot, M)$, the $\Gamma_1$-region, denoted as $\Gamma_1^i(\cdot, M) := \Gamma_1^{N_{PE}}(\cdot, M)$, is constructed as follows; see also Figure 8:

Let $\xi_j(r, x)$ be the line parallel to $e_j$ such that $\xi_j(r, x) \cap T(\mathcal{Y}_3) \neq \emptyset$ and $r d(y_j, \xi_j(r, x)) = d(y_j, \ell(y_j, x))$ for $j \in \{1, 2, 3\}$. Then

$$\Gamma_1^i(x, M) = \bigcup_{j=1}^3 [\Gamma_1^i(x, M) \cap R_M(y_j)]$$

where $\Gamma_1^i(x, M) \cap R_M(y_j) = \{z \in R_M(y_j) : d(y_j, \ell(y_j, z)) \geq d(y_j, \xi_j(r, x))\}$ for $j \in \{1, 2, 3\}$.

Notice that $r \geq 1$ implies that $x \in \Gamma_1^i(x, M)$. Furthermore, $\lim_{r \to \infty} \Gamma_1^i(x, M) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$ and so we define $\Gamma_1^{=\infty}(x, M) = T(\mathcal{Y}_3)$ for all such $x$. For $x \in \mathcal{Y}_3$, $\Gamma_1^i(x, M) = \{x\}$ for all $r \in [1, \infty]$.

![Figure 8: Construction of the $\Gamma_1$-region, $\Gamma_1^{=2}(x, M_C)$ (shaded region).](image)

Notice that $\Gamma_1^i(x, M_C)$ is a convex hexagon for all $r \geq 2$ and $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, (since for such an $x$, $\Gamma_1^i(x, M_C)$ is bounded by $\xi_j(r, x)$ and $e_j$ for all $j \in \{1, 2, 3\}$, see also Figure 8) else it is either a convex hexagon or a non-convex but star-shaped polygon depending on the location of $x$ and the value of $r$.

### 4.3 The $\Gamma_1$-Region of a Multiple Point Data Set for $N_{PE}$

So far, we have described the $\Gamma_1$-region for a point in $x \in T(\mathcal{Y}_3)$. For a set $\mathcal{X}_n$ of size $n$ in $T(\mathcal{Y}_3)$, the region $\Gamma_1^i(\mathcal{X}_n, M)$ can be specified by the edge extrema only. The (closest) edge extrema of a set $B$ in $T(\mathcal{Y}_3)$ are the points closest to the edges of $T(\mathcal{Y}_3)$, denoted $x_{e_j}$ for $j \in \{1, 2, 3\}$; that is, $x_{e_j} \in \arg\min_{x \in B} d(x, e_j)$. Note that if $B = \mathcal{X}_n$ is a set of iid random variables of size $n$ from $F$ then the edge extrema, denoted $x_{e_j}(n)$, are random variables. Below, we show that the edge extrema are the active points in defining $\Gamma_1^i(\mathcal{X}_n, M)$.
The Asymptotic Distribution of \( \gamma \)

Below, we demonstrate that edge extrema are distinct with probability 1 as \( n \to \infty \). Hence in the limit three distinct points suffice to determine the \( \Gamma_1 \)-region.

**Theorem 4.4.** Let \( X_n \) be a set of iid random variables from \( U(T(Y_3)) \) and let \( E_{c,3}(n) \) be the event that (closest) edge extrema are distinct. Then \( P(E_{c,3}(n)) \to 1 \) as \( n \to \infty \).

We can also define the regions associated with \( \gamma(X_n, N) = k \) for \( k \leq n \) called \( \Gamma_k \)-region for proximity map \( N_{Y_3}(\cdot) \) and set \( B \subseteq \Omega \) for \( k = 1, \ldots, n \) (see Ceyhan (2004)).

## 5 The Asymptotic Distribution of \( \gamma_n(r, M) \)

In this section, we first present a finite sample upper bound for \( \gamma_n(r, M) \), then present the degenerate cases, and the nondegenerate case of the asymptotic distribution of \( \gamma_n(r, M) \) given in Equation 3.
5.1 An Upper Bound for $\gamma_n(r, M)$

Recall that by definition, $\gamma(X_n, N) \leq n$. We will seek an a.s. least upper bound for $\gamma(X_n, N)$. Let $X_n$ be a set of iid random variables from $F$ on $T(\mathcal{Y}_3)$ and let $\gamma(X_n, N)$ be the domination number for the PCD based on a proximity map $N$. Denote the general a.s. least upper bound for $\gamma(X_n, N)$ that works for all $n \geq 1$ and is independent of $n$ (which is called $\kappa$-value in [Cevher 2004]) as $\kappa(N) := \min \{ k : \gamma(X_n, N) \leq k \text{ a.s. for all } n \geq 1 \}$.

In $\mathbb{R}$ with $\mathcal{Y}_2 = \{0, 1\}$, for $X_n$ a set of iid random variables from $\mathcal{U}(0, 1)$, $\gamma(X_n, N_S) \leq 2$ with equality holding with positive probability. Hence $\kappa(N_S) = 2$.

**Theorem 5.1.** Let $X_n$ be a set of iid random variables from $\mathcal{U}(T(\mathcal{Y}_3))$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$. Then $\kappa(N_{PE}(\cdot, M)) = 3$ for some $N_{PE}(\cdot, M)$.

**Proof:** For $N_{PE}(\cdot, M)$, pick the point closest to edge $e_j$ in vertex region $R_M(y_j)$; that is, pick $U_j \in \text{argmin}_{X \in \mathcal{X}_n \cap R_M(y_j)} d(X, e_j) = \text{argmax}_{X \in \mathcal{X}_n \cap R_M(y_j)} d(t(y, X), y_j)$ in the vertex region for which $X \cap R_M(y_j) \neq \emptyset$.

Then $X \cap R_M(y_j) \subset N_{PE}(U_j, M)$. Hence $X_n \cap \bigcup_{j=1}^3 N_{PE}(U_j, M)$. So $\gamma_n(r, M_C) \leq 3$ with equality holding with positive probability. Hence $\kappa(N_{PE}) = 3$. ■

Below is a general result for the limiting distribution of $\gamma(X_n, N)$ for $X_n$ from a very broad family of distributions and for general $N(\cdot)$.

**Lemma 5.2.** Let $\mathcal{R}_S(N)$ be the superset region for the proximity map $N(\cdot)$ and $X_n$ be a set of iid random variables from $F$ with $P_F(X \in \mathcal{R}_S(N)) > 0$. Then $\lim_{n \to \infty} P_F(\gamma(X_n, N) = 1) = 1$.

**Proof:** Suppose $P_F(X \in \mathcal{R}_S(N)) > 0$. Recall that for any $x \in \mathcal{R}_S(N)$, we have $N(x) = \Omega$, so $X_n \subset N(x)$, hence if $X_n \cap \mathcal{R}_S(N) \neq \emptyset$ then $\gamma(X_n, N) = 1$. Then $P(X_n \cap \mathcal{R}_S(N) \neq \emptyset) \leq P(\gamma(X_n, N) = 1)$. But $P(X_n \cap \mathcal{R}_S(N) \neq \emptyset) = 1 - P(X_n \cap \mathcal{R}_S(N) = \emptyset) = 1 - [1 - P_F(X \in \mathcal{R}_S(N))]^n \to 1$ as $n \to \infty$, since $P_F(X \in \mathcal{R}_S(N)) > 0$. Hence $\lim_{n \to \infty} P(\gamma(X_n, N) = 1) = 1$. ■

**Remark 5.3.** In particular, for $F = \mathcal{U}(T(\mathcal{Y}_3))$, the inequality $P_F(X \in \mathcal{R}_S(N)) > 0$ holds iff $A(\mathcal{R}_S(N)) > 0$, then $P(X_n \cap \mathcal{R}_S(N) \neq \emptyset) \to 1$.

For $\mathcal{Y}_2 = \{0, 1\} \subset \mathbb{R}$, $\mathcal{R}_S(N_S) = \{1/2\}$, so Lemma 5.2 does not apply to $N_S$ in $\mathbb{R}$.

Recall that $\kappa(N_{PE}) = 3$, then

$$1 \leq \mathbb{E} [\gamma_n(r, M)] \leq 3 \text{ and } 0 \leq \mathbb{V}ar [\gamma_n(r, M)] \leq 9/4.$$  

Furthermore, there is a stochastic ordering for $\gamma_n(r, M)$.

**Theorem 5.4.** Suppose $X_n$ is a set of iid random variables from a continuous distribution $F$ on $T(\mathcal{Y}_3)$. Then for $r_1 < r_2$, we have $\gamma_n(r_2, M) \leq ST \gamma(X_n, N_{PE}^2, M)$.

**Proof:** Suppose $r_1 < r_2$. Then $P(\gamma_n(r_2, M) \leq 1) > P(\gamma_n(r_1, M) \leq 1)$ since $\Gamma_1^1(X_n, M) \subset \Gamma_1^2(X_n, M)$ for any realization of $X_n$ and by a similar argument $P(\gamma_n(r_2, M) \leq 2) > P(\gamma_n(r_1, M) \leq 2)$ so $P(\gamma_n(r_2, M) \leq 3) = P(\gamma_n(r_1, M) \leq 3)$. Hence the desired result follows. ■

5.2 Geometry Invariance

We present a “geometry invariance” result for $N_{PE}(\cdot, M)$ where $M$-vertex regions are constructed using the lines joining $\mathcal{Y}_3$ to $M$, rather than the orthogonal projections from $M$ to the edges. This invariance property will simplify the notation in our subsequent analysis by allowing us to consider the special case of the equilateral triangle.

**Theorem 5.5.** (Geometry Invariance Property) Suppose $X_n$ is a set of iid random variables from $\mathcal{U}(T(\mathcal{Y}_3))$. Then for any $r \in [1, \infty]$ the distribution of $\gamma_n(r, M)$ is independent of $\mathcal{Y}_3$ and hence the geometry of $T(\mathcal{Y}_3)$.

**Proof:** Suppose $X \sim \mathcal{U}(T(\mathcal{Y}))$. A composition of translation, rotation, reflections, and scaling will take any given triangle $T(\mathcal{Y}) = T(y_1, y_2, y_3)$ to the basic triangle $T_0 = T((0,0), (1,0), (c_1, c_2))$ with $0 < c_1 \leq 1/2$, $c_2 > 0$, and $(1 - c_1)^2 + c_2^2 \leq 1$. Furthermore, when $X$ is also transformed in the same manner, say to
Then for \( r > \), by Lemma 5.2. \( \blacksquare \)

Suppose \( \gamma(r,M) \) involves only probability content of unions and intersections of regions bounded by precisely such lines and the probability content of such regions is preserved since uniformity is preserved; the desired result follows. \( \blacksquare \)

Note that geometry invariance of \( \gamma(\mathcal{X}_n, N_{PE}^{r=\infty}, M) \) also follows trivially, since for \( r = \infty \), we have \( \gamma_n(r = \infty, M) = 1 \) a.s. for all \( \mathcal{X}_n \) from any \( F \) with support in \( T(\mathcal{Y}_3) \setminus \mathcal{Y}_3 \).

Based on Theorem 5.6, we may assume that \( T(\mathcal{Y}_3) \) is a standard equilateral triangle with \( \mathcal{Y}_3 = \{ (0,0), (1,0), (1/2, \sqrt{3}/2) \} \) for \( N_{PE}(r,M) \) with \( M \)-vertex regions.

Notice that, we proved the geometry invariance property for \( N_{PE} \) where \( M \)-vertex regions are defined with the lines joining \( \mathcal{Y}_3 \) to \( M \). On the other hand, if we use the orthogonal projections from \( M \) to the edges, the vertex regions, hence \( N_{PE} \) will depend on the geometry of the triangle. That is, the orthogonal projections from \( M \) to the edges will not be mapped to the orthogonal projections in the standard equilateral triangle. Hence with the choice of the former type of \( M \)-vertex regions, it suffices to work on the standard equilateral triangle. On the other hand, with the orthogonal projections, the exact and asymptotic distribution of \( \gamma_n \) will depend on \( c_1, c_2 \), so one needs to do the calculations for each possible combination of \( c_1, c_2 \).

### 5.3 The Degenerate Case with \( \gamma_n(r,M) \rightarrow 3 \)

Below, we prove that \( \gamma_n(r,M) \) is degenerate in the limit for \( r > 3/2 \).

**Theorem 5.6.** Suppose \( \mathcal{X}_n \) is a set of iid random variables from a continuous distribution \( F \) on \( T(\mathcal{Y}_3) \). If \( M \notin \mathcal{T}_r \) (see Figure 4 and Equation 4 for \( \mathcal{T}_r \)), then \( \lim_{n \to \infty} P(\gamma_n(r,M) = 1) = 1 \) for all \( M \in \mathbb{R}^2 \setminus \mathcal{Y}_3 \).

**Proof:** Suppose \( M \notin \mathcal{T}_r \). Then \( \mathcal{S}(N_{PE}, M) \) is nonempty with positive area. Hence the result follows by Lemma 5.2. \( \blacksquare \)

**Corollary 5.7.** Suppose \( \mathcal{X}_n \) is a set of iid random variables from a continuous distribution \( F \) on \( T(\mathcal{Y}_3) \). Then for \( r > 3/2 \), \( \lim_{n \to \infty} P(\gamma_n(r,M) = 1) = 1 \) for all \( M \in \mathbb{R}^2 \setminus \mathcal{Y}_3 \).

**Proof:** For \( r > 3/2 \), \( \mathcal{T}_r = \emptyset \), so \( M \notin \mathcal{T}_r \). Hence the result follows by Theorem 5.5. \( \blacksquare \)

We estimate the distribution of \( \gamma_n(r,M) \) with \( r = 2 \) and \( M = M_C \) for various \( n \) empirically. In Table 1 (left), we present the empirical estimates of \( \gamma_n(r,M) \) with \( n = 10, 20, 30, 50, 100 \) based on 1000 Monte Carlo replicates in \( T_r \). Observe that the empirical estimates are in agreement with the asymptotic distribution given in Corollary 5.7.

| \( k \setminus n \) | 10  | 20  | 30  | 50  | 100 |
|----------------|-----|-----|-----|-----|-----|
| 1              | 961 | 1000| 1000| 1000| 1000|
| 2              | 34  | 0   | 0   | 0   | 0   |
| 3              | 5   | 0   | 0   | 0   | 0   |

Table 1: The number of \( \gamma_n(r,M) = k \) out of \( N = 1000 \) Monte Carlo replicates with \( M = M_C \) and \( r = 2 \) (left) and \( r = 5/4 \) (right).

The asymptotic distribution of \( \gamma_n(r,M) \) for \( r < 3/2 \) depends on the relative position of \( M \) with respect to the triangle \( \mathcal{T}_r \).

### 5.4 The Degenerate Case with \( \gamma_n(r,M) \rightarrow 3 \)

**Theorem 5.8.** Suppose \( \mathcal{X}_n \) is a set of iid random variables from a continuous distribution \( F \) on \( T(\mathcal{Y}_3) \). If \( M \in (\mathcal{T}_r)^o \), then \( P(\gamma_n(r,M) = 3) \rightarrow 1 \) as \( n \to \infty \).
We estimate the distribution of \( \gamma_n(r, M) \) with \( r = 5/4 \) and \( M = M_C \) for various \( n \) values empirically. In Table 1(right), we present the empirical estimates of \( \gamma_n(r, M) \) with \( n = 10, 20, 30, 50, 100 \) based on 1000 Monte Carlo replicates in \( T_e \). Observe that the empirical estimates are in agreement with our result in Theorem 5.8.

**Theorem 5.9.** Suppose \( \mathcal{X}_n \) is a set of iid random variables from \( \mathcal{U}(T(Y_3)) \). If \( M \in \partial(T) \), then \( P(\gamma_n(r, M) > 1) \to 1 \) as \( n \to \infty \).

For \( M \in \partial(T) \), there are two separate cases:

(i) \( M \in \partial(T) \setminus \{t_1(r), t_2(r), t_3(r)\} \) where \( t_j(r) \) with \( j \in \{1, 2, 3\} \) are the vertices of \( T \) whose explicit forms are given in Equation (2).

(ii) \( M \in \{t_1(r), t_2(r), t_3(r)\} \).

**Theorem 5.10.** Suppose \( \mathcal{X}_n \) is a set of iid random variables from \( \mathcal{U}(T(Y_3)) \). If \( M \in \partial(T) \setminus \{t_1(r), t_2(r), t_3(r)\} \), then \( P(\gamma_n(r, M) = 3) \to 1 \) as \( n \to \infty \).

We estimate the distribution of \( \gamma_n(r, M) \) with \( r = 5/4 \) and \( M = (3/5, \sqrt{3}/10) \) in Table 1 with \( r, M \) for various \( n \) empirically. In Table 2 we present empirical estimates of \( \gamma_n(r, M) \) with \( n = 10, 20, 30, 50, 100, 500, 1000, 2000 \) based on 1000 Monte Carlo replicates in \( T_e \). Observe that the empirical estimates are in agreement with our result in Theorem 5.10.

| \( k \backslash n \) | 10 | 20 | 30 | 50 | 100 | 500 | 1000 | 2000 |
|---------------------|----|----|----|----|-----|-----|------|------|
| 1                   | 118| 60 | 51 | 39 | 15  | 2   | 1    | 1    |
| 2                   | 402| 409| 361| 299| 258 | 100 | 57   | 29   |
| 3                   | 420| 531| 588| 662| 727 | 899 | 941  | 970  |

Table 2: The number of \( \gamma_n(r, M) = k \) out of \( N = 1000 \) Monte Carlo replicates with \( r = 5/4 \) and \( M = (3/5, \sqrt{3}/10) \).

### 5.5 The Nondegenerate Case

**Theorem 5.11.** Suppose \( \mathcal{X}_n \) is a set of iid random variables from \( \mathcal{U}(T(Y_3)) \). If \( M \in \{t_1(r), t_2(r), t_3(r)\} \), then \( P(\gamma_n(r, M) = 2) \to p_r \) as \( n \to \infty \) where \( p_r \in (0, 1) \) is provided in Equation (4) but only numerically computable.

For example, \( p_{r=5/4} \approx .6514 \) and \( p_{r=\sqrt{2}} \approx .4826 \).

So the asymptotic distribution of \( \gamma_n(r, M) \) with \( r \in [1, 3/2) \) and \( M \in \{t_1(r), t_2(r), t_3(r)\} \) is given by

\[
\gamma_n(r, M) \sim 2 + \text{BER}(1 - p_r).
\]

(7)

We estimate the distribution of \( \gamma_n(r, M) \) with \( r = 5/4 \) and \( M = (7/10, \sqrt{3}/10) \) for various \( n \) empirically. In Table 3 we present the empirical estimates of \( \gamma_n(r, M) \) with \( n = 10, 20, 30, 50, 100, 500, 1000, 2000 \) based on 1000 Monte Carlo replicates in \( T_e \). Observe that the empirical estimates are in agreement with our result \( p_{r=5/4} \approx .6514 \).

| \( k \backslash n \) | 10 | 20 | 30 | 50 | 100 | 500 | 1000 | 2000 |
|---------------------|----|----|----|----|-----|-----|------|------|
| 1                   | 174| 118| 82 | 61 | 22  | 5   | 1    | 1    |
| 2                   | 532| 526| 548| 561| 611 | 617 | 633  | 649  |
| 3                   | 294| 356| 370| 378| 367 | 378 | 366  | 330  |

Table 3: The number of \( \gamma_n(r, M) = k \) out of \( N = 1000 \) Monte Carlo replicates with \( r = 5/4 \) and \( M = (7/10, \sqrt{3}/10) \).
Remark 5.12. For $r = 3/2$, as $n \to \infty$, $P(\gamma_n(r, M_C) > 1) \to 1$ at rate $O(n^{-1})$. □

**Theorem 5.13.** Suppose $X_n$ is a set of iid random variables from $\mathcal{U}(T(3))$. Then for $r = 3/2$, as $n \to \infty$,
\[
\gamma_n(3/2, M_C) \sim 2 + \text{BER}(p \approx .2487)
\]  
(8)

For the proof of Theorem 5.13 see Ceyhan and Priebe (2004, 2005).

Using Theorem 5.13
\[
\lim_{n \to \infty} \mathbb{E} [\gamma_n(3/2, M_C)] = 3 - p_{3/2} \approx 2.2587
\]  
(9)

and
\[
\lim_{n \to \infty} \text{Var} [\gamma_n(3/2, M_C)] = 6 + p_{3/2}^2 - p_{3/2}^2 \approx .1917.
\]  
(10)

Indeed, the finite sample distribution of $\gamma_n(3/2, M_C)$ hence the finite sample mean and variance can also be obtained by numerical methods.

We also estimate the distribution of $\gamma_n(3/2, M_C)$ for various $n$ values empirically. The empirical estimates for $n = 10, 20, 30, 50, 100, 500, 1000, 2000$ based on 1000 Monte Carlo replicates are given in Table 4; estimates are in agreement with our result $p_r = 3/2 \approx .7413$.

| $k \backslash n$ | 10 | 20 | 30 | 50 | 100 | 500 | 1000 | 2000 |
|-----------------|----|----|----|----|-----|-----|------|------|
| 1               | 151| 82 | 61 | 50 | 27  | 2   | 3    | 1    |
| 2               | 602| 630| 688| 693| 718 | 753 | 729  | 749  |
| 3               | 247| 282| 251| 257| 255 | 245 | 268  | 250  |

Table 4: The number of $\gamma_n(3/2, M_C) = k$ out of $N = 1000$ Monte Carlo replicates.

### 5.6 Distribution of the $\gamma_n(r, M)$ in Multiple Triangles

So far we have worked with data in one Delaunay triangle, i.e., $m = 3$ or $J_3 = 1$. In this section, we present the asymptotic distribution of the domination number of $r$-factor PCDs in multiple Delaunay triangles. Suppose $Y_m = \{y_1, y_2, \ldots, y_m\} \subset \mathbb{R}^2$ be a set of $m$ points in general position with $m > 3$ and no more than 3 points are cocircular. Then there are $J_m > 1$ Delaunay triangles each of which is denoted as $T_j$. Let $M^j$ be the point in $T_j$ that corresponds to $M$ in $T_c$, $\mathcal{T}^j$ be the triangle that corresponds to $T_j$ in $T_c$, and $t^j_i(r)$ be the vertices of $\mathcal{T}^j$ that correspond to $t_i(r)$ in $T_c$ for $i \in \{1, 2, 3\}$. Moreover, let $n_j = |X_n \cap T_j|$, the number of $X$ points in Delaunay triangle $T_j$. For $X_n \subset \mathcal{U}(Y_m)$, let $\gamma_{nj}(r, M^j)$ be the domination number of the digraph induced by vertices of $T_j$ and $X_n \cap T_j$. Then the domination number of the $r$-factor PCD in $J_m$ triangles is
\[
\gamma_n(r, M, J_m) = \sum_{j=1}^{J_m} \gamma_{nj}(r, M^j).
\]

See Figure 10 (left) for the 77 $A'$ points that are in $\mathcal{C}_H(Y_m)$ out of the 200 $A$ points plotted in Figure 1. Observe that 10 points yield $J_{10} = 13$ Delaunay triangles. In Figure 10 (right) are the corresponding arcs for $M = M_C$ and $r = 3/2$. The corresponding $\gamma_n = 22$. Suppose $X_n$ is a set of iid random variables from $\mathcal{U}(\mathcal{C}_H(Y_m))$, the uniform distribution on convex hull of $Y_m$, and we construct the $r$-factor PCDs using the points $M^j$ that correspond to $M$ in $T_c$. Then for fixed $m$ (or fixed $J_m$), as $n \to \infty$, so does each $n_j$. Furthermore, as $n \to \infty$, each component $\gamma_{nj}(r, M^j)$ become independent. Therefore using Equation 3, we can obtain the asymptotic distribution of $\gamma_n(r, M, J_m)$ as $n \to \infty$, for fixed $J_m$,
\[
\gamma_n(r, M, J_m) \sim \begin{cases} 
2J_m + \text{BIN}(J_m, 1-p_r), & \text{for } M^j \in \{t^j_1(r), t^j_2(r), t^j_3(r)\} \text{ and } r \in [1,3/2], \\
J_m, & \text{for } r > 3/2, \\
3J_m, & \text{for } M \in \mathcal{T}^j \setminus \{t^j_1(r), t^j_2(r), t^j_3(r)\} \text{ and } r \in [1,3/2],
\end{cases}
\]  
(11)

where $\text{BIN}(n, p)$ stands for binomial distribution with $n$ trials and probability of success $p$, for $r \in [1,3/2]$ and $M \in \{t_1(r), t_2(r), t_3(r)\}$, $p_r$ is given in Equation 3 and for $r = 3/2$ and $M = M_C$, $p_r \approx .7413$ (see Equation 5).
5.7 Extension of $N_{PE}^r$ to Higher Dimensions

The extension to $\mathbb{R}^d$ for $d > 2$ with $M = M_C$ is provided in Ceyhan and Priebe (2005), but the extension for general $M$ is similar.

Let $\gamma_n(r, M, d) := \gamma(\mathcal{X}_n, N_{PE}^r, M, d)$ be the domination number of the PCD based on the extension of $N_{PE}^r(\cdot, M)$ to $\mathbb{R}^d$. Then it is easy to see that $\gamma_n(r, M, 3)$ is nondegenerate as $n \to \infty$ for $r = 4/3$. In $\mathbb{R}^d$, it can be seen that $\gamma_n(r, M, d)$ is nondegenerate in the limit only when $r = (d + 1)/d$. Furthermore, for large $d$, asymptotic distribution of $\gamma_n(r, M, d)$ is nondegenerate at values of $r$ closer to 1. Moreover, it can be shown that $\lim_{n \to \infty} P(2 \leq \gamma_n(r = (d + 1)/d, M, d) \leq d + 1) = 1$ and we conjecture the following.

Conjecture 5.14. Suppose $\mathcal{X}_n$ is set of iid random variables from the uniform distribution on a simplex in $\mathbb{R}^d$. Then the domination number $\gamma_n(r, M)$ in the simplex satisfies

$$
\lim_{n \to \infty} P(d \leq \gamma_n((d + 1)/d, M, d) \leq d + 1) = 1.
$$

For instance, with $d = 3$ we estimate the empirical distribution of $\gamma(\mathcal{X}_n, 4/3)$ for various $n$. The empirical estimates for $n = 10, 20, 30, 40, 50, 100, 200, 500, 1000, 2000$ based on 1000 Monte Carlo replicates for each $n$ are given in Table 5.

| $k \setminus n$ | 10  | 20  | 30  | 40  | 50  | 100 | 200 | 500 | 1000 | 2000 |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-------|-------|
| 1                | 52  | 18  | 5   | 5   | 4   | 0   | 0   | 0   | 0     | 0     |
| 2                | 385 | 308 | 263 | 221 | 219 | 155 | 88  | 41  | 31    | 19    |
| 3                | 348 | 455 | 557 | 609 | 621 | 725 | 773 | 831 | 845   | 862   |
| 4                | 215 | 219 | 175 | 165 | 156 | 120 | 139 | 128 | 124   | 119   |

Table 5: The number of $\gamma_n(4/3, M_C) = k$ out of $N = 1000$ Monte Carlo replicates.

6 Discussion

The $r$-factor proportional-edge proximity catch digraphs (PCDs), when compared to class cover catch digraphs (CCCDs), have some advantages. The asymptotic distribution of the domination number $\gamma_n(r, M)$
of the $r$-factor PCDs, unlike that of CCCDs, is mathematically tractable (computable by numerical integration). A minimum dominating set can be found in polynomial time for $r$-factor PCDs in $\mathbb{R}^d$ for all $d \geq 1$, but finding a minimum dominating set is an NP-hard problem for CCCDs (except for $\mathbb{R}$). These nice properties of $r$-factor PCDs are due to the geometry invariance of distribution of $\gamma_n(r, M)$ for uniform data in triangles.

On the other hand, CCCDs are easily extendable to higher dimensions and are defined for all $X_n \subset \mathbb{R}^d$, while $r$-factor PCDs are only defined for $X_n \subset CH(Y_m)$. Furthermore, the CCCDs based on balls use proximity regions that are defined by the obvious metric, while the PCDs in general do not suggest a metric. In particular, our $r$-factor PCDs are based on some sort of dissimilarity measure, but no metric underlying this measure exists.

The finite sample distribution of $\gamma_n(r, M)$, although computationally tedious, can be found by numerical methods, while that of CCCDs can only be empirically estimated by Monte Carlo simulations. Moreover, we had to introduce many auxiliary tools to compute the distribution of $\gamma_n(r, M)$ in $\mathbb{R}^2$. Same tools will work in higher dimensions, perhaps with more complicated geometry.

The $r$-factor PCDs have applications in classification and testing spatial patterns of segregation or association. The former can be performed building discriminant regions for classification in a manner analogous to the procedure proposed in [Priebe et al. (2003a)]; and the latter can be performed by using the asymptotic distribution of $\gamma_n(r, M)$ similar to the procedure used in [Ceyhan and Priebe (2003)].

Acknowledgements

This work was partially by the Defense Advanced Research Projects Agency as administered by the Air Force Office of Scientific Research under contract DOD F49620-99-1-0213 and by Office of Naval Research Grant N00014-95-1-0777. We also thank anonymous referees, whose constructive comments and suggestions greatly improved the presentation and flow of this article.

References

Ceyhan, E. (2004). An Investigation of Proximity Catch Digraphs in Delaunay Tessellations. PhD thesis, 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. (2004). On the distribution of the domination number of random $r$-factor proportional-edge proximity catch digraphs. Technical Report 651, Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218.

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.

Ceyhan, E., Priebe, C. E., and Marchette, D. J. (2005). A new family of random graphs for testing spatial segregation. Submitted for publication. (Available as Technical Report No. 644, with title Relative density of random $r$-factor proximity catch digraph for testing spatial patterns of segregation and association. Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD 21218.)

Ceyhan, E., Priebe, C. E., and Wierman, J. C. (2006). Relative density of the random $r$-factor proximity catch digraphs for testing spatial patterns of segregation and association. Computational Statistics & Data Analysis, 50(8):1925–1964.

Chartrand, G. and Lesniak, L. (1996). Graphs & Digraphs. Chapman & Hill.

DeVinney, J. (2003). The Class Cover Problem and its Applications in Pattern Recognition. PhD thesis, The Johns Hopkins University, Baltimore, MD, 21218.
Appendix

First, we begin with a remark that introduces some terminology which we will use for asymptotics throughout this appendix.

Remark 6.1. Suppose $X_n$ is a set of iid random variables from $F$ with support $S(F) \subseteq \Omega$. If over a sequence $\Omega_n \subseteq \Omega$, $n = 1, 2, 3, \ldots$, $X$ restricted to $\Omega_n$, $X|_{\Omega_n}$, has distribution $F_n$ with $F_n(x) = F(x)/P_F(X \in \Omega_n)$ and $P_F(X \in \Omega_n) \to 1$ as $n \to \infty$, then we call $F_n$ the asymptotically accurate distribution of $X$ and $\Omega_n$ the asymptotically accurate support of $F$. If $F$ has density $f$, then $f_n = f(x)/P_F(X \in \Omega_n)$ is called the asymptotically accurate pdf of $X$. In both cases, if we are concerned with asymptotic results, for simplicity we will, respectively, use $F$ and $f$ for asymptotically accurate distribution and pdf. Conditioning will be implied by stating that $X \in \Omega_n$ with probability 1, as $n \to \infty$ or for sufficiently large $n$. □

Proof of Theorem 4.4

Without loss of generality, assume $T(Y_b) = T_6 = T((0, 0), (1, 0), (c_1, c_2))$. Note that the probability of edge extrema all being equal to each other is $P(X_{e_1}(n) = X_{e_2}(n) = X_{e_3}(n)) = I(n = 1)$. Let $E_{e_1}(n)$ be the event that there are only two distinct (closest) edge extrema. Then for $n > 1$,

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

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

By simple integral calculus, we can show that $P(E_{e_3}(n)) \to 0$ as $n \to \infty$, which will imply the desired result. □
Proof of Theorem 5.8

Note that $(\mathcal{F}, \gamma) \neq \emptyset$ iff $r < 3/2$. Suppose $M \in (\mathcal{F})^\gamma$. Then for any point $u \in R_M(y_j), N_{P_E}(u, M) \subseteq T(\gamma_3)$, because there is a tiny strip adjacent to edge $e_j$ not covered by $N_{P_E}(u, M)$, for each $j \in \{1, 2, 3\}$. Then, $N_{P_E}(u, M) \cup N_{P_E}(v, M) \subseteq T(\gamma_3)$ for all $(u, v) \in R_M(y_1) \times R_M(y_2)$. Pick $\sup_{(u,v)\in R_M(y_1)\times R_M(y_2)} N_{P_E}(u, M) \cup N_{P_E}(v, M) \subseteq T(\gamma_3)$. Then $T(\gamma_3) \setminus \{\sup_{(u,v)\in R_M(y_1)\times R_M(y_2)} N_{P_E}(u, M) \cup N_{P_E}(v, M)\}$ has positive area. So $X_n \cap [T(\gamma_3) \setminus \{\sup_{(u,v)\in R_M(y_1)\times R_M(y_2)} N_{P_E}(u, M) \cup N_{P_E}(v, M)\}] \neq \emptyset$

with probability 1 for sufficiently large $n$. (The supremum of a set functional $A(x)$ over a range $B$ is defined as the set $S := \sup_{x \in B} A(x)$ such that $S$ is the smallest set satisfying $A(x) \subseteq S$ for all $x \in B$.) Then at least three points—one for each vertex region—are required to dominate $X_n$. Hence for sufficiently large $n$, $\gamma_n(r, M) \geq 3$ with probability 1, but $\kappa(N_{P_E}) = 3$ by Theorem 5.4. Then $\lim_{n \to \infty} P(\gamma_n(r, M) = 3) = 1$ for $r < 3/2$. ■

Proof of Theorem 5.9

Let $M = (m_1, m_2) \in \partial(\mathcal{F})$, say $M \subseteq q_3(r, x)$ (recall that $q_3(r, x)$ are defined such that $d(y_j, e_j) = r \cdot d(q_j(r, x), y_j)$ for $j \in \{1, 2, 3\}$, then $m_2 = \frac{\sqrt{3}(2-r)}{2}$ and $m_1 = \frac{(3r-1)}{2}$. Let $X_n(n)$ be one of the closest point(s) to the edge $e_j$; i.e., $X_n(n) \in \arg\min_{X \in X_n} d(X, e_j)$ for $j \in \{1, 2, 3\}$. Let $X_n(n)$ be the circle a.s. for each $j$.

Notice that for all $j \in \{1, 2, 3\}$, $X_n(n) \subseteq N_{P_E}(X)$ for all $X \in X_n \cap R_M(y_j)$ implies that $\gamma_n(r, M) > 1$ with probability 1. For sufficiently large $n$, $X_n(n) \subseteq N_{P_E}(X)$ for all $X \in X_n \cap R_M(y_j)$ with probability 1, for $j \in \{1, 2\}$, by the choice of $M$. Hence we consider only $X_n(n)$. The asymptotically accurate pdf of $X_n(n)$ is

$f_{x_3}(x, y) = n \left(\frac{A(S_U(x, y)}{A(T(\gamma_3))}\right)^{n-1} \frac{1}{A(T(\gamma_3))},$

where $S_U(x, y)$ is the unshaded region in Figure 11 (left) (for a given $X_n(n) = x_{x_3} = (x, y)$) whose area is $A(S_U(x, y)) = \sqrt{3} \left(2y - \sqrt{3}\right)^2/12$. Note that $X_{x_3}(n) \not\subseteq N_{P_E}(X)$ for all $X \subseteq X_n \cap R_M(y_3)$ if $X_n \cap \Gamma_1^1 \cap R_M(y_3) \neq \emptyset$. Then given $X_{x_3}(n) = (x, y), P(X_n \cap \Gamma_1^1 \cap R_M(y_3) = \emptyset) = \left(\frac{A(S_U(x, y)) - A \left(\Gamma_1^1 \cap R_M(y_3)\right)}{A(S_U(x, y))}\right)^{n-1}.$

where $A(\Gamma_1^1 \cap R_M(y_3)) = \frac{\sqrt{3} \pi^2}{2(r-1)^2}$ (see Figure 11 (right) where the shaded region is $\Gamma_1^1 \cap R_M(y_3)$ for a given $X_{x_3}(n) = (x, y)$), then for sufficiently large $n$

$P(X_n \cap \Gamma_1^1 \cap R_M(y_3) = \emptyset) \approx \int \left(\frac{A(S_U(x, y)) - A \left(\Gamma_1^1 \cap R_M(y_3)\right)}{A(S_U(x, y))}\right)^{n-1} f_{x_3}(x, y) \, dy \, dx = \frac{n}{A(T(\gamma_3))} \left(\frac{\sqrt{3} \left(2y - \sqrt{3}\right)^2}{12} - \frac{\sqrt{3} y^2}{3(r-1)r}\right)^{n-1} \, dy \, dx.$

Let

$G(x, y) = \frac{A(S_U(x, y)) - A \left(\Gamma_1^1 \cap R_M(y_3)\right)}{A(T(\gamma_3))} = \frac{4}{\sqrt{3}} \left(\frac{\sqrt{3} \left(2y - \sqrt{3}\right)^2}{12} - \frac{\sqrt{3} y^2}{3(r-1)r}\right),$

which is independent on $x$, so we denote it as $G(y)$.

Let $\varepsilon > 0$ be sufficiently small, then for sufficiently large $n$,

$P(X_n \cap \Gamma_1^1 \cap R_M(y_3) = \emptyset) \approx \int_0^\varepsilon \int_0^{1-y/\sqrt{3}} n \, G(y) \, n^{-1} 4/\sqrt{3} \, dy \, dx = \left(1 - 2y/\sqrt{3}\right) \int_0^{\varepsilon} n \, G(y) \, n^{-1} 4/\sqrt{3} \, dy.$

The integrand is critical at $y = 0$, since $G(0) = 1$ (i.e., when $x_{x_3} \in e_3$). Furthermore, $G(y) = 1 - 4y/\sqrt{3} + O \left(y^2\right)$ around $y = 0$. Then letting $y = w/n$, we get

$P(X_n \cap \Gamma_1^1 \cap R_M(y_3) = \emptyset) \approx \left(1 - \frac{2w}{\sqrt{3} n}\right) \frac{4}{\sqrt{3}} \int_0^{n \varepsilon} \left(1 - \frac{4w}{\sqrt{3} n} + O \left(n^{-2}\right)\right)^{n-1} \, dw.$

letting $n \to \infty$, \[\approx 4/\sqrt{3} \int_0^\infty \exp \left(-4w/\sqrt{3}\right) \, dw = 1.

Hence $\lim_{n \to \infty} P(\gamma_n(r, M) > 1) = 1$. For $M \subseteq q_3(r, x) \cap \mathcal{F}$ with $j \in \{1, 2\}$ the result follows similarly. ■
Figure 11: A figure for the description of the pdf of $X_{c_3}(n)$ (left) and $\Gamma_1'(X_n, M)$ (right) given $X_{c_3}(n) = x_{c_3} = (x, y)$.  

**Proof of Theorem 5.10**

Let $M = (m_1, m_2) \in \partial(\mathcal{T}) \setminus \{t_1(r), t_2(r), t_3(r)\}$, say $M \in q_3(r, x)$. Then $m_2 = \frac{\sqrt{r(n-1)}}{2r}$. Without loss of generality, assume $\frac{1}{2} \leq m_1 < \frac{3}{2r}$. See also Figure 12.

Figure 12: A figure for the description of the pdf of $\bar{Q}_1(n)$ and $\bar{Q}_3(n)$ (left) and the unshaded region is $N_{PE}^r(\bar{q}_1, M) \cup N_{PE}^r(\bar{q}_3, M)$ (right).

Whenever $X_n \cap R_M(y_j) \neq \emptyset$, let

\[ \bar{Q}_j(n) \in \arg\max_{X \in X_n \cap R_M(y_j)} d(X, e_j) = \arg\max_{X \in X_n \cap R_M(y_j)} d(\ell(y_j, X), y_j) \text{ for } j \in \{1, 2, 3\}. \]

Note that at least one of the $\bar{Q}_j(n)$ uniquely exists w.p. 1 for finite $n$ and as $n \to \infty$, $\bar{Q}_j(n)$ are unique w.p. 1. Then

\[ \gamma_n(r, M) \leq 2 \text{ iff } X_n \subset \left[ N_{PE}^r(\bar{Q}_1(n), M) \cup N_{PE}^r(\bar{Q}_2(n), M) \right] \text{ or } X_n \subset \left[ N_{PE}^r(\bar{Q}_2(n), M) \cup N_{PE}^r(\bar{Q}_3(n), M) \right] \text{ or } X_n \subset \left[ N_{PE}^r(\bar{Q}_1(n), M) \cup N_{PE}^r(\bar{Q}_3(n), M) \right]. \]

Let $E_{n,j}^{1,3}$ be the event that $X_n \subset N_{PE}^r(\bar{Q}_i, M) \cup N_{PE}^r(\bar{Q}_j(n), M)$ for $(i, j) \in \{(1, 2), (1, 3), (2, 3)\}$. Then

\[ P(\gamma_n(r, M) \leq 2) = P(E_{n,1}^{1,2}) + P(E_{n,2}^{2,3}) + P(E_{n,3}^{3,1}) - P(E_{n,1}^{1,2} \cap E_{n,2}^{2,3}) - P(E_{n,1}^{1,2} \cap E_{n,3}^{3,1}) - P(E_{n,2}^{2,3} \cap E_{n,3}^{3,1}) + P(E_{n,1}^{1,2} \cap E_{n,2}^{2,3} \cap E_{n,3}^{3,1}). \]

But note that $P(E_{n,1}^{1,2}) \to 0$ as $n \to \infty$ by the choice of $M$ since

\[ \sup_{u \in R_M(y_1)} N_{PE}^r(u, M) \cup N_{PE}^r(v, M) \subset T(Y_3), \]
and
\[
P \left( \mathcal{A} \cap T(\mathcal{Y}_3) \backslash \left\{ \sup_{u \in R_M(y_3)} N_{PE}(u, M) \cup N_{PE}(v, M) \right\} = \emptyset \right) \to 1 \text{ as } n \to \infty.
\]

Then,
\[
P \left( E_n^{1,2} \right) = P \left( E_n^{1,2} \cap E_n^{2,3} \right) + P \left( E_n^{1,2} \cap E_n^{2,3} \cap E_n^{3,4} \right) \leq 4 P \left( E_n^{1,2} \right) \to 0 \text{ as } n \to \infty.
\]

Therefore,
\[
\lim_{n \to \infty} P \left( \gamma_n(r, M) \leq 2 \right) = \lim_{n \to \infty} \left( P \left( E_n^{2,3} \right) + P \left( E_n^{3,4} \right) \right).
\]

Furthermore, observe that \( P \left( E_n^{3,4} \right) \geq P \left( E_n^{2,3} \right) \) by the choice of \( M \). Then we first find \( \lim_{n \to \infty} P \left( E_n^{3,4} \right) \). Given a realization of \( \mathcal{X}_n \) with \( \tilde{Q}_1(n) = \tilde{q}_1 = (x_1, y_1) \) and \( \tilde{Q}_3(n) = \tilde{q}_3 = (x_3, y_3) \), the remaining \( n - 2 \) points should fall, for example, in the undshaded region in Figure 12 (left). Then the asymptotically accurate joint pdf of \( \tilde{Q}_1(n), \tilde{Q}_3(n) \) is
\[
f_{13} (\tilde{z}) = \frac{n(n - 1)}{A(T(\mathcal{Y}_3))^2} \left( \frac{A(T(\mathcal{Y}_3)) - A(S_R(\tilde{z}))}{A(T(\mathcal{Y}_3))} \right)^{n-2}
\]
where \( \tilde{z} = (x_1, y_1, x_3, y_3) \). \( S_R(\tilde{z}) \) is the shaded region in Figure 12 (left) whose area is \( A(S_R(\tilde{z})) = \frac{\sqrt{3} (2 r y_3 - \sqrt{3} (r - 1))^2}{3 r (r - 1)} \). Given \( \tilde{Q}_j(n) = \tilde{q}_j = (x_j, y_j) \) for \( j \in \{1, 3\} \),
\[
P \left( E_n^{3,4} \right) = \left( \frac{A(N_{PE}(\tilde{q}_1, M) \cup N_{PE}(\tilde{q}_3, M)) - A(S_R(\tilde{z}))}{A(T(\mathcal{Y}_3)) - A(S_R(\tilde{z}))} \right)^{n-2}
\]
then for sufficiently large \( n \)
\[
P \left( E_n^{3,4} \right) \approx \int f_{13} (\tilde{z}) d\tilde{z},
\]
where
\[
A(N_{PE}(\tilde{q}_1, M) \cup N_{PE}(\tilde{q}_3, M)) = \frac{\sqrt{3}}{4} - \left( \frac{\sqrt{3} r y_1 + 3 r x_1 - 3 (\sqrt{3} r - 1) - 2 r y_3}{6} \right).
\]
See Figure 12 (right) for \( N_{PE}(\tilde{q}_1, M) \cup N_{PE}(\tilde{q}_3, M) \). Let
\[
G(\tilde{z}) = \frac{A(N_{PE}(\tilde{q}_1, M) \cup N_{PE}(\tilde{q}_3, M)) - A(S_R(\tilde{z}))}{A(T(\mathcal{Y}_3))}.
\]
Note that the integral is critical at \( x_1 = x_3 = m_1 \) and \( y_1 = y_3 = m_2 \), since \( G(\tilde{z}) = 1 \). Since \( N_{PE}(x, M_C) \) depends on the distance \( d(x, e_j) \) for \( x \in R_M(y_j) \), we make the change of variables \( (x_1, y_1) \to (d(M, e_1) + z_1, y_1) \) where
\[
d(M, e_1) = \frac{\sqrt{3} (r + 1 - 2 r m_1)}{4 r} \text{ and } (x_3, y_3) \to (x_3, m_2 + z_3) \text{ then } G(z_1, z_3) \text{ depends only on } z_1, z_3, \text{ we denote it } G(z_1, z_3) \text{ which is}
\]
\[
G(z_1, z_3) = 1 - \frac{2 \sqrt{3} z_3^2}{3} = \frac{4 r z_3^2}{3 (r - 1)} - \frac{2 r z_3 \sqrt{3} (3 r - 3 r) + r (4 z_1 - 2 \sqrt{3} m_1)}{3}.
\]
The new integrand is \( \frac{n(n - 1)}{A(T(\mathcal{Y}_3))^2} G(z_1, z_3)^{n-2} \). Integrating with respect to \( x_3 \) and \( y_1 \) yields \( \frac{\sqrt{3} (r + 1 - 2 r m_1)}{6} \) and \( \frac{4 \sqrt{3} r z_1}{3 (2 r m_1 + r - 1)} \), respectively. Hence for sufficiently large \( n \)
\[
P \left( E_n^{3,4} \right) \approx \int_0^\infty \int_0^\infty \left( \frac{n(n - 1)}{A(T(\mathcal{Y}_3))^2} \left( \frac{2 \sqrt{3} r z_3^2}{3 (r - 1)} \right) \frac{4 \sqrt{3} r z_1}{3 (2 r m_1 + r - 1)} \right) G(z_1, z_3)^{n-2} dz_1 dz_3.
\]
Note that the new integral is critical when \( z_1 = z_3 = 0 \), so we make the change of variables \( z_1 = w_1 / \sqrt{3} \) and \( z_3 = w_3 / n \) then \( G(z_1, z_3) \) becomes
\[
G(w_1, w_3) = 1 - \left( \frac{2 \sqrt{3} r (r - 3 + 2 r m_1)}{3} + \frac{8 r}{3 (r + 1 - 2 r m_1)} w_1^2 \right) + O \left( \frac{n^{-3/2}}{20} \right),
\]

so for sufficiently large \( n \)

\[
P (E_{1,3}^n) \approx \int_0^{\sqrt{n}} \int_0^{n - \epsilon} (n - 1) \frac{16}{3} \left( 2 \sqrt{3} r \right) \left( \frac{4 \sqrt{3} r}{3 (2 m_1 - r - 1)} \right) (-4 m_1 + 2 + \sqrt{2}) w_1 w_3
\]

\[
\left[ 1 - \frac{1}{n} \left( \frac{2 \sqrt{3} r (3 + 2 r m_1)}{3} w_3 + \frac{8 r}{3 (r + 1 - 2 r m_1)} w_3^2 \right) + O \left( n^{-3/2} \right) \right]^{n-2} dw_3 w_1,
\]

\[
\approx O \left( n^{-1} \right) \int_0^{\infty} w_1 w_3 \exp \left( -2 \sqrt{3} r (3 + 2 r m_1) w_3 - \frac{8 r w_3^2}{3 (r + 1 - 2 r m_1)} \right) \] dw_3 w_1 = O \left( n^{-1} \right)
\]

since \( \int_0^{\infty} w_1 w_3 \exp \left( \frac{-2 \sqrt{3} r (3 + 2 r m_1)}{3} w_3 - \frac{8 r w_3^2}{3 (r + 1 - 2 r m_1)} \right) \) \( dw_3 w_1 = \frac{3}{8 r (3 - r (m_2 + 1))} \), which is a finite constant. Then \( P (E_{1,3}^n) \to 0 \) as \( n \to \infty \), which also implies \( P (E_{2,3}^n) \to 0 \) as \( n \to \infty \). Then \( P (\gamma_n(r, M) \leq 2) \to 0 \).

Hence the desired result follows. \( \blacksquare \)

**Proof of Theorem 5.11**

Let \( M = (m_1, m_2) \in \{ t_1(r), t_2(r), t_3(r) \} \). Without loss of generality, assume \( M = t_2(r) \) then \( m_1 = \frac{2 - r + c_1 (r - 1)}{r} \) and \( m_2 = \frac{c_2 (r - 1)}{r} \). See Figure 13.

Figure 13: A figure for the description of the pdf of \( \hat{Q}_1(n) \) and \( \hat{Q}_3(n) \) (left) and the unshaded region is \( N_{\hat{Y}} (\hat{q}_1, M) \cup N_{\hat{Y}} (\hat{q}_3, M) \) (right) given \( \hat{Q}_j(n) = \hat{q}_j \) for \( j \in \{1, 3\} \).

Let \( \hat{Q}_j(n) \) and the events \( E_{i,j}^n \) be defined as in the proof of Theorem 5.10 for \( (i,j) \in \{ (1,2), (1,3), (2,3) \} \). Then as in the proof of Theorem 5.10

\[
P (\gamma_n(r, M) \leq 2) = P (E_{1,3}^n) + P (E_{2,3}^n) + P (E_{1,3}^n - P (E_{1,2}^n \cap E_{2,3}^n) -
\]

\[
P (E_{1,2}^n \cap E_{2,3}^n) - P (E_{1,3}^n \cap E_{2,3}^n) + P (E_{1,2}^n \cap E_{2,3}^n \cap E_{1,3}^n).
\]

Observe that the choice of \( M \) implies that \( P (E_{1,3}^n) \geq P (E_{2,3}^n) \) and by symmetry (in \( T_i \)) \( P (E_{1,2}^n) = P (E_{2,3}^n) \). So first we find \( P (E_{1,3}^n) \). As in the proof of Theorem 5.10 asymptotically accurate joint pdf of \( \hat{q}_1(n), \hat{q}_3(n) \) is

\[
f_{13} (\zeta) = \frac{n (n - 1)}{A (T (3 \zeta))} \left( A (T (3 \zeta)) - A (S_R (\zeta)) \right)^{n-2}
\]

where \( \zeta = (x, y_1, x_3, y_3) \) and \( S_R (\zeta) \) is the shaded region in Figure 13 (left) whose area is

\[
A (S_R (\zeta)) = \frac{\sqrt{3} (2 r y_3 - \sqrt{3} (r - 1)^2)}{12 (r - 1) r} + \frac{\sqrt{3} \left( \sqrt{3} r y_1 + 3 x_1 r - 3 \right)^2}{36 (r - 1) r}.
\]

Given \( \hat{Q}_j(n) = \hat{q}_j = (x_j, y_j) \) for \( j \in \{1, 3\} \),

\[
P (E_{1,3}^n) = \left( \frac{A (N_{\hat{Y}} (\hat{q}_1, M) \cup N_{\hat{Y}} (\hat{q}_3, M)) - A (S_R (\zeta))}{A (T (3 \zeta))} \right)^{n-2},
\]
then for sufficiently large \( n \)

\[
P(E_n^{1,3}) \approx \int \left( \frac{A \left( N_{PF}(\tilde{Q} = \tilde{q}, M) \cup N_{PF}(\tilde{q}, M) \right) - A(S_R(\tilde{z}))}{A(T(\tilde{y}_3))} \right)^{n-2} f_{z_3}(\tilde{z}) \, d\tilde{z},
\]

\[
= \int \frac{n(n-1)}{A(T(\tilde{y}_3))^2} \left( \frac{A \left( N_{PF}(\tilde{q}, M) \cup N_{PF}(\tilde{q}, M) \right) - A(S_R(\tilde{z}))}{A(T(\tilde{y}_3))} \right)^{n-2} d\tilde{z},
\]

where

\[
A(N_{PF}(\tilde{q}, M) \cup N_{PF}(\tilde{q}, M)) = \frac{\sqrt{3}}{4} \left( \frac{2 r y_3 - \sqrt{3}(r - 1)}{6} \right) (3 - \sqrt{3} r y_1 - 3 r x_1).
\]

See Figure 13 (right) for \( N_{PF}(\tilde{q}, M) \cup N_{PF}(\tilde{q}, M) \). Let

\[
G(\tilde{z}) = \frac{A \left( N_{PF}(\tilde{q}, M) \cup N_{PF}(\tilde{q}, M) \right) - A(S_R(\tilde{z}))}{A(T(\tilde{y}_3))}.
\]

Note that the integral is critical when \( x_1 = x_3 = m_1 \) and \( y_1 = y_3 = m_2 \), since \( G(\tilde{z}) = 1 \).

As in the proof of Theorem 5.10, we make the change of variables \( (x, y_1) \rightarrow (d(M, e_1) + z, y) \) where \( d(M, e_1) = \frac{\sqrt{3}(r - 1)}{2 r} \) and \( (x, y_3) \rightarrow (x, m_2 + z) \). Then \( G(\tilde{z}) \) becomes

\[
G(z_1, z_3) = 1 - \frac{4 r}{3 (r - 1)} z_1 - \frac{4 r}{3 (r - 1)} z_3^2 - \frac{8 r^2}{3} z_1 z_3.
\]

The new integral is

\[
\int \left( \frac{n(n-1)}{A(T(\tilde{y}_3))^2} G(z_1, z_3)^{n-2} d\tilde{x_3} d\tilde{y}_3 d\tilde{z}_3 d\tilde{z}_1.
\]

Note that \( G(z_1, z_3) \) is independent of \( y_1, y_3 \), so integrating with respect to \( x_3 \) and \( y_1 \) yields \( 2 \sqrt{\pi} r^{\frac{3}{2} (r - 1)} \) and \( 2 \sqrt{\pi} r^{\frac{3}{2} (r - 1)} \), respectively. The new integral is critical at \( z_1 = z_3 = 0 \). Hence, for sufficiently large \( n \) and sufficiently small \( \varepsilon > 0 \), the integral becomes,

\[
P(E_n^{1,3}) \approx \int_{\varepsilon}^{\sqrt{\pi} \varepsilon} \int_{\varepsilon}^{\sqrt{\pi} \varepsilon} \left( \frac{n(n-1)}{A(T(\tilde{y}_3))^2} \left( \frac{12 r^2}{9 (r - 1)^2} \right) z_1 z_3 \right)^{n-2} d\tilde{x_3} d\tilde{y}_3 d\tilde{z}_3 d\tilde{z}_1.
\]

Since the new integral is critical when \( z_1 = z_3 = 0 \), we make the change of variables \( z_j = w_j / \sqrt{n} \) for \( j \in \{1, 3\} \); then \( G(z_1, z_3) \) becomes

\[
G(w_1, w_3) = 1 - \frac{4 r}{3 n (r - 1)} (w_1^2 + w_3^2 + 2 r (r - 1) w_1 w_3),
\]

so

\[
p_r := P(E_n^{1,3}) \approx \int_{\varepsilon}^{\sqrt{\pi} \varepsilon} \left( \frac{n(n-1)}{A(T(\tilde{y}_3))^2} \left( \frac{12 r^2}{9 (r - 1)^2} \right) w_1 w_3 \right)^{n-2} d\tilde{x_3} d\tilde{y}_3 d\tilde{z}_3 d\tilde{z}_1.
\]

\[
\approx \int_{0}^{\infty} \int_{0}^{\infty} \left( \frac{4 r}{3 n (r - 1)} (w_1^2 + w_3^2 + 2 r (r - 1) w_1 w_3) \right)^{n-2} d\tilde{x_3} d\tilde{y}_3 d\tilde{z}_3 d\tilde{z}_1.
\]

which is not analytically integrable, but \( p_r \) can be obtained by numerical integration, e.g., \( p_r = 0.4826 \) and \( p_r = 0.6514 \).

Next, we find \( \lim_{n \to \infty} P(E_n^{2,3}) \). The asymptotically accurate joint pdf of \( \tilde{Q}_2(n), \tilde{Q}_3(n) \) is

\[
f_{z_3} (\tilde{z}) = \frac{n(n-1)}{A(T(\tilde{y}_3))^2} \left( \frac{A(T(\tilde{y}_3)) - A(S_R(\tilde{z}))}{A(T(\tilde{y}_3))} \right)^{n-2},
\]

where \( \tilde{z} = (x_2, y_2, x_3, y_3) \) and \( S_R(\tilde{z}) \) is the shaded region in Figure 14 (left) whose area is

\[
A(S_R(\tilde{z})) = \frac{\sqrt{3} (2 r y_3 + \sqrt{3}(1 - r))}{12 r (r - 1)} + \frac{\sqrt{3} (\sqrt{3} r y_2 - 3 r x_2 - 3 r + 6)}{36 (2 - r) r},
\]

22
Figure 14: A figure for the description of the pdf of $\hat{q}_2(n)$ and $\hat{q}_3(n)$ (left) and unshaded region is $N^\dagger_{PE}(\hat{q}_2) \cup N^\dagger_{PE}(\hat{q}_3)$ (right) given $\hat{q}_j(n) = q_j$ for $j \in \{2, 3\}$.

As before,

$$P \left( E_{n}^{2,3} \right) = \int \frac{A(N^\dagger_{PE}(\hat{q}_2, M) \cup N^\dagger_{PE}(\hat{q}_3, M)) - A(S_R(\zeta))}{A(T(Y_3)) - A(S_R^2(\zeta))} \frac{n-2}{n} f_{23}(\zeta) d\zeta,$$

where $A(N^\dagger_{PE}(\hat{q}_2, M) \cup N^\dagger_{PE}(\hat{q}_3, M)) = \sqrt{\frac{2}{2\pi}} - \frac{2}{6} \sqrt{r_y - (r - 1)} \left( \frac{\sqrt{r_y} + 2 + r_y - 3r}{6} \right)$. See Figure 14 (right) for $N^\dagger_{PE}(\hat{q}_2) \cup N^\dagger_{PE}(\hat{q}_3, M)$. Let

$$G(\zeta) = \frac{A(N^\dagger_{PE}(\hat{q}_2, M) \cup N^\dagger_{PE}(\hat{q}_3, M)) - A(S_R(\zeta))}{A(T(Y_3))}.$$

Note that the integral is critical when $x_2 = x_3 = m_1$ and $y_2 = y_3 = m_2$, since $G(\zeta) = 1$.

We make the change of variables $(x_3, y_3) \rightarrow (x_3, m_2 + z_3)$ and $(x_2, y_2) \rightarrow (m_1, e_2 + z_2, y_2)$ where $d(M, e_2) = \sqrt{\frac{2}{2\pi r_y}}$. Then $G(\zeta)$ becomes

$$G(z_2, z_3) = 1 - \frac{4r z_2^2}{3(2-r)} - \frac{4r z_2^2}{3(r-2)} - \frac{4\sqrt{3}r z_3 (3 - 2r)}{3} - \frac{8r^2 z_2 z_3}{3}.$$

The new integral is

$$\int \frac{n(n-1)}{A(T(Y_3))^2} G(z_2, z_3)^{n-2} d\zeta.$$

The integrand is independent of $x_3$ and $y_2$, so integrating with respect to $x_3$ and $y_2$ yields $\frac{2 \sqrt{r_y z_3}}{z_3(2-r)^2}$ and $\frac{2 \sqrt{r_y z_3}}{z_3(2-r)}$, respectively. Hence, for sufficiently large $n$

$$P \left( E_{n}^{2,3} \right) \approx \int_0^1 \int_0^1 \frac{n(n-1)}{A(T(Y_3))^2} \left( \frac{4r^2}{3(r-1)(2-r)} \right) z_3 z_2 G(z_2, z_3)^{n-2} d\zeta.$$

Note that the new integral is critical when $z_2 = z_3 = 0$, so we make the change of variables $z_2 = w_2/\sqrt{n}$ and $z_3 = w_3/n$ then $G(z_2, z_3)$ becomes

$$G(w_2, w_3) = 1 - \frac{1}{n} \left[ \frac{4r w_2^2}{3(2-r)} - \frac{4\sqrt{3} r w_3 (3 - 2r)}{3} \right] + O \left( \frac{1}{n^{3/2}} \right),$$
so for sufficiently large $n$

$$P(E_{n}^{2,3}) \approx \int_{0}^{\infty} \int_{0}^{n} \frac{(n-1) - 64 r^2}{9 (r-1) (2-r) n^2} w_2 w_3 \left[ 1 - \frac{1}{n} \left( \frac{4 r w_2^2}{3 (2-r)} - \frac{4 \sqrt{3} r w_3 (3-2r)}{3} \right) + O \left( n^{-\frac{3}{2}} \right) \right] w_3 dw_2,$$

$$\approx O \left( n^{-1} \right) \int_{0}^{\infty} \int_{0}^{\infty} \frac{w_2 w_3}{3 (2-r)} \exp \left( -\frac{4 r w_2^2}{3 (2-r)} - \frac{4 \sqrt{3} r u_3 (3-2r)}{3} \right) dw_3 dw_2 = O \left( n^{-1} \right)$$

since

$$\int_{0}^{\infty} \int_{0}^{\infty} \frac{w_2 w_3}{3 (2-r)} \exp \left( -\frac{4 r w_2^2}{3 (2-r)} - \frac{4 \sqrt{3} r u_3 (3-2r)}{3} \right) dw_3 dw_2 = \frac{27 (2-r)}{384 r^3 (3-2r)^2}$$

which is a finite constant.

Thus we have shown that $P(E_{n}^{2,3}) \to 0$ as $n \to \infty$, which implies that as $n \to \infty$,

$$P(E_{n}^{2,3}) + P(E_{n}^{1,2}) - P(E_{n}^{1,2} \cap E_{n}^{2,3}) - P(E_{n}^{1,2} \cap E_{n}^{1,3}) - P(E_{n}^{1,3} \cap E_{n}^{2,3}) + P(E_{n}^{1,2} \cap E_{n}^{2,3} \cap E_{n}^{1,3}) \leq 5 P(E_{n}^{2,3}) \to 0.$$

Hence $\lim_{n \to \infty} P(\gamma_n(r, M) \leq 2) = \lim_{n \to \infty} P(E_{n}^{1,3})$ and $\lim_{n \to \infty} P(\gamma_n(r, M) > 1) = 1$ together imply that

$$\lim_{n \to \infty} P(\gamma_n(r, M) = 2) = p_r.$$

\[\square\]