An Investigation of Proximity Catch Digraphs in Delaunay Tessellations

Elvan Ceyhan

ceyhan@ams.jhu.edu

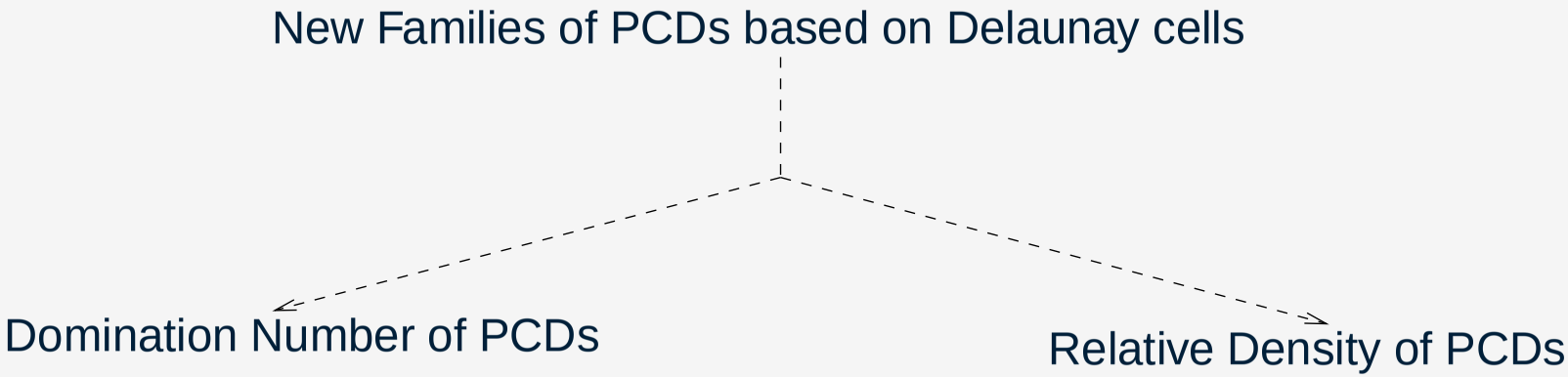
Department of Applied Mathematics and Statistics

Johns Hopkins University

Baltimore, Maryland

Result Categories

- Three major result categories:



- Investigate the one triangle case and then the multiple triangle case.

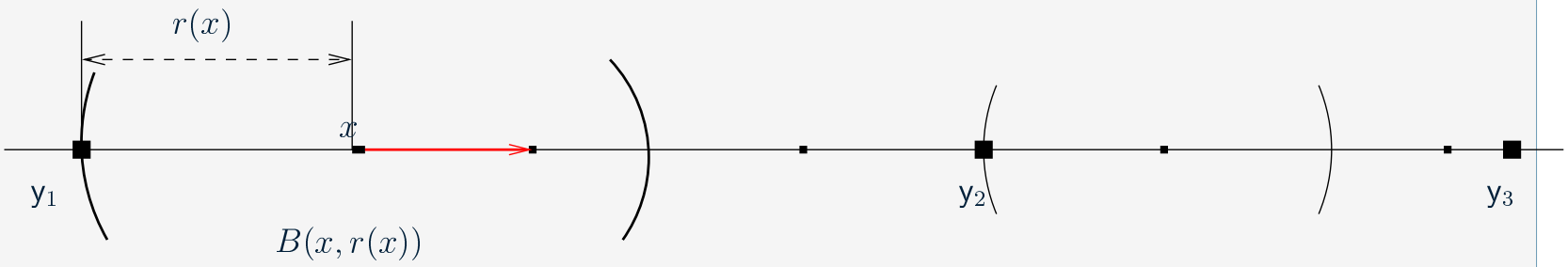
Motivation

- **Motivation:**
  - Mathematical tractability
  - Geometry invariance for uniform data.
  - applicability to hypothesis testing and pattern classification.

  (A direct extension of CCCD to $\mathbb{R}^d$ lacks the first two.)
- devise a test for spatial point patterns

Spherical Proximity Maps

- Spherical Proximity Maps are the building blocks of CCCDs.
- Given two sets of points, $\mathcal{X}_n$ and $\mathcal{Y}_m$, the spherical proximity region of an $x \in \mathcal{X}_n$, $N_S(x) := B(x, r(x))$ where $B(x, r(x))$ is the open ball centered at x with radius $r(x) := \min_{\mathsf{y} \in \mathcal{Y}_m} d(x, \mathsf{y})$.



- In $\mathbb{R}$ with $\mathcal{X}_n$ and $\mathcal{Y}_m$ sets of iid random variables from $\mathcal{U}(0, 1)$, Priebe et al. give the exact and asymptotic distribution of the domination number.

$\mathcal{X}$ points and the Delaunay Triangulation

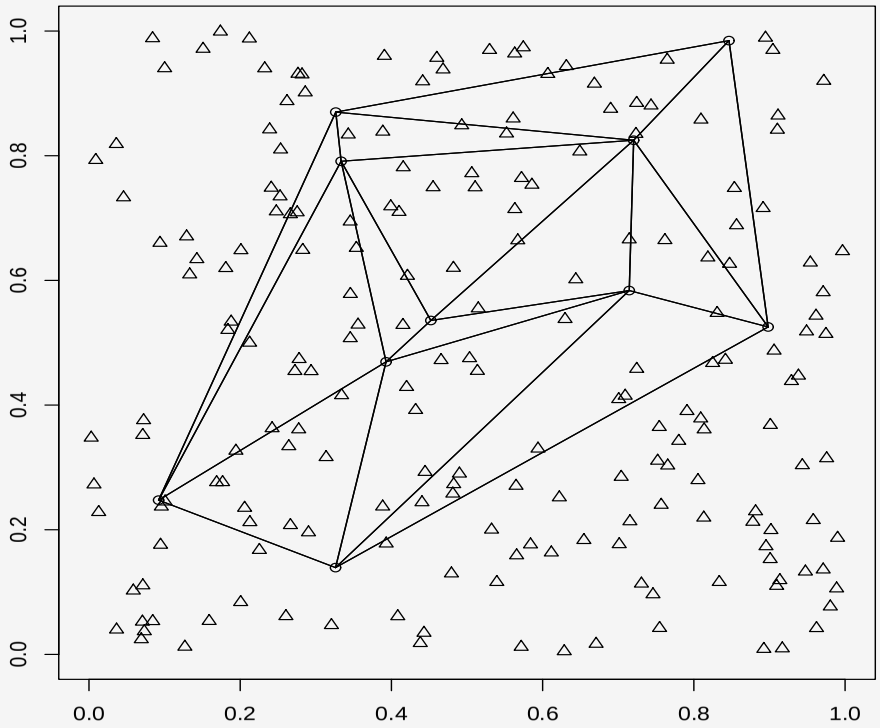


Figure 1: 200 $\mathcal{X}$ points and the Delaunay triangulation of 10 $\mathcal{Y}$ points.

$\mathcal{X}$ points in convex hull of $\mathcal{Y}_m$

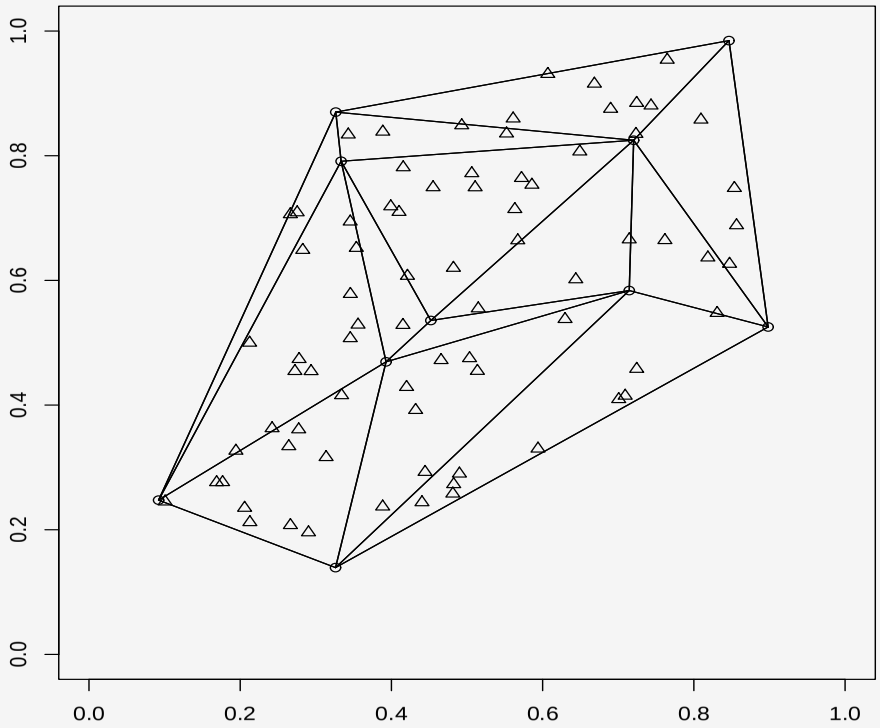


Figure 2: 77 $\mathcal{X}$ points in the Delaunay triangulation of 10 $\mathcal{Y}$ points.

Delaunay Triangulation

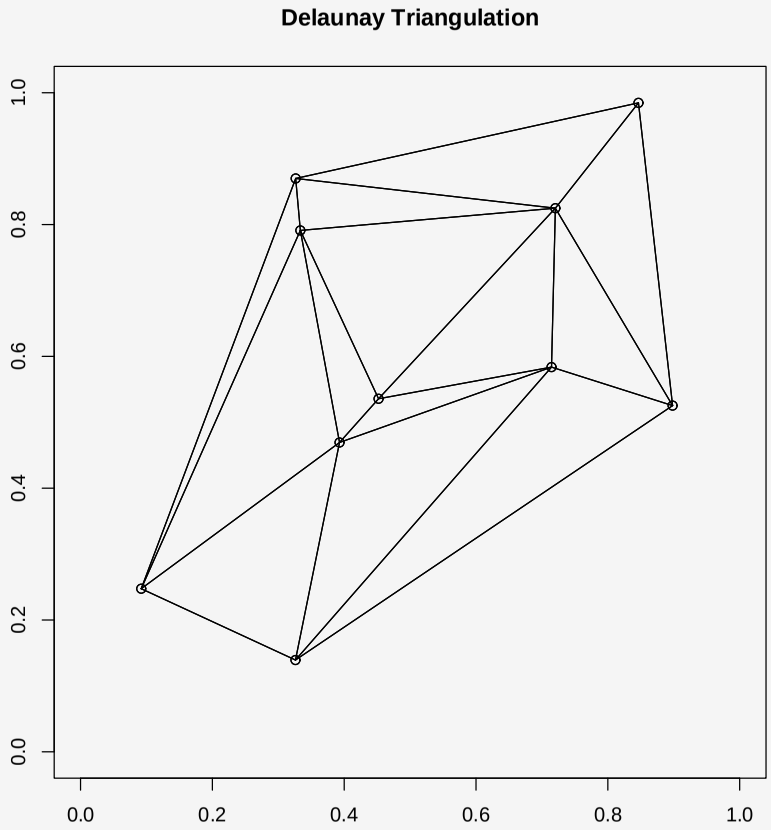


Figure 3: The Delaunay triangulation of the 10 $\mathcal{Y}$ points.

Voronoi Diagram

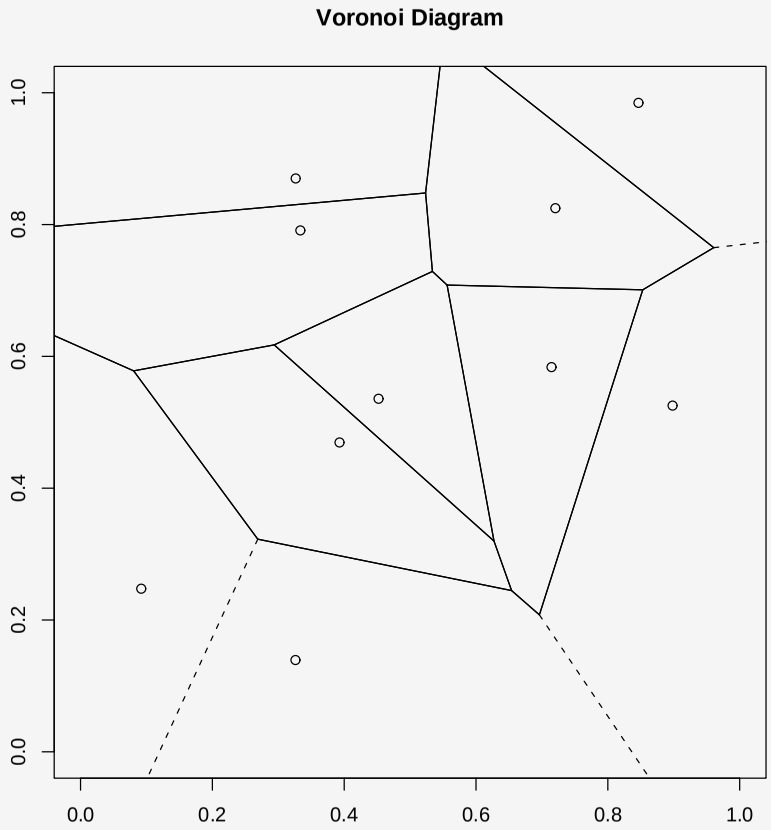


Figure 4: The Voronoi diagram based on the 10 $\mathcal{Y}$ points.

Spherical Proximity Regions

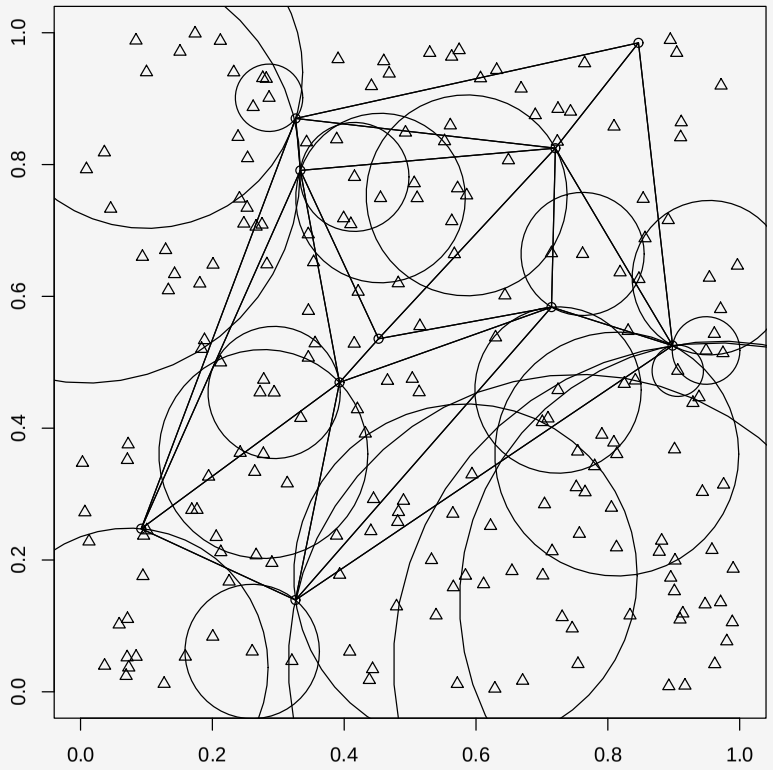


Figure 5: The spherical proximity regions for 20 arbitrarily selected $\mathcal{X}$ points.

Arcs in CCCD (or spherical PCD)

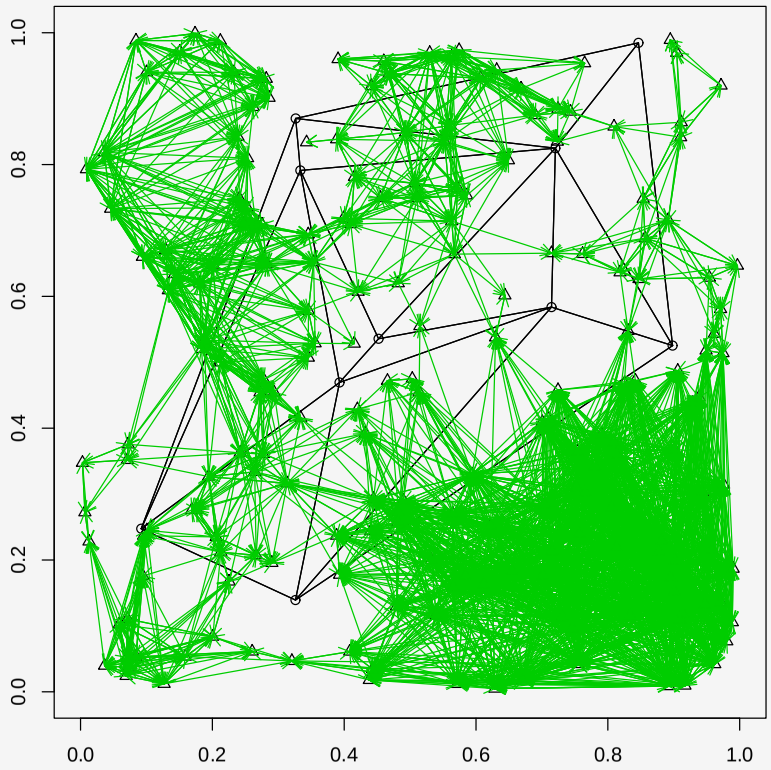


Figure 6: The arcs in the CCCD constructed with the 200 $\mathcal{X}$ points and 10 $\mathcal{Y}$ points.

Arcs in CCCD (or spherical PCD)

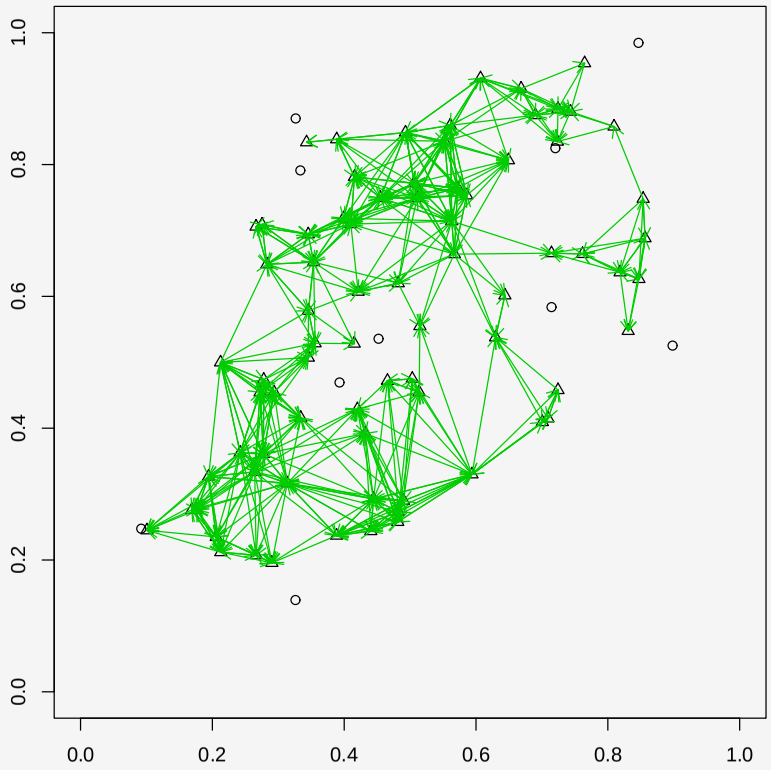


Figure 7: The arcs in the CCCD constructed with the 77 $\mathcal{X}$ points in convex hull of the $\mathcal{Y}$ points.

Arcs in CCCD (or spherical PCD)

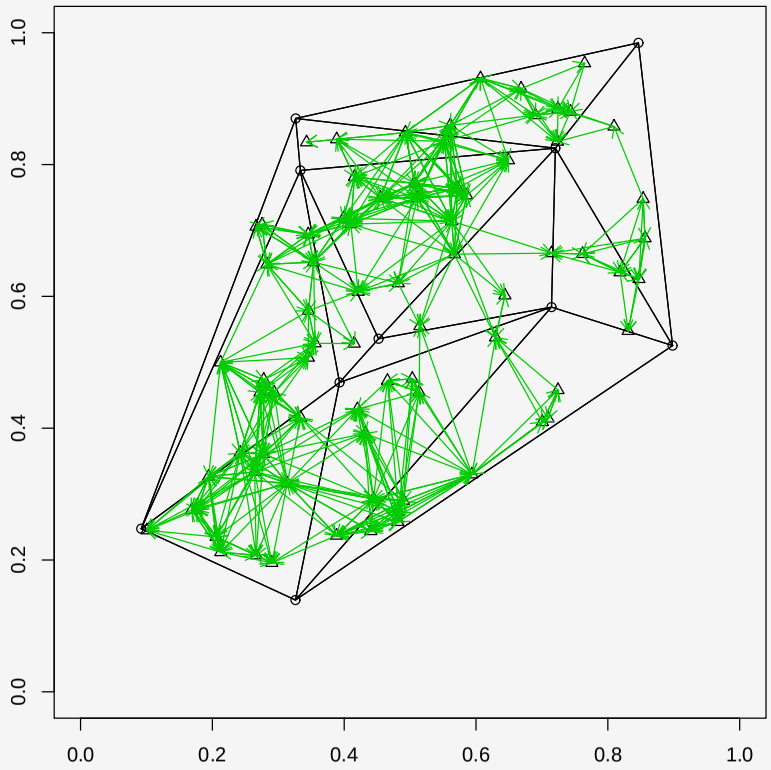


Figure 8: The arcs in the CCCD constructed with the 77 $\mathcal{X}$ points in convex hull of the $\mathcal{Y}$ points.

Definitions

- Let $\mathcal{Y} = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\} \subset \mathbb{R}^2$ be three non-collinear points, $T(\mathcal{Y})$ be the triangle based on $\mathcal{Y}$, and $\mathcal{X}_n$ be a random sample from uniform distribution, $\mathcal{U}(T(\mathcal{Y}))$.
- **Arc-Slice Proximity Maps**
  $N_{AS}(x) := \overline{B}(x, r(x)) \cap T(\mathcal{Y})$ where $\overline{B}(x, r(x))$ is the closed ball centered at x with radius $r(x) := \min_{\mathsf{y} \in \mathcal{Y}} d(x, \mathsf{y})$.

Arc-Slice Proximity Maps

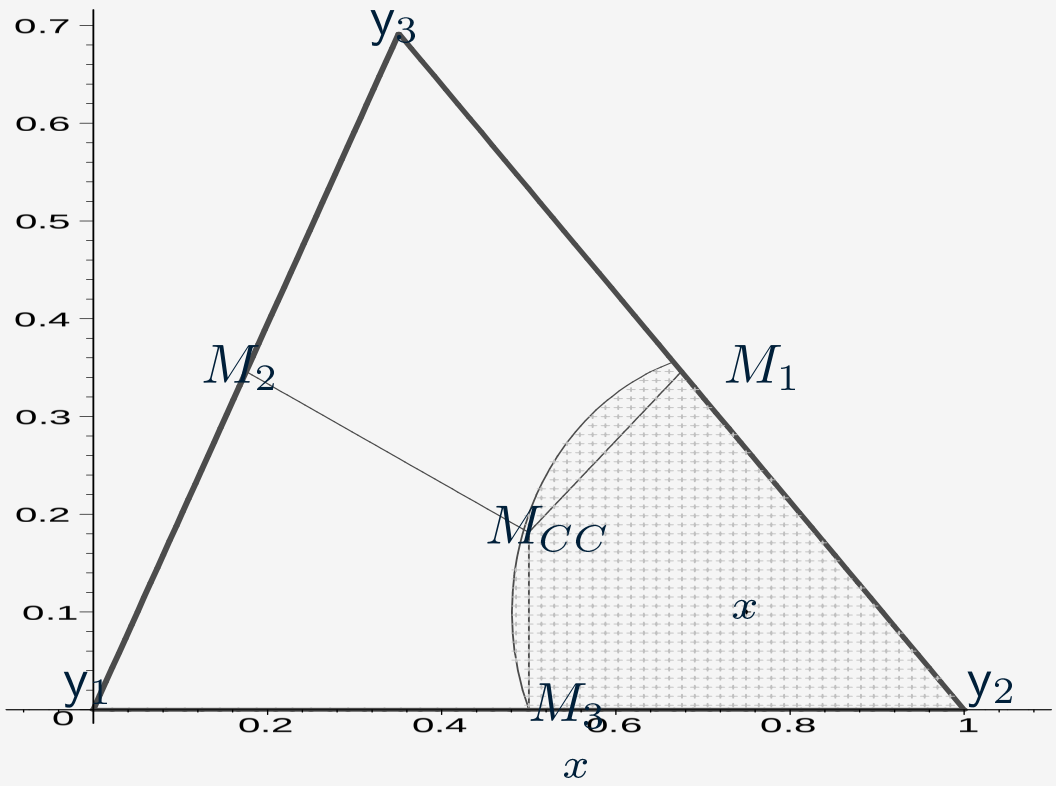


Figure 9: $N_{AS}(x, M_{CC})$ with an $x \in R_{CC}(\mathsf{y}_2)$.

Arcs in A-S PCD in one triangle

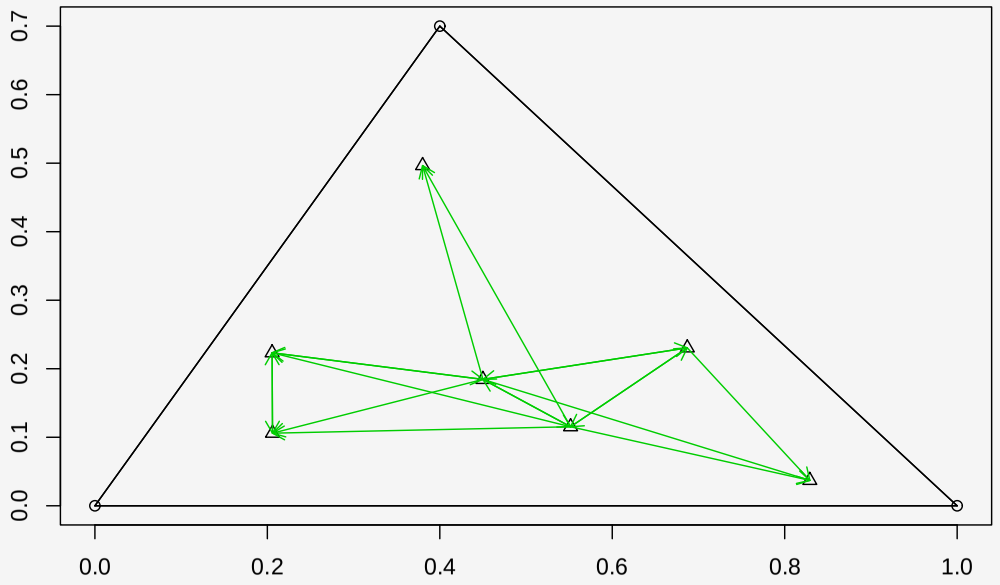


Figure 10: The arcs in the A-S PCD constructed with the 7 $\mathcal{X}$ points in the triangle $T(\mathcal{Y})$.

Arcs in the Arc-Slice PCD

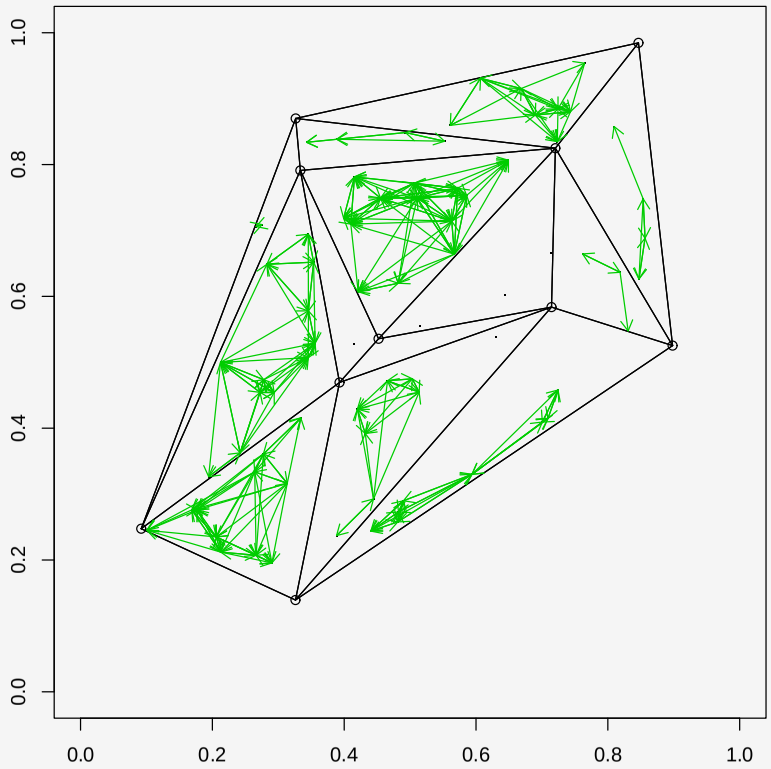


Figure 11: The arcs in the arc-slice PCD constructed with the 77 $\mathcal{X}$ points in $C_H(\mathcal{Y})$.

Γ_1-Region for N_{AS}

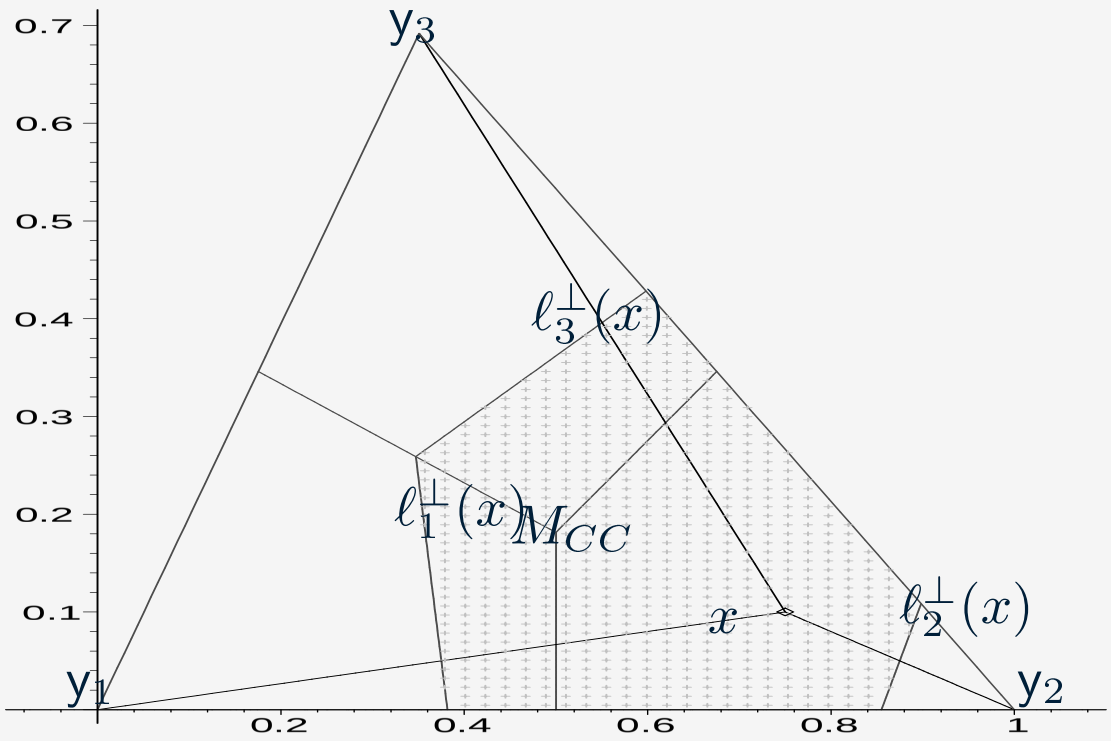


Figure 12: $\Gamma_1(x, N_{AS}, M_{CC})$ and $x \in R_{CC}(\mathsf{y}_2)$.

Γ_1-Region for N_{AS}

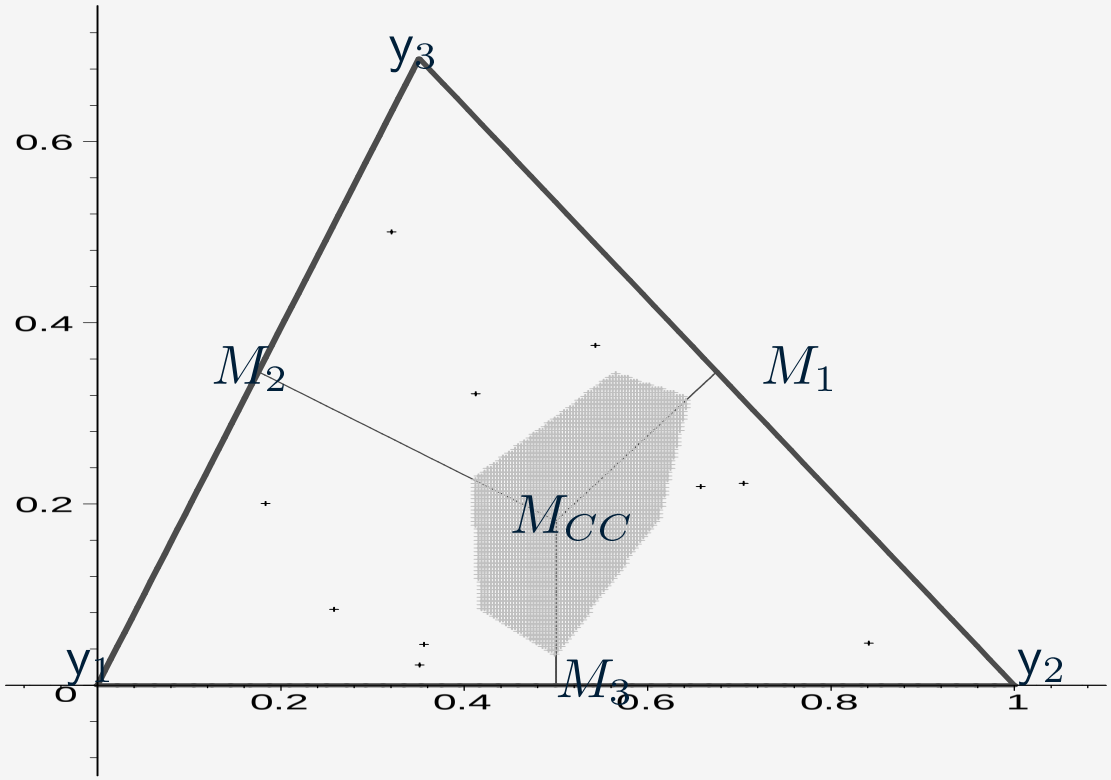


Figure 13: An empirical Γ_1-region, $\Gamma_1(\mathcal{X}_n, N_{AS}, M_{CC})$ with $n = 10$

r-Factor Proportional-Edge Proximity Maps

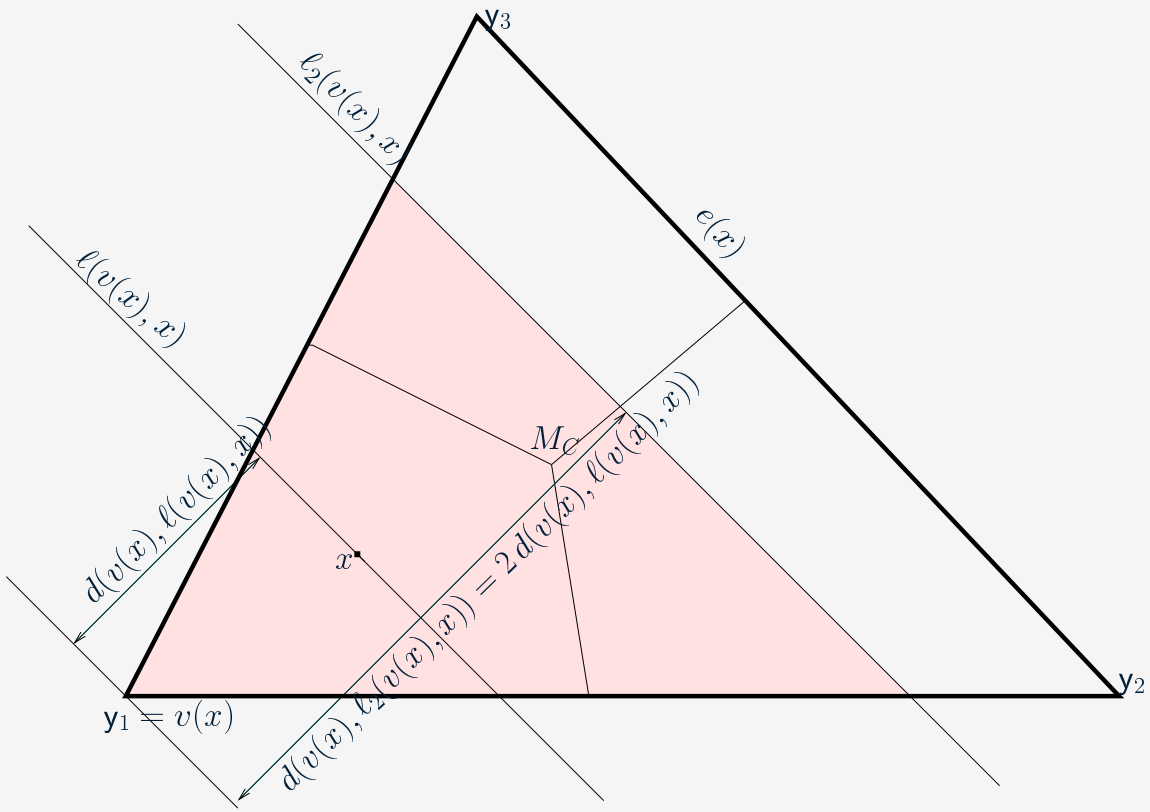


Figure 14: Construction of $N^2_{PE}(x)$ (shaded region).

Arcs in r-factor P-E PCD in one triangle

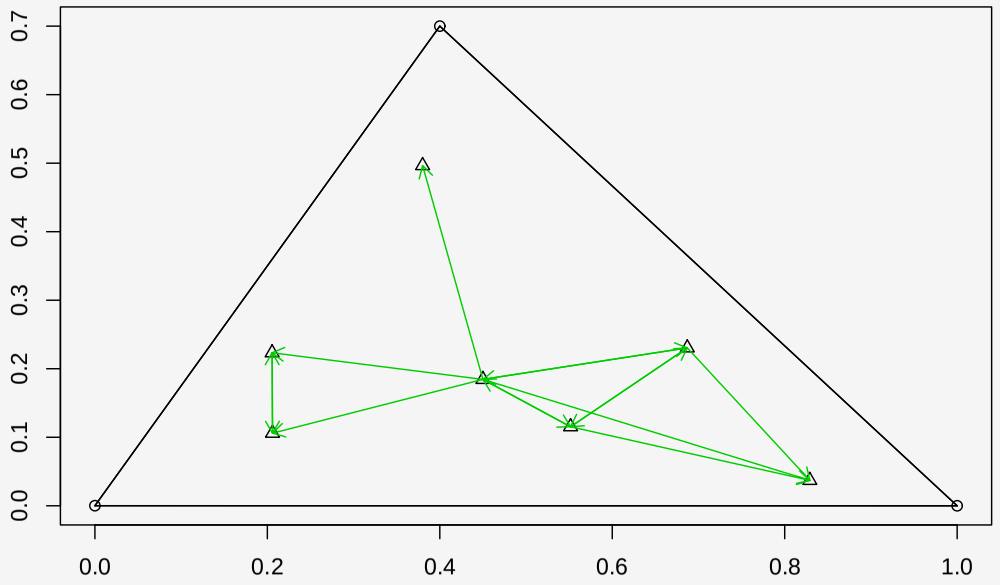


Figure 15: The arcs in the r-factor P-E PCD with $r = 1.5$, 7 $\mathcal{X}$ points in the triangle $T(\mathcal{Y})$.

Arcs in r-factor P-E PCD in multiple triangles

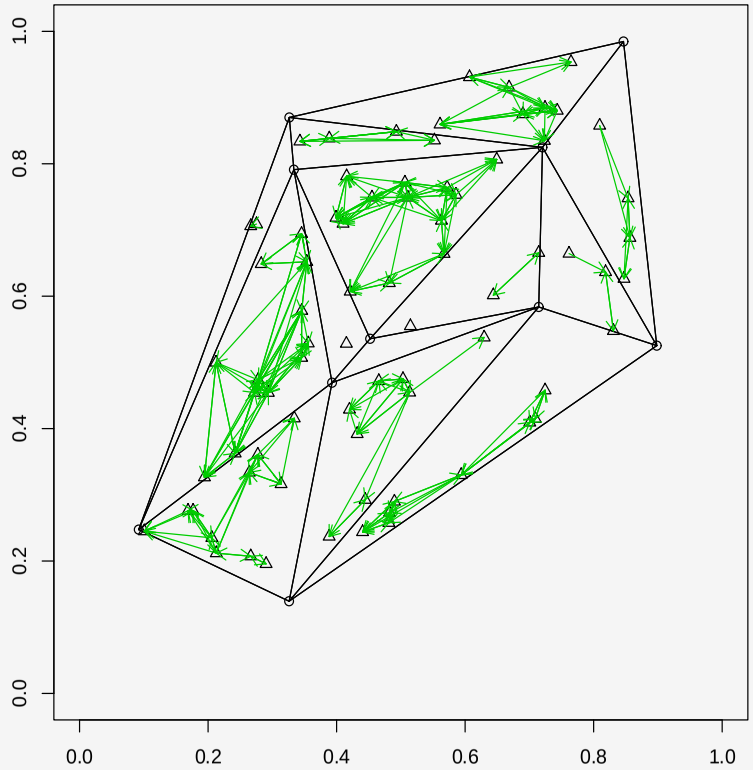


Figure 16: The arcs in the r-factor P-E PCD with $r = 1.5$ and the 77 $\mathcal{X}$ points in the $C_H(\mathcal{Y})$.

Types of Γ_1-regions for N^r_{PE}

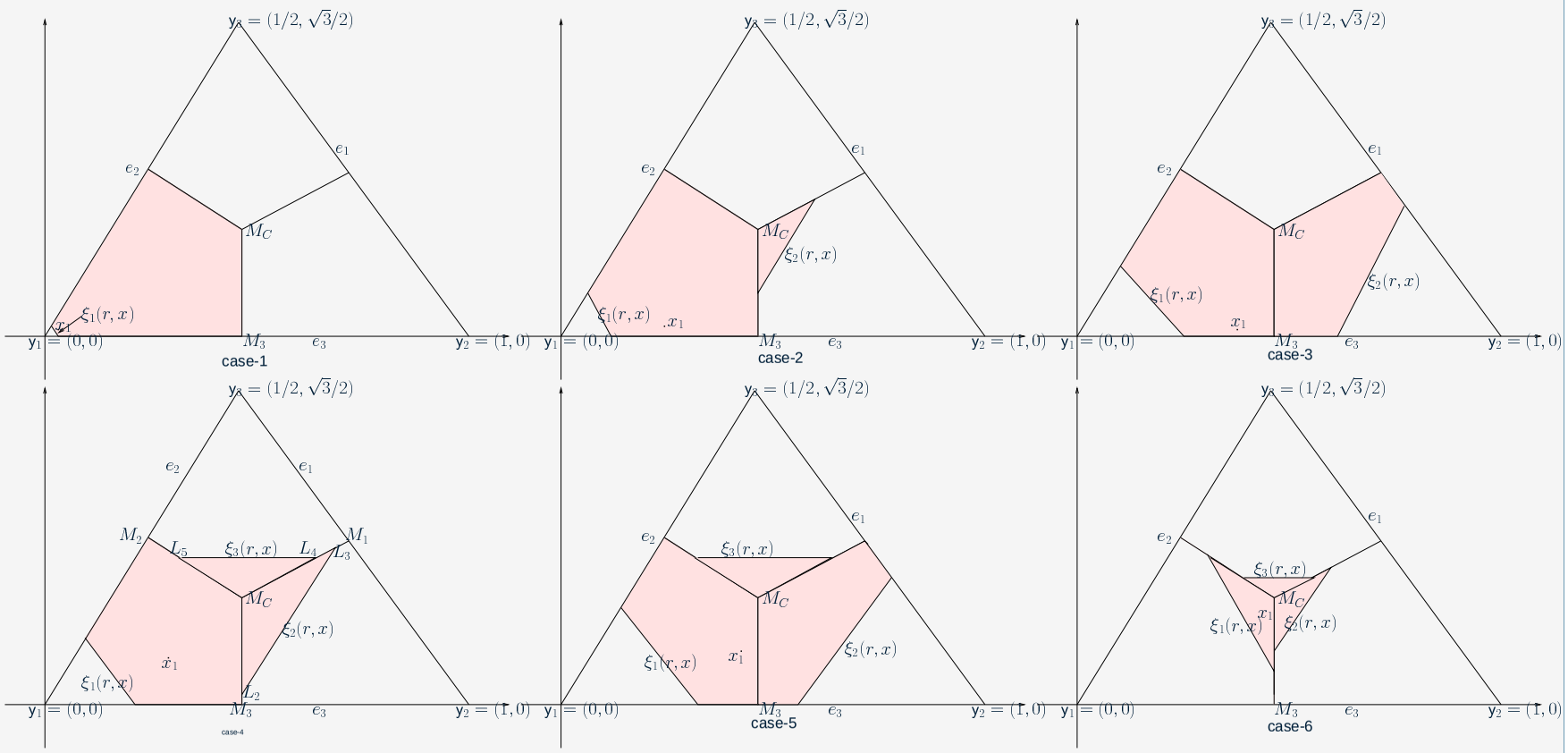


Figure 17: The prototypes of the six cases of $\Gamma_1\left(x_1, N^r_{PE}\right)$ for $x_1 \in T_s$ for $r \in [1, 4/3)$.

τ-Factor Central Similarity Proximity Maps

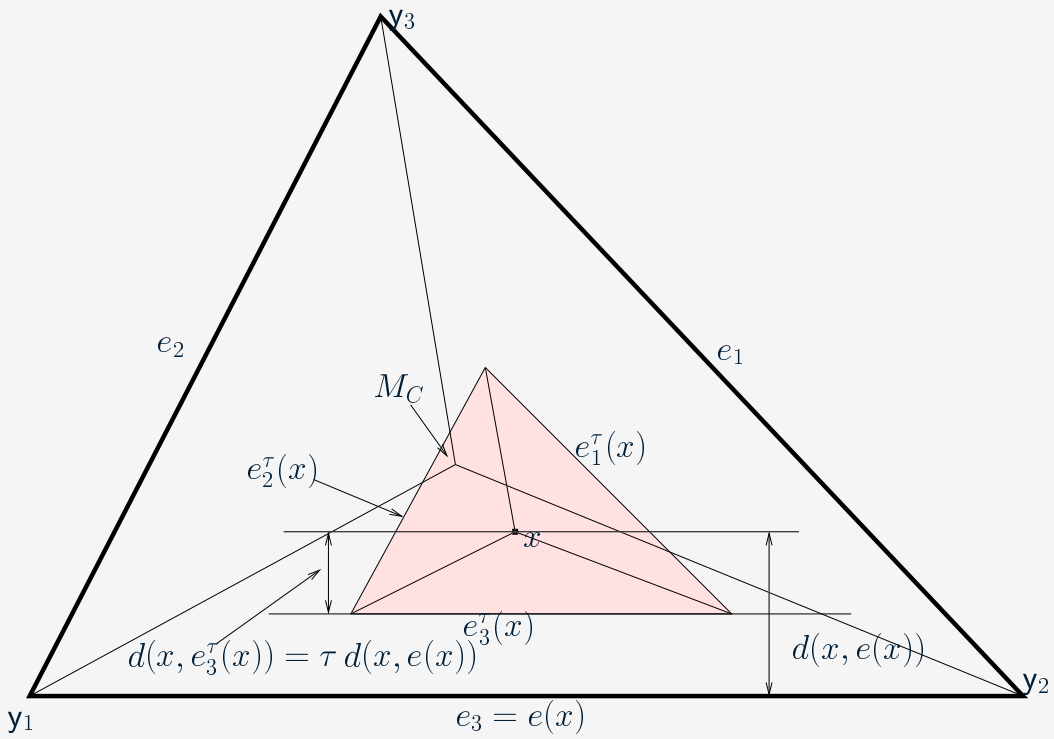


Figure 18: Construction of $N_{CS}^{\tau=1/2}(x, M_C)$ (shaded region).

Central Similarity Proximity Map with $\tau = 1$

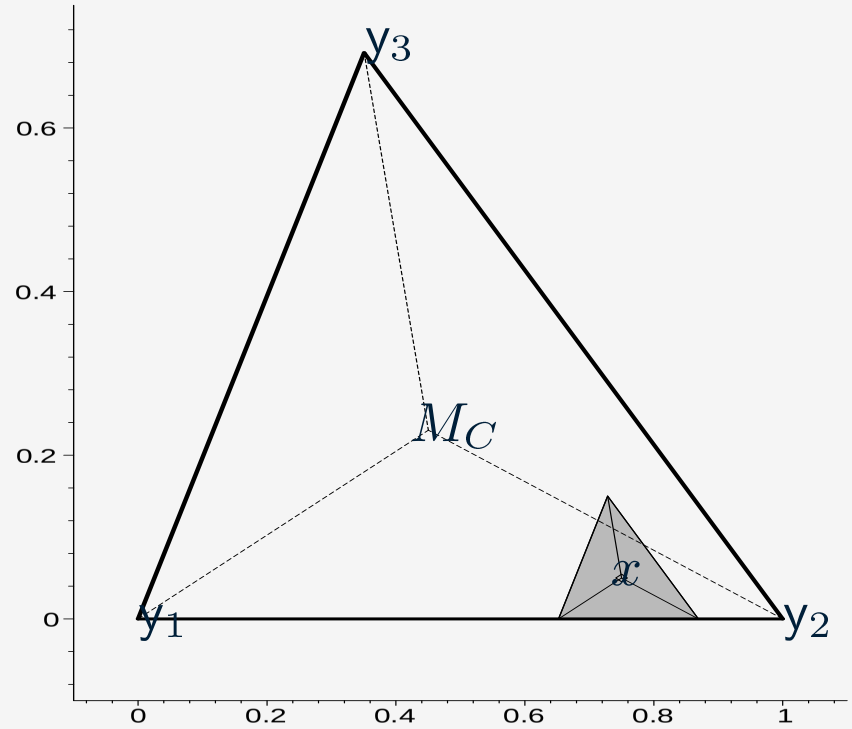


Figure 19: $N_{CS}^{\tau=1}(x, M_C)$ with an $x \in R_{M_C}(e_3)$

Arcs in τ-factor C-S PCD in one triangle

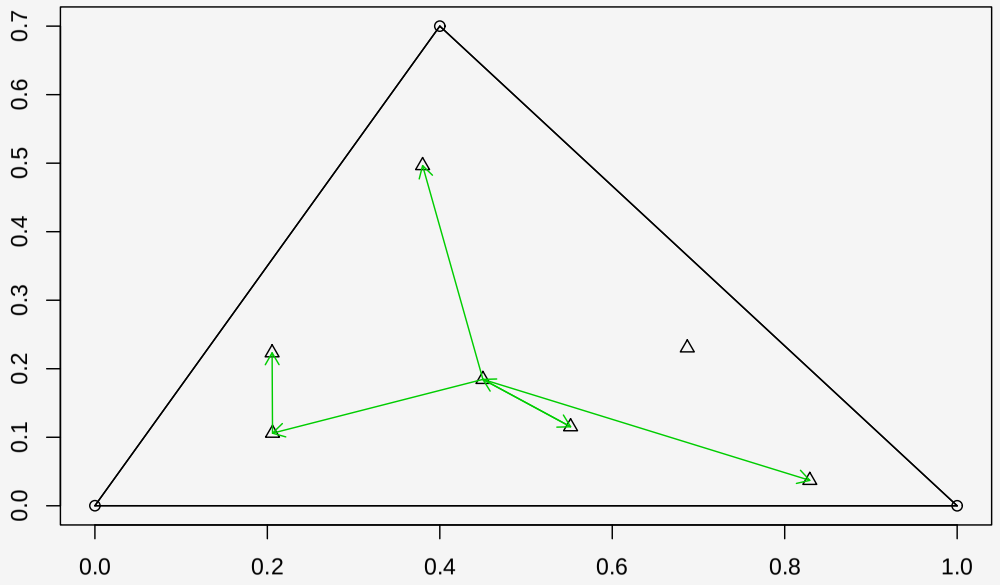


Figure 20: The arcs in the τ-factor C-S PCD with $\tau = 1$ and the 7 $\mathcal{X}$ points in the triangle $T(\mathcal{Y})$.

Arcs in τ-factor C-S PCD in multiple triangles

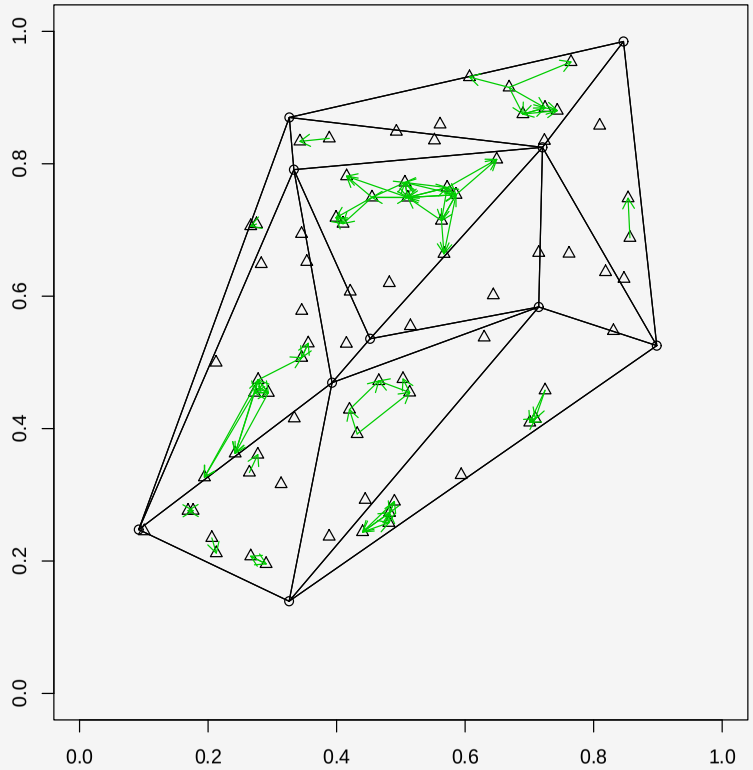


Figure 21: The arcs in the τ-factor C-S PCD with $\tau = 1$ and the 77 $\mathcal{X}$ points in the $C_H(\mathcal{Y})$.

Types of Γ_1-regions for N^{τ}_{CS}

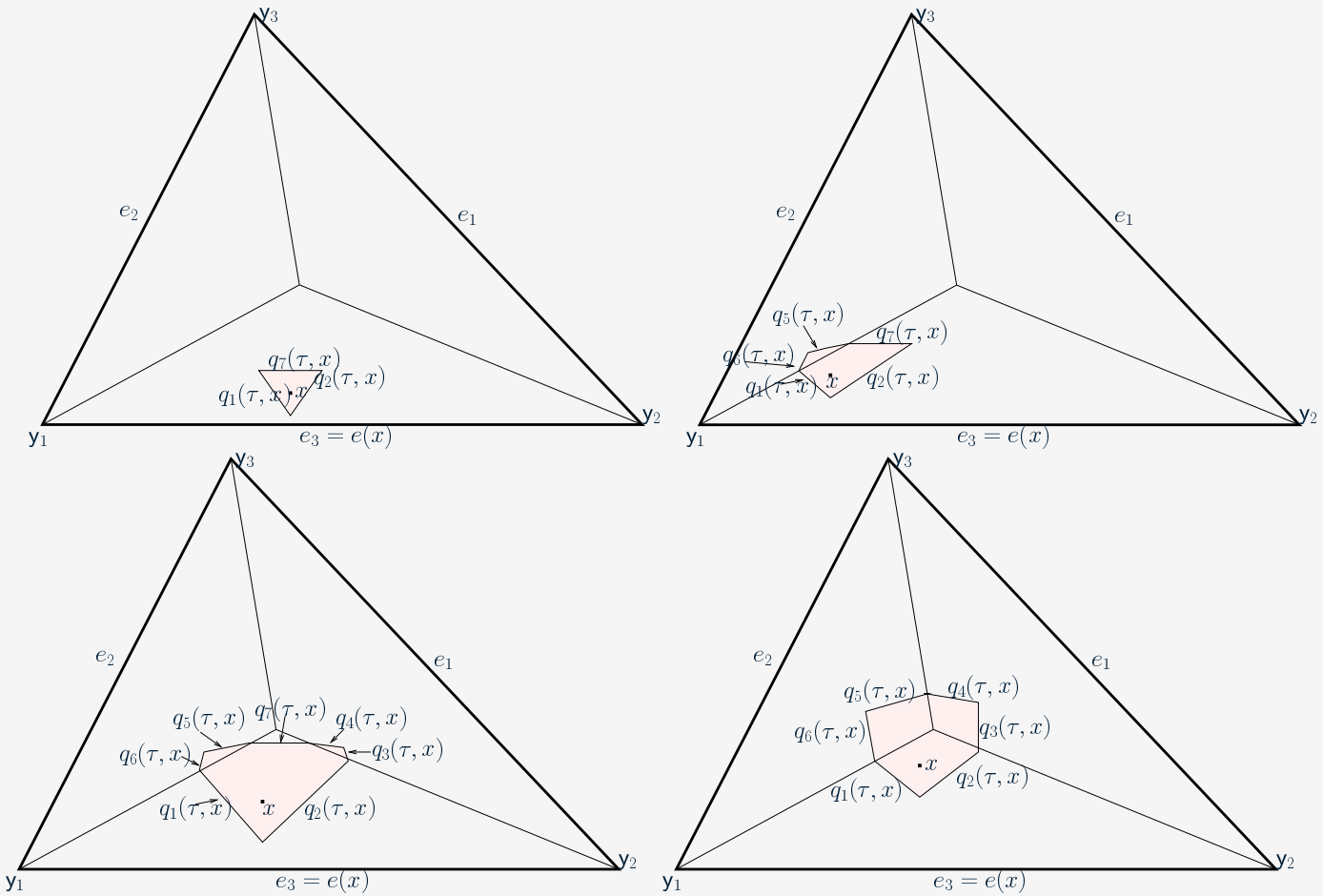


Figure 22: Examples of the Γ_1-region, $\Gamma_1\left(x, N_{CS}^{\tau=1/2}, M_C\right)$ with four distinct $x \in R_{CM}(e_3)$

Γ_1-regions for $N_{CS}^{\tau=1}$

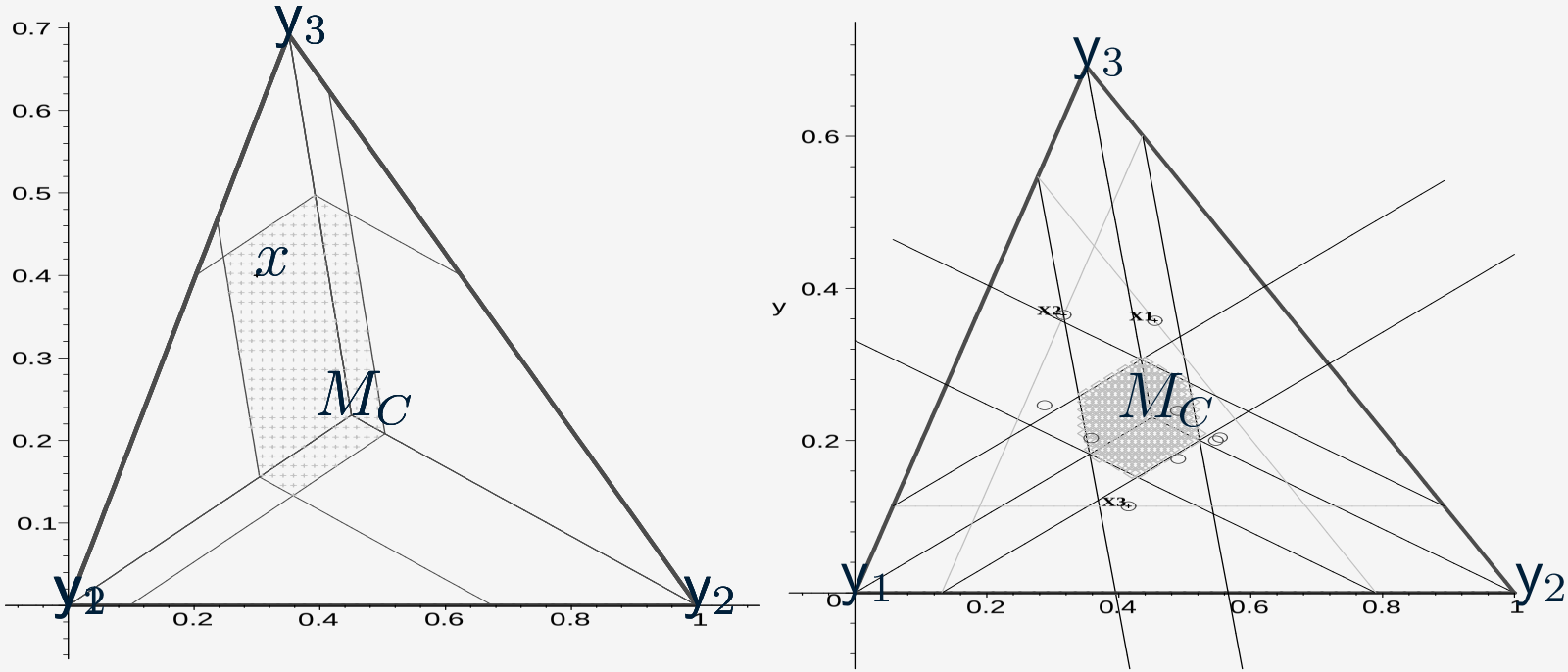


Figure 23: The Γ_1-region $\Gamma_1\left(x, N_{CS}^{\tau}, M_C\right)$ with $x \in R_M(e_2)$ (left) and $\Gamma_1\left(X_n, N_{CS}^{\tau}, M_C\right)$ with $n > 1$ (right).

Distribution of Domination Number of PCDs

- The main results on the asymptotic distribution of domination number of the arc-slice PCD based on the center M, $\gamma_n(N_{AS}, M)$
- As $n \to \infty$,

$\gamma_n(N_{AS}, M) = 1$, w.p. 1, for all $M \in T(\mathcal{Y}) \setminus \{M_{CC}\}$

$1 \leq \gamma_n(N_{AS}, M) \leq 2$, w.p. 1, for $M = M_{CC}$ and acute $T(\mathcal{Y})$

$1 \leq \gamma_n(N_{AS}, M) \leq 3$, w.p. 1, for $M = M_{CC}$ and right $T(\mathcal{Y})$

$2 \leq \gamma_n(N_{AS}, M) \leq 3$, w.p. 1, for $M = M_{CC}$ and obtuse $T(\mathcal{Y})$

Distribution of the Domination Number of PCDs

- The Asymptotic Distribution of the Domination Number for r-Factor PCDs, $\gamma_n\left(N_{PE}^r, M\right)$ is given below.
- Let $d(\cdot,\cdot)$ be the Euclidean distance. In $T(\mathcal{Y})$, drawing the lines $q_j(r,x)$ such that $d(\mathsf{y}_j, e_j) = r\, d(q_j(r,x), \mathsf{y}_j)$ for $j = 1,2,3$ yields a triangle, $\mathscr{T}^r$, for $r < 3/2$. Let t_j, $j = 1,2,3$ be the vertices of $\mathscr{T}^r$.
- For $\mathcal{X}_n$ a set of iid $\mathcal{U}(T(\mathcal{Y}))$ random variables, as $n \to \infty$,

$$\gamma_n\left(N_{PE}^r, M\right) \to \begin{cases} \text{n.d.} & \text{for } M \in \{t_1,t_2,t_3\} \text{ and } r \in [1,3/2] \\ 1 & \text{for } r > 3/2 \\ 3 & \text{for } M \in \mathscr{T}^r \setminus \{t_1,t_2,t_3\} \text{ and } r \in [1,3/2) \end{cases}$$

Distribution of the Domination Number of PCDs

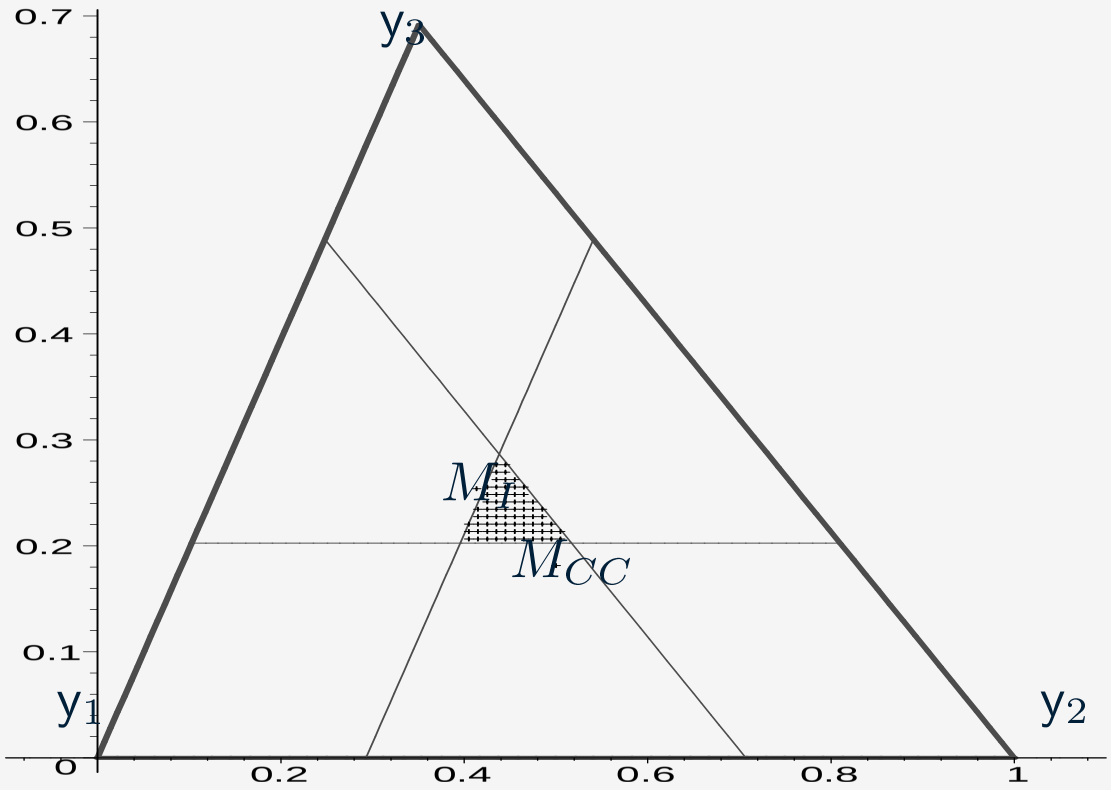


Figure 24: The triangle $\mathscr{T}^{r=\sqrt{2}}$.

Distribution of the Domination Number of PCDs

- For $M \in \{t_1, t_2, t_3\}$ and $r \in [1, 3/2)$

$$\gamma_n\left(N_{PE}^r, M\right) = \begin{cases} 2 & \text{wp} \quad p_r \\ 3 & \text{wp} \quad 1 - p_r \end{cases} \qquad \text{as } n \to \infty \tag{1}$$

  where p_r as a function of $r \in (1, 1.5)$ is numerically available.

- For $r = 3/2$,

$$\gamma_n\left(N_{PE}^{3/2}, M_C\right) = \begin{cases} 2 & \text{wp} \approx .7413 \\ 3 & \text{wp} \approx .2487 \end{cases} \qquad \text{as } n \to \infty. \tag{2}$$

- The distribution of the domination number of r-factor PCD with $r = 3/2$ and M_C is used in testing spatial patterns of segregation and association.

p_r as a function of r

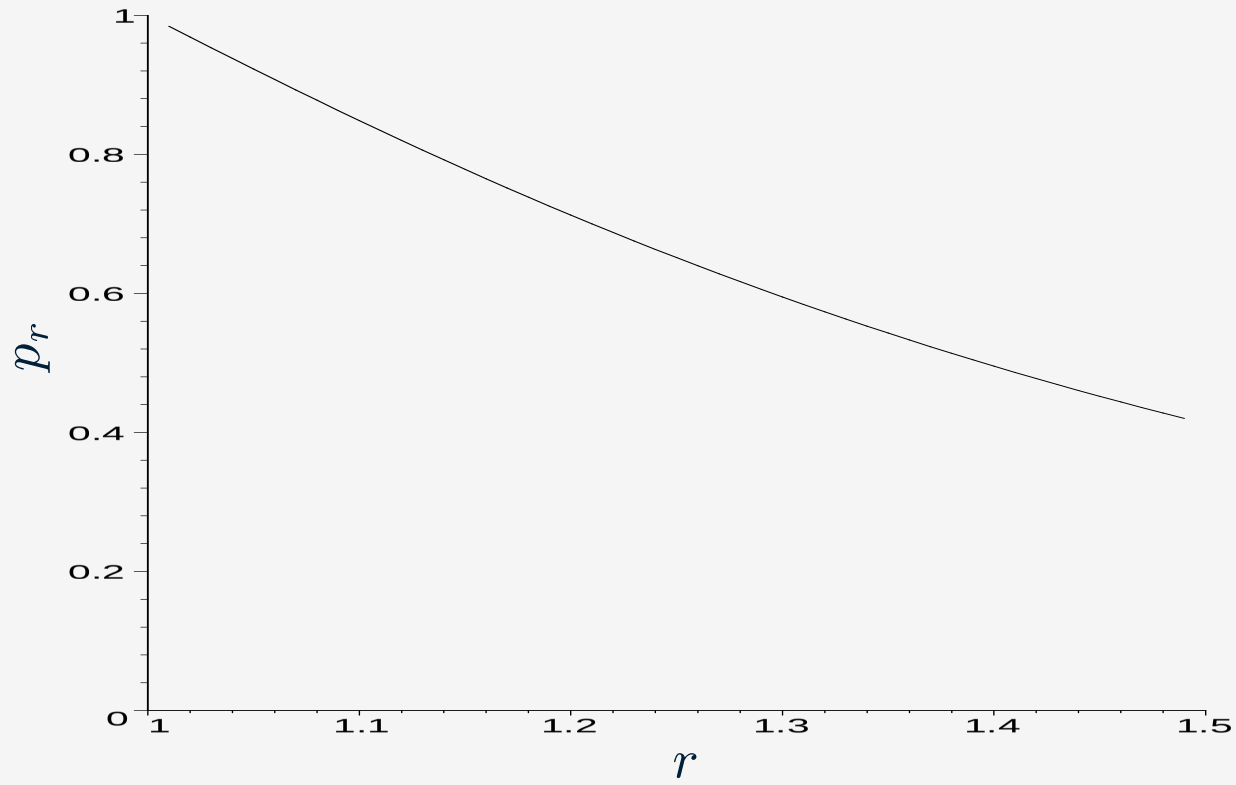


Figure 25: The probability p_r as a function of r for $r \in [1, 3/2)$ where $p_r = \lim_{n\to\infty} P\left(\gamma\left(\mathcal{X}_n, N_{PE}^r, M\right) = 2\right)$.

Distribution of the Domination Number of PCDs

- The Asymptotic Distribution of the Domination Number for τ-Factor Central Similarity PCDs
- The asymptotic distribution of $\gamma_n(N_{CS}^{\tau}, M)$: Not available but as $n \to \infty$

$$2 \leq \gamma_n(N_{CS}^{\tau=1}, M) \leq 6, \text{ w.p. 1, for all } M \in T(\mathcal{Y})^o$$

$$\gamma_n(N_{CS}^{\tau_1}, M) \geq 2, \text{ w.p. 1, for all } M \in T(\mathcal{Y})^o \text{ and } \tau_1 \in [0, 1)$$

- Also, $P\left(\gamma_n\left(N_{CS}^{\tau}, M\right) = n\right) > 0$ for all $n < \infty$.

Relative Density of PCDs

- For digraph $D = (\mathcal{V}, \mathcal{A})$ with vertex set of size $|\mathcal{V}| = n$ and arc set $\mathcal{A}$. The *relative density* of D, is defined as

$$\rho(D) := \frac{|\mathcal{A}|}{n\,(n-1)}.$$

  That is, $\rho(D)$ is the ratio of number of arcs in D to the number of arcs in a *complete symmetric digraph* of order n, which is $n\,(n-1)$.

- For $X_i \overset{iid}{\sim} F$, relative density $\rho(D)$ of the PCD, $D = D(\mathcal{X}_n, \mathcal{A})$, associated with the proximity map $N(\cdot)$ is denoted as $\rho(\mathcal{X}_n; h, N)$ and is a U-statistic;

$$\rho(\mathcal{X}_n; h, N) = \frac{1}{n\,(n-1)} \sum\sum_{i<j} h\big(X_i, X_j; N\big)$$

where

Relative Density of PCDs

- $h\big(X_i, X_j; N\big) = \mathbf{I}\big(X_i X_j \in \mathcal{A}\big) + \mathbf{I}\big(X_j X_i \in \mathcal{A}\big) = \mathbf{I}\big(X_i \in N(X_j)\big) + \mathbf{I}\big(X_j \in N(X_i)\big)$ is the number of arcs between points X_i and X_j.
- A central limit theorem for U-statistics yields

$$\sqrt{n}\big(\rho_n(N) - \mathbf{E}\left[\rho_n(N)\right]\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \mathbf{Cov}\left[h_{12}(N), h_{13}(N)\right]\big)$$

provided that $\mathbf{Cov}\left[h_{12}(N), h_{13}(N)\right] > 0$.

Relative Density of PCDs

- Relative density is used to test
  complete spatial randomness
  against
  spatial point patterns of segregation and association.
- The power is investigated by using
  Pitman asymptotic efficacy (PAE),
  Hodges-Lehman asymptotic efficacy (HLAE),
  asymptotic power function,
  and Monte Carlo simulations.

Segregation, CSR, and Association

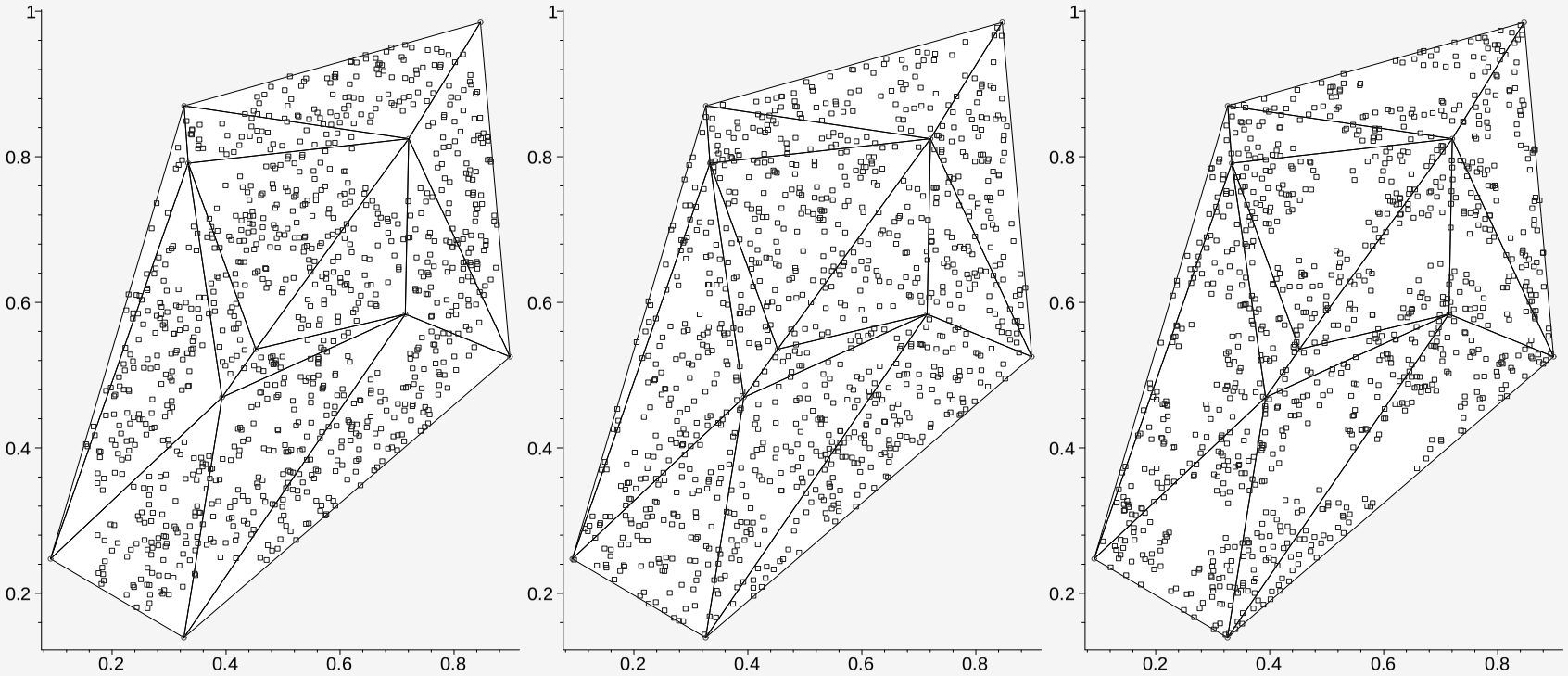


Figure 26: Realizations of segregation (left), null case (middle), and association (right) for $|\mathcal{Y}| = 10$, $J = 13$, and $n = 1000$.

Relative Density

- The mean and asymptotic variance of the relative density of arc-slice proximity catch digraphs are not available.
- The mean and asymptotic variance of the relative density of r-factor proximity catch digraphs are presentend below together with the asymptotic normality under H_0.
  Let $\rho_n(r)$ be the relative density of r-factor PCD based on $\mathcal{X}_n$, $\mu(r) = \mathbf{E}\left[\rho_n(r)\right]$ and $\nu(r) := \mathbf{Cov}\left[h_{12}, h_{13}\right]$.
- For $r = \infty$, $\rho_n(r)$ is degenerate.
- For $r \in [1, \infty)$,

$$\frac{\sqrt{n}\left(\rho_n(r) - \mu(r)\right)}{\sqrt{\nu(r)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

  where

Relative Density

$$\mu(r) = \begin{cases} \frac{37}{216}\, r^2 & \text{for} \quad r \in [1, 3/2) \\ -\frac{1}{8}\, r^2 + 4 - 8r^{-1} + \frac{9}{2}\, r^{-2} & \text{for} \quad r \in [3/2, 2) \\ 1 - \frac{3}{2}\, r^{-2} & \text{for} \quad r \in [2, \infty) \end{cases}$$

and

$$\nu(r) = \nu_1(r)\, \mathbf{I}(r \in [1, 4/3)) + \nu_2(r)\, \mathbf{I}(r \in [4/3, 3/2)) + \nu_3(r)\, \mathbf{I}(r \in [3/2, 2)) + \nu_4(r)\, \mathbf{I}(r \in [2, \infty])$$

with

Relative Density

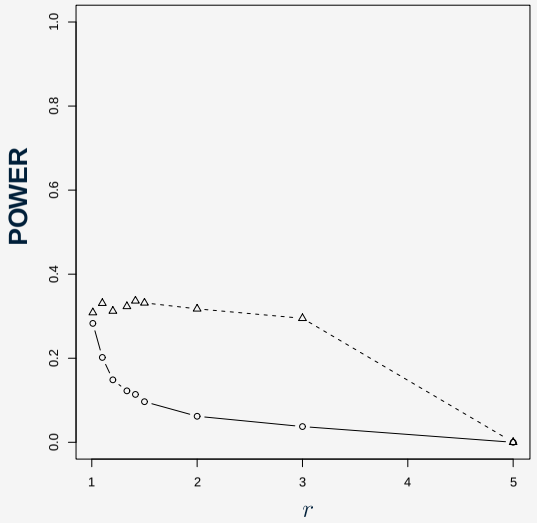
$$
\begin{aligned}
\nu_1(r) &= \Big[3007\,r^{10} - 13824\,r^9 + 898\,r^8 + 77760\,r^7 - 117953\,r^6 + 48888\,r^5 - \\
&\quad 24246\,r^4 + 60480\,r^3 - 38880\,r^2 + 3888\Big] \Big/ \Big[58320\,r^4\Big], \\
\nu_2(r) &= \Big[5467\,r^{10} - 37800\,r^9 + 61912\,r^8 + 46588\,r^6 - 191520\,r^5 + 13608\,r^4 + \\
&\quad 241920\,r^3 - 155520\,r^2 + 15552\Big] \Big/ \Big[233280\,r^4\Big] \\
\nu_3(r) &= -\Big[7\,r^{12} - 72\,r^{11} + 312\,r^{10} - 5332\,r^8 + 15072\,r^7 + 13704\,r^6 - 139264\,r^5 + \\
&\quad 273600\,r^4 - 242176\,r^3 + 103232\,r^2 - 27648\,r + 8640\Big] \Big/ \Big[960\,r^6\Big], \\
\nu_4(r) &= \frac{15\,r^4 - 11\,r^2 - 48\,r + 25}{15\,r^6}.
\end{aligned}
$$

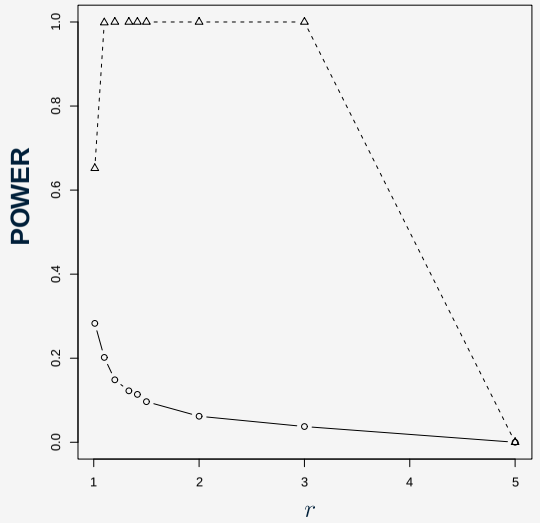
Relative Density

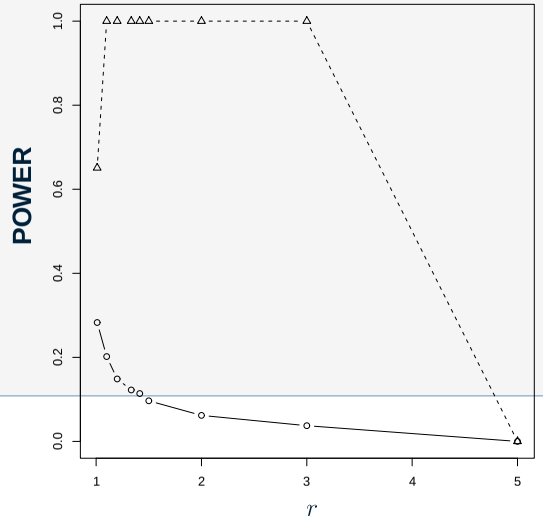
- Asymptotic normality under the alternatives is straightforward, provided $\nu(r,\varepsilon) > 0$.
- Let $R = \frac{\sqrt{n}\big(\rho_n(r) - \mu(r)\big)}{\sqrt{\nu(r)}}$.

  Then the test against H_ε^S which rejects for $R > z_{1-\alpha}$
  and
  the test against H_ε^A which rejects for $R < z_\alpha$ are consistent for $r \in [1,\infty)$ and $\varepsilon \in \big(0, \sqrt{3}/3\big)$.

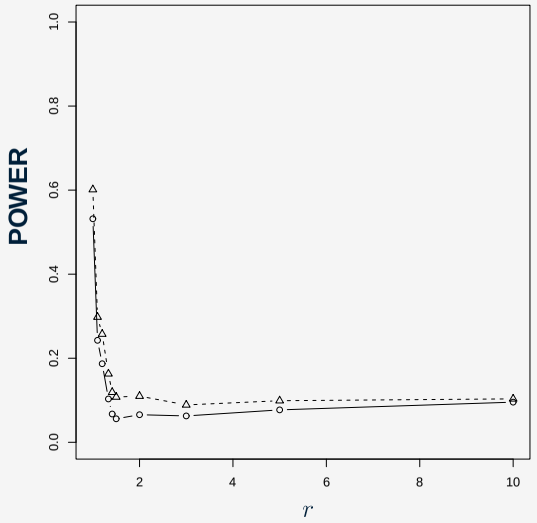
MC Power Analysis-Segregation

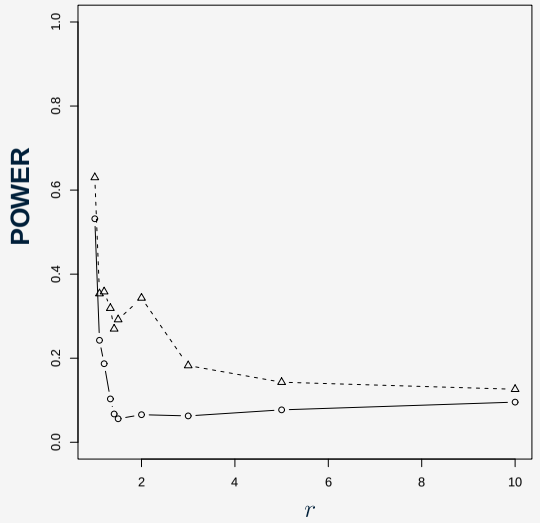


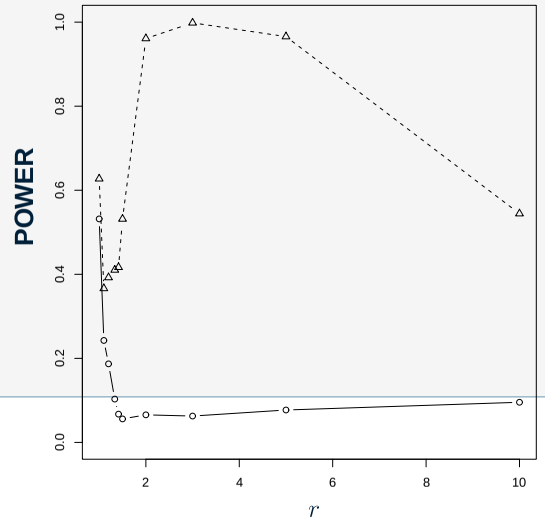




MC Power Analysis-Association







MC Power Analysis

- Monte Carlo power analysis $\Rightarrow$
  moderate values of r ($r \in [3/2, 3]$) are
  more appropriate for normal approximation,
  for both segregation and association alternatives.

PAE Analysis

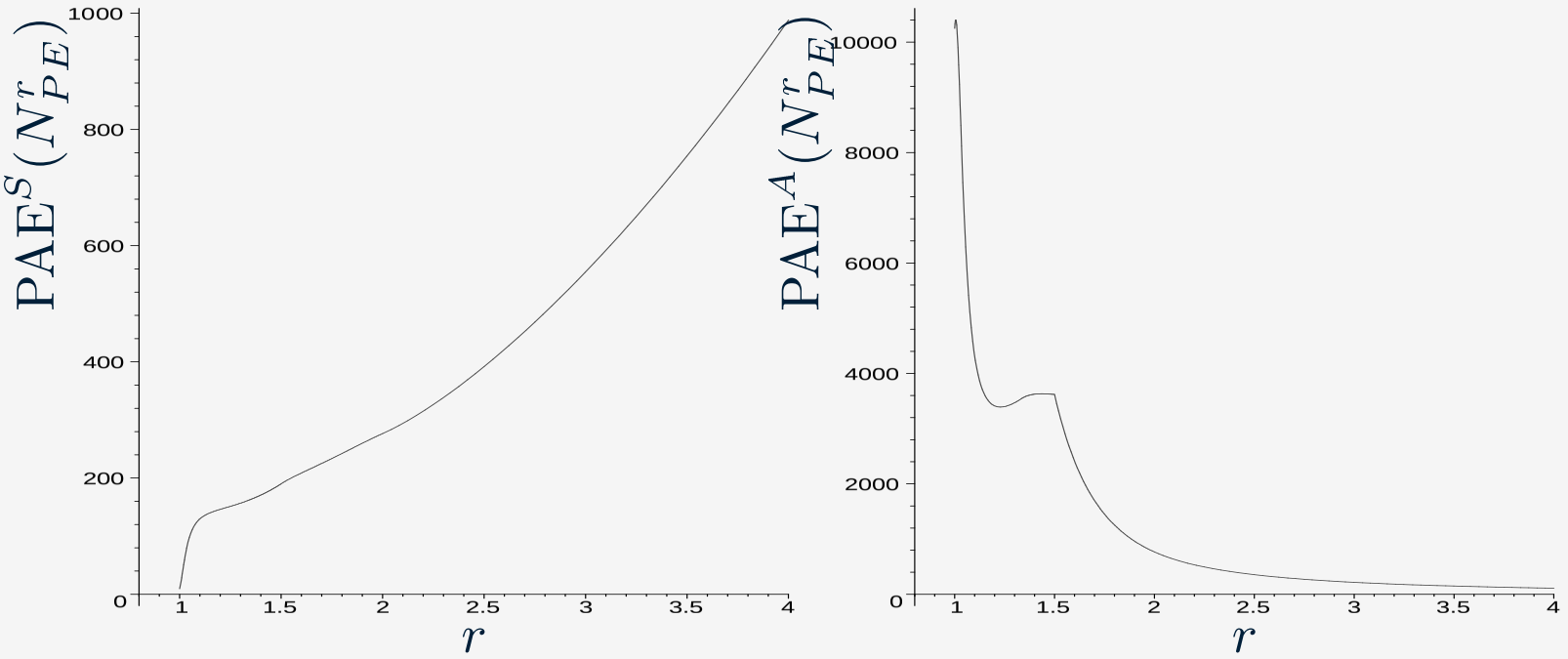
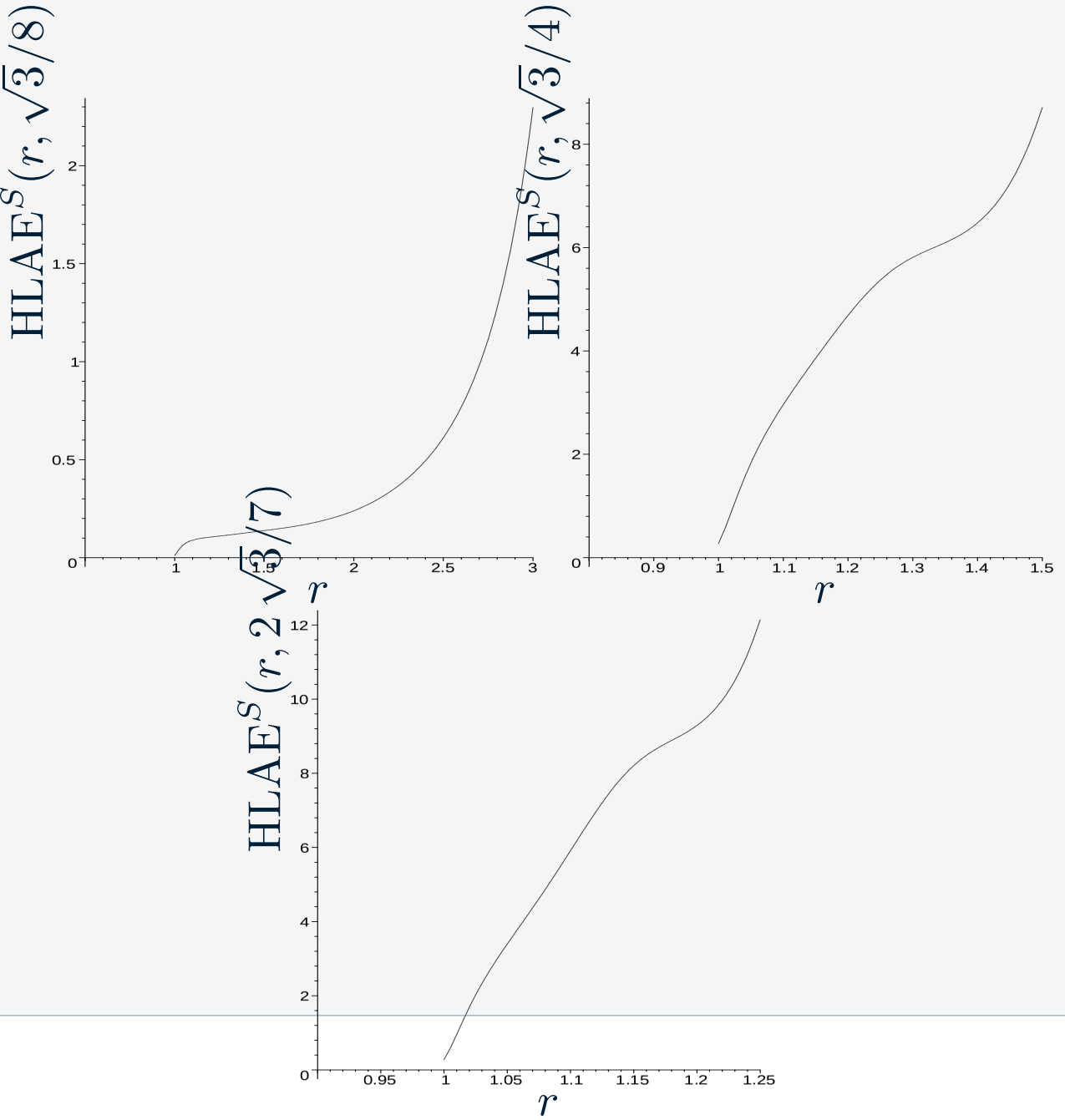


Figure 29: Pitman asymptotic efficacy against segregation (left) and against association (right) as a function of r.

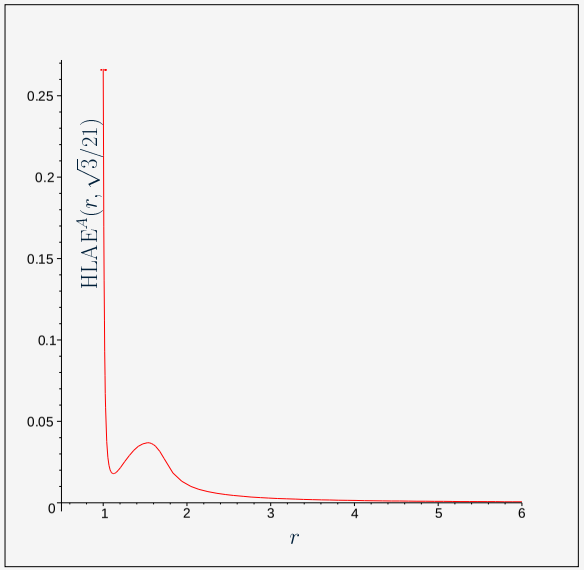
PAE Analysis

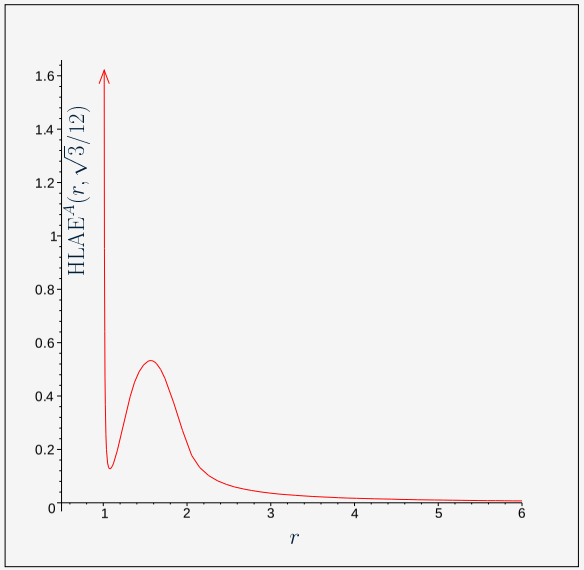
- The PAE analysis suggests, (for large n and small ε), choosing large r for testing against segregation and small r against association.

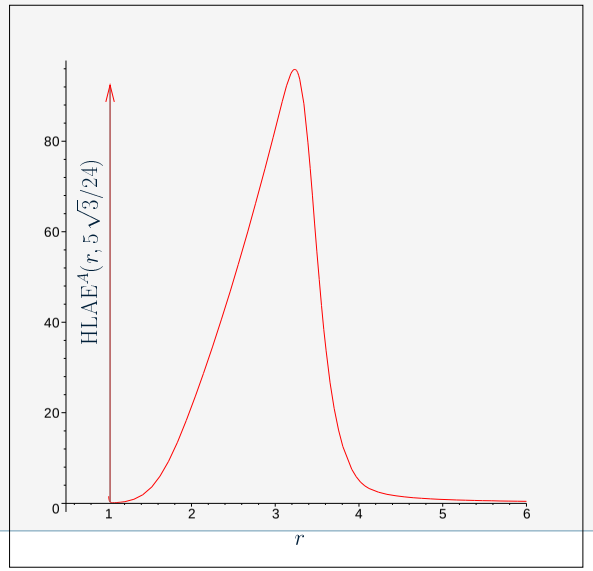
HLAE Analysis-Segregation



HLAE Analysis-Association







HLAE Analysis

- Against segregation, the HLAE analysis suggests choosing r larger as the segregation gets more severe;
- and choosing moderate r against association.

Relative Density of the τ-Factor Central Similarity PCDs

- For $\tau \in (0, 1]$,

$$\sqrt{n}\left(\frac{\rho_n(\tau) - \mu(\tau)}{\sqrt{\nu(\tau)}}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

where

$$\mu(\tau) = \tau^2/6 \quad \text{and} \quad \nu(\tau) = \frac{\tau^4(6\,\tau^5 - 3\,\tau^4 - 25\,\tau^3 + \tau^2 + 49\,\tau + 14)}{45\,(\tau+1)(2\,\tau+1)(\tau+2)}.$$

For $\tau = 0$, $\rho_n(\tau)$ is degenerate.

Relative Density of the τ-Factor Central Similarity PCDs

- Asymptotic normality under the alternatives and the consistency of the proposed test are presented below.
- Let $\mu_S(\tau,\varepsilon)$ be the mean, $\mathbf{E}_{\varepsilon}^{S}[\rho_n(\tau)]$ and $\nu_S(\tau,\varepsilon)$ be the covariance, $\mathbf{Cov}_{\varepsilon}^{S}[h_{12},h_{13}]$ for $\tau\in(0,1]$ and $\varepsilon\in\left(0,\sqrt{3}/3\right)$ under H_{ε}^{S}, then $\sqrt{n}\big(\rho_n(\tau)-\mu_S(\tau,\varepsilon)\big)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\nu_S(\tau,\varepsilon)\right)$ for the values of the pair (τ,ε) for which $\nu_S(\tau,\varepsilon)>0$. Likewise for H_{ε}^{A}.
- The test against H_{ε}^{S} which rejects for $R>z_{\alpha}$ and the test against H_{ε}^{A} which rejects for $R<z_{1-\alpha}$ are consistent for $\tau\in(0,1]$ and $\varepsilon\in\left(0,\sqrt{3}/3\right)$.

Power Analysis

- Monte Carlo power analysis implies that, against segregation, large τ values, ($\tau = .8$), yield the better performance for the normal approximation and the empirical power;
- against association, for small n, moderate τ values, ($\tau \in [.4, .6]$), and for large n, large τ values ($\tau = 1.0$) yield better performance for the normal approximation and the empirical power.
- The PAE analysis suggests, choosing τ large against segregation;
- and choosing τ small against association.
- Due to skewness, for small n, we suggest large τ values.

Power Analysis

- Based on the PAE analysis, for large n and small ε, we suggest the use of
  $\rho_n(N^r_{PE})$ with moderate r values ($r \in [3/2, 3]$) against segregation
  and use of
  $\rho_n(N^\tau_{CS})$ with large τ values ($\tau \approx 1$) against association.

References