Seminario de Teoría de la Información y Aplicaciones 2010
Valparaiso, April 2010

Mutual Information Estimation based Data-Driven Partitions

Jorge Silva and Shrikanth S. Narayanan

Universidad de Chile
Department of Electrical Engineering

University of Southern California
Department of Electrical Engineering

Introduction

- Mutual information (MI) fundamental quantity in information theory
  - (Shannon's coding theorems): channel capacity, achievable rate-distortion curve,....
- Statistical learning-decision problems,
  - fidelity indicator for image registration, image segmentation, feature extraction (MPE-SR), detection,...
  - performance limits in Pattern recognition (*Westover et al. 2008*)

Introduction

- Mutual information (MI) fundamental quantity in information theory
  - (Shannon's coding theorems): channel capacity, achievable rate-distortion curve,....
- Statistical learning-decision problems,
  - fidelity indicator for image registration, image segmentation, feature extraction (MPE-SR), detection,...
  - performance limits in Pattern recognition (*Westover et al. 2008*)

Mutual Information Estimation

- estimation based on empirical data
- a distribution-free framework is fundamental

Introduction

Extensive work on related differential entropy estimation (*Beirlant et al. 1997*)

- non-adaptive product type of histogram-based approaches
- kernel plug-in estimates
  - ✓ strong consistency is well understood

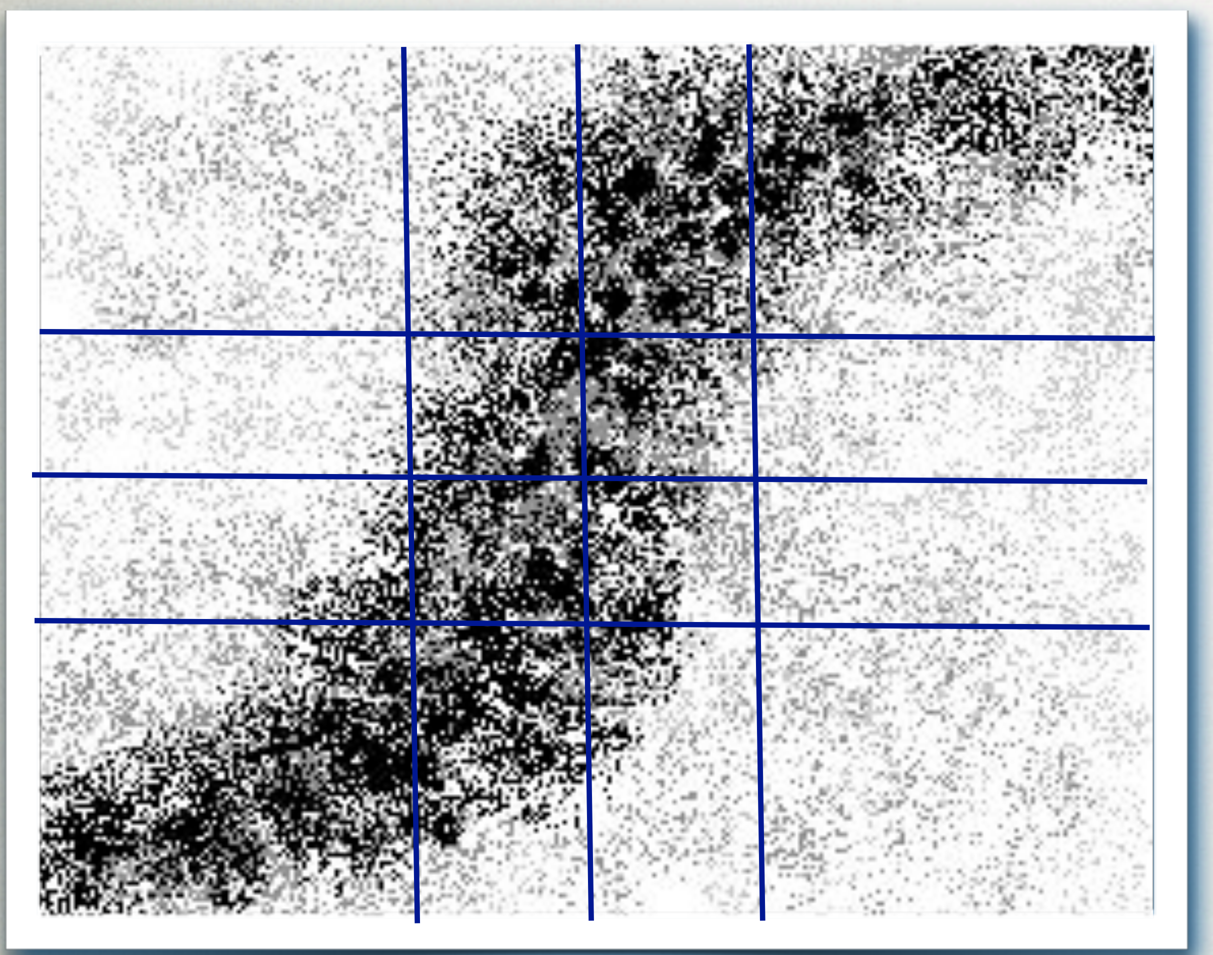
Introduction

Extensive work on related differential entropy estimation (*Beirlant et al. 1997*)

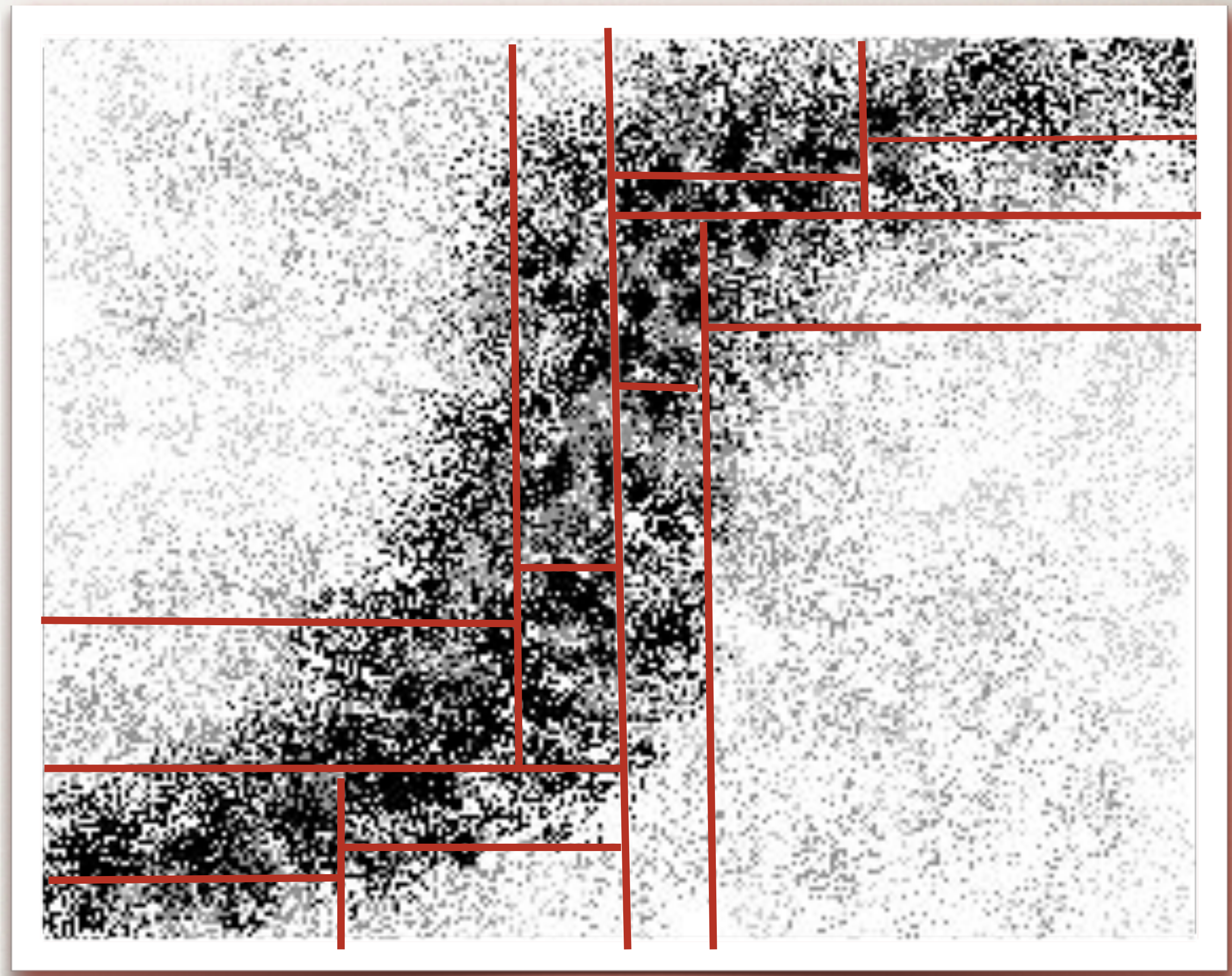
- non-adaptive product type of histogram-based approaches
- kernel plug-in estimates
  - ✓ strong consistency is well understood

Focus and Motivation of this Work

- ✤ Explore the role of data-dependent partitions -> non-product structure
- ✤ Hypothesis: better adaptation to the data -> better approximation properties -> better estimates
  - ✓ effectiveness demonstrated in other stat. learn. settings (classification, density estimation, *Lugosi et al. 1996*)



PRODUCT PARTITION SCHEME



NON-PRODUCT DATA-DRIVEN PARTITION

Introduction

Work in this direction....

Darbellay et al. 1999 proposed tree-structured data-driven construction

- ✓ based on a local splitting process to adapt to the data
- ✤ consistency an open problem for this setting
  - design parameters need to be triggered empirically

Introduction

Work in this direction....

Darbellay et al. 1999 proposed tree-structured data-driven construction

- ✓ based on a local splitting process to adapt to the data
- ✤ consistency an open problem for this setting
  - design parameters need to be triggered empirically

...... on the KL Divergence

- Data-dependent histogram-based approach (Wang et al. IEEE Trans. IT 2005)

Introduction

Proposed directions

- Theory: Find consistency conditions for general partition schemes
- Applications: Propose new concrete data-driven estimates

Content Outline

Main Result

- The Estimator:
  - Approximation Error Analysis
  - Estimation error Analysis
- The Result: Conditions for Strong Consistency

Applications

- Multivariate Statistically Equivalent Block
- Tree-structured Partitions
- Consistency

- Experiments

- Some Extensions

Preliminaries

Let X, Y two random variables with joint distribution $P_{X,Y}$.
The mutual information,

$$I(X;Y) = D(P_{X,Y} || P_X \times P_Y)$$

where $D(P||Q)$ is the Kullback-Leibler divergence,

$$D(P||Q) = \int \log \frac{\partial P}{\partial Q}(x) \cdot \partial P(x).$$

Preliminaries

Let X, Y two random variables with joint distribution $P_{X,Y}$.
The mutual information,

$$I(X;Y) = D(P_{X,Y} || P_X \times P_Y)$$

where $D(P||Q)$ is the Kullback-Leibler divergence,

$$D(P||Q) = \int \log \frac{\partial P}{\partial Q}(x) \cdot \partial P(x).$$

Main Problem

- joint distribution is unknown
- only i.i.d. realization $(X_1, Y_1), \cdots, (X_n, Y_n)$ available
- a distribution free estimator of $I(X,Y)$ is needed !

$\hat{I}_n(X;Y)$

Preliminaries

Requirements for the Estimate

- ✦ Large sample regime:

  -Universal (distribution free) strongly consistent

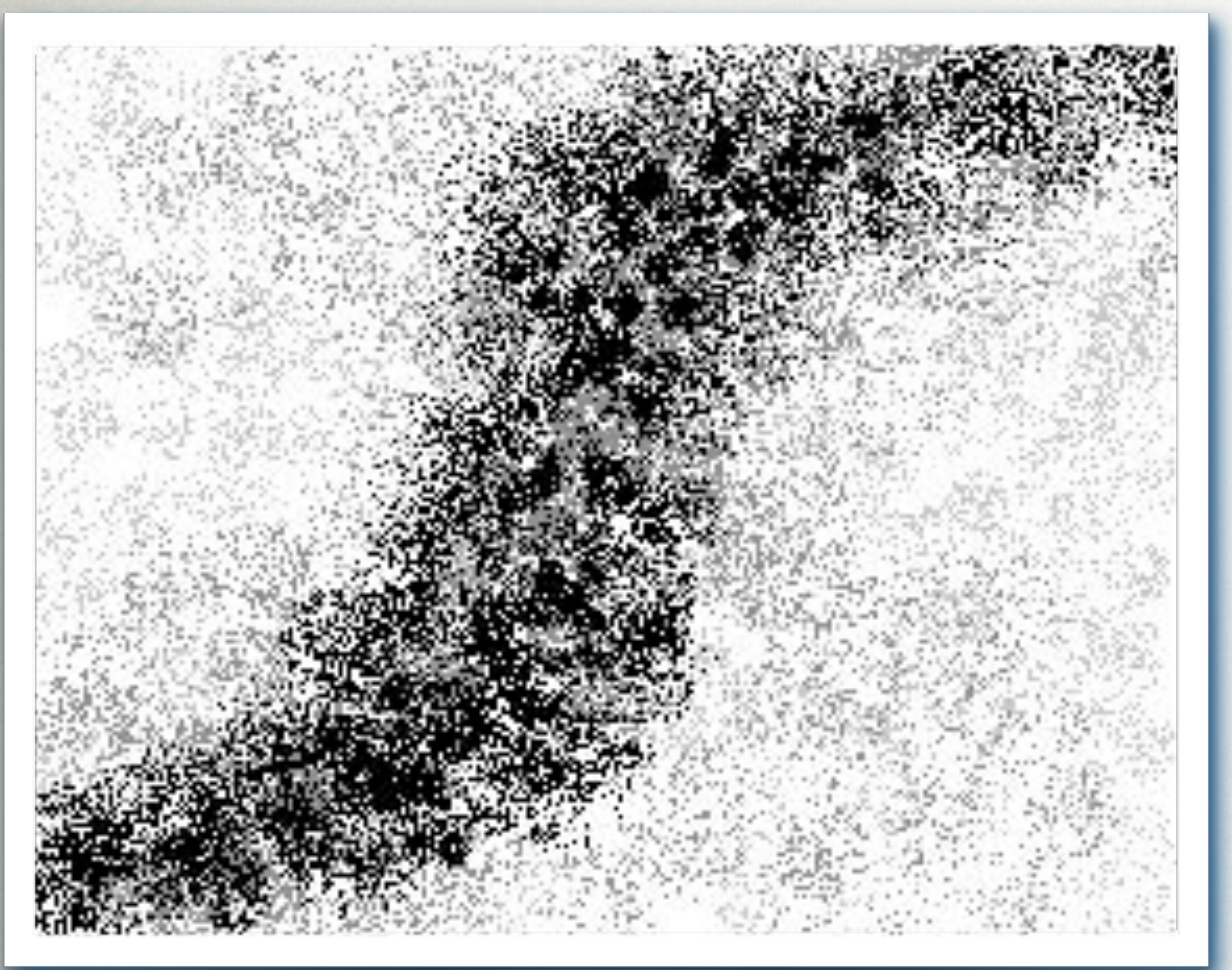
  $$\lim_{n\to\infty} \hat{I}_n(X;Y) = I(X;Y) \text{ almost surely}$$

- ✦ Small sample regime:

  -good approximation to the empirical data

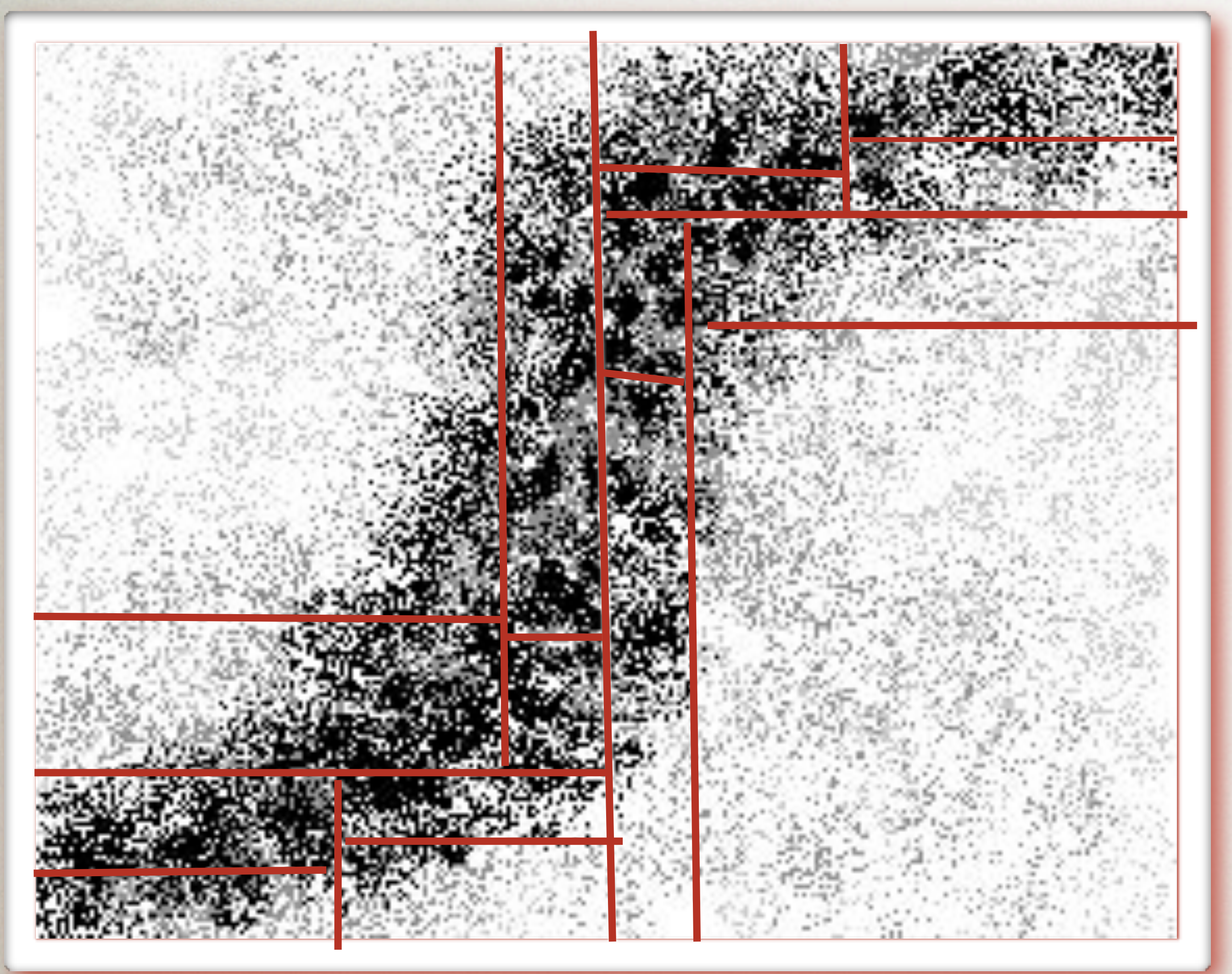
  Bias and stand. dev. of $\left|\hat{I}_n(X;Y) - I(X;Y)\right|$, for finite n

Basic Construction



$$Z_1 = (X_1, Y_1), .., Z_n = (X_n, Y_n)$$

Empirical Data from $P_{X,Y}$



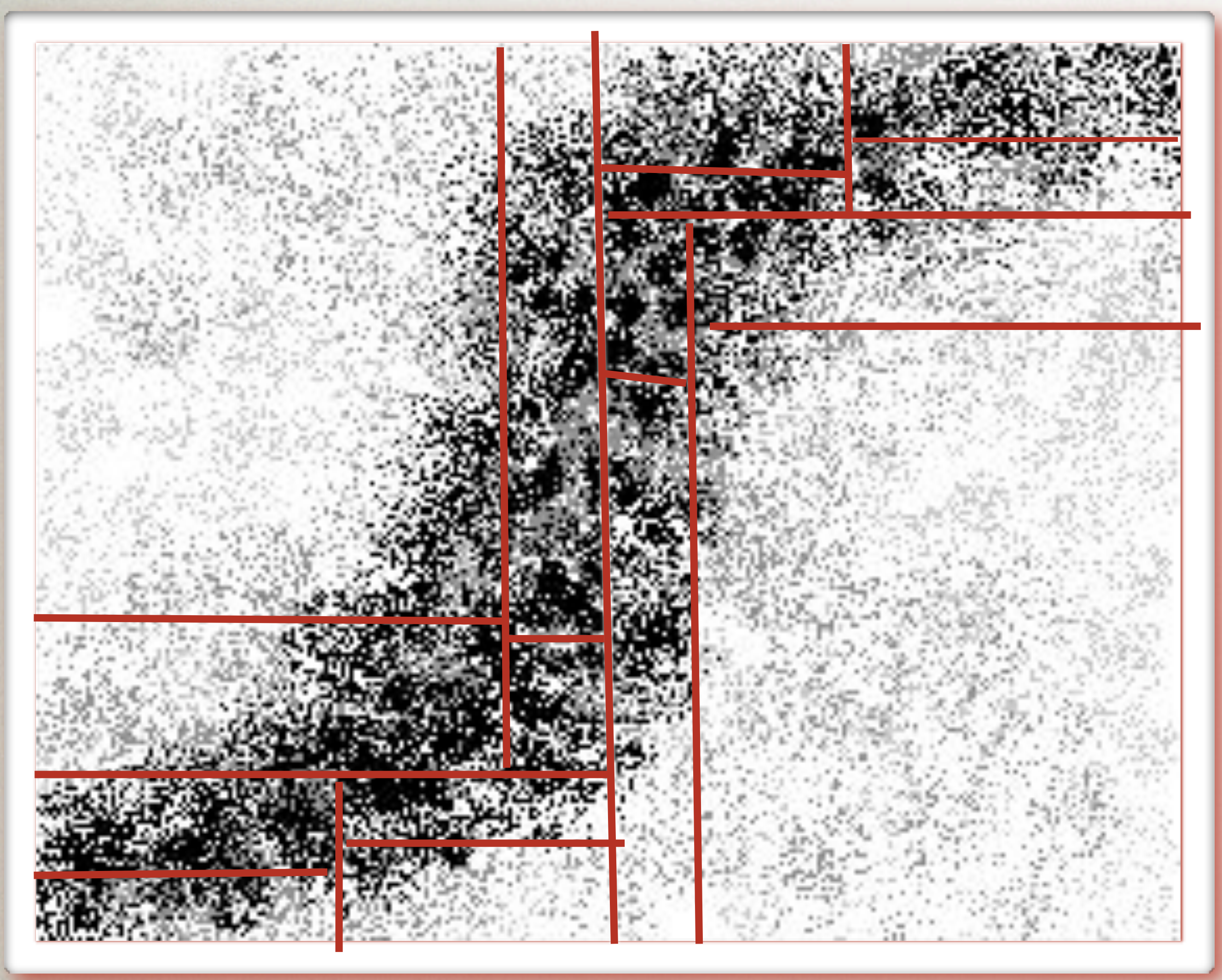
Basic Construction

$$Z_1 = (X_1, Y_1), .., Z_n = (X_n, Y_n)$$

Empirical Data from $P_{X,Y}$

$$\pi_n(Z_1, .., Z_n) = \{A_1, .., A_k\} \in \mathcal{B}\left(\mathbb{R}^d\right)$$

Data-dependent partition rule



Basic Construction

$$Z_1 = (X_1, Y_1), .., Z_n = (X_n, Y_n)$$

Empirical Data from $P_{X,Y}$

$$\pi_n(Z_1, .., Z_n) = \{A_1, .., A_k\} \in \mathcal{B}\left(\mathbb{R}^d\right)$$

Data-dependent partition rule

$$\mathrm{P}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_A(Z_i),\ \forall A \in \pi_n(Z_1, .., Z_n)$$

Computation of empirical frequencies

Basic Construction

$$Z_1 = (X_1, Y_1), .., Z_n = (X_n, Y_n)$$

Empirical Data from $P_{X,Y}$

$$\pi_n(Z_1, .., Z_n) = \{A_1, .., A_k\} \in \mathcal{B}\left(\mathbb{R}^d\right)$$

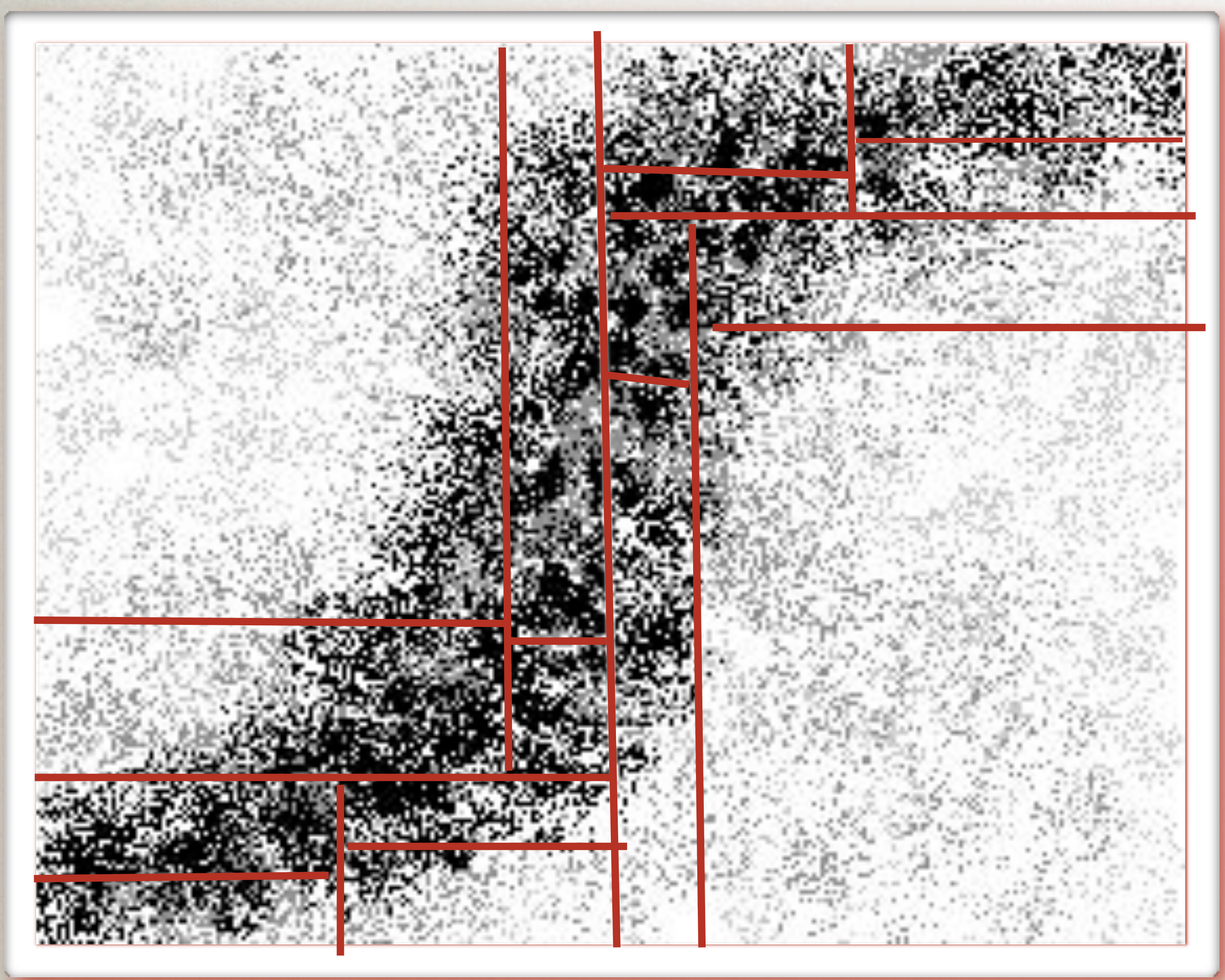
Data-dependent partition rule

$$\mathrm{P}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_A(Z_i), \ \forall A \in \pi_n(Z_1, .., Z_n)$$

Computation of empirical frequencies

$$\hat{I}_n(X;Y) = \sum_{A\in\pi_n(Z_1^n)} P_n(A) \cdot \log \frac{P_n(A)}{P_n(A_1 \times \mathbb{R}^q) \cdot P_n(\mathbb{R}^p \times A_2)};$$

Empirical mutual information



Product Bin Structure

All $A \in \pi_n(Z_1^n)$ have a product structure $A = A_1 \times A_2$ with $A_1 \in \mathbb{R}^p, A_2 \in \mathbb{R}^q$.
to estimate the product of marginals !
$P_X \times P_Y$

Consistency

Problem

$\Pi = \{\pi_1(\cdot) \cdots \pi_n(\cdot) \cdots\}$

Find the sufficient conditions on the partition scheme to get,

$$\lim_{n\to\infty} \hat{I}_n(X;Y) = I(X;Y) \text{ almost surely}$$

distribution free.

Consistency

Problem

Find the sufficient conditions on the partition scheme to get,

$\Pi = \{\pi_1(\cdot) \cdots \pi_n(\cdot) \cdots\}$

$$\lim_{n\to\infty} \hat{I}_n(X;Y) = I(X;Y) \text{ almost surely}$$

distribution free.

Main Inequality

$$\left|\hat{I}_n(X;Y) - I(X;Y)\right| \leq$$

estimation error

$$\left|\sum_{A\in\pi_n(Z_1^n)} P_n(A)\cdot\log\frac{P_n(A)}{P_n(A_1\times\mathbb{R}^q)\cdot P_n(\mathbb{R}^p\times A_2)} - \sum_{A\in\pi_n(Z_1^n)} P(A)\cdot\log\frac{P(A)}{P(A_1\times\mathbb{R}^q)\cdot P(\mathbb{R}^p\times A_2)}\right|$$

$$+\left|\sum_{A\in\pi_n(Z_1^n)} P(A)\cdot\log\frac{P(A)}{P(A_1\times\mathbb{R}^q)\cdot P_n(\mathbb{R}^p\times A_2)} - I(X,Y)\right|$$

approximation error

Consistency

Problem

Find the sufficient conditions on the partition scheme $\Pi = \{\pi_1(\cdot) \cdots \pi_n(\cdot) \cdots\}$ to get,

$$\lim_{n\to\infty} \hat{I}_n(X;Y) = I(X;Y) \text{ almost surely}$$

distribution free.

Main Inequality

$$\left|\hat{I}_n(X;Y) - I(X;Y)\right| \leq$$

estimation error

$$\left| \sum_{A \in \pi_n(Z_1^n)} P_n(A) \cdot \log \frac{P_n(A)}{P_n(A_1 \times \mathbb{R}^q) \cdot P_n(\mathbb{R}^p \times A_2)} - \sum_{A \in \pi_n(Z_1^n)} P(A) \cdot \log \frac{P(A)}{P(A_1 \times \mathbb{R}^q) \cdot P(\mathbb{R}^p \times A_2)} \right|$$

Estimation Error:

Deviation of empirical measures from probabilities in the MI functional

approximation error

Consistency

Problem

Find the sufficient conditions on the partition scheme to get,

$\Pi = \{\pi_1(\cdot) \cdots \pi_n(\cdot) \cdots\}$

$$\lim_{n\to\infty} \hat{I}_n(X;Y) = I(X;Y) \text{ almost surely}$$

distribution free.

Main Inequality

$\left|\hat{I}_n(X;Y) - I(X;Y)\right| \le$ estimation error

Approximation error:
The effect of quantization reduces magnitude of information theoretic quantities.

$$+\left|\sum_{A\in\pi_n(Z_1^n)} P(A)\cdot\log\frac{P(A)}{P(A_1\times\mathbb{R}^q)\cdot P_n(\mathbb{R}^p\times A_2)} - I(X,Y)\right|$$

approximation error

Consistency

Our approach considers statistical learning tools and inequalities to bound these two sources of error asymptotically.

Consistency

Our approach considers statistical learning tools and inequalities to bound these two sources of error asymptotically.

Estimation Error Inequality

Let $\mathcal{A}$ collection of measurable partitions for $\mathbb{R}^d$

Definitions: (*Lugosi et al. 1996*)

- The maximum cell count of $\mathcal{A}$ is given by, $$\mathcal{M}(\mathcal{A}) = \sup_{\pi \in \mathcal{A}} |\pi|$$
- Let $\{x_1, .., x_n\} \subset \mathbb{R}^d$ then we can define $$\Delta(\mathcal{A}, x_1, .., x_n) = |\{\{x_1, .., x_n\} \cup \pi : \pi \in \mathcal{A}\}|$$

  and the growth function of $\mathcal{A}$ by $$\Delta_n^*(\mathcal{A}) = \max_{x_1^n \in \mathbb{R}^{d \cdot n}} \Delta(\mathcal{A}, x_1, .., x_n)$$

Theorem: (*Lugosi et al. 1996*)

Let $\mathcal{A}$ be a collection of measurable partitions for $\mathbb{R}^d$, then $\forall \epsilon > 0, \forall n,$

$$\mathbb{P}\left(\sup_{\pi \in \mathcal{A}} \sum_{A \in \pi} |P_n(A) - P(A)| > \epsilon \right) \leq 4 \Delta_{2n}^*(\mathcal{A}) 2^{\mathcal{M}(\mathcal{A})} \exp^{-\frac{n\epsilon^2}{32}},$$

where $\mathbb{P}$ the distribution of the empirical process.

Consistency

Our approach considers statistical learning tools and inequalities to bound these two sources of error asymptotically.

Approximation error Result

Definitions:

- Let $A \in \mathbb{R}^d$ a measurable event, then its diameter is

$$diam(A) = \sup_{x,y \in A} ||x - y||,$$

Theorem: (*Silva et al. 2007*)

Let $\Pi = \{\pi_1(\cdot), \pi_2(\cdot), \cdots\}$ our data-dependent partition scheme if, driven by iid realizations $Z_1 = (X_1, Y_1), .., Z_n = (X_n, Y_n)$ of $\mathrm{P}_{X,Y}$

$$\lim_{n \to \infty} P_{X,Y}\left(\left\{z \in \mathbb{R}^d : diam(\pi_n(z|Z_1^n)) > \epsilon\right\}\right) \to 0,$$

SHRINKING CELL CONDITION

then,

$$\lim_{n \to \infty} \sum_{A \in \pi_n(Z_1^n)} P_{X,Y}(A) \cdot \log \frac{P_{X,Y}(A)}{P_X \times P_Y(A)} = I(X, Y),$$

$\mathbb{P}$- almost surely.

THE MAIN RESULT

THEOREM

Let $\Pi = \{\pi_1(\cdot), \pi_2(\cdot), \cdots\}$ a partition scheme driven by the i.i.d. realizations $Z_1 = (X_1, Y_1), .., Z_n = (X_n, Y_n)$ of $\mathrm{P}_{X,Y}$. If there exist $\tau \in (0,1)$ such that,

i) $\lim_{n\to\infty} \frac{1}{n^\tau} \log \mathcal{S}_n(\mathcal{C}_{[1-p],n}) = 0,$

$\lim_{n\to\infty} \frac{1}{n^\tau} \log \mathcal{S}_n(\mathcal{C}_{[p+1-d],n}) = 0,$

ii) $$\lim_{n\to\infty} \frac{1}{n^\tau} \log \Delta_n^*(\mathcal{A}_n) = 0$$

iii) $$\lim_{n\to\infty} \frac{1}{n^\tau} \mathcal{M}(\mathcal{A}_n) = 0,$$

iv) $\exists\ (k_n)_{n\in\mathbb{N}}$ a sequence of non-negative numbers, with $(k_n) \approx (n^{0.5+\tau/2})$ such that, $\forall A \in \pi_n(Z_1, .., Z_n)$

$$\mathrm{P}_n(A) \geq \frac{k_n}{n},$$

v) $\forall \epsilon > 0$ (shrinking cell condition)

$$\lim_{n\to\infty} P_{X,Y}\left(\left\{z \in \mathbb{R}^d : diam(\pi_n(z|Z_1^n)) > \epsilon\right\}\right) \to 0,$$

then,

$$\lim_{n\to\infty} \hat{I}_n(X;Y) = I(X,Y),$$

$\mathbb{P}$ (empirical process distribution) - almost surely.

Content Outline

Main Result

- The Estimator
- Approximation Error results
- Estimation error results
- Conditions for Strong Consistency

Applications

- Multivariate Statistically Equivalent Block
- Tree-structured Partitions
- Consistency

- Experiments

- Future Work

Statistically equivalent blocks (Gessaman)

Idea: use $Z_1, .., Z_n$ to partition the space in equal empirical mass bins

Pseudo-Algorithm

Let $k_n > 0$ the minimum number of sampled points per bin.

★ $T_n = \lfloor (n/k_n)^{1/d} \rfloor$ # axis parallel partition per-coordinate

✤ $\pi_n(Z_1^n)$ induced by T_n statistically equivalent axis - parallel splits per coordinate => $(T_n)^d$ bins

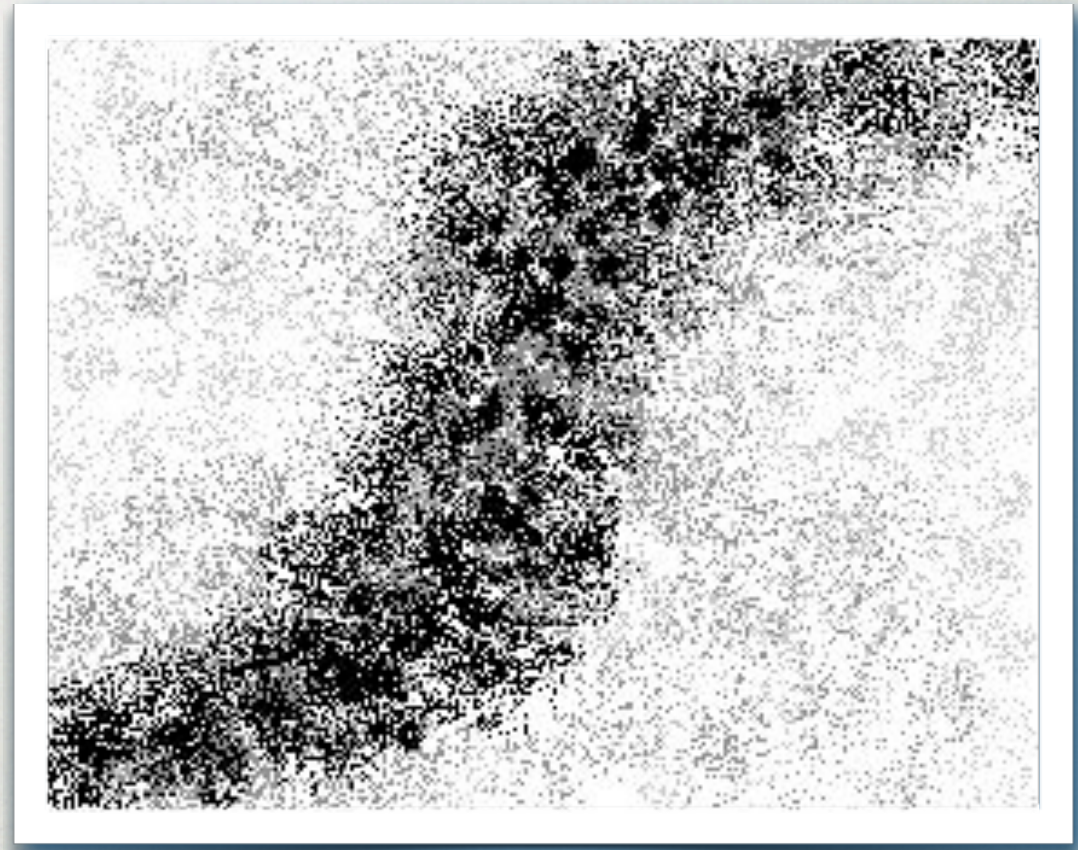


Fig.1: Statistically equivalent blocks

Statistically equivalent blocks (Gessaman)

Idea: use $Z_1, .., Z_n$ to partition the space in equal empirical mass bins

Pseudo-Algorithm

Let $k_n > 0$ the minimum number of sampled points per bin.

★$\mathrm{T}_n = \lfloor (n/k_n)^{1/d} \rfloor$ # axis parallel partition per-coordinate

✤ $\pi_n(Z_1^n)$ induced by T_n statistically equivalent axis - parallel splits per coordinate => $(\mathrm{T}_n)^d$ bins

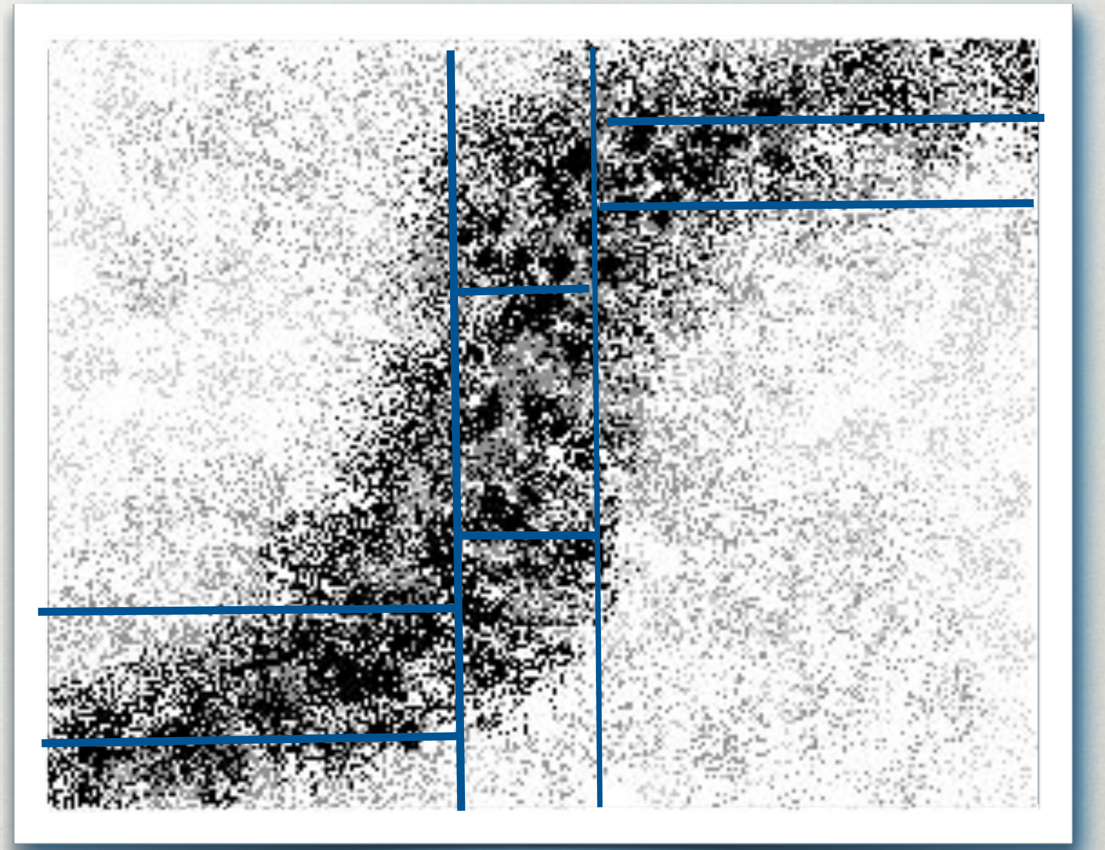


Fig.1: Statistically equivalent blocks

Idea: use $Z_1, .., Z_n$ to partition the space in equal empirical mass bins

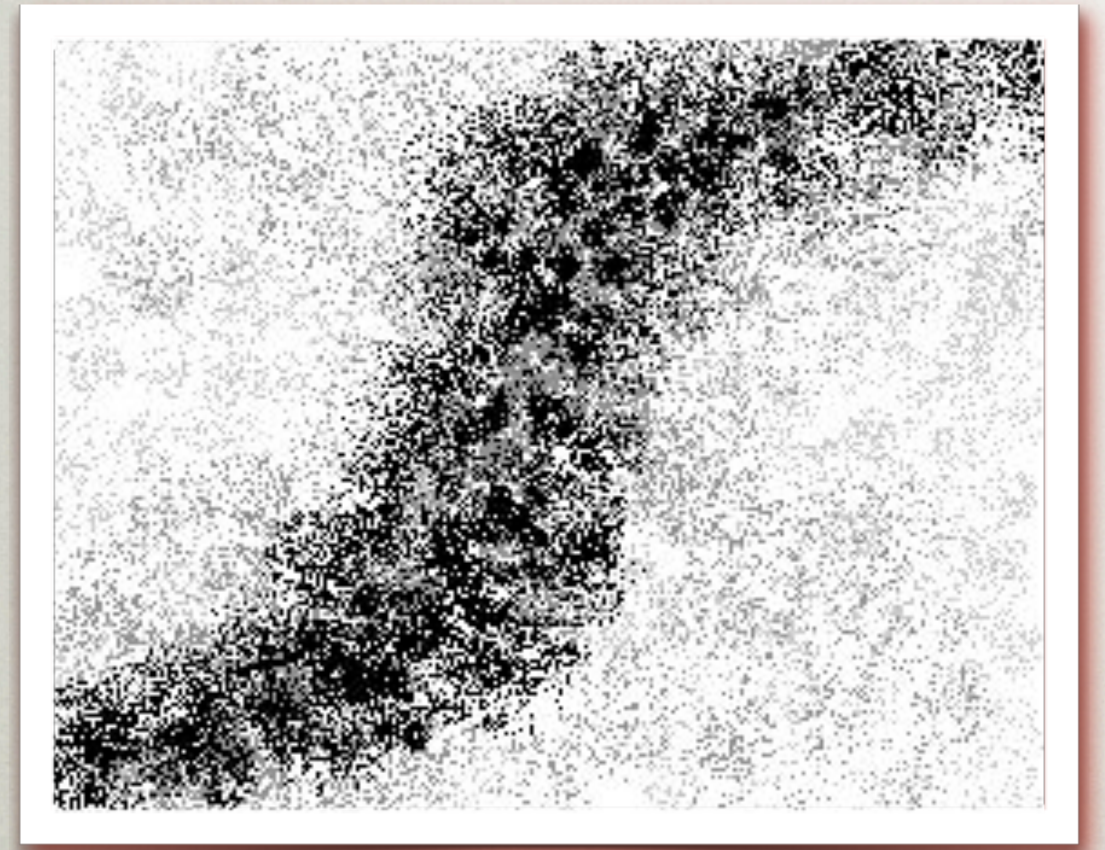


Fig.2: Tree-Structured Partitions

Pseudo-Algorithm

Let $k_n > 0$ the minimum number of sampled points per bin.

★ initialization $U = \{\mathbb{R}^d\}$

✤ For each $A \in U$

- if $P_n(A) \geq \frac{2 \cdot k_n}{n}$,
  - chose a dimension to partition the space in equal emp. mass, axis-parallel => $A \mapsto B_1, B_2$
  - update $U = U \setminus \{A\} \cup \{B_1, B_2\}$

✤ end For each, $P_n(A) < \frac{2 \cdot k_n}{n}, \forall A \in \mathcal{U}$

✤ finally $\pi_n(Z_1^n) = \mathcal{U}$.

Tree-structured (dyadic) Partitions

Idea: use $Z_1, .., Z_n$ to partition the space in equal empirical mass bins

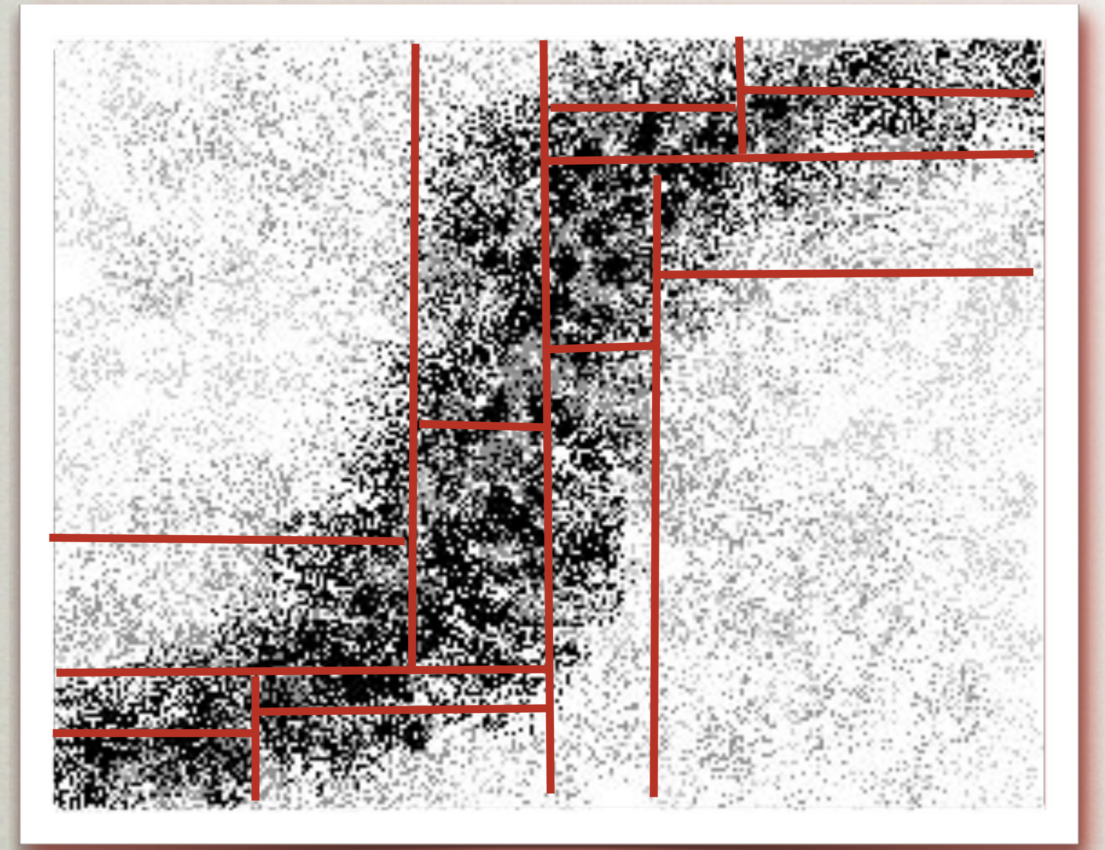


Fig.2: Tree-Structured Partitions

Pseudo-Algorithm

Let $k_n > 0$ the minimum number of sampled points per bin.

★ initialization $U = \{\mathbb{R}^d\}$

✤ For each $A \in U$

- if $P_n(A) \geq \frac{2 \cdot k_n}{n}$,
  - chose a dimension to partition the space in equal emp. mass, axis-parallel => $A \mapsto B_1, B_2$
  - update $U = U \setminus \{A\} \cup \{B_1, B_2\}$

✤ end For each, $P_n(A) < \frac{2 \cdot k_n}{n}, \forall A \in \mathcal{U}$

✤ finally $\pi_n(Z_1^n) = \mathcal{U}$.

Statistically equivalent blocks (Gessaman)

Tree-structured (dyadic) Partitions

Idea: use $Z_1, .., Z_n$ to partition the space in equal empirical mass bins

Pseudo-Algorithm

Let $k_n > 0$ the minimum number of sampled points per bin.

★ $T_n = \lfloor (n/k_n)^{1/d} \rfloor$ # axis parallel partition per-coordinate

✤ $\pi_n(Z_1^n)$ induced by T_n statistically equivalent axis - parallel splits per coordinate => $(T_n)^d$ bins

Pseudo-Algorithm

Let $k_n > 0$ the minimum number of sampled points per bin.

★ initialization $U = \{\mathbb{R}^d\}$

✤ For each $A \in U$

- if $P_n(A) \geq \frac{2 \cdot k_n}{n}$,
  - chose a dimension to partition the space in equal emp. mass, axis-parallel => $A \mapsto B_1, B_2$
  - update $U = U \setminus \{A\} \cup \{B_1, B_2\}$

✤ end For each, $P_n(A) < \frac{2 \cdot k_n}{n}, \forall A \in \mathcal{U}$

✤ finally $\pi_n(Z_1^n) = \mathcal{U}$.

Key question

What is the asymptotic behavior of k_n that guarantee consistency?

Consistency

Statistically equivalent blocks

Tree-structured (dyadic) Partitions k_n

THEOREM

Under the problem setting of Theorem 1, (and respective construction), if $(k_n) \approx (n^{0.5+\tau/2})$ for $\tau \in (1/3, 1)$ then,

$$\lim_{n\to\infty} \hat{I}_n(X;Y) = I(X,Y),$$

$\mathbb{P}$ - almost surely.

Remarks

✳ consistency requires a sub-linear behavior on k_n

✳ the results are stronger than the conditions for L1 consistency

❖ $(a_n) \approx (b_n)$ if an only if $\lim_{n\to\infty} \frac{a_n}{b_n} = C$, for some $C > 0$.

Outline

Main Result

- The Estimator
- Approximation Error results
- Estimation error results
- Conditions for Strong Consistency

Applications

- Multivariate Statistically Equivalent Block
- Tree-structured Partitions
- Consistency

- Experiments

- Future Work

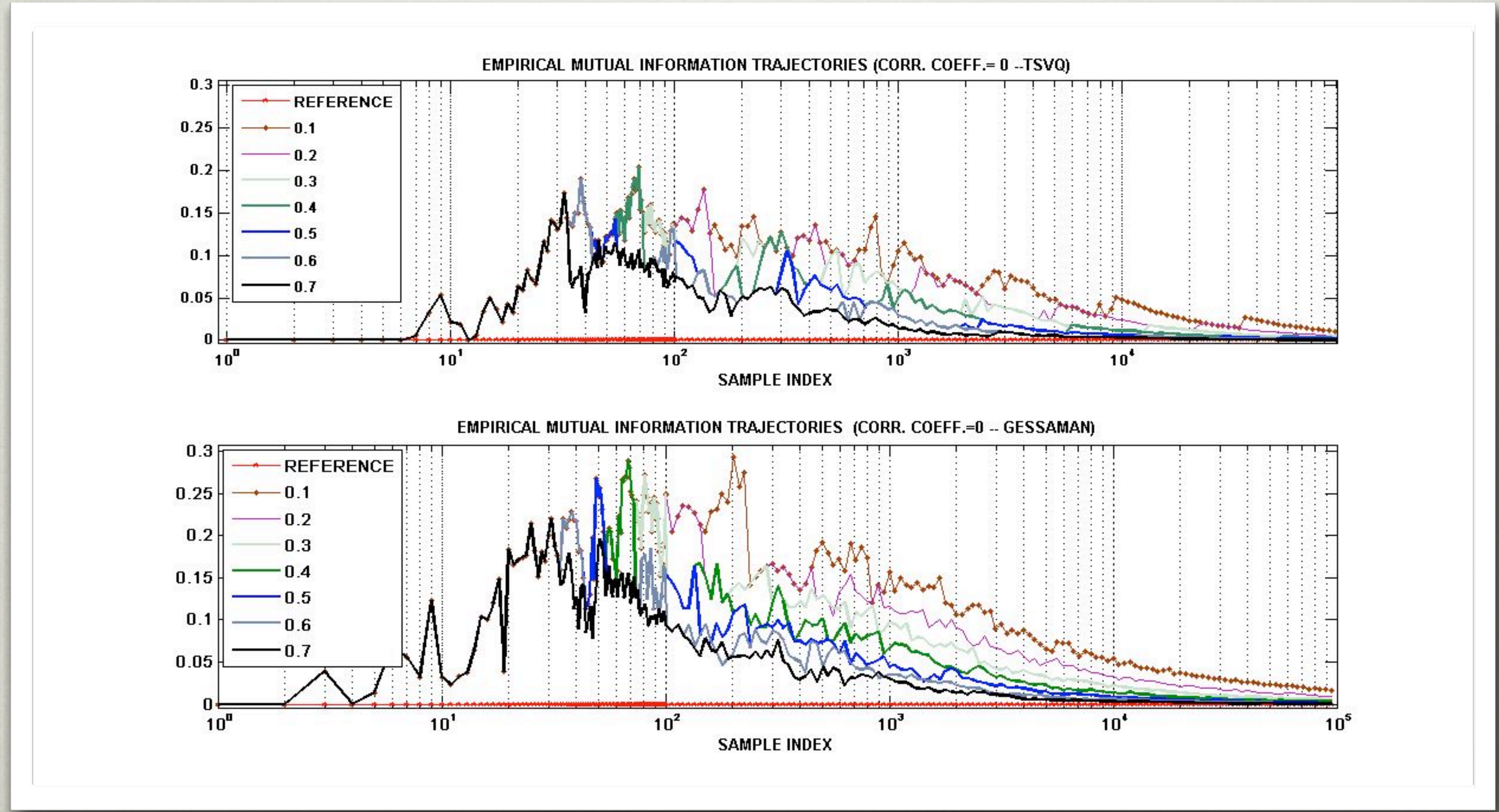
Simulation Analysis

General Setting

- joint Gaussian distribution simulated (two dimensional case)
  - closed form expression for MI
  - ML estimation available (used as a benchmark!)
- evaluation of different level of statistical dependencies
  - correlation coefficient in the range of [0-1]
- across different sampling lengths of the empirical process
- Bias and standard deviations computed using 1000 realizations
- kernel plug-in estimates and non-product histogram based estimates evaluated

Simulation Analysis

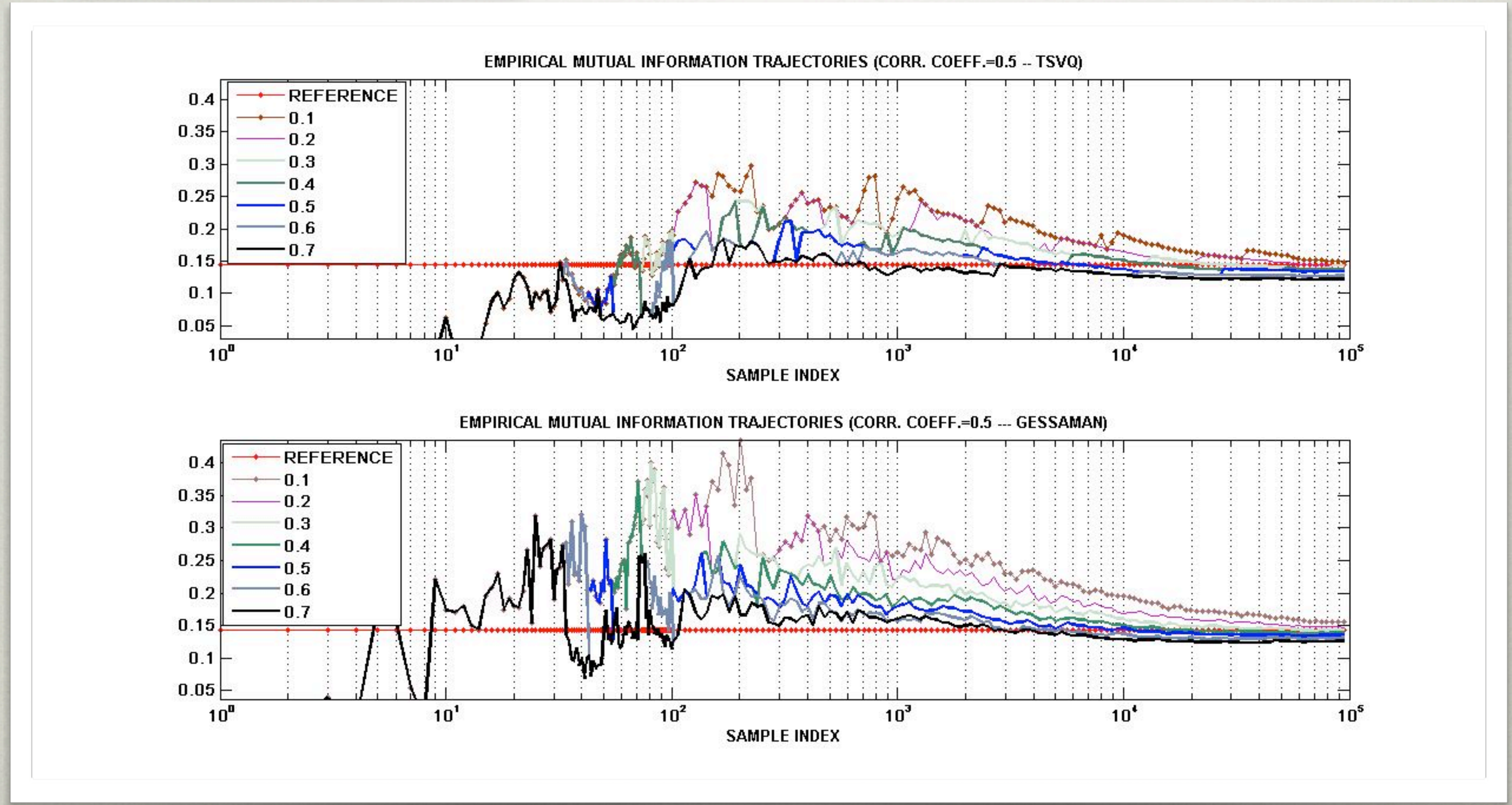
Setting i)



Empirical mutual Information trajectories for TSVQ and Gessaman scheme.
Data simulated with correlation coefficient r=0

Simulation Analysis

Setting i)



Empirical mutual Information trajectories for TSVQ and Gessaman scheme.
Data simulated with correlation coefficient r=0.5

Simulation Analysis

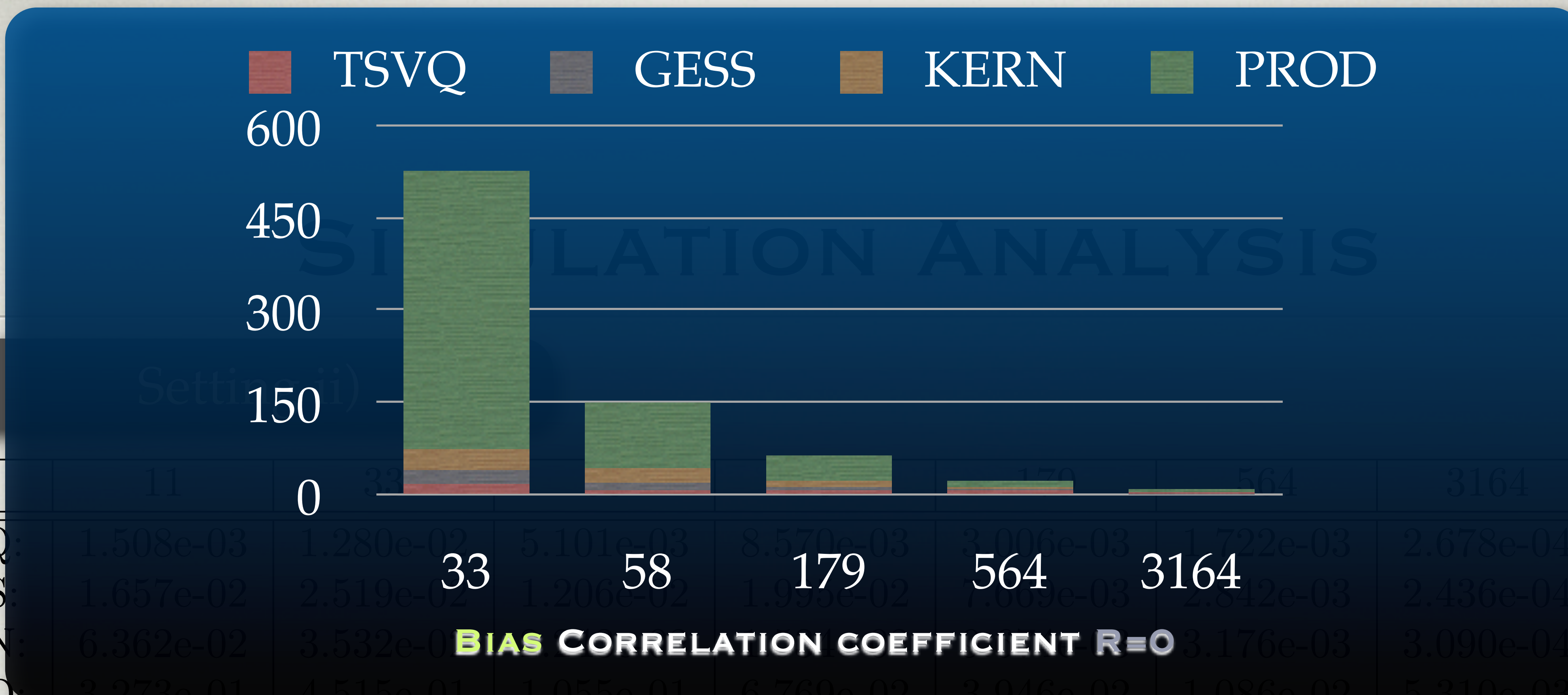
Setting ii)

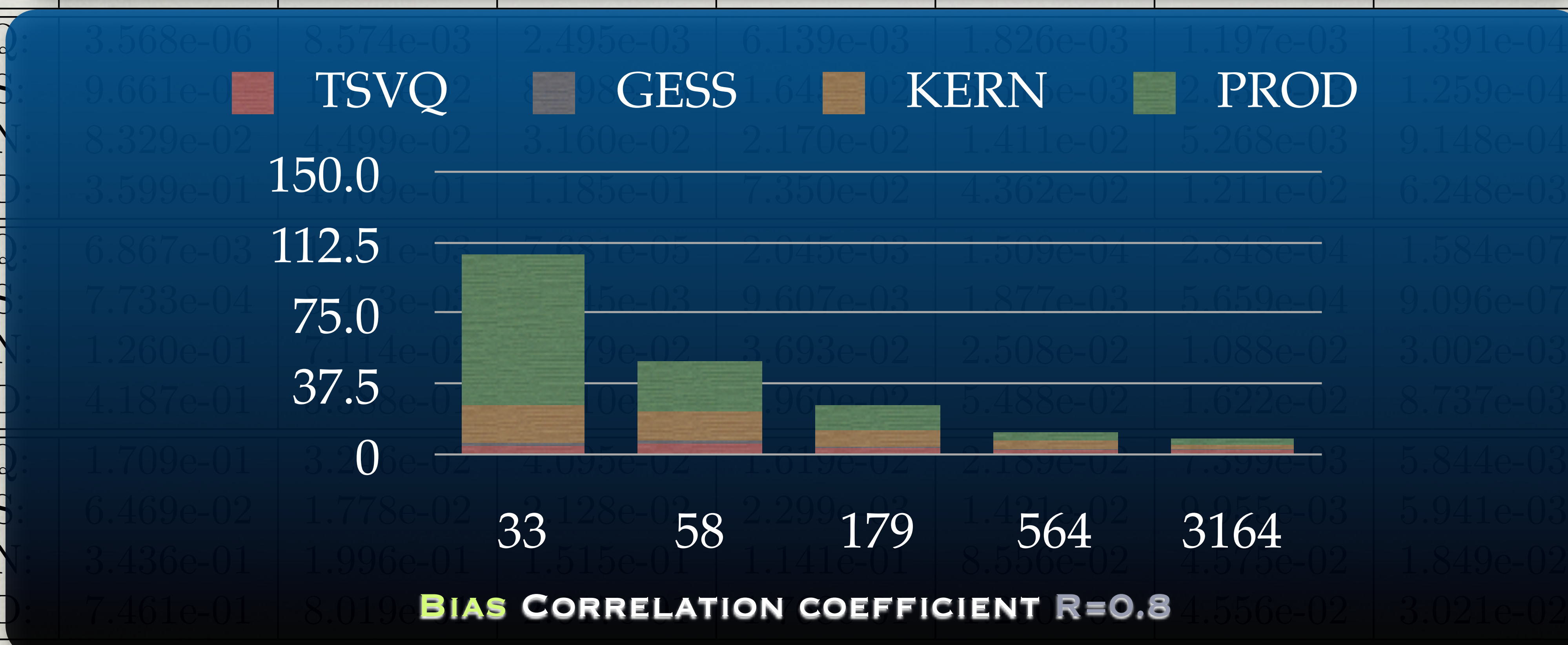
		11	33	58	101	179	564	3164	5626
0	TSVQ:	1.508e-03	1.280e-02	5.101e-03	8.570e-03	3.006e-03	1.722e-03	2.678e-04	8.479e-05
	GESS:	1.657e-02	2.519e-02	1.206e-02	1.995e-02	7.669e-03	2.842e-03	2.436e-04	1.594e-04
	KERN:	6.362e-02	3.532e-02	2.243e-02	1.534e-02	9.674e-03	3.176e-03	3.090e-04	9.035e-05
	PROD:	3.273e-01	4.515e-01	1.055e-01	6.769e-02	3.946e-02	1.086e-02	5.310e-03	5.652e-03
0.3	TSVQ:	3.568e-06	8.574e-03	2.495e-03	6.139e-03	1.826e-03	1.197e-03	1.391e-04	2.359e-05
	GESS:	9.661e-03	1.914e-02	8.498e-03	1.645e-02	5.515e-03	2.037e-03	1.259e-04	8.186e-05
	KERN:	8.329e-02	4.499e-02	3.160e-02	2.170e-02	1.411e-02	5.268e-03	9.148e-04	4.355e-04
	PROD:	3.599e-01	4.769e-01	1.185e-01	7.350e-02	4.362e-02	1.211e-02	6.248e-03	6.680e-03
0.5	TSVQ:	6.867e-03	1.951e-03	7.681e-05	2.045e-03	1.509e-04	2.848e-04	1.584e-07	5.105e-05
	GESS:	7.733e-04	8.473e-03	2.645e-03	9.607e-03	1.877e-03	5.659e-04	9.096e-07	1.270e-06
	KERN:	1.260e-01	7.114e-02	5.079e-02	3.693e-02	2.508e-02	1.088e-02	3.002e-03	1.854e-03
	PROD:	4.187e-01	5.388e-01	1.410e-01	8.960e-02	5.488e-02	1.622e-02	8.737e-03	9.381e-03
0.8	TSVQ:	1.709e-01	3.246e-02	4.095e-02	1.619e-02	2.189e-02	7.399e-03	5.844e-03	6.679e-03
	GESS:	6.469e-02	1.778e-02	2.128e-02	2.299e-03	1.421e-02	9.955e-03	5.941e-03	4.351e-03
	KERN:	3.436e-01	1.996e-01	1.515e-01	1.141e-01	8.556e-02	4.575e-02	1.849e-02	1.354e-02
	PROD:	7.461e-01	8.019e-01	2.617e-01	1.768e-01	1.290e-01	4.556e-02	3.021e-02	3.275e-02

Bias for the non-parametric mutual information estimates (Gessaman partition scheme (GESS), tree-structured partition (TSVQ), classical product partition (PROD) and a kernel plug-in estimate (KERN))

Simulation Analysis

Setting (ii)





R		5626
0	TSVQ:	8.479e-05
	GESS:	1.594e-04
	KERN:	9.035e-05
	PROD:	5.652e-03
0.3	TSVQ:	2.359e-05
	GESS:	8.186e-05
	KERN:	4.355e-04
	PROD:	6.680e-03
0.5	TSVQ:	5.105e-05
	GESS:	1.270e-06
	KERN:	1.854e-03
	PROD:	9.381e-03
0.8	TSVQ:	6.679e-03
	GESS:	4.351e-03
	KERN:	1.354e-02
	PROD:	3.275e-02

Bias for the non-parametric mutual information estimates (Gessaman partition scheme (GESS), tree-structured partition (TSVQ), classical product partition (PROD) and a kernel plug-in estimate (KERN)

Simulation Analysis

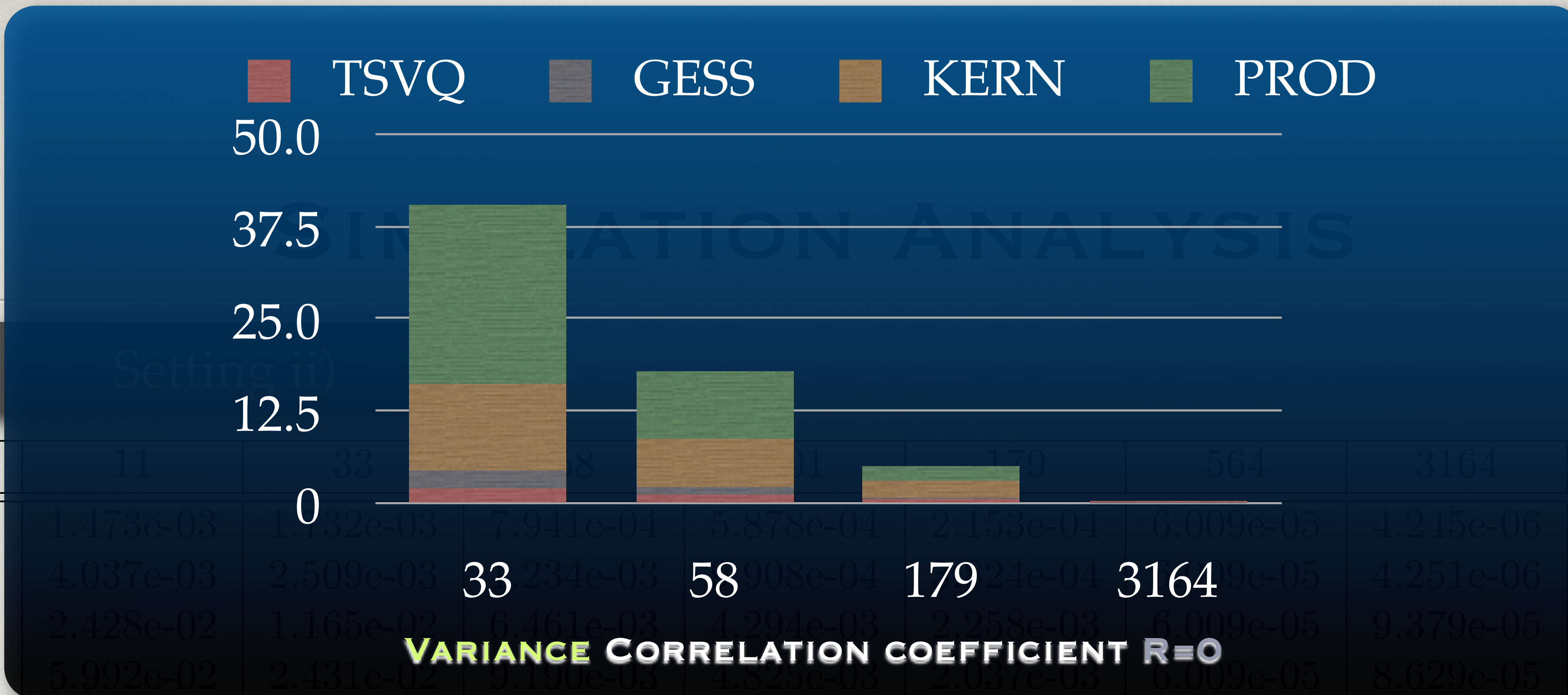
Setting ii)

		11	33	58	101	179	564	3164	5626
	TSVQ:	1.473e-03	1.732e-03	7.941e-04	5.878e-04	2.153e-04	6.009e-05	4.245e-06	1.534e-06
0	GESS:	4.037e-03	2.509e-03	1.234e-03	9.908e-04	3.824e-04	6.009e-05	4.251e-06	1.992e-06
	KERN:	2.428e-02	1.165e-02	6.461e-03	4.294e-03	2.258e-03	6.009e-05	9.379e-05	4.941e-05
	PROD:	5.992e-02	2.431e-02	9.190e-03	4.825e-03	2.037e-03	6.009e-05	8.629e-05	6.193e-05
	TSVQ:	1.406e-03	2.563e-03	1.418e-03	1.022e-03	5.302e-04	1.926e-04	3.049e-05	1.437e-05
0.3	GESS:	5.565e-03	3.369e-03	1.729e-03	1.384e-03	6.555e-04	1.926e-04	3.006e-05	1.632e-05
	KERN:	2.565e-02	1.198e-02	6.465e-03	4.127e-03	2.461e-03	1.926e-04	1.145e-04	6.314e-05
	PROD:	6.055e-02	2.276e-02	9.664e-03	5.346e-03	2.454e-03	1.926e-04	1.212e-04	7.653e-05
	TSVQ:	1.563e-03	3.355e-03	2.268e-03	1.697e-03	9.138e-04	3.759e-04	6.863e-05	3.585e-05
0.5	GESS:	6.514e-03	4.386e-03	2.632e-03	2.079e-03	9.792e-04	3.759e-04	7.069e-05	3.894e-05
	KERN:	2.820e-02	1.322e-02	7.290e-03	4.674e-03	2.497e-03	3.759e-04	1.494e-04	7.831e-05
	PROD:	6.687e-02	2.631e-02	1.327e-02	6.010e-03	2.894e-03	3.759e-04	1.769e-04	1.146e-04
	TSVQ:	1.007e-03	3.621e-03	2.464e-03	2.227e-03	1.333e-03	6.586e-04	1.386e-04	7.679e-05
0.8	GESS:	6.426e-03	4.900e-03	3.601e-03	3.257e-03	1.521e-03	6.586e-04	1.418e-04	8.362e-05
	KERN:	3.132e-02	1.534e-02	1.042e-02	6.225e-03	3.412e-03	6.586e-04	2.345e-04	1.244e-04
	PROD:	7.025e-02	2.808e-02	1.602e-02	8.991e-03	4.766e-03	6.586e-04	4.581e-04	3.277e-04

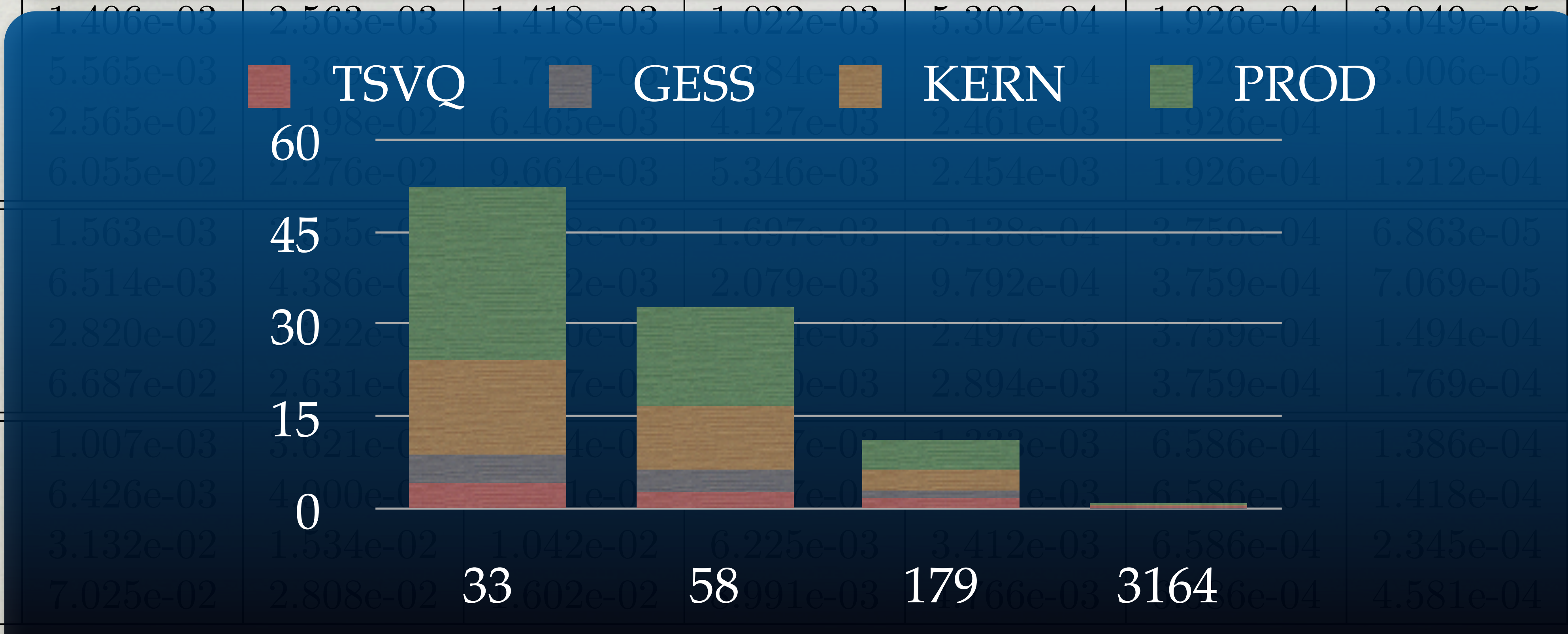
Variance for the non-parametric mutual information estimates (Gessaman partition scheme (GESS), tree-structured partition (TSVQ), classical product partition (PROD) and a kernel plug-in estimate (KERN))

Simulation Analysis

Setting ii)



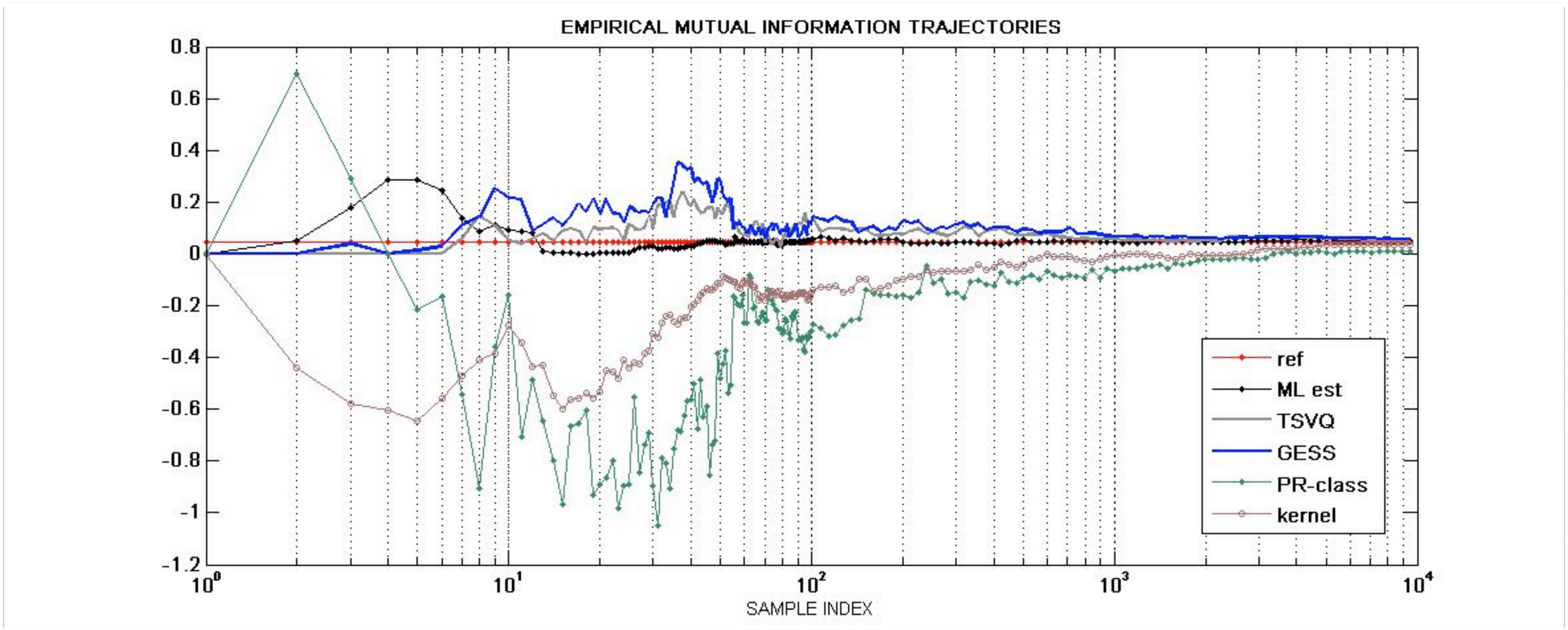
R	Method	5626
0	TSVQ:	1.534e-06
	GESS:	1.992e-06
	KERN:	4.941e-05
	PROD:	6.193e-05
0.3	TSVQ:	1.437e-05
	GESS:	1.632e-05
	KERN:	6.314e-05
	PROD:	7.653e-05
0.5	TSVQ:	3.585e-05
	GESS:	3.894e-05
	KERN:	7.831e-05
	PROD:	1.146e-04
0.8	TSVQ:	7.679e-05
	GESS:	8.362e-05
	KERN:	1.244e-04
	PROD:	3.277e-04



Varia... estimates (Gessaman partition scheme (GESS), tree-structured partition (TSVQ), classical product partition (PROD) and a kernel plug-in estimate (KERN)

Simulation Analysis

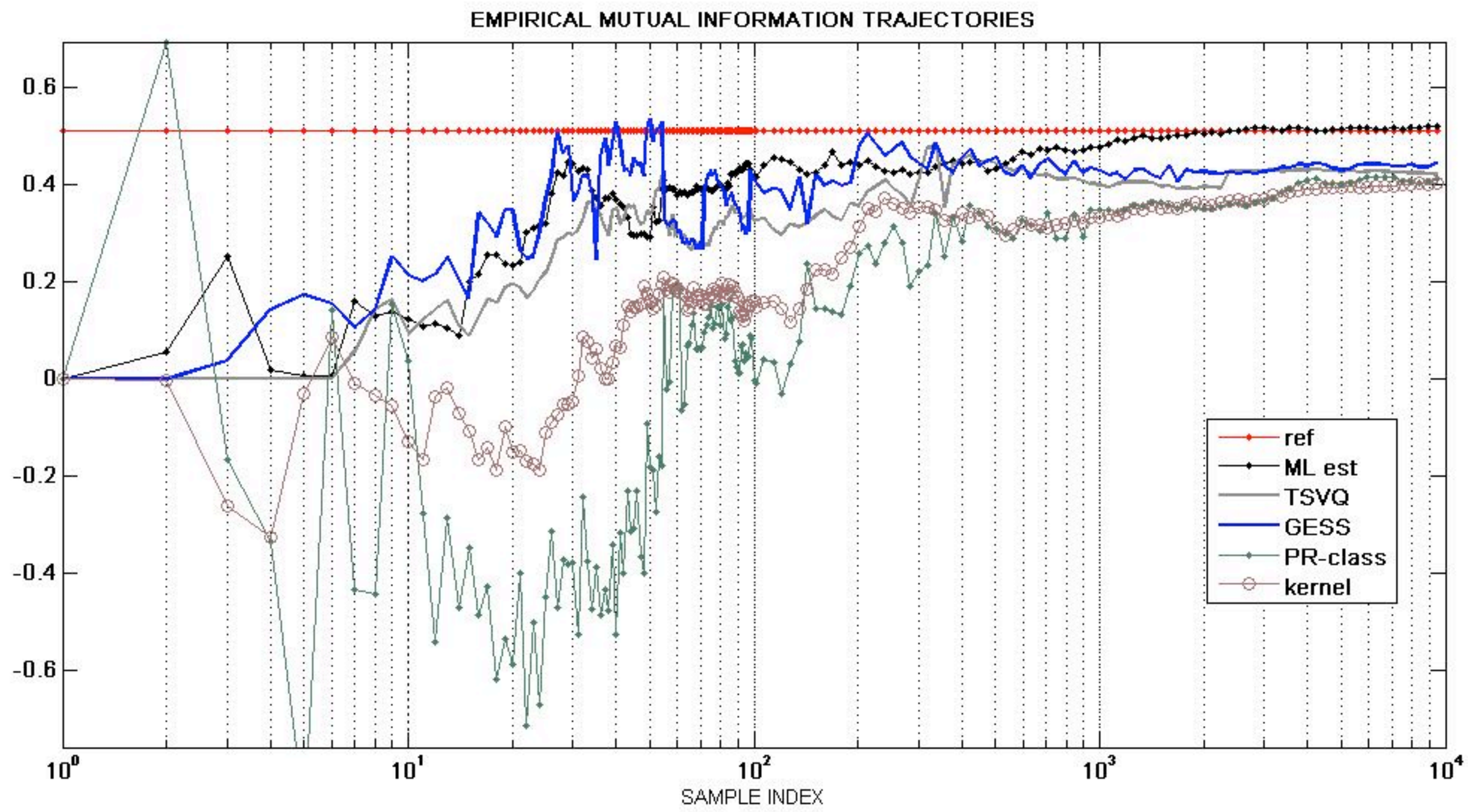
Setting ii)



Empirical mutual Information trajectories for different estimation techniques. Data simulated with correlation coefficient r=0.3

Simulation analysis

Setting ii)



Empirical mutual Information trajectories for different estimation techniques. Data simulated with correlation coefficient r=0.8

Extensions

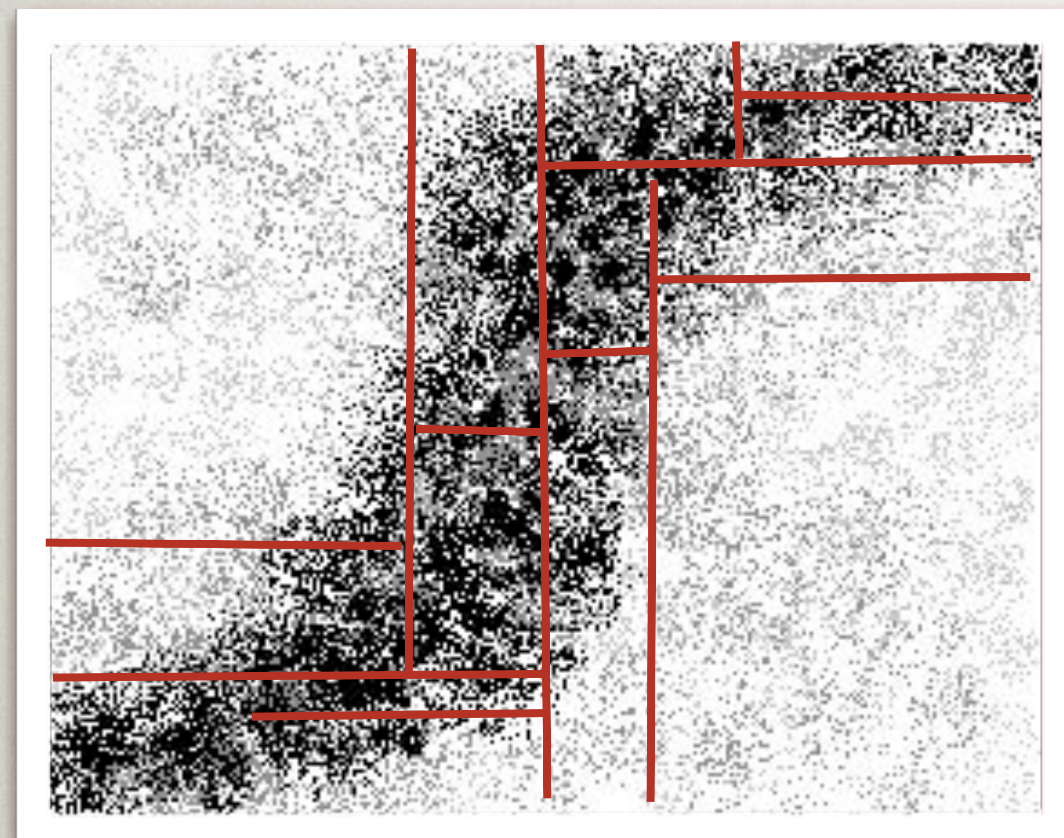
- Working in improvements for the case of TSP
- Applications in test of independence
- Similar results for the case of the KLD

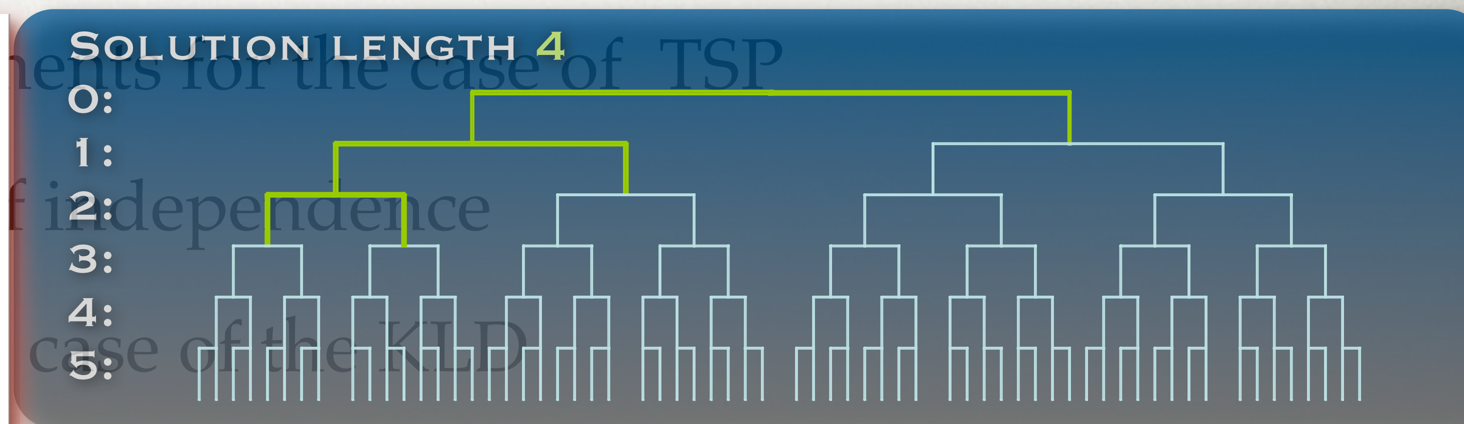
Extensions

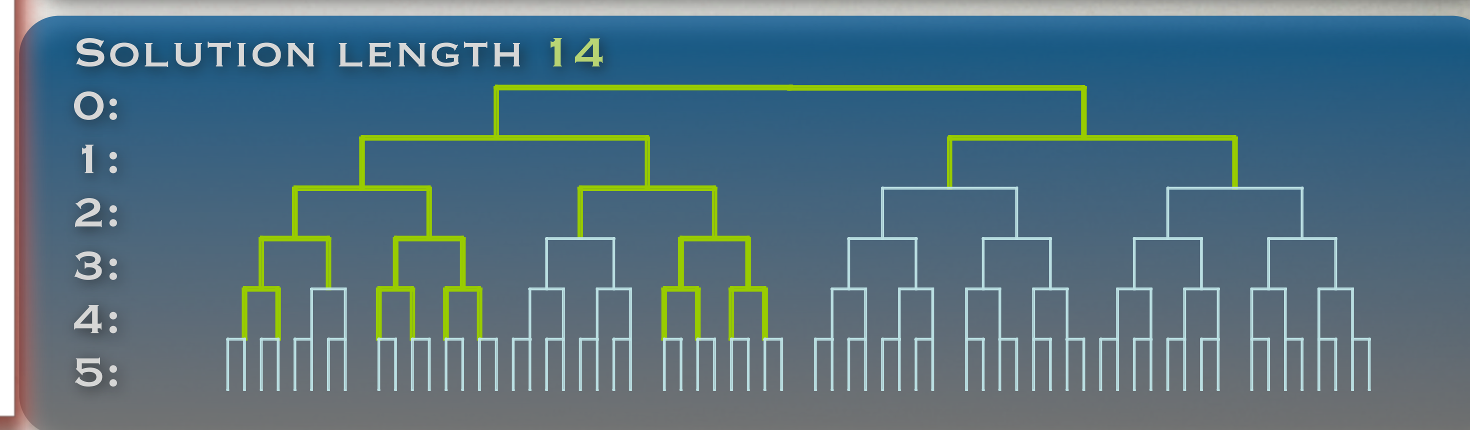
$$\hat{\mathbf{T}}^* = \arg \min_{T \ll \mathbf{T}_{full}} \left[L_{\mathcal{X}_T} - L_{\mathcal{X}} \right] + \Delta(\hat{P}_{X_T,Y}, P_{X_T,Y}).$$

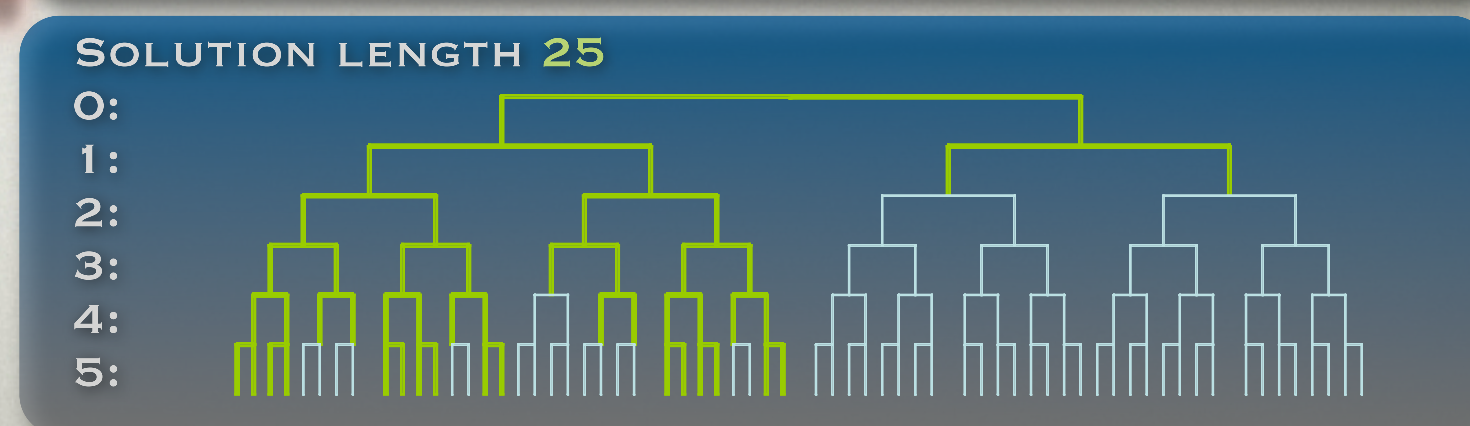
Approximation error

Estimation error









Extensions

- Working in improvements for the case of TSP
- Applications in test of independence
- Similar results for the case of the KLD

Thank you!