Entropic-graphs: Theory

Alfred O. Hero

Dept. EECS, Dept Biomed. Eng., Dept. Statistics

University of Michigan - Ann Arbor

hero@eecs.umich.edu

http://www.eecs.umich.edu/~hero

Collaborators: H. Heemuchwala, J. Costa, B. Ma, O. Michel

- Rényi α-Entropy and Rényi α-Divergence
- Minimal graphs and entropic graphs
- Entropic graphs: asymptotic convergence results
- Extensions to partitioning approximations
- Open problems

Rényi Entropy and Rényi Divergence

- $X \sim f(x)$ a d-dimensional random vector.

- Rényi Entropy of order α

$$H_\alpha(f) = \frac{1}{1-\alpha} \ln \int f^\alpha(x) dx \tag{1}$$

- The Rényi α-divergence of fractional order $\alpha \in [0,1]$ [Rényi:61,70]

$$\begin{aligned} D_\alpha(f_1 \parallel f_0) &= \frac{1}{\alpha - 1} \ln \int f_1 \left(\frac{f_1}{f_0}\right)^\alpha dx \\ &= \frac{1}{\alpha - 1} \ln \int f_1^\alpha f_0^{1-\alpha} dx \end{aligned}$$

– α-Divergence vs α-Entropy

$$H_\alpha(f_1) = \frac{1}{1-\alpha} \ln \int f_1^\alpha dx = -D_\alpha(f_1 \parallel f_0)|_{f_0 = U([0,1]^d)}$$

– α-Divergence vs. Batthacharyya-Hellinger distance

$$\begin{aligned} D_{BH}^2(f_1 \parallel f_0) &= \int \left(\sqrt{f_1} - \sqrt{f_0}\right)^2 dx \\ &= 2\left(1 - \exp\left(\tfrac{1}{2} D_{\frac{1}{2}}(f_1 \parallel f_0)\right)\right) \end{aligned}$$

– α-Divergence vs. Kullback-Liebler divergence (Shannon MI)

$$\lim_{\alpha \to 1} D_\alpha(f_1 \| f_0) = \int f_1 \ln \frac{f_1}{f_0} dx.$$

Entropic Graphs

A graph G of degree l consists of vertices and edges

- vertices are subset of $\mathcal{X}_n = \{x_i\}_{i=1}^n$: n points in $\mathbf{R}^d$
- edges are denoted $\{e_{ij}\}$
- for any i: $card\{e_{ij}\}_j \leq l$

Weight (with power exponent γ) of G

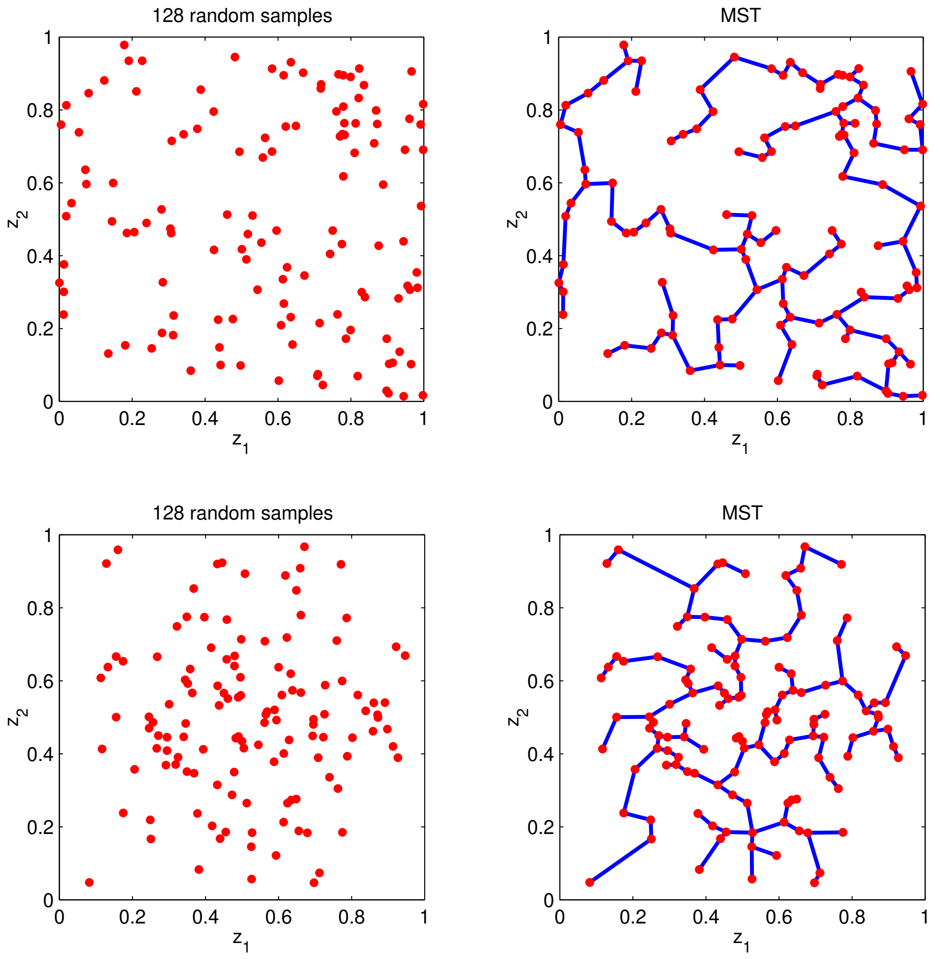
$$L_\gamma^{\mathrm{G}}(\mathcal{X}_n) = \sum_{e \in \mathrm{G}} \|e\|^\gamma$$

Minimal Spanning Tree (MST)

Let $\mathrm{T}_n = \mathrm{T}(\mathcal{X}_n)$ denote the possible sets of edges in the class of acyclic graphs spanning $\mathcal{X}_n$ (spanning trees).

The Euclidean Power Weighted MST achieves

$$L_\gamma^{\mathrm{MST}}(\mathcal{X}_n) = \min_{\mathrm{T}_n} \sum_{e \in \mathrm{T}_n} \|e\|^\gamma.$$



128 random samples
MST
128 random samples
MST
z_1
z_2
0
0.2
0.4
0.6
0.8
1

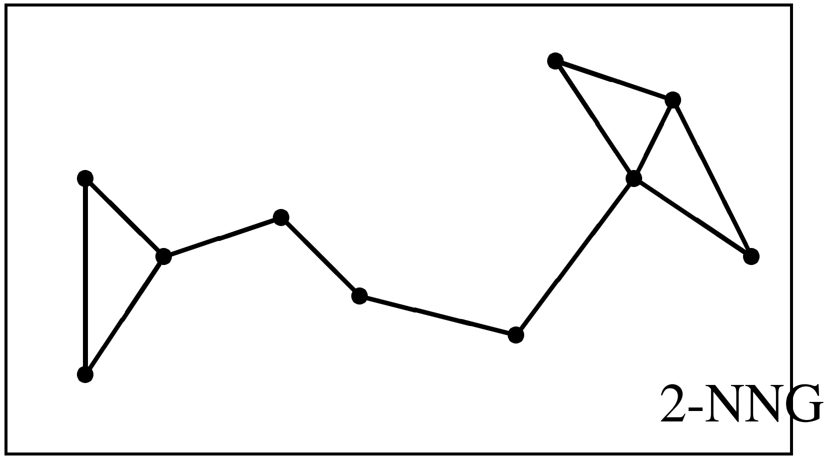
Figure 1:

Minimal Euclidean graphs: k-NNG

Example: k-**Nearest Neighbors Graph** (k-**NNG**)

Let $\mathcal{N}_{k,i}(\mathcal{X}_n)$ denote the possible sets of k edges connecting point x_i to all other points in $\mathcal{X}_n$.
The Euclidean Power Weighted k-NNG is

$$L_\gamma^{k-NNG}(\mathcal{X}_n) = \sum_{i=1}^{n} \min_{\mathcal{N}_{k,i}(\mathcal{X}_n)} \sum_{e \in \mathcal{N}_{k,i}(\mathcal{X}_n)} |e|^\gamma$$



Large n behavior of MST

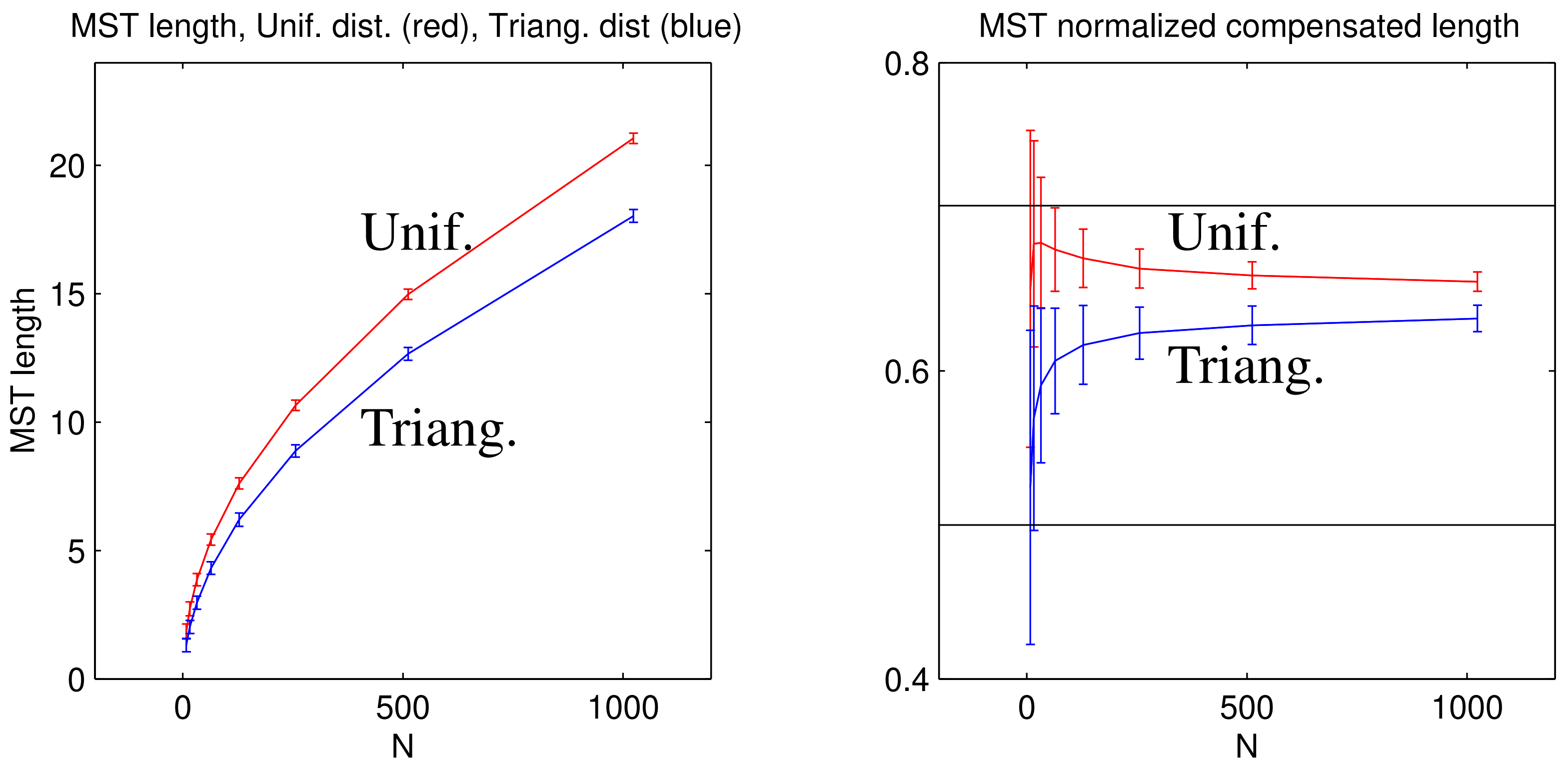


Figure: MST and log MST weights as function of the number of samples.

Asymptotics: the BHH Theorem

Define the MST length functional

$$L_\gamma(\mathcal{X}_n) = \min_{\mathrm{T}_n} \sum_{e \in \mathrm{T}_n} \|e\|^\gamma.$$

Theorem 1 *(**Beardwood&etal:Camb59**) Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be an i.i.d. realization from a Lebesgue density f on $[0,1]^d$.*

$$\lim_{n\to\infty} L_\gamma(\mathcal{X}_n)/n^{(d-\gamma)/d} = \beta_{L_\gamma,d} \int f(x)^{(d-\gamma)/d} dx, \qquad (a.s.)$$

Or, letting $\alpha = (d-\gamma)/d$

$$\lim_{n\to\infty} L_\gamma(\mathcal{X}_n)/n^\alpha = \beta_{L_\gamma,d} \exp\left((1-\alpha) H_\alpha(f)\right), \qquad (a.s.)$$

Question: What is r.m.s. rate of convergence?

Find constant r such that

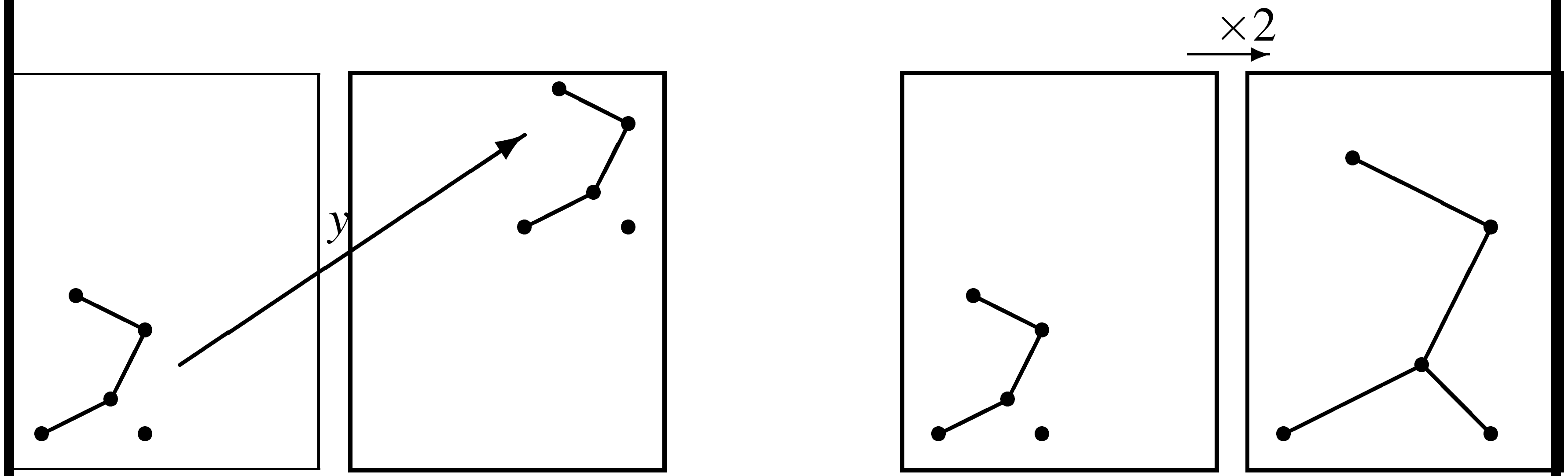
$$E^{1/2}\left[\left|L_\gamma(\mathcal{X}_n)/n^{(d-\gamma)/d} - \beta_{L_\gamma,d}\int f(x)^{(d-\gamma)/d}dx\right|^2\right] \leq O(n^{-r})$$

Method: adopt Yukich's general setting of quasi-additive continuous Euclidean functionals

Quasi-additive continuous Euclidean functionals

L_γ is a Euclidean functional over $\mathbf{R}^d$ if for every finite subset F of $[0,1]^d$

$$\begin{aligned} \forall\, \mathbf{y} \in \mathbf{R}^d,\ L_\gamma(F+\mathbf{y}) &= L_\gamma(F), \quad \text{(translation invariance)} \\ \forall\, c>0,\ L_\gamma(cF) &= c^\gamma L_\gamma(F), \quad \text{(homogeneity)} \end{aligned}$$



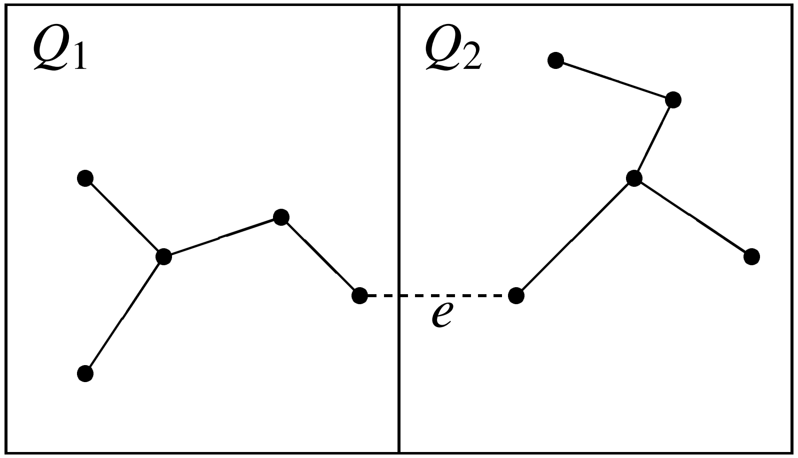
Quasi-additive continuous Euclidean functionals

Let L_γ be a Euclidean functional. Define

• **Null Condition**: $L_\gamma(\phi) = 0$, where ϕ is the null set.

• **Subadditivity**: There exists a constant C_1 with the following property: For any uniform resolution $1/m$-partition Q^m

$$L_\gamma(F) \le m^{-1} \sum_{i=1}^{m^d} L_\gamma(m[(F \cap Q_i) - q_i]) + C_1 m^{d-\gamma}$$



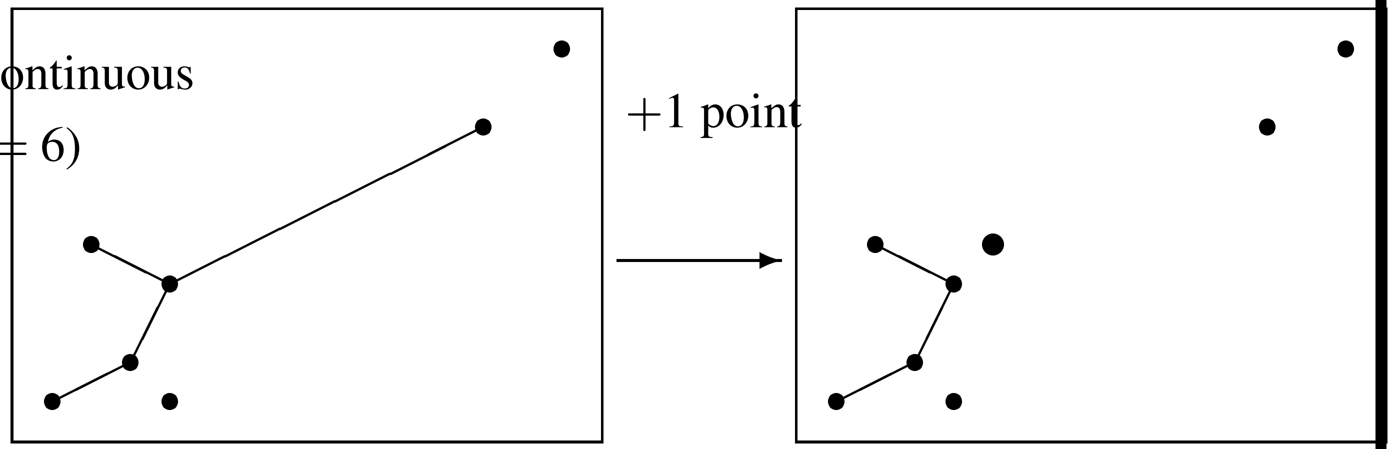
• **Superadditivity**: For the same conditions as above, there exists a constant C_2 s.t.

$$L_\gamma(F) \geq m^{-1} \sum_{i=1}^{m^d} L_\gamma(m[(F \cap Q_i) - q_i]) - C_2 m^{d-\gamma}$$

• **Continuity**: There exists a constant C_3 such that for all finite subsets F and G of $[0,1]^d$

$$|L_\gamma(F \cup G) - L_\gamma(F)| \leq C_3 \left(card(G)\right)^{(d-\gamma)/d}$$

Example of non-continuous graph: k-MST ($k = 6$)



Definition 1 *A continuous subadditive functional L_γ is said to be a quasi-additive functional when there exists a continuous superadditive functional L^*_γ which satisfies $L_\gamma(F) + 1 \geq L^*_\gamma(F)$ and the approximation property*

$$|E[L_\gamma(U_1, \ldots, U_n)] - E[L^*_\gamma(U_1, \ldots, U_n)]| \leq C_4 n^{(d-\gamma-1)/d} \tag{2}$$

where $U_1, \ldots, U_n$ are i.i.d. uniform random vectors in $[0,1]^d$.

Another smoothness condition

Definition 2 *L_γ is said to satisfy the add-one bound when*

$$|E[L_\gamma(U_1,\ldots,U_{n+1})] - E[L_\gamma(U_1,\ldots,U_n)]| \le C_4 n^{-\gamma/d} \tag{3}$$

where $U_1,\ldots,U_{n+1}$ are i.i.d. uniform random vectors in $[0,1]^d$.

Convergence rate for uniform f

Theorem 2 *(***Thm 5.2 Yukich:1998***) Let L_γ be a quasi-additive continuous Euclidean functional which satisfies the add-one bound. Assume that $f(\mathbf{x})$ is uniform over $[0,1]^d$. Then for all $d \geq 2$ and $1 \leq \gamma < d$*

$$\left| E[L_\gamma(\mathcal{X}_n)]/n^{(d-\gamma)/d} - \beta_{L_\gamma,d} \int f(x)^{(d-\gamma)/d} dx \right|^2 \leq O(n^{-1/d})$$

Question: How to extend to non-uniform f?

1. Extend to piecewise constant "block densities" over a uniform partition Q^m:

$$f(x) = \sum_{i=1}^{m^d} \phi_i 1_{Q_i}(x)$$

2. Extend to space of densities sufficiently well approximated by block densities.

3. Obtain worst-case bound on rate over this space of densities.

Block densities

For a set of non-negative constants $\{\phi_i\}_{i=1}^{m^d}$ satisfying $\sum_{i=1}^{m^d} \phi_i = m^d$, define

$$\phi(x) = \sum_{i=1}^{m^d} \phi_i 1_{Q_i}(x)$$

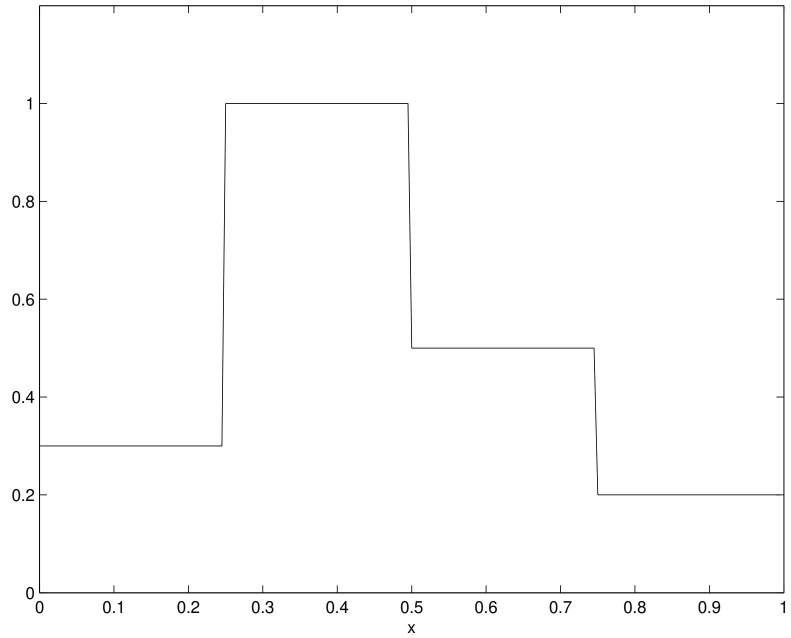


Figure: Block density over the unit interval.

A rate result for block densities

Proposition 1 *Let $d \geq 2$ and $1 \leq \gamma \leq d-1$. Assume $X_1, \ldots, X_n$ are i.i.d. sample points over $[0,1]^d$ whose marginal is a block density f with m^d levels and support $\mathcal{S} \subset [0,1]^d$. Then for any continuous quasi-additive Euclidean functional L_γ of order γ which satisfies the add-one bound*

$$\left| E[L_\gamma(X_1, \ldots, X_n)]/n^{(d-\gamma)/d} - \beta_{L_\gamma,d} \int_{\mathcal{S}} f^{(d-\gamma)/d}(x)\,\mathrm{d}x \right| \leq O\left((nm^{-d})^{-1/d} \right).$$

Extension to general densities

Define the resolution-m block density approximation of f by

$$\phi(x) = \sum_{i=1}^{m^d} \phi_i 1_{Q_i}(x),$$

where $\phi_i = m^d \int_{Q_i} f(x)\mathrm{d}x$.

Figure: Block density approximation over the unit interval.

Three term bound

By triangle inequality

$$
\begin{aligned}
&\left| E[L_\gamma(X_1,\ldots,X_n)]/n^{\frac{d-\gamma}{d}} - \beta_{L_\gamma,d} \int_S f^{\frac{d-\gamma}{d}}(x)\mathrm{d}x \right| \\
&\quad \leq \quad \left| E[L_\gamma(\tilde{X}_1,\ldots,\tilde{X}_n)]/n^{\frac{d-\gamma}{d}} - \beta_{L_\gamma,d} \int_S \phi^{\frac{d-\gamma}{d}}(x)\mathrm{d}x \right| \qquad (I) \\
&\qquad + \beta_{L_\gamma,d} \left| \int_S \phi^{\frac{d-\gamma}{d}}(x)\mathrm{d}x - \int_S f^{\frac{d-\gamma}{d}}(x)\mathrm{d}x \right| \qquad (II) \\
&\qquad\quad + \left| E[L_\gamma(X_1,\ldots,X_n)] - E[L_\gamma(\tilde{X}_1,\ldots,\tilde{X}_n)] \right| /n^{\frac{d-\gamma}{d}} \quad (III)
\end{aligned}
$$

1. Bound on I directly follows from Proposition 1

2. Bound on II is block density approximation error

3. Bound on III is error due to block realizations instead of true realizations of X

Sobolev Spaces

Consider the Sobolev space of $\mathcal{L}_p$ functions on $\mathbf{R}^d$

$$W^{1,p}(\mathbb{R}^d) = \mathcal{L}_p(\mathbb{R}^d) \cap \{f : \mathrm{D}_{x_j} f \in \mathcal{L}_p(\mathbb{R}^d),\ 1 \le j \le d\} \ .$$

• $\mathrm{D}_{x_j} f$ is the x_j-th *weak derivative* of f which satisfies

$$\int_{\mathbb{R}^d} f(x) \frac{\partial}{\partial x_j} \varphi(x) \mathrm{d}x = - \int_{\mathbb{R}^d} \mathrm{D}_{x_j} f(x) \varphi(x) \mathrm{d}x$$

for any function φ infinitely differentiable with compact support.

• $W^{1,p}$ is equipped with a norm

$$\|f\|_{1,p} = \|f\|_p + \|\mathrm{D}f\|_p \ .$$

Approximation Lemma

Lemma 1 *For $1 \leq p < \infty$, let $f \in W^{1,p}(\mathbb{R}^d)$ have support $\mathcal{S} \subset [0,1]^d$. Then there exists a constant $C_6 > 0$, independent of m, such that*

$$\int_{\mathcal{S}} |\phi(x) - f(x)| \mathrm{d}x \leq C_6 m^{-\lambda(p)} (\|\mathrm{D}f\|_p + o(1)), \tag{4}$$

where

$$\lambda(p) = \begin{cases} 1, & 1 \leq p \leq d \\ d + 1 - d/p, & d < p < \infty \end{cases}$$

An m-dependent bound

Using the Lemma to bound II and III we obtain

$$
\begin{aligned}
&\left| E[L_\gamma(X_1, \ldots, X_n)]/n^{(d-\gamma)/d} - \beta_{L_\gamma,d} \int_S f(x)^{(d-\gamma)/d} \mathrm{d}x \right| \\
&\quad \le \frac{K_1 + C_4}{(nm^{-d})^{1/d}} \left(\int_S f^{\frac{d-1-\gamma}{d}}(x) \mathrm{d}x + o(1) \right) \\
&\quad + \frac{\beta_{L_\gamma,d}}{(nm^{-d})^{1/2}} \left(\int_S f^{\frac{1}{2}-\frac{\gamma}{d}}(x) \mathrm{d}x + o(1) \right) \\
&\quad + \frac{K_2}{(nm^{-d})^{(d-\gamma)/d}} \\
&\quad + (\beta_{L_\gamma,d} + C_3')\, C_6' m^{-\lambda(p)(d-\gamma)/d} \left(\|\mathrm{D}f\|_p^{(d-\gamma)/d} + o(1) \right) \\
&\quad + \frac{1}{n^{(d-\gamma)/d}}
\end{aligned}
$$

An m-independent bound

By selecting m as the function of n that minimizes this bound, we obtain

Proposition 2 *Let $d \geq 2$ and $1 \leq \gamma \leq d-1$. Assume $X_1, \ldots, X_n$ are i.i.d. random vectors over $[0,1]^d$ with density $f \in W^{1,p}(\mathbb{R}^d)$, $1 \leq p < \infty$, having support $\mathcal{S} \subset [0,1]^d$. Assume also that $f^{\frac{1}{2}-\frac{\gamma}{d}}$ is integrable over $\mathcal{S}$. Then, for any continuous quasi-additive Euclidean functional L_γ of order γ that satisfies the add-one bound*

$$\left| E[L_\gamma(X_1, \ldots, X_n)]/n^{(d-\gamma)/d} - \beta_{L_\gamma,d} \int_{\mathcal{S}} f^{(d-\gamma)/d}(x)\mathrm{d}x \right| \leq O\left(n^{-r_1(d,\gamma,p)}\right),$$

where

$$r_1(d,\gamma,p) = \frac{\alpha\lambda(p)}{\alpha\lambda(p)+1}\,\frac{1}{d}$$

and $\alpha = \frac{d-\gamma}{d}$ and $\lambda(p)$ is defined in Lemma 1.

Concentration inequality

Rhee's inequality for Euclidean functionals (AnnAppProb:93):

$$P\left(\left|L_\gamma(X_1,\ldots,X_n)-E[L_\gamma(X_1,\ldots,X_n)]\right|>t\right)\le C\exp\left(\frac{-(t/C_3)^{2d/(d-\gamma)}}{Cn}\right),$$

Hence,

$$\begin{aligned}
&E\left[\left|L_\gamma(X_1,\ldots,X_n)-E[L_\gamma(X_1,\ldots,X_n)]\right|^\kappa\right]\\
&\qquad=\int_0^\infty P\left(\left|L_\gamma(X_1,\ldots,X_n)-E[L_\gamma(X_1,\ldots,X_n)]\right|>t^{1/\kappa}\right)\mathrm{d}t\\
&\qquad\le C_3C\int_0^\infty \exp\left(\frac{-t^{2d/[\kappa(d-\gamma)]}}{Cn}\right)\mathrm{d}t\\
&\qquad=A_\kappa n^{\kappa(d-\gamma)/(2d)}
\end{aligned}$$

Main convergence result

Corollary 1 *Let $d \geq 2$ and $1 \leq \gamma \leq d-1$. Assume $X_1, \ldots, X_n$ are i.i.d. random vectors over $[0,1]^d$ with density $f \in W^{1,p}(\mathbb{R}^d)$, $1 \leq p < \infty$, having support $\mathcal{S} \subset [0,1]^d$. Assume also that $f^{\frac{1}{2}-\frac{\gamma}{d}}$ is integrable over $\mathcal{S}$. Then, for any continuous quasi-additive Euclidean functional L_γ of order γ that satisfies the add-one bound*

$$E\left[\left|L_\gamma(X_1,\ldots,X_n)/n^{(d-\gamma)/d} - \beta_{L_\gamma,d}\int_{\mathcal{S}} f^{(d-\gamma)/d}(x)\mathrm{d}x\right|^\kappa\right]^{1/\kappa} \leq O\left(n^{-r_1(d,\gamma,p)}\right), \quad (5)$$

where

$$r_1(d,\gamma,p) = \begin{cases} \frac{\alpha}{d(\alpha+1)} & , \quad 1 \leq p \leq d \\ \frac{\alpha}{d(\alpha+\frac{1}{d+1-d/p})} & , \quad d < p < \infty \end{cases}$$

where $\alpha = \frac{d-\gamma}{d}$.

Extension to partition approximations

$$L_\gamma^m(\mathcal{X}_n) = \sum_{i=1}^{m^d} L_\gamma(\mathcal{X}_n \cap Q_i) + b(m),$$

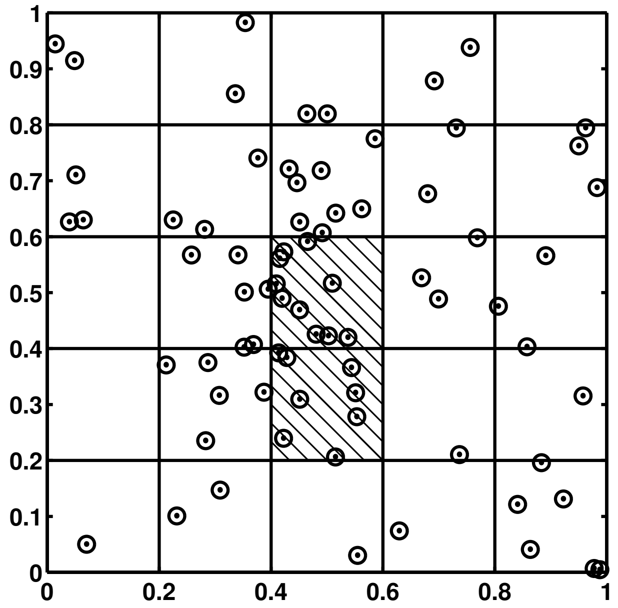


Figure 2: *Partition approximation.*

Pointwise closeness bound

Two Euclidean functionals L_γ and L^*_γ ae said to satisfy the *pointwise closeness bound* if

$$\left|L_\gamma(F) - L^*_\gamma(F)\right| \leq \begin{cases} C[card(F)]^{(d-\gamma-1)/(d-1)}, & 1 \leq \gamma < d-1 \\ C \log card(F), & \gamma = d-1 \neq 1 \\ C, & d-1 < \gamma < d \end{cases},$$

for any finite $F \subset [0,1]^d$. This condition is satisfied by the MST, TSP and minimal matching function (Lemma 3.7 Yukich:98).

Corollary 2 *Let $d \geq 2$ and $1 \leq \gamma < d-1$. Assume that the Lebesgue density $f \in W^{1,p}(\mathbb{R}^d)$, $1 \leq p < \infty$ has support $\mathcal{S} \subset [0,1]^d$. Assume also that $f^{1/2-\gamma/d}$ is integrable over $\mathcal{S}$. Let $L_\gamma^m(\mathcal{X}_n)$ be a partition approximation to $L_\gamma(\mathcal{X}_n)$ where L_γ is a continuous quasi-additive functional of order γ which satisfies the pointwise closeness bound and the add-one bound. Then if $b(m) = O(m^{d-\gamma})$*

$$E\left[\left|L_\gamma^{m(n)}(X_1,\ldots,X_n)/n^{(d-\gamma)/d} - \beta_{L_\gamma,d}\int_{\mathcal{S}} f^{(d-\gamma)/d}(x)\mathrm{d}x\right|\right] \leq O\left(n^{-r_2(d,\gamma,p)}\right), \tag{6}$$

where

$$r_2(d,\gamma,p) = \frac{\alpha\lambda(p)}{\frac{d-1}{\gamma}\ \alpha\lambda(p)+1}\ \frac{1}{d}\ ,$$

where $\alpha = \frac{d-\gamma}{d}$ and $\lambda(p)$ is defined in Lemma 1. This rate is attained by choosing the progressive-resolution sequence $m = m(n) = n^{1/[d(\frac{d-1}{\gamma}\ \alpha\lambda(p)+1)]}$.

Application to Entropy Estimation

Consider entropy estimates

$$\hat{H}_\alpha = (1-\alpha)^{-1} \log \hat{I}_\alpha,$$

where $\hat{I}_\alpha$ is a consistent estimator of the integral

$$I_\alpha(f) = \int f^\alpha(x)\mathrm{d}x$$

- $\hat{I}_\alpha = L_\gamma(X_1, \ldots, X_n)/(\beta_{L_\gamma,d} n^\alpha)$ is a strongly consistent estimator of $I_\alpha(f)$;
- indirect estimator: given non-parametric function estimates $\hat{f}$ of f define the function plug-in estimator $\hat{I}_\alpha = I_\alpha(\hat{f})$.

Convergence rate comparisons

Question: How fast is

$$|E[\hat{H}_\alpha] - H_\alpha(f)| \to 0, \text{ when } n \to \infty?$$

Find $r > 0$ such that

$$|E[\hat{H}_\alpha] - H_\alpha(f)| = O\left(n^{-r}\right)$$

For both cases

$$|\hat{H}_\alpha - H_\alpha(f)| = \frac{1}{1-\alpha} \frac{|\hat{I}_\alpha - I_\alpha(f)|}{I_\alpha(f)} + o(|\hat{I}_\alpha - I_\alpha(f)|)$$

$\Rightarrow$ convergence rate of $|E[\hat{H}_\alpha] - H_\alpha(f)|$ is identical to that of $|E[\hat{I}_\alpha] - I_\alpha(f)|$.

Multivariate Besov function space

Definition 3 *(Nikolskii:75) Let* $1 \leq p, q < \infty$*,* $\sigma > 0$ *and* k, ρ *be nonnegative integers satisfying the inequalities* $k > \sigma - \rho > 0$*. The function* f *belongs to the class* $B^{\sigma}_{p,q}(\mathbb{R}^d)$ *if* $f \in \mathcal{L}_p(\mathbb{R}^d)$ *and there exist partial weak derivatives* $\mathrm{D}^{(s)} f = \partial^{\rho} f / \partial x_1^{s_1} \ldots \partial x_d^{s_d}$ *of order* $s = (s_1, \ldots, s_d)$ *(*$|s| = s_1 + \ldots + s_d = \rho$*) such that the following seminorm is finite:*

$$\|f\|_{b^{\sigma}_{p,q}} = \sum_{|s|=\rho} \left\{ \int_{\mathbb{R}^d} \left(\frac{\|\Delta^k_{\tau} \mathrm{D}^{(s)} f\|_p}{|\tau|^{\sigma-\rho}} \right)^q \frac{\mathrm{d}\tau}{|\tau|^d} \right\} ,$$

where Δ^k_{τ} *is an operator which takes the k-th order finite difference in the direction of* τ*.*

Soboley vs Besov

Lemma 2 *(Besov&etal:79) Let $p > d$ and let σ be a positive integer. Then*

$$B^{\sigma}_{p,1}(\mathbb{R}^d) \subset W^{\sigma,p}(\mathbb{R}^d)$$

Asymptotic convergence of entropy estimators

Proposition 3 *Let* $p > d \geq 2$ *and* $\alpha = (d-\gamma)/d \in [1/2, (d-1)/d]$

$$\sup_{f^\alpha \in B_p^{1,1}} E^{1/\kappa}\left[\left|\int \widehat{f^\alpha}(x)dx - \int f^\alpha(x)dx\right|^\kappa\right] \geq O\left(n^{-1/(2+d)}\right)$$

while,

$$\sup_{f^\alpha \in B_p^{1,1}} E^{1/\kappa}\left[\left|\frac{L_\gamma(X_1,\ldots,X_n)}{n^\alpha} - \beta_{L_\gamma,d}\int f^\alpha(x)dx\right|^\kappa\right] \leq O\left(n^{-\frac{\alpha\lambda(p)}{1+\alpha\lambda(p)}\frac{1}{d}}\right)$$

where $\lambda(p) = d + 1 - d/p$.

Note: minimal-graph estimator converges faster for

$$\alpha \geq \frac{1}{2}\frac{d}{d+1-d/p}$$

Open problems

- ▷ Extend convergence rates to:
  - – smoother densities, e.g. in $W^{\sigma,p}(\mathbf{R}^d)$
  - – densities with unbounded support.
  - – weaker quasi-additive continuity conditions
- ▷ Analysis of clustering algorithms in entropic graph setting.
- ▷ Analysis of distributional properties of entropic graphs.