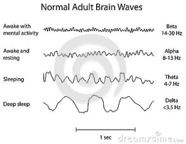
Sample Complexity of Estimating Entropy

Himanshu Tyagi
Indian Institute of Science, Bangalore

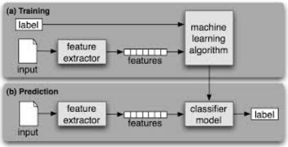
Joint work with Jayadev Acharya, Ananda Theertha Suresh, and Alon Orlitsky

Measuring Randomness in Data

Estimating randomness of the observed data:



Neural signal processing



Feature selection for machine learning

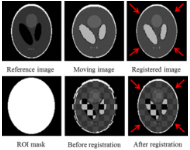
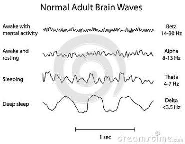


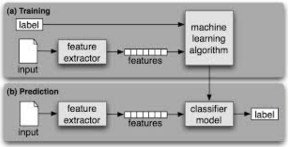
Image Registration

Measuring Randomness in Data

Estimating randomness of the observed data:



Neural signal processing



Feature selection for machine learning

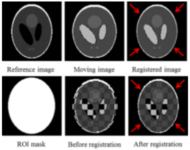
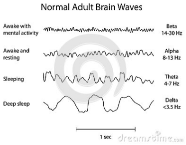


Image Registration

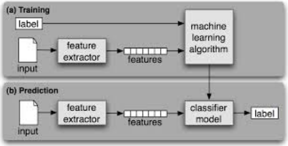
Approach: Estimate the "entropy" of the generating distribution

Measuring Randomness in Data

Estimating randomness of the observed data:



Neural signal processing



Feature selection for machine learning

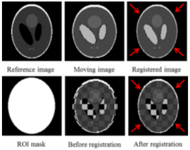


Image Registration

Approach: Estimate the "entropy" of the generating distribution

Shannon entropy $H(\mathrm{p}) \stackrel{\text{def}}{=} \sum_x -\mathrm{p}_x \log \mathrm{p}_x$

Estimating Shannon Entropy

For an (unknown) distribution p with a (unknown) support-size k,

How many samples are needed for estimating $H(\mathrm{p})$?

Estimating Shannon Entropy

For an (unknown) distribution p with a (unknown) support-size k,

How many samples are needed for estimating $H(\mathrm{p})$?

PAC Framework or Large Deviation Guarantees

Let $X^n = X_1, ..., X_n$ denote n independent samples from p

Performance of an estimator $\hat{H}$ is measured by

$$S^{\hat{H}}(\delta, \epsilon, k) \stackrel{\text{def}}{=} \min \left\{ n : \mathrm{p}^n \left(|\hat{H}(X^n) - H(\mathrm{p})| < \delta \right) > 1 - \epsilon, \quad \forall \ \mathrm{p} \text{ with support-size } k \right\}$$

The sample complexity of estimating Shannon Entropy is defined as

$$S(\delta, \epsilon, k) \stackrel{\text{def}}{=} \min_{\hat{H}} S^{\hat{H}}(\delta, \epsilon, k)$$

Sample Complexity of Estimating Shannon Entropy

Focus only on the dependence of $S(\delta, \epsilon, k)$ on k

- Asymptotically consistent and normal estimators: [Miller55], [Mokkadem89], [AntosK01]
- [Paninski03] For the empirical estimator $\hat{H}_e$, $S^{\hat{H}_e}(k) \leq O(k)$
- [Paninski04] There exists an estimator $\hat{H}$ s.t. $S^{\hat{H}}(k) \leq o(k)$
- [ValiantV11] $S(k) = \Theta(k/\log k)$
  - The proposed estimator is constructive and is based on a LP
  - See, also, [WuY14], [JiaoVW14] for new proofs

Sample Complexity of Estimating Shannon Entropy

Focus only on the dependence of $S(\delta, \epsilon, k)$ *on* k

- Asymptotically consistent and normal estimators:
  [Miller55], [Mokkadem89], [AntosK01]
- [Paninski03] For the empirical estimator $\hat{H}_e$, $S^{\hat{H}_e}(k) \leq O(k)$
- [Paninski04] There exists an estimator $\hat{H}$ s.t. $S^{\hat{H}}(k) \leq o(k)$
- [ValiantV11] $S(k) = \Theta(k/\log k)$
  - The proposed estimator is constructive and is based on a LP
  - See, also, [WuY14], [JiaoVW14] for new proofs

But we can estimate the distribution itself using $O(k)$ *samples.*

Sample Complexity of Estimating Shannon Entropy

Focus only on the dependence of $S(\delta, \epsilon, k)$ on k

- Asymptotically consistent and normal estimators: [Miller55], [Mokkadem89], [AntosK01]
- [Paninski03] For the empirical estimator $\hat{H}_e$, $S^{\hat{H}_e}(k) \leq O(k)$
- [Paninski04] There exists an estimator $\hat{H}$ s.t. $S^{\hat{H}}(k) \leq o(k)$
- [ValiantV11] $S(k) = \Theta(k/\log k)$
  - The proposed estimator is constructive and is based on a LP
  - See, also, [WuY14], [JiaoVW14] for new proofs

But we can estimate the distribution itself using $O(k)$ samples.

Is it easier to estimate some other entropy??

Estimating Rényi Entropy

Definition. The *Rényi entropy* of order α, $0 < \alpha \neq 1$, for a distribution p is given by

$$H_\alpha(\mathrm{p}) = \frac{1}{1-\alpha} \log \sum_x \mathrm{p}_x^\alpha$$

Sample Complexity of Estimating Rényi Entropy

Performance of an estimator $\hat{H}$ is measured by

$$S_\alpha^{\hat{H}}(\delta, \epsilon, k) \stackrel{\text{def}}{=} \min \left\{ n : \mathrm{p}^n \left(|\hat{H}(X^n) - H_\alpha(\mathrm{p})| < \delta \right) > 1 - \epsilon, \forall \ \mathrm{p} \in \mathcal{P}_k \right\}$$

The sample complexity of estimating Rényi Entropy of order α is given by

$$S_\alpha(\delta, \epsilon, k) \stackrel{\text{def}}{=} \min_{\hat{H}} S_\alpha^{\hat{H}}(\delta, \epsilon, k)$$

Estimating Rényi Entropy

Definition. The *Rényi entropy* of order α, $0 < \alpha \neq 1$, for a distribution p is given by

$$H_\alpha(\mathrm{p}) = \frac{1}{1-\alpha} \log \sum_x \mathrm{p}_x^\alpha$$

Sample Complexity of Estimating Rényi Entropy

Performance of an estimator $\hat{H}$ is measured by

$$S_\alpha^{\hat{H}}(\delta, \epsilon, k) \stackrel{\text{def}}{=} \min \left\{ n : \mathrm{p}^n \left(|\hat{H}(X^n) - H_\alpha(\mathrm{p})| < \delta \right) > 1 - \epsilon, \forall\ \mathrm{p} \in \mathcal{P}_k \right\}$$

The sample complexity of estimating Rényi Entropy of order α is given by

$$S_\alpha(\delta, \epsilon, k) \stackrel{\text{def}}{=} \min_{\hat{H}} S_\alpha^{\hat{H}}(\delta, \epsilon, k)$$

We mainly seek to characterize the dependence of $S_\alpha(\delta, \epsilon, k)$ on k and α

Which Rényi Entropy is the Easiest to Estimate?

Notations:

$S_\alpha(k) \geq \tilde{\tilde{\Omega}}(k^\beta) \Rightarrow$ for every $\eta > 0$ and for all δ, ϵ small,

$$S_\alpha(\delta, \epsilon, k) \geq k^{\beta - \eta}, \quad \text{for all } k \text{ large}$$

$S_\alpha(k) \leq O(k^\beta) \Rightarrow$ there is a constant c depending on δ, ϵ s.t.

$$S_\alpha(\delta, \epsilon, k) \leq ck^{\beta}, \quad \text{for all } k \text{ large}$$

$S_\alpha(k) = \Theta(k^\beta) \Rightarrow \Omega(k^\beta) \leq S_\alpha(k) \leq O(k^\beta)$

Which Rényi Entropy is the Easiest to Estimate?

Notations:

$S_\alpha(k) \geq \tilde{\tilde{\Omega}}(k^\beta) \Rightarrow$ for every $\eta > 0$ and for all δ, ϵ small,

$$S_\alpha(\delta, \epsilon, k) \geq k^{\beta - \eta}, \quad \text{for all } k \text{ large}$$

$S_\alpha(k) \leq O(k^\beta) \Rightarrow$ there is a constant c depending on δ, ϵ s.t.

$$S_\alpha(\delta, \epsilon, k) \leq ck^\beta, \quad \text{for all } k \text{ large}$$

$S_\alpha(k) = \Theta(k^\beta) \Rightarrow \Omega(k^\beta) \leq S_\alpha(k) \leq O(k^\beta)$

Theorem

For every $0 < \alpha < 1$*:* $\tilde{\tilde{\Omega}}(k^{1/\alpha}) \leq S_\alpha(k) \leq O(k^{1/\alpha}/\log k)$

For every $1 < \alpha \notin \mathbb{N}$*:* $\tilde{\tilde{\Omega}}(k) \leq S_\alpha(k) \leq O(k/\log k)$

Which Rényi Entropy is the Easiest to Estimate?

Notations:

$S_\alpha(k) \geq \tilde{\tilde{\Omega}}(k^\beta) \Rightarrow$ for every $\eta > 0$ and for all δ, ϵ small,

$$S_\alpha(\delta, \epsilon, k) \geq k^{\beta - \eta}, \quad \text{for all } k \text{ large}$$

$S_\alpha(k) \leq O(k^\beta) \Rightarrow$ there is a constant c depending on δ, ϵ s.t.

$$S_\alpha(\delta, \epsilon, k) \leq c k^{\beta}, \quad \text{for all } k \text{ large}$$

$S_\alpha(k) = \Theta(k^\beta) \Rightarrow \Omega(k^\beta) \leq S_\alpha(k) \leq O(k^\beta)$

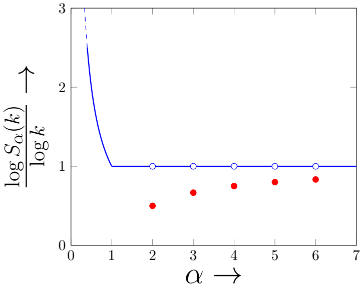
Theorem

For every $0 < \alpha < 1$*:* $\tilde{\tilde{\Omega}}(k^{1/\alpha}) \leq S_\alpha(k) \leq O(k^{1/\alpha}/\log k)$

For every $1 < \alpha \notin \mathbb{N}$*:* $\tilde{\tilde{\Omega}}(k) \leq S_\alpha(k) \leq O(k/\log k)$

For every $1 < \alpha \in \mathbb{N}$*:* $S_\alpha(k) = \Theta(k^{1-1/\alpha})$

Which Rényi Entropy is the Easiest to Estimate?



Theorem

For every $0 < \alpha < 1$: $\quad \tilde{\tilde{\Omega}}(k^{1/\alpha}) \le S_\alpha(k) \le O(k^{1/\alpha}/\log k)$

For every $1 < \alpha \notin \mathbb{N}$: $\quad \tilde{\tilde{\Omega}}(k) \le S_\alpha(k) \le O(k/\log k)$

For every $1 < \alpha \in \mathbb{N}$: $\quad S_\alpha(k) = \Theta(k^{1-1/\alpha})$

Related Work

The αth power sum of a distribution p is given by

$$P_\alpha(\mathrm{p}) \stackrel{\mathrm{def}}{=} \sum_x \mathrm{p}_x^\alpha$$

Estimating Rényi entropy with small additive error is the same as estimating power sum with small multiplicative error

- [Bar-YossefKS01] Integer moments of frequencies in a sequence with multiplicative and additive accuracies
- [JiaoVW14] Estimating power sums with small additive error

Related Work

The αth power sum of a distribution p is given by

$$P_\alpha(\mathrm{p}) \stackrel{\text{def}}{=} \sum_x \mathrm{p}_x^\alpha$$

Estimating Rényi entropy with small additive error is the same as estimating power sum with small multiplicative error

- [Bar-YossefKS01] Integer moments of frequencies in a sequence with multiplicative and additive accuracies
- [JiaoVW14] Estimating power sums with small additive error

For $\alpha < 1$: Additive and multiplicative accuracy estimation have roughly the same sample complexity

For $\alpha > 1$: Additive accuracy estimation requires only a constant number of samples

The Estimators

Empirical or Plug-in Estimator

Given n samples $X_1, ..., X_n$,

Let N_x denote the empirical frequency of x.

$$\hat{\mathrm{p}}_n(x) \stackrel{\mathrm{def}}{=} \frac{N_x}{n}$$

$$\hat{H}^e_\alpha \stackrel{\mathrm{def}}{=} \frac{1}{1-\alpha} \log \sum \hat{\mathrm{p}}_n(x)^\alpha$$

Empirical or Plug-in Estimator

Given n samples $X_1, ..., X_n$,

Let N_x denote the empirical frequency of x.

$$\hat{\mathrm{p}}_n(x) \stackrel{\mathrm{def}}{=} \frac{N_x}{n}$$

$$\hat{H}^e_\alpha \stackrel{\mathrm{def}}{=} \frac{1}{1-\alpha} \log \sum \hat{\mathrm{p}}_n(x)^\alpha$$

Theorem

For $\alpha > 1$: $S_\alpha^{\hat{H}^e_\alpha}(\delta, \epsilon, k) \leq O\left(\frac{k}{\delta^{\max\{4, 1/(\alpha-1)\}}} \log \frac{1}{\epsilon}\right)$

For $\alpha < 1$: $S_\alpha^{\hat{H}^e_\alpha}(\delta, \epsilon, k) \leq O\left(\frac{k^{1/\alpha}}{\delta^{\max\{4, 2/\alpha\}}} \log \frac{1}{\epsilon}\right)$

Empirical or Plug-in Estimator

Given n samples $X_1, ..., X_n$,

Let N_x denote the empirical frequency of x.

$$\hat{\mathrm{p}}_n(x) \stackrel{\text{def}}{=} \frac{N_x}{n}$$

$$\hat{H}^e_\alpha \stackrel{\text{def}}{=} \frac{1}{1-\alpha} \log \sum \hat{\mathrm{p}}_n(x)^\alpha$$

Theorem

For $\alpha > 1$: $S_\alpha^{\hat{H}^e_\alpha}(\delta, \epsilon, k) \leq O\left(\frac{k}{\delta^{\max\{4, 1/(\alpha-1)\}}} \log \frac{1}{\epsilon}\right)$

For $\alpha < 1$: $S_\alpha^{\hat{H}^e_\alpha}(\delta, \epsilon, k) \leq O\left(\frac{k^{1/\alpha}}{\delta^{\max\{4, 2/\alpha\}}} \log \frac{1}{\epsilon}\right)$

Proof??

Rényi Entropy Estimation to Power Sum Estimation

Estimating Rényi entropy with small additive error

is the same as

estimating power sum with small multiplicative error

Rényi Entropy Estimation to Power Sum Estimation

Estimating Rényi entropy with small additive error

is the same as

estimating power sum with small multiplicative error

Using a well-known sequence of steps,

suffices to show that bias and variance of $\hat{\mathrm{p}}_n$ are multiplicatively small

Poisson Sampling

The empirical frequencies N_x are correlated.

Suppose $N \sim \mathrm{Poi}(n)$ and $X_1, ..., X_N$ be independent samples from p.

Then,

1. $N_x \sim \mathrm{Poi}(n\mathrm{p}_x)$
2. $\{N_x, x \in \mathcal{X}\}$ are mutually independent
3. For each estimator $\hat{H}$, there is a modified estimator $\hat{H}'$ such that

$$\mathbb{P}\left(|H_\alpha(\mathrm{p}) - \hat{H}'(X^n)| > \delta\right) \leq \mathbb{P}\left(|H_\alpha(\mathrm{p}) - \hat{H}(X^N)| > \delta\right) + \frac{\epsilon}{2},$$

where $N \sim \mathrm{Poi}(n/2)$ and $n \geq 8\log(2/\epsilon)$.

It suffices to bound the error probability under Poisson sampling

Performance of the Empirical Estimator

For the empirical estimator $\hat{\mathrm{p}}_n$:

$$\frac{1}{P_\alpha(\mathrm{p})}\left|\mathbb{E}\left[\frac{\sum_x N_x^\alpha}{n^\alpha}\right]-P_\alpha(\mathrm{p})\right| \leq \begin{cases} c_1 \max\left\{\left(\frac{k}{n}\right)^{\alpha-1}, \sqrt{\frac{k}{n}}\right\}, & \alpha > 1, \\ c_2\left(\frac{k^{1/\alpha}}{n}\right)^{\alpha}, & \alpha < 1 \end{cases}$$

$$\frac{1}{P_\alpha(\mathrm{p})^2}\mathbb{V}\mathrm{ar}\left[\sum_x \frac{N_x^\alpha}{n^\alpha}\right] \leq \begin{cases} c_1' \max\left\{\left(\frac{k}{n}\right)^{2\alpha-1}, \sqrt{\frac{k}{n}}\right\}, & \alpha > 1, \\ c_2' \max\left\{\left(\frac{k^{1/\alpha}}{n}\right)^{\alpha}, \sqrt{\frac{k}{n}}, \frac{1}{n^{2\alpha-1}}\right\}, & \alpha < 1 \end{cases}$$

Theorem

For $\alpha > 1$: $S_\alpha^{\hat{H}_\alpha^e}(\delta, \epsilon, k) \leq O\left(\frac{k}{\delta^{\max\{4, 1/(\alpha-1)\}}}\log\frac{1}{\epsilon}\right)$

For $\alpha < 1$: $S_\alpha^{\hat{H}_\alpha^e}(\delta, \epsilon, k) \leq O\left(\frac{k^{1/\alpha}}{\delta^{\max\{4, 2/\alpha\}}}\log\frac{1}{\epsilon}\right)$

A Bias-Corrected Estimator

Consider an integer $\alpha > 1$

$n^{\underline{\alpha}} = n(n-1)\ldots(n-\alpha+1) =$ αth falling power of n

Claim: For $X \sim \mathrm{Poi}(\lambda)$, $\mathbb{E}[X^{\underline{\alpha}}] = \lambda^{\alpha}$

Under Poisson sampling, an unbiased estimator of $P_\alpha(\mathrm{p})$ is

$$\hat{P}_n^u \stackrel{\mathrm{def}}{=} \sum_x \frac{N_x^{\underline{\alpha}}}{n^{\alpha}}$$

Our estimator for $H_\alpha(\mathrm{p})$ is $\hat{H}_n^u \stackrel{\mathrm{def}}{=} \frac{1}{1-\alpha} \log \hat{P}_n^u$

Performance of the Bias-Corrected Estimator

For the bias-corrected estimator $\hat{\mathrm{p}}_n^u$ and an integer $\alpha > 1$

$$\frac{1}{P_\alpha(\mathrm{p})^2} \mathbb{V}\mathrm{ar}[\hat{\mathrm{p}}_n^u] \leq \sum_{r=0}^{\alpha-1} \left(\frac{\alpha^2 k^{1-1/\alpha}}{n} \right)^{\alpha-r}$$

Theorem

For integer $\alpha > 1$:

$$S_\alpha^{\hat{H}_n^u}(\delta, \epsilon, k) \leq O\left(\frac{k^{1-1/\alpha}}{\delta^2} \log \frac{1}{\epsilon} \right)$$

Performance of the Bias-Corrected Estimator

For the bias-corrected estimator $\hat{\mathrm{p}}_n^u$ and an integer $\alpha > 1$

$$\frac{1}{P_\alpha(\mathrm{p})^2} \mathbb{V}\mathrm{ar}[\hat{\mathrm{p}}_n^u] \leq \sum_{r=0}^{\alpha-1} \left(\frac{\alpha^2 k^{1-1/\alpha}}{n} \right)^{\alpha-r}$$

Theorem

For integer $\alpha > 1$:

$$S_\alpha^{\hat{H}_n^u}(\delta, \epsilon, k) \leq O\left(\frac{k^{1-1/\alpha}}{\delta^2} \log \frac{1}{\epsilon} \right)$$

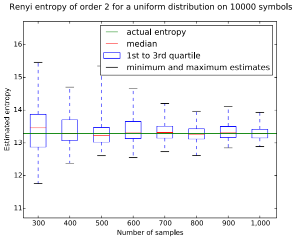
To summarize:

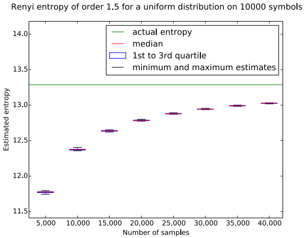
For every $0 < \alpha < 1$: $S_\alpha(k) \leq O(k^{1/\alpha})$

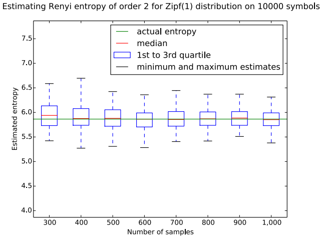
For every $1 < \alpha \notin \mathbb{N}$: $S_\alpha(k) \leq O(k)$

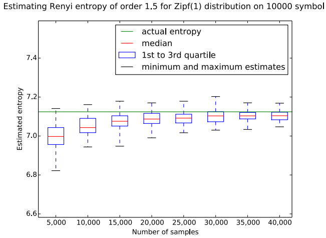
For every $1 < \alpha \in \mathbb{N}$: $S_\alpha(k) \leq O(k^{1-1/\alpha})$

Constants are Small in Practice









Lower Bounds

The General Approach

$S_\alpha(\delta, \epsilon, k) \geq g(k)$ for all δ, ϵ sufficiently small:

Show that there exist two distributions p and q such that

1. Support-size of both p and q is k;
2. $|H_\alpha(\mathrm{p}) - H_\alpha(\mathrm{q})| > \delta$;
3. For all $n < g(k)$, the variation distance $\|\mathrm{p}^n - \mathrm{q}^n\|$ is small.

The General Approach

$S_\alpha(\delta, \epsilon, k) \geq g(k)$ for all δ, ϵ sufficiently small:

Show that there exist two distributions p and q such that

1. Support-size of both p and q is k;
2. $|H_\alpha(\mathrm{p}) - H_\alpha(\mathrm{q})| > \delta$;
3. For all $n < g(k)$, the variation distance $\|\mathrm{p}^n - \mathrm{q}^n\|$ is small.

We can replace X^n with a sufficient statistic $\psi(X^n)$ to replace (3) with:

For all $n < g(k)$, the variation distance $\|\mathrm{p}_{\psi(X^n)} - \mathrm{q}_{\psi(X^n)}\|$ is small.

Distance between Profile Distributions

Definition. *Profile* of X^n refers $\Phi = (\Phi_1, ..., \Phi_n)$ where

$$\begin{aligned}\Phi_i &= \text{ number of symbols appearing } i \text{ times in } X^n \\ &= \sum_x \mathbb{1}(N_x = i)\end{aligned}$$

Two simple observations:

1. Profile is a sufficient statistic for the probability multiset of p
2. We can assume Poisson sampling without loss of generality

Let p_Φ and q_Φ denote the distribution of profiles under Poisson sampling

Distance between Profile Distributions

Definition. *Profile* of X^n refers $\Phi = (\Phi_1, ..., \Phi_n)$ where

$$\Phi_i = \text{ number of symbols appearing } i \text{ times in } X^n$$
$$= \sum_x \mathbb{1}(N_x = i)$$

Two simple observations:

1. Profile is a sufficient statistic for the probability multiset of p
2. We can assume Poisson sampling without loss of generality

Let p_Φ and q_Φ denote the distribution of profiles under Poisson sampling

Theorem (Valiant08)

Given distributions p *and* q *such that* $\max_x \max\{\mathrm{p}_x; \mathrm{q}_x\} \leq \frac{\epsilon}{40n}$, *for Poisson sampling with* $N \sim \mathrm{Poi}(n)$, *it holds that*

$$\|\mathrm{p}_\Phi - \mathrm{q}_\Phi\| \leq \frac{\epsilon}{2} + 5 \sum_a n^a |P_a(\mathrm{p}) - P_a(\mathrm{q})|.$$

Derivation of our Lower Bounds

For distributions p and q:

- $\|\mathrm{p}_\Phi - \mathrm{q}_\Phi\| \lesssim 5\sum_a n^a |P_a(\mathrm{p}) - P_a(\mathrm{q})|$
- $|H_\alpha(\mathrm{p}) - H_\alpha(\mathrm{q})| = \frac{1}{1-\alpha}\left|\log \frac{P_\alpha(\mathrm{p})}{P_\alpha(\mathrm{q})}\right|$

Choose p and q to be mixtures of d uniform distributions as follows:

$$\mathrm{p}_{ij} = \frac{|x_i|}{k\|x\|_1}, \quad 1 \le i \le d, 1 \le j \le k$$

$$\mathrm{q}_{ij} = \frac{|y_i|}{k\|y\|_1}, \quad 1 \le i \le d, 1 \le j \le k$$

Derivation of our Lower Bounds

For distributions p and q:

- $\|\mathrm{p}_\Phi - \mathrm{q}_\Phi\| \lesssim 5\sum_a n^a |P_a(\mathrm{p}) - P_a(\mathrm{q})|$
- $|H_\alpha(\mathrm{p}) - H_\alpha(\mathrm{q})| = \frac{1}{1-\alpha}\left|\log\frac{P_\alpha(\mathrm{p})}{P_\alpha(\mathrm{q})}\right|$

Choose p and q to be mixtures of d uniform distributions as follows:

$$\mathrm{p}_{ij} = \frac{|x_i|}{k\|x\|_1}, \quad 1 \le i \le d, 1 \le j \le k$$

$$\mathrm{q}_{ij} = \frac{|y_i|}{k\|y\|_1}, \quad 1 \le i \le d, 1 \le j \le k$$

Thus,

$$\|\mathrm{p}_\Phi - \mathrm{q}_\Phi\| \lesssim 5\sum_a \left(\frac{n}{k^{1-1/a}}\right)^a \left|\left(\frac{\|x\|_a}{\|x\|_1}\right)^a - \left(\frac{\|y\|_a}{\|y\|_1}\right)^a\right|$$

$$|H_\alpha(\mathrm{p}) - H_\alpha(\mathrm{q})| = \frac{\alpha}{(1-\alpha)k^{\alpha-1}}\left|\log\frac{\|x\|_\alpha}{\|y\|_\alpha}\cdot\frac{\|x\|_1}{\|y\|_1}\right|$$

Derivation of our Lower Bounds: Key Construction

Distributions with $\|\mathbf{x}|_r = \|\mathbf{y}|_r$, $\forall 1 \le r \le m-1$ cannot be distinguished with fewer than $k^{1-1/m}$ samples

Distributions with $\|\mathbf{x}|_\alpha \neq \|\mathbf{y}|_\alpha$ have different H_α

Derivation of our Lower Bounds: Key Construction

Distributions with $\|\mathbf{x}|_r = \|\mathbf{y}|_r, \forall 1 \leq r \leq m-1$ cannot be distinguished with fewer than $k^{1-1/m}$ samples

Distributions with $\|\mathbf{x}|_\alpha \neq \|\mathbf{y}|_\alpha$ have different H_α

Lemma

For every $d \in \mathbb{N}$ and α not integer, there exist positive vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that

$$\begin{aligned} \|\mathbf{x}\|_r &= \|\mathbf{y}\|_r, \quad 1 \leq r \leq d-1, \\ \|\mathbf{x}\|_d &\neq \|\mathbf{y}\|_d, \\ \|\mathbf{x}\|_\alpha &\neq \|\mathbf{y}\|_\alpha. \end{aligned}$$

Derivation of our Lower Bounds: Key Construction

Lemma

For every $d \in \mathbb{N}$ and α not integer, there exist positive vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that

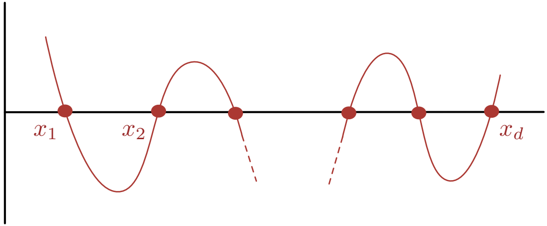
$$\begin{aligned} \|\mathbf{x}\|_r &= \|\mathbf{y}\|_r, \quad 1 \le r \le d-1, \\ \|\mathbf{x}\|_d &\neq \|\mathbf{y}\|_d, \\ \|\mathbf{x}\|_\alpha &\neq \|\mathbf{y}\|_\alpha. \end{aligned}$$

Derivation of our Lower Bounds: Key Construction

Lemma

For every $d \in \mathbb{N}$ and α not integer, there exist positive vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that

$$
\begin{aligned}
\|\mathbf{x}\|_r &= \|\mathbf{y}\|_r, \quad 1 \le r \le d-1, \\
\|\mathbf{x}\|_d &\neq \|\mathbf{y}\|_d, \\
\|\mathbf{x}\|_\alpha &\neq \|\mathbf{y}\|_\alpha.
\end{aligned}
$$

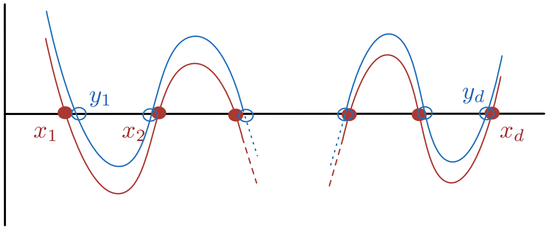


Derivation of our Lower Bounds: Key Construction

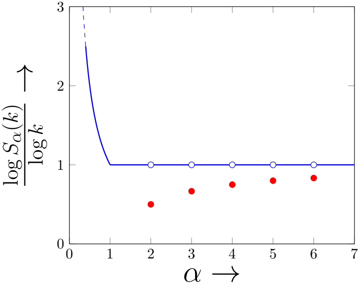
Lemma

For every $d \in \mathbb{N}$ and α not integer, there exist positive vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ such that

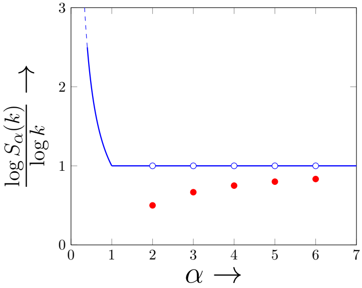
$$\begin{aligned} \|\mathbf{x}\|_r &= \|\mathbf{y}\|_r, \quad 1 \leq r \leq d-1, \\ \|\mathbf{x}\|_d &\neq \|\mathbf{y}\|_d, \\ \|\mathbf{x}\|_\alpha &\neq \|\mathbf{y}\|_\alpha. \end{aligned}$$



In Closing ...



Rényi entropy of order 2 is the “easiest” entropy to estimate, requiring only $O(\sqrt{k})$ samples



Rényi entropy of order 2 is the "easiest" entropy to estimate, requiring only $O(\sqrt{k})$ samples

Sample complexity of estimating other information measures