Weighted k-NN graphs for entropy estimation in high dimensions*

Kumar Sricharan, Alfred O. Hero III, Department of EECS, University of Michigan

*This work is partially funded by the Air Force Office of Scientific Research, grant number FA9550-09-1-0471.

Motivation

- To estimate the Rényi α-entropy $H_\alpha(f) = (1-\alpha)^{-1}\int f^\alpha(x)dx$ from M, d-dimensional i.i.d. samples $\mathbf{X}_1, \ldots, \mathbf{X}_M \sim f$
- We seek to estimate this for high dimensional data; i.e. $d >> 5$
- Several asymptotically consistent estimators have been proposed.
- Performance for finite sample size M not known
- Recently, nearest neighbor methods have become popular. Advantages:
  - Rates of convergence known
  - Circumvent density estimation
  - k-NN graph estimators, entropic graph estimators
- Bias is of order $O((1/M)^{1/d})$; Variance is of order $O(1/M)$
- Problem: curse of dimensionality. Bias is very large for large dimensions
- Can we do better in high dimensions?

Previous work

- $H_\alpha(f) = (1-\alpha)^{-1} I_\alpha(f)$ where $I_\alpha(f) = \int f^\alpha(x)dx$
- Leonenko's estimator [1] for $I_\alpha(f)$

$$\hat{I}_{M,k,\alpha} = \frac{1}{M}\sum_{i=1}^{M} \frac{\Gamma(k)}{\Gamma(k+1-\alpha)}(c_d(M-1)(r^{(i)}_{k,M-1})^d)^{1-\alpha}$$

- $r^{(i)}_{k,M-1}$ is k-th nearest neighbor distance from $\mathbf{X}_i$ to some other sample $\mathbf{X}_j$
- Leonenko showed that the estimator is consistent
- Liitiäinen et.al. [2] showed that
  - $Bias(\hat{I}_{M,k,\alpha}) = r_k M^{-1/d} + o(M^{-1/d})$
  - $Var(\hat{I}_{M,k,\alpha}) = O(M^{-1})$

Weighted estimator

- For a weight vector $w = \{w(l)\}, l = \{1, \ldots, k\}$ with $\sum w(l) = 1$

$$\hat{I}^w_{M,k,\alpha} = \sum_{l=1}^{k} w(l)\hat{I}_{M,l,\alpha}$$

- $Bias(\hat{I}^{(w)}_{M,k,\alpha}) = (\sum r_l w(l))M^{-1/d} + o(M^{-1/d})$
- Liitiäinen et.al.'s [2] first order correction: choose w so that $\sum r_l w(l) = 0$
- Bias reduces to $o(M^{-1/d})$; $Var(\hat{I}^{(w)}_{M,k,\alpha}) = O(M^{-1})$
- In theory, bias is reduced to $o(M^{-1/d})$, can continue to be quite large
- In simulations, bias was found to increase in comparison to unweighted estimator for small to moderate sample sizes

Our contribution

- Higher order analysis of bias
- Provide a choice of weights which reduces bias to $O(M^{-1/2})$
- MSE convergence rate of $1/M$
- Performs well with small sample sizes as well

Higher order bias analysis

- For s-times differentiable bounded densities, with boundary corrected k-NN distances, we show [3] for constants c_i, $\{i = 1, .., s\}$, c_v

$$Bias(\hat{I}_{M,k,\alpha}) = \sum_{i=1}^{s} c_i \left(\frac{k}{M}\right)^{i/d} + o\left(\sqrt{\frac{k}{M}}\right)$$

$$Var(\hat{I}_{M,k,\alpha}) = c_v\left(\frac{1}{M}\right) + o\left(\frac{1}{M}\right)$$

- Define $\gamma_w(i) = \sum_{l=1}^{k} w(l) l^{i/d}$. Bias:

$$Bias(\hat{I}^w_{M,k,\alpha}) = \sum_{i=1}^{s} c_i\gamma_w(i)M^{-i/d} + o\left(M^{-1/2}\right)$$

- Bound on variance:

$$Var(\hat{I}^w_{M,k,\alpha}) \le \frac{||w||_1^2 c_v}{M} + o\left(\frac{1}{M}\right)$$

- If magnitude of the weight coefficients $\{w_f(l)\}$ are large, then the variance and the coefficients in the bias expansion $\gamma_{w_f}(i), i \ge 1$ will be large.

Weight selection via convex optimization

- Seek weight w_o that
  - reduces bias by minimizing coefficients $\gamma_w(i)$, $i = \{1, .., t\}$ for some $t \le s$
  - has minimum possible l_1 norm $||w||_1$ to reduce variance
- Convex optimization for optimal weight w_o (solved using interior point methods)

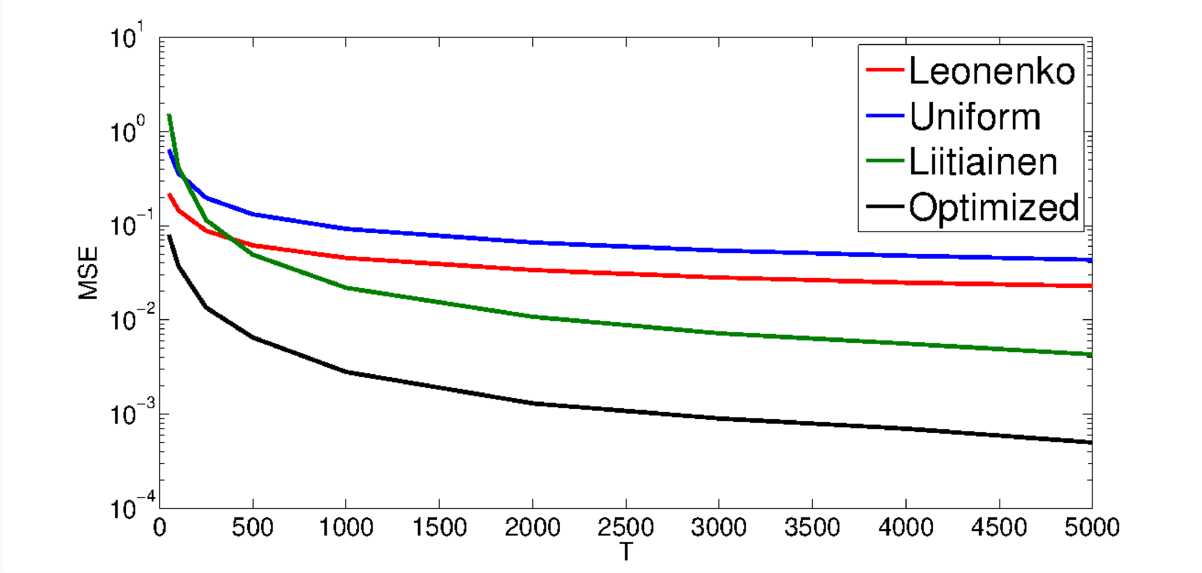
$$\begin{aligned} \underset{w}{\text{minimize}} \quad & ||w||_1 \\ \text{subject to} \quad & \gamma_w(0) = 1, \\ & |\gamma_w(i)M^{-i/d}| \le \epsilon,\ i \in \{1, 2, \ldots, t\} \end{aligned}$$

- The solution $||w||_1$ increases with number of constraints t and decreases with ϵ. For large k, the solution is shown to be sparse [4]
- Best possible MSE rate of weighted estimator is $O(1/M)$ because variance is $O(1/M)$. Therefore seek to reduce bias to $O(1/\sqrt{M})$.
- Choose t to be as small as possible; Assume that density is sufficiently smooth, i.e., $s > \lceil d/2 \rceil$ and fix $t = \lceil d/2 \rceil$.
- Choose ϵ as large as possible; $\epsilon = \sqrt{c_0/M}$ where c_0 is a bound on c_v/c_i^2
- $Bias(\hat{I}^w_{M,k,\alpha}) = \sum_{i=1}^{\lceil d/2 \rceil} c_i\gamma_w(i)M^{-i/d} + o\left(M^{-1/2}\right) = O\left(M^{-1/2}\right)$

Simulations

Four different choices of weight vectors:
(1) Leonenko et.al.'s estimator: $w_s = [1, 0, \ldots, 0]$, (2) uniform weighted estimator: $w_u = (1/k)[1, \ldots, 1]$, (3) First-order correction estimator of Liitiäinen et.al. : w_f, (4) Proposed optimized weighted estimator: w_o

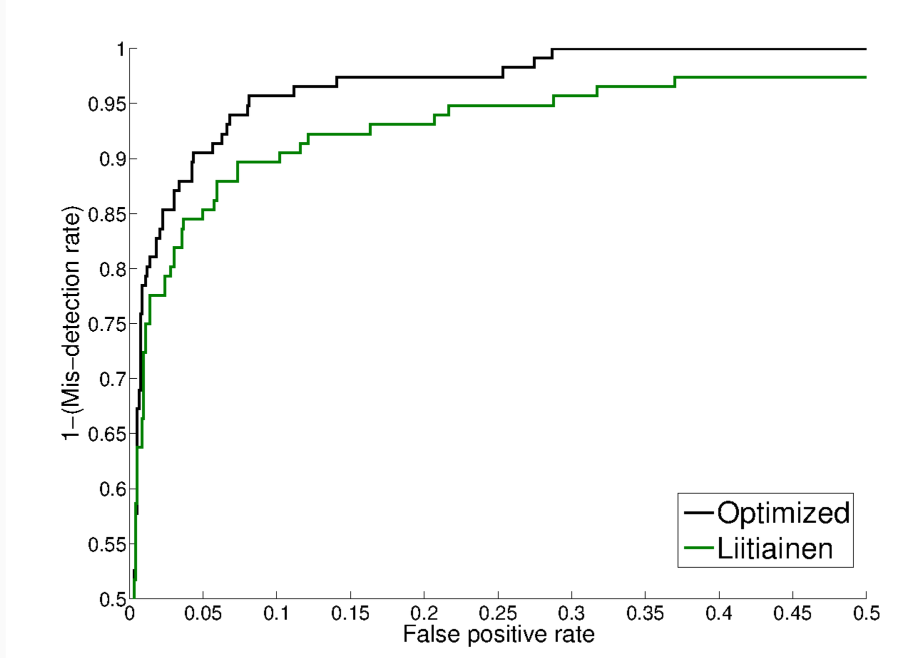
- Dimension $d = 6$
- Density $0.8f_\beta(1.5, 1.5) + .2f_u$
- Simulation shows that optimized weighted estimator outperforms other estimators



Anomaly detection

Mission: To use RSS measurements to detect intruders. 14 sensor nodes randomly deployed inside and outside a lab room. $14 \times 13 = 182$ RSS measurements recorded every 0.5 secs for 30 mins

- Form a temporal dependency discriminant by considering vectors of $d = 3$ successive time samples at each sensor
- Estimating the Rényi entropy by averaging over $M = 182$ spatial samples
- Perform anomaly detection by thresholding the entropy estimate
- Optimized weighted estimator outperforms first order correction estimator



Conclusions

- k-NN estimators suffer from curse of dimensionality; Bias is of order $O((1/M)^{1/d})$
- Higher order analysis of bias reveals basis functions to be $(k/M)^{i/d}$
- Can use simple weighted linear combination of estimators to reduce bias from $O((1/M)^{1/d})$ to $O((1/M)^{1/2})$; RMS rate of convergence of $1/\sqrt{M}$

References

[1] N. Leonenko, L. Prozanto, and V. Savani, *"A class of Rényi information estimators for multidimensional densities,"* Annals of Statistics, vol. 36, pp. 2153–2182, 2008.

[2] E. Liitiäinen, A. Lendasse, and F. Corona, *"A boundary corrected expansion of the moments of nearest neighbor distributions,"* Random Structures and Algorithms, vol. 37, pp. 223–247, September 2010.

[3] K. Sricharan, R. Raich, and A. O. Hero III, *"Empirical estimation of entropy functionals with confidence,"* ArXiv e-prints, Dec. 2010.

[4] D. Donoho, *"For Most Large Underdetermined Systems of Linear Equations the Minimal l1-norm Solution is also the Sparsest Solution,"* Comm. Pure Appl. Math, vol. 59, pp. 797-829, 2004.