Numerical Estimation of Information Measures

Haoran Ni (haoran.ni@warwick.ac.uk)

Supervisors: Martin Lotz (Warwick), Keith Briggs(BT)

Mathematics for Real-World Systems CDT, Mathematics Institute, University of Warwick

Mutual Information Estimation

The Mutual Information (MI) between two random variables measures the reduction in uncertainty of one random variable due to information obtained from the other. Its applications include decision trees in machine learning, independent component analysis (ICA), gene detection and expression, link prediction, topic discovery, image registration, feature selection and transformations, and channel capacity. We develop efficient mutual information estimation systems with a set of estimators that produce accurate results irrespective of sample size, data dimension, and correlation. The systems include bias correction approaches, new approximate k-th Nearest Neighbour (k-NN) estimators, and new algorithms based on fast and sparse Johnson-Lindenstrauss transforms.

Mutual Information and kNN Estimators

Mutual information is the reduction in uncertainty about a variable X after observing another variable Y. Mutual information can be calculated using entropy, referred to as 3H-principle:

$$I(X;Y) := H(X) + H(Y) - H(X,Y),$$

where $H(X)$ and $H(X,Y)$ are the entropy and joint entropy, defined as

$$H(X) := -\mathbb{E}[\log f_X(x)], \quad H(X,Y) := -\mathbb{E}[\log f_{X,Y}(x,y)],$$

with f_X the density of X, and $f_{X,Y}$ the joint density of (X,Y).

Among the various methods to estimate MI, the k-th nearest neighbour (k-NN) estimators are singled out due to their superior theoretical and practical performance. The idea of k-NN is to estimate the density $f_X(x)$ locally at $x = x_i$ for $i = 1, 2, \cdots, N$ points in $\mathbb{R}^d$ by finding the distance $\epsilon_i^{k,p}$ from x_i to its k-th nearest neighbour in ℓ_p $(p \geq 1)$ space. The local estimate $\hat{f}_X(x_i)$ is then given by:

$$\hat{f}_X(x_i) = \frac{k/N}{V^{p,d}(\epsilon_i^{k,p})^d},$$

where $V^{p,d}(r)$ denotes the volume of an ℓ_p-ball of radius r in $\mathbb{R}^d$. From this, we obtain an estimate of entropy by:

$$\hat{H}(X) = -\frac{1}{N}\sum_{i=1}^{N}[\log \hat{f}_X(x_i)]$$

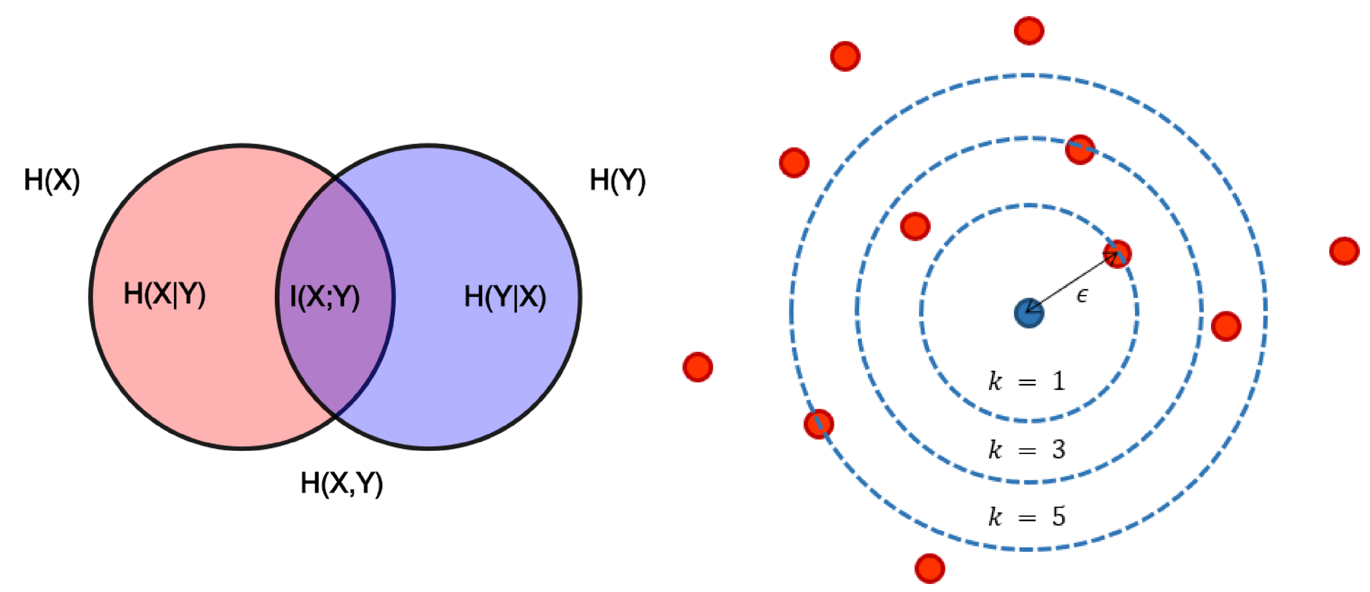

Fig. 1: (Left) Mutual Information (Right) k-NN Estimator

Existing Methods

A k-NN entropy estimator with bias correction was introduced by Kozachenko and Leonenko in 1987 [4] (KL). In 2004, Kraskov et.al. [1] proposed an improved k-NN estimator of mutual information (KSG). In 2017, Gao et.al. [7] realised that the better performance of KSG is due to a *correlation boosting effect*. They then introduced a bias-improved KSG (BI-KSG) estimator. Lord et.al. [8] noticed that most k-NN methods use regular local volume elements, which allows the estimators to be asymptotically consistent. They introduced the geometric k-NN estimators (G-kNN) based on elliptical local volume elements. Their method outperforms KSG for small sample size and highly dependent samples, but it is not corrected for asymptotic bias. We tested these methods on Gaussian and uniform distributions, see Figure 2.

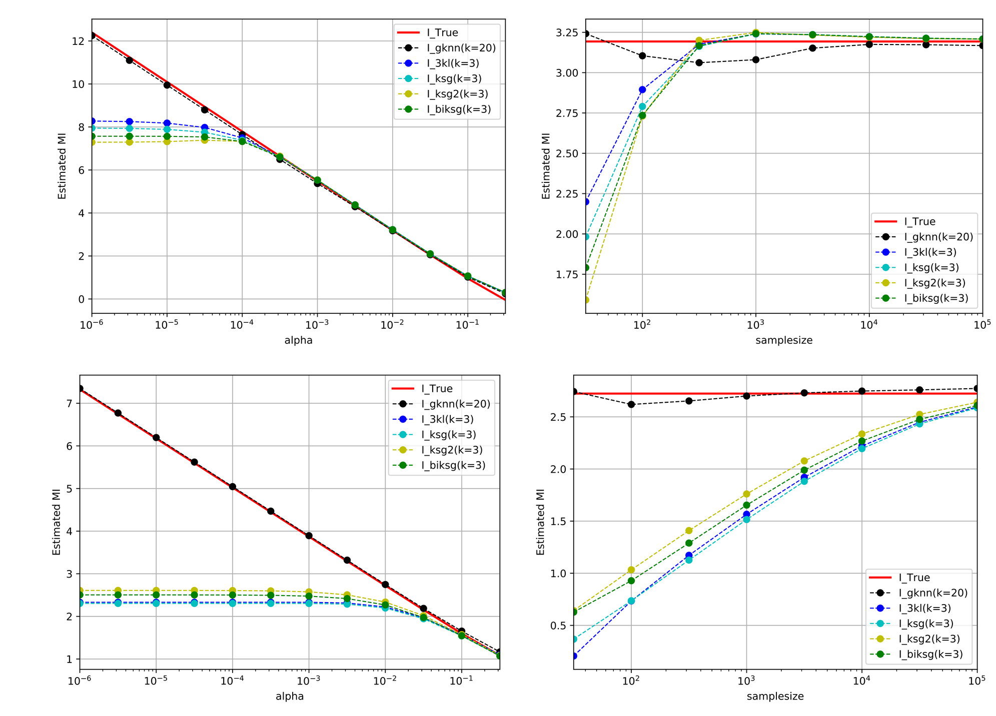

Fig. 2: (Top)Uniform distribution with Gaussian noise (Bottom)Gaussian distribution

Bias Improvement and Approximate kNN

We can reduce the bias of an estimator in various ways:
1. By excluding the center when estimating the local densities, we keep the virtue of G-kNN while improving the bias.
2. Since the estimation of mutual information in G-kNN is based on three entropy estimates, cancelling out k-NN distance terms can improve bias;
3. By using approximations of the expectation, we could add correction terms to improve the bias;
4. Choosing a large k for generating local geometry and a small k which depends on the generated local geometry to estimate local density.

Inspired by approximate k-NN search algorithms, we believe that there will be an improvement of efficiency if we find a point x'_k which is close to the k^{th} nearest neighbour x_k of x_i with distance $\|x'_k - x_k\|_p \leqslant \tau\epsilon_i^{k,p}$ in each local estimate. Introducing small uncertainty may also improve the performance of the estimations. This idea was shown in the bottom two plots of Figure 3.

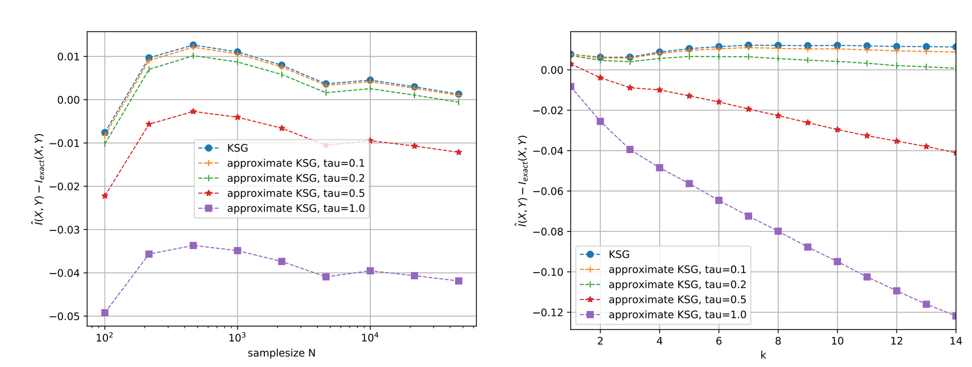

Fig. 3: Approximate k-NN

Dimensionality Reduction

All considered MI estimators introduce bias when the dimension increases and lack efficiency on high-dimensional data and large sample sizes (Figure 4).
Lombardi and Pant [5] introduced the non-parametric k-NN entropy estimator (kpN), which uses the Gaussian distribution as the assumption of the local density distribution, which improves the estimates for both high-dimensional and highly dependent data. However, it is not developed to be a MI estimator, introduces bias when dimension increases and lacks efficiency.
A different approach relies on random dimensionality reduction methods. Ailon and Chazelle [2] showed that the fast and sparse Johnson-Lindenstrauss transforms can be implemented on k-NN related algorithms. An extension of recent results by Lotz [6] and Arya et al [3] shows that features of data that are directly relevant to the estimation of entropy and mutual information, are preserved under randomized dimensionality reduction. Following the conclusions of these two papers and aforementioned approximate k-NN estimator, we aim to design new k-NN estimators based on fast and sparse Johnson-Lindenstrauss transforms in order to reduce the dimension and hence computational cost.

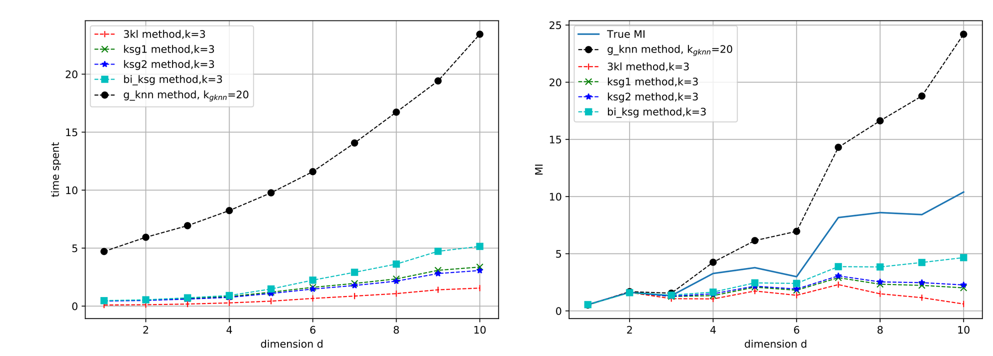

Fig. 4: (Left)Time spent (Right)MI estimation

Acknowledgements and References

The authors would like to acknowledge the aforementioned supervisors for the help and assistance in creating this poster.

[1] A. Kraskov, H. Stögbauer, and P. Grassberger. "Estimating mutual information". In: *Phys. Rev. E* 69.6 (2004).

[2] N. Ailon and B. Chazelle. "Approximate Nearest Neighbors and the Fast Johnson-Lindenstrauss Transform". In: *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*. 2006, pp. 557–563.

[3] S. Arya et al. "Dimensionality reduction for k-distance applied to persistent homology". In: *Proceedings of the 36th Symposium on Computational Geometry* (2020).

[4] L. F. Kozachenko and N. N. Leonenko. "Sample Estimate of the Entropy of a Random Vector". In: *Probl. Peredachi Inf.* 23.2 (1987), pp. 95–101.

[5] D. Lombardi and S. Pant. "A non-parametric k-nearest neighbor entropy estimator". In: *Physical Review E* 93 (June 2015). DOI: 10.1103/PhysRevE.93.013310.

[6] M. Lotz. "Persistent homology for low-complexity models." In: *Proc. R. Soc.* 475.2230 (2015). DOI: 10.1098/rspa.2019.0081.

[7] W. Gao, S. Oh, and P. Viswanath. "Demystifying fixed k-nearest neighbor information estimators". In: *IEEE International Symposium on Information Theory (ISIT)* (2017), pp. 1267–1271.

[8] W. M. Lord, J. Sun, and E. M. Bollt. "Geometric k-nearest neighbor estimation of entropy and mutual information". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.3 (2018).