A Users' Guide for the Python 2.7/3 Package of 'Minimax Estimation of Divergences between Discrete Distributions' by Yanjun Han, Jiantao Jiao and Tsachy Weissman, arXiv preprint arXiv:1605.09124, 2016
Version 1.0

January 22, 2017

Abstract

It is the users' guide for version 1.0 of the Python 2.7/3 package of paper 'Minimax Estimation of Divergences between Discrete Distributions' by Yanjun Han, Jiantao Jiao and Tsachy Weissman, arXiv preprint arXiv:1605.09124, 2016. It demonstrates how to use the KL divergence estimator developed in the paper in practice.

1 What is KL divergence?

The Kullback–Leibler divergence (or the KL divergence, the relative entropy) between discrete distributions P and Q is introduced by Kullback and Leibler [1],

$$D(P\|Q) \triangleq \begin{cases} \sum_{i=1}^{S} p_i \ln \frac{p_i}{q_i} & \text{if } P \ll Q, \\ +\infty & \text{otherwise.} \end{cases} \tag{1}$$

which is closely related to the entropy and mutual information introduced by Shannon [2], and plays significant roles in information theory and various disciplines such as statistics, machine learning, physics, neuroscience, computer science, linguistics, etc. Here distributions P, Q have support size S, which is assumed to be unknown.

In applications, the true distributions P, Q are usually unknown, hence we cannot compute this information-theoretic measure directly by definition. Instead, we consider the model where we obtain jointly independent m samples following distribution P and n samples following distribution Q, and would like to estimate the KL divergence $D(P\|Q)$.

However, unlike some other information-theoretic measures such as entropy and mutual information, the KL divergence can be infinity in certain scenarios. Moreover, even if $P \ll Q$ holds, it was shown in Han, Jiao and Weissman [3] that no estimator can achieve a vanishing L_2 risk for any (P, Q) if no additional assumption is made. As a result, we assume a bounded likelihood ratio between P and Q, i.e., $\frac{p_i}{q_i} \le u(S)$ for some upper bound $u(S) \ge 1$ and any $i = 1, \cdots, S$.

Recently, Han, Jiao and Weissman [3] derived the first explicit KL divergence estimator that achieves the minimax rates in the widest range of $(m, n, S, u(S))$ (sample size from P, sample size from Q, support size and likelihood ratio bound) pairs. Moreover, this estimator is adaptive in the sense that it does not require the knowledge of the support size S nor the likelihood ratio bound $u(S)$. This Python package provides an efficient implementation of the estimator in [3], which is based on the entropy estimator in [4].

2 How to use the estimator?

In the Python implementation, there is one main function that users may use: est_rel_entro_HJW.py. We explain its usage here:

est = est_rel_entro_HJW(sampP, sampQ)

This function returns the HJW estimate of the KL divergence (in bits) between each column of sampP and sampQ. The inputs can be lists, tuples, arrays and matrices, but all types will be converted into arrays in the code and the output will also be an array. Moreover, inputs sampP and sampQ must only contain integers, but need not be of the same length. If sampP, sampQ are 1D vectors (row vectors and column vectors are both acceptable), the function returns a scalar HJW estimate of the KL divergence between these two random variables. If sampP and sampQ are 2D matrices, the function will return a row vector with each entry containing the HJW estimate of the KL divergence between the corresponding columns in sampP and sampQ (and in this case the inputs must have the same number of columns).

The readers are welcomed to run the file test_rel_entro.py to test the performance of the HJW KL divergence estimator. For comparison, we also provide the implementation of the maximum likelihood estimator of the KL divergence (where the empirical distribution Q_n is replaced by $Q'_n = \max\{Q_n, 1/n\}$), in function est_rel_entro_MLE(sampP, sampQ). The way to use function est_rel_entro_MLE(sampP, sampQ) is exactly the same as that of est_rel_entro_HJW(sampP, sampQ).

This estimator can be used under both Python 2.7 and Python 3.

3 Acknowledgment

We are grateful to Ethan Ordentlich for his significant work in implementing this Python package and optimizing the performance of the estimator.

References

[1] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[2] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[3] Y. Han, J. Jiao, and T. Weissman, "Minimax estimation of KL divergence between discrete distributions," *arXiv preprint arXiv:1605.09124*, 2016.

[4] J. Jiao, K. Venkat, Y. Han, and T. Weissman, "Minimax estimation of functionals of discrete distributions," *Information Theory, IEEE Transactions on*, vol. 61, no. 5, pp. 2835–2885, 2015.