A Users Guide for the MATLAB Package of Estimation of KL Divergence: Optimal Minimax Rate

Yuheng Bu Shaofeng Zou Yingbin Liang Venugopal V. Veeravalli

April 21, 2017

1 Introduction

Consider the estimation of KL divergence between two probability distributions P and Q defined as

$$D(P\|Q) = \sum_{i=1}^{k} P_i \log \frac{P_i}{Q_i}, \tag{1}$$

where P and Q are supported on a common alphabet set $[k] \triangleq \{1, \ldots, k\}$, and P is absolutely continuous with respect to Q, i.e., if $Q_i = 0$, $P_i = 0$, for $1 \leq i \leq k$.

In practice, P and Q are unknown, but m independent and identically distributed (i.i.d.) samples $X_1, \ldots, X_m$ drawn from P and n i.i.d. samples $Y_1, \ldots, Y_n$ drawn from Q are available for estimation.

However, the KL divergence can be infinity if $Q_i = 0, P_i > 0$ holds for some i. Moreover, even if P and Q are supported on a common alphabet, it was shown in [1] that there is no consistent estimator can achieve a vanishing quadratic risk for any (P, Q) pair, if no additional assumption is made. Thus, we assume a bounded density ratio constraint between P and Q, i.e. $\frac{P_i}{Q_i} \leq f(k), \forall\ 1 \leq i \leq k$ for some function $f(k) \geq 1$.

We proposed the minimax optimal KL divergence estimator under the density ratio constraint in [2], and this MATLAB package provides an efficient implementation of the BZLV optimal estimator in [2].

2 Usage of the package

In the MATLAB package, the main function that users may use is: `est = est_KL_opt(samp_P,samp_Q,k)`. We explain its usage here:

This function returns a scalar estimation of KL divergence $D(P\|Q)$ when `samp_P` and `samp_Q` are vectors, or returns a (row-) vector consisting of the estimation of KL divergence of each column of samples when `samp_P` and `samp_Q` are matrix.

Input:

`samp_P` : a vector or matrix which can only contain integers. The input data type can be any interger classes such as uint8/int8/uint16/int16/uint32/int32/uint64/int64, or floating-point such as single/double.

`samp_Q` :same conditions as `samp_P`. Must have the same number of columns if a matrix.

`k` :An estimate of the alphabet size of the distribution pair p and q.

Output:

`est` :the KL divergence (in nats) of the input vector or that of each column of the input matrix. The output data type is double.

The users can also run the test files `KLdiv_change_k.m` and `KLdiv_change_n.m` to test the performance of the our KL divergence estimator under different settings. For comparison, we also provide the implementation of a plug-in based estimator of the KL divergence in function `est_KL_Aplugin(samp_P,samp_Q)` (where the empirical distribution $\hat{Q}_i$ is replaced by $\hat{Q}'_i = \hat{Q}_i + \frac{1}{n}$). The input and output of the function

`est_KL_Aplugin(samp_P,samp_Q)` is basically the same as that of `est = est_KL_opt(samp_P,samp_Q,k)`, except that `est_KL_Aplugin(samp_P,samp_Q)` does not require the information of `k`.

References

[1] Y. Bu, S. Zou, Y. Liang, and V. V. Veeravalli, "Universal outlying sequence detection for continuous observations," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4254–4258.

[2] Y. Bu, S. Zou, Y. Liang, and V. V Veeravalli, "Estimation of kl divergence: Optimal minimax rate," *arXiv preprint arXiv:1607.02653*, 2016.