Draft version June 5, 2020
Typeset using LaTeX **preprint** style in AASTeX62

k-Nearest Neighbor Estimator for Bayesian Evidence

ChangHoon Hahn[1, 2, *]

[1] *Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley CA 94720, USA*
[2] *Berkeley Center for Cosmological Physics, University of California, Berkeley CA 94720, USA*

(Dated: `DRAFT --- 53a1352 --- 2020-06-03 --- NOT READY FOR DISTRIBUTION`)

ABSTRACT

abstract

Keywords: keyword1 – keyword2 – keyword3

1. ESTIMATOR

Consider the commonly used Kullback-Leibler (KL) divergence between two distributions $p(x)$ and $g(x)$:

$$D_{\rm KL}\left(p\,||\,q\right) = \int p(x)\;\log\frac{p(x)}{q(x)}\;{\rm d}x. \tag{1}$$

The KL divergence between the posterior $p = p(\theta\,|\,D, m)$ and prior $\pi = p(\theta|m)$ is then

$$D_{\rm KL}\left(p\,||\,\pi\right) = \int p(\theta\,|\,D, m)\;\log\frac{p(\theta\,|\,D, m)}{p(\theta|m)}\;{\rm d}\theta. \tag{2}$$

If we use Bayes' Theorem to substitute for $p(\theta\,|\,D, m)$ in the numerator, we can rewrite this as

$$D_{\rm KL}\left(p\,||\,\pi\right) = \int p(\theta\,|\,D, m)\;\log\frac{p(D\,|\,\theta, m)}{p(D|m)}\;{\rm d}\theta \tag{3}$$

$$= -\log p(D\,|\,m) + \int p(\theta\,|\,D, m)\;\log p(D\,|\,\theta, m)\;{\rm d}\theta. \tag{4}$$

The first term of the right hand side is the log evidence and the second term is the expectation value of the log likelihood under the posterior, $\big\langle\log\mathcal{L}\big\rangle$. In standard Bayesian analyses in astronomy where we use MCMC to sample the posterior, $\big\langle\log\mathcal{L}\big\rangle$ can be easily derived with these samples using Monte Carlo integration:

$$\big\langle\log\mathcal{L}\big\rangle \approx \frac{1}{n}\sum_{i=1}^{n}\log\mathcal{L}(\theta^{(i)}) \tag{5}$$

Therefore, if we can estimate the KL divergence between the posterior and prior we can also estimate the evidence:

$$\log p(D\,|\,m) = \big\langle\log\mathcal{L}\big\rangle - D_{\rm KL}\left(p(D\,|\,\theta, m)\,||\,\pi\right). \tag{6}$$

* hahn.changhoon@gmail.com

To estimate the divergence, we can make use of non-parametric divergence estimators, which have been applied to Support Distribution Machines and used in the machine learning and astronomical literature (*e.g.* Póczos et al. 2011; Poczos et al. 2012a,b; Xu et al. 2013; Ntampaka et al. 2015, 2016; Ravanbakhsh et al. 2017; **?**). These estimators allow us to estimate the divergence between distributions p and q, $D(p\,||\,q)$ using samples drawn from them. In our case, p and q are the posterior and prior distributions. Again, standard analyses already sample the posterir distribution; the prior distribution is typcially straightforward to sample.

For the KL divergence, we use the k-Nearest Neighbor (NN) estimator presented in Wang et al. (2009). Let $X_{1:n} = \{X_1, ...X_n\}$ and $Y_{1:m} = \{Y_1, ...Y_m\}$ be n and m samples drawn from the p and q d-dimensional distributions, respectively. Then the divergence between p and q can be estimated as:

$$D_{\mathrm{KL}}\,(p\,||\,q) \approx \widehat{D_{\mathrm{KL}}}(p\,||\,q) = \frac{d}{n}\sum_{i=1}^{n}\Big[\log\frac{\nu_{\ell_i}(i)}{\rho_{k_i}(i)}\Big] + \frac{1}{n}\sum_{i=1}^{n}\Big[\psi(\ell_i) - \psi(k_i)\Big] + \log\frac{m}{n-1}. \tag{7}$$

In the first term, $\rho_k(i)$ denotes the Euclidean distance between X_i and the k^{th}-NN of X_i in sample $\{X_j\}_{i\neq j}$. $\nu_\ell(i)$ denotes the Euclidean distance between X_i and the ℓ^{th} NN of X_i in the sample $Y_{1:m}$. In the second term, ψ is the Digamma function: $\psi(k) = \Gamma'(k)/\Gamma(k)$. This term corrects for the estimation bias that comes from nonuniformity of the distribution near each sample point and guarantees that the estimator is asymptotically unbiased. Furthermore, while some k-NN estimators require chome choice in k and ℓ (*e.g.* Poczos et al. 2012a), the Wang et al. (2009) estimator adaptively determines ℓ_i and k_i as the number of samples $X_{1:n}$ and $Y_{1:m}$), respectively, contained in $B(X_i, \epsilon(i))$, a Euclidean ball centered at X_i with radius $\epsilon(i)$.

$$\epsilon(i) = \max(\rho(i), \nu(i)) \tag{8}$$

where

$$\rho(i) = \min_{j\neq i} ||X_i - X_j|| \tag{9}$$

$$\nu(i) = \min_{j\neq i} ||X_i - Y_j||. \tag{10}$$

For further details on the estimator and proofs that the estimator is asymptotically unbiased and mean-square consistent we refer readers to Wang et al. (2009).

ACKNOWLEDGEMENTS

It's a pleasure to thank ... This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under contract No. DE-AC02-05CH11231. This project used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

APPENDIX

REFERENCES

Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Poczos B., Schneider J., 2015, The Astrophysical Journal, 803, 50

Ntampaka M., Trac H., Sutherland D. J., Fromenteau S., Poczos B., Schneider J., 2016, The Astrophysical Journal, 831, 135

Póczos B., Szabó Z., Schneider J., 2011, in 2011 19th European Signal Processing Conference. pp 1718–1722

Poczos B., Xiong L., Schneider J., 2012a, arXiv:1202.3758 [cs, stat]

Poczos B., Xiong L., Sutherland D., Schneider J., 2012b, Machine Learning Department

Ravanbakhsh S., Lanusse F., Mandelbaum R., Schneider J., Poczos B., 2017, in Thirty-First AAAI Conference on Artificial Intelligence.

Wang Q., Sanjeev K., Sergio V., 2009, IEEE TRANSACTIONS ON INFORMATION THEORY, 55, 2392

Xu X., Ho S., Trac H., Schneider J., Poczos B., Ntampaka M., 2013, The Astrophysical Journal, 772, 147