Parametric estimation using a Kullback-Leibler divergence estimator based on nearest neighbours

Carlos Aya-Moreno

Abstract

This paper introduces a new parametric estimator based on Kullback-leibler divergence and nearest neighbours. This because the Kullback-Leibler divergence is estimated using nearest neighbour balls. Under some general conditions, the estimator is shown to be asymptotically consistent.

Parametric family $\mathcal{F} = \{f_\theta \,|\, \theta \in \Theta\}$. We are interested in finding θ given an i.i.d. sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ distributed according to F_θ, with density f_θ.

Given a proposed parameter θ', the Kullback-Leibler divergence between $F_{\theta'}$ and F_θ is, for fixed θ,

$$\mathcal{D}(\theta') = \mathcal{D}_{KL}(F'_\theta | F_\theta) = \int f_{\theta'}(x) \log\left(\frac{f_{\theta'}(x)}{f_\theta(x)}\right) \mathrm{d}x. \tag{1}$$

Expanding the log, we obtain

$$\mathcal{D}(\theta') = \int f_{\theta'}(x) \log\left(f_{\theta'}(x)\right) \mathrm{d}x - \int f_{\theta'}(x) \log\left(f_\theta(x)\right) \mathrm{d}x. \tag{2}$$

The first integral depends on the current estimate θ' can have an exact formula. For example, for the normal distribution with parameters μ and σ, it is $-\left(1 + \log 2\pi\sigma^2\right)/2$. The second integral depends on the value to be estimated, θ, and it is not available. Here we propose to use a technique similar to [1, Sec. 2.1], but instead of using the square root of the volume of the nearest neighbour ball, we again lean on [2, Prop. 2] which asserts that

$$nV_{(1);i} \rightsquigarrow \mathrm{Exp}\left\{f(X_i)\right\}, \tag{3}$$

to find a suitable transformation to approximate $\log f_\theta$.

The reader can verify that if Z is exponentially distributed, i.e. $Z \sim \mathrm{Exp}(\lambda)$, then

$$\mathbb{E}\left[Z \log Z\right] = \frac{1 - \gamma - \log(\lambda)}{\lambda},$$

where γ is Euler's gamma constant. A simple transformation on above gives

$$\mathbb{E}\left[Z \log\left(e^{\gamma-1} Z\right)\right] = \frac{\log(\lambda)}{\lambda}, \tag{4}$$

which allows us to estimate the second integral using a sum similar to [1, (8)]. Define

$$S_n \doteq \sum_{i=1}^{n} f_{\theta'}(X_i) \log\left(e^{\gamma-1} n\, V_{(1);i}\right) V_{(1);i}. \tag{5}$$

By the Law of Iterated Expectations, and treating $nV_{(1);i}$ as an exponentially distributed random variable (3) and using (4), we get

$$\begin{aligned}
\mathbb{E}\left[S_n\right] &= \mathbb{E}\left[f_{\theta'}(X_i)\mathbb{E}\left[\log\left(e^{\gamma-1} n\, V_{(1);i}\right) nV_{(1);i} | X_i\right]\right] \\
&\rightsquigarrow \mathbb{E}\left[f_{\theta'}(X_i) \frac{\log f_\theta(X_i)}{f_\theta(X_i)}\right] \\
&= \int f_{\theta'}(x) \frac{\log f_\theta(x)}{f_\theta(x)} f_\theta(x)\ \mathrm{d}x \\
&= \int f_{\theta'}(x) \log\left(f_\theta(x)\right) \mathrm{d}x,
\end{aligned}$$

the second integral in (2).

Therefore, define the estimated KL divergence as

$$\widehat{\mathcal{D}}(\theta') \doteq K(\theta') - S_n, \tag{6}$$

where $K(\theta')$ is calculated directly based on the family f_θ according to

$$K(\theta') \doteq \int f_{\theta'}(x) \log\left(f_{\theta'}(x)\right) \, \mathrm{d}x. \tag{7}$$

Finally, we choose for parameter θ' the one that minimizes the estimated KL divergence:

$$\widehat{\theta} \doteq \arg\min_{\theta' \in \Theta} \widehat{\mathcal{D}}(\theta'). \tag{8}$$

TODO: write asymptotic consistency proof
TODO: plot split vs non-split trials as in Hellinger distance paper
TODO: can we use this KL in other settings where KL is used (?!)
This estimator was implemented in Rust, using the Burn framework.
It can found at https://github.com/carlosayam/mkle

References

[1] Carlos Aya-Moreno, Gery Geenens, and Spiridon Penev. “Shape-preserving wavelet-based multivariate density estimation”. In: *Journal of Multivariate Analysis* 168 (2018), pp. 30–47.

[2] Bo Ranneby, S Rao Jammalamadaka, and Alex Teterukovskiy. “The maximum spacing estimation for multivariate observations”. In: *Journal of statistical planning and inference* 129.1-2 (2005), pp. 427–446.