Parametric estimation using a Hellinger distance estimator

Carlos Aya-Moreno

Abstract

This paper introduces a new parametric estimator based on Hellinger distance. The squared Hellinger distance 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.

Given a proposed parameter θ', the Hellinger distance between $f_{\theta'}$ and f_θ (θ fixed) is

$$\mathcal{H}^2(\theta') = \frac{1}{2} \int \left(\sqrt{f_{\theta'}}(x) - \sqrt{f_\theta}(x) \right)^2 \, \mathrm{d}x. \tag{1}$$

Expanding the square, we obtain

$$\mathcal{H}^2(\theta') = 1 - \int \sqrt{f_{\theta'}}(x) \sqrt{f_\theta}(x) \, \mathrm{d}x \doteq 1 - \mathcal{B}(\theta'), \tag{2}$$

owing to $\int f_{\theta'} = \int f_\theta = 1$. The integral $\mathcal{B}(\theta')$ is the 'Bhattacharyya affinity coefficient' [2] between the distribution $f_{\theta'}$ and ground truth f_θ. As $\mathcal{B}(\theta')$ involves the unknown f_θ, it needs to be estimated.

Here we propose to use [1, Lemma 1], allowing consistent estimation of quantities of type $\int \phi \sqrt{f}$. Specifically, we estimate these by

$$\widehat{\mathcal{B}}(\theta') \doteq \frac{2}{\sqrt{\pi}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{f_{\theta'}}(X_i) \sqrt{V_i}, \tag{3}$$

$$\widehat{\mathcal{H}^2}(\theta') \doteq 1 - \widehat{\mathcal{B}}(\theta') \tag{4}$$

where, V_i is the volume of the ball centred at X_i whose radius is the distance from X_i to its closest neighbour in the sample.

We choose for parameter θ' the one that minimizes the estimated squared Hellinger distance:

$$\widehat{\theta} \doteq \arg \min_{\theta' \in \Theta} \widehat{\mathcal{H}^2}(\theta'). \tag{5}$$

TODO: write asymptotic consistency proof

The terms $\sqrt{f_{\theta'}}(X_i)\sqrt{V_i}$ are asymptotically independent [3] but still correlated. Given a sample, one can split this in two disjoint parts $\mathcal{X} = \mathcal{X}_1 \sqcup \mathcal{X}_2$ if so desired to have i.i.d terms. The sum in (3) is calculated over the first half, whereas nearest neighbours and ball volumes V_i are calculated from $X_i \in \mathcal{X}_1$ distances to points in $\mathcal{X}_2$. This setup makes the terms uncorrelated but halving the sample size has a bigger impact and hence not worth doing

TODO: plot split vs non-split trials

TODO: can we use for hierarchical models (?)

This estimator was implemented in Rust, using the Burn framework.

It can found at https://github.com/carlosayam/mhde.

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] Anil Bhattacharyya. "On a measure of divergence between two multinomial populations". In: *Sankhyā: the indian journal of statistics* (1946), pp. 401–406.

[3] 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.