Title: Efficient integral functional estimation via k-nearest neighbour distances

Authors: Thomas Berrett

In many statistical problems it of interest to estimate *integrals functionals*, namely functionals of the form

$$T(f) = \int_{\mathcal{X}} f(x)\phi(x, f(x))\, dx,$$

for some known ϕ, given an i.i.d. sample $X_1, \ldots, X_m$ with density function f. These functionals often have information-theoretic interpretations, such as the differential entropy with $\phi(x, y) = -\log y$, or the Rényi entropy of order α with $\phi(x, y) = y^{\alpha-1}$. We may alternatively be interested in the estimation of two-sample integral functionals of the form

$$T(f, g) = \int_{\mathcal{X}} f(x)\phi(x, f(x), g(x))\, dx,$$

given an additional i.i.d. sample $Y_1, \ldots, Y_n$ with density function g. These include well known quantities such as the Kullback–Leibler divergence, Rényi divergence and Hellinger distance.

Our proposed estimation procedure is based on k-nearest neighbour distances of our samples $\rho_{(k),i,X} = \|X_i - X_{(k),i}\|$ and $\rho_{(k),i,Y} = \|X_i - Y_{(k),i}\|$ where, for each i, we find the orderings $X_{(1),i}, \ldots, X_{(m-1),i}$ and $Y_{(1),i}, \ldots Y_{(n),i}$ of $\{X_1, \ldots, X_m\} \setminus \{X_i\}$ and $\{Y_1, \ldots, Y_n\}$ respectively such that

$$\|Y_{(1),i} - X_i\| \leq \ldots \leq \|Y_{(n),i} - X_i\| \quad \text{and} \quad \|X_{(1),i} - X_i\| \leq \ldots \leq \|X_{(m-1),i} - X_i\|.$$

Our starting point is to consider estimators

$$\hat{T}_m = \frac{1}{m}\sum_{i=1}^m \phi\biggl(X_i, \frac{k}{mV_d\rho_{(k),i,X}^d}\biggr) \quad \text{and} \quad \hat{T}_{m,n} = \frac{1}{m}\sum_{i=1}^m \phi\biggl(X_i, \frac{k_X}{mV_d\rho_{(k_X),i,X}^d}, \frac{k_Y}{nV_d\rho_{(k_Y),i,Y}^d}\biggr),$$

which we show can be modified, in some cases, so that the bias is of order $o(m^{-1/2})$. We then show that these modified estimators achieve the local asymptotic minimax lower bound, in that

$$m\mathrm{Var}\hat{T}_m = \mathrm{Var}\bigl(\phi(X, f(X)) + f(X)\phi'(X, f(X))\bigr) + o(1)$$

in the one-sample case, and that, in the two-sample case,

$$m\mathrm{Var}\hat{T}_{m,n} = \mathrm{Var}_f(\phi + f\phi_1) + \frac{\zeta}{1-\zeta}\mathrm{Var}_g(f\phi_2) + o(1),$$

where $\zeta = \lim_{m\to\infty} m/(m+n)$, and we write $\phi_1(x, y, z) = \partial\phi/\partial y$ and $\phi_2(x, y, z) = \partial\phi/\partial z$.

Suprisingly, when we are interested in estimating the Rényi entropy $T_\alpha(f) = \int_{\mathcal{X}} f(x)^\alpha\, dx$, we find that the variance of our efficient estimator is given by $\alpha^2\mathrm{Var}\{f(X)^{\alpha-1}\}$. When $\alpha \in (1/2, 1)$, this is strictly smaller than the variance of the 'oracle' estimator $T_m^* = \frac{1}{m}\sum_{i=1}^m f(X_i)^{\alpha-1}$, which uses knowledge of the unknown density f.