KERNEL DENSITY ESTIMATION USING LOW DISCREPANCY SAMPLING

FRED J. HICKERNELL

Abstract. This project is where all of the files and commands go that are needed elsewhere

Let $Y = f(\boldsymbol{X})$, where $\boldsymbol{X} \sim \mathcal{U}[0,1]^d$. Low discrepancy sequences are used for computing $\mu = \mathbb{E}(Y)$. Can they be used for estimation of ϱ, the probability density function of Y?

Let ν be a probability mass or density function. A generalized kernel density estimator (KDE), $\widehat{\varrho}(\cdot, \nu, k)$, can be defined as

$$\widehat{\varrho}(y, \nu, k) := \int_{-\infty}^{\infty} k(z, y)\, \nu(z)\, \mathrm{d}z, \qquad y \in \mathbb{R}, \tag{1}$$

where we call the $k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ the *density kernel*. Here, $k(\cdot, y)$ is a weight function used to estimate the density at y, so $k(\cdot, y)$ should typically be larger near y than away from y. Moreover, we assume that

$$\int_{-\infty}^{\infty} k(z, y)\, \mathrm{d}z = 1 \qquad \forall y \in \mathbb{R}. \tag{2}$$

For (1) to be a good estimator, the density ν should approximate ϱ well.

Let $\varrho_{\boldsymbol{y}}$ be the empirical probability mass function of a vector of samppled Y values, $\boldsymbol{y} = (y_1, \ldots, y_n)$. Then a *practical KDE* is

$$\widehat{\varrho}(y, \varrho_{\boldsymbol{y}}, k) = \int_{-\infty}^{\infty} k(z, y)\, \varrho_{\boldsymbol{y}}(z)\, \mathrm{d}z = \frac{1}{n} \sum_{i=1}^{n} k(y_i, y) = \frac{1}{n} \sum_{i=1}^{n} k(f(\boldsymbol{x}_i), y).$$

Moreover, a *smoothed density* can be defined as

$$\widehat{\varrho}(y, \varrho, k) = \int_{-\infty}^{\infty} k(z, y)\, \varrho(z)\, \mathrm{d}z = \int_{[0,1]^d} k(f(\boldsymbol{x}), y)\, \mathrm{d}\boldsymbol{x}.$$

The absolute error of the practical KDE can be bounded as the sum of two terms:

$$\begin{aligned} |\varrho(y) - \widehat{\varrho}(y, \varrho_{\boldsymbol{y}}, k)| &= |\varrho(y) - \widehat{\varrho}(y, \varrho, k) + \widehat{\varrho}(y, \varrho, k) - \widehat{\varrho}(y, \varrho_{\boldsymbol{y}}, k)| \\ &\le |\varrho(y) - \widehat{\varrho}(y, \varrho, k)| + |\widehat{\varrho}(y, \varrho, k) - \widehat{\varrho}(y, \varrho_{\boldsymbol{y}}, k)|\,. \end{aligned}$$

The first term depends on the density kernel and measures how well the smoothed density estimates the true density. The second term measures how well the practical KDE estimates the smoothed density. We will analyze the two terms separately assuming that ϱ and $k(f(\cdot), y)$ lie in reproducing kernel Hilbert spaces.

Let $K^y : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a reproducing kernel for a Hilbert space containing ϱ. Then an upper bound on the first term in the expression for the error can be

bounded via the Cauchy-Schwarz inequality and the Riesz representation theorm:

$$
\begin{aligned}
|\varrho(y) - \widehat{\varrho}(y,\varrho,k)| &= \left|\varrho(y) - \int_{-\infty}^{\infty} k(z,y)\,\varrho(z)\,\mathrm{d}z\right| \\
&= \left|\left\langle K^y(\cdot,y) - \int_{-\infty}^{\infty} k(z,y)\,K^y(z,\cdot)\,\mathrm{d}z, \varrho\right\rangle_{K^y}\right| \\
&\le \left\|K^y(\cdot,y) - \int_{-\infty}^{\infty} k(z,y)\,K^y(z,\cdot)\,\mathrm{d}z\right\|_{K^y} \|\varrho\|_{K^y}\,,
\end{aligned} \tag{3}
$$

where the part of the error bound depending on density kernel can be rewritten in terms of integrals:

$$
\begin{aligned}
&\left\|K^y(\cdot,y) - \int_{-\infty}^{\infty} k(z,y)\,K^y(z,\cdot)\,\mathrm{d}z\right\|_{K^y}^2 \\
&\quad = K^y(y,y) - 2\int_{-\infty}^{\infty} k(z,y)\,K^y(z,y)\,\mathrm{d}z \\
&\qquad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(z,y)\,K^y(z,t)\,k(t,y)\,\mathrm{d}z\,\mathrm{d}t.
\end{aligned} \tag{4}
$$

Equation (3) is an upper bound on the first term in the expression for the error that separates the part depending on the choice of the density kernel from the part depending on the probability density.

Let $K^{\boldsymbol{x}} : [0,1]^d \times [0,1]^d \to \mathbb{R}$ be a reproducing kernel for a Hilbert space containing $k(y, f(\cdot))$ for all $y \in \mathbb{R}$ and all f of interest. Then

$$
\begin{aligned}
|\widehat{\varrho}(y,\varrho,k) - \widehat{\varrho}(y,\varrho_{\boldsymbol{y}},k)| &= \left|\int_{-\infty}^{\infty} k(z,y)\,\varrho(z)\,\mathrm{d}z - \frac{1}{n}\sum_{i=1}^{n} k(y_i,y)\right| \\
&= \left|\int_{[0,1]^d} k(f(\boldsymbol{x}),y)\,\mathrm{d}\boldsymbol{x} - \frac{1}{n}\sum_{i=1}^{n} k(f(\boldsymbol{x}_i),y)\right| \\
&= \left|\left\langle \int_{[0,1]^d} K^{\boldsymbol{x}}(\cdot,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{n}\sum_{i=1}^{n} K^{\boldsymbol{x}}(\cdot,\boldsymbol{x}_i), k(f(\cdot),y)\right\rangle_{K^y}\right| \\
&\le \left\|\int_{[0,1]^d} K^{\boldsymbol{x}}(\cdot,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{n}\sum_{i=1}^{n} K^{\boldsymbol{x}}(\cdot,\boldsymbol{x}_i)\right\|_{K^{\boldsymbol{x}}} \|k(f(\cdot),y)\|_{K^{\boldsymbol{x}}}\,,
\end{aligned}
$$

where

$$
\begin{aligned}
&\left\|\int_{[0,1]^d} K^{\boldsymbol{x}}(\cdot,\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{n}\sum_{i=1}^{n} K^{\boldsymbol{x}}(\cdot,\boldsymbol{x}_i)\right\|_{K^{\boldsymbol{x}}}^2 \\
&\quad = \int_{[0,1]^d\times[0,1]^d} K^{\boldsymbol{x}}(\boldsymbol{x},\boldsymbol{t})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{t} - \frac{2}{n}\sum_{i=1}^{n}\int_{[0,1]^d} K^{\boldsymbol{x}}(\boldsymbol{x},\boldsymbol{x}_i)\,\mathrm{d}\boldsymbol{x} \\
&\qquad + \frac{1}{n^2}\sum_{i,j=1}^{n} K^{\boldsymbol{x}}(\boldsymbol{x}_i,\boldsymbol{x}_j).
\end{aligned}
$$

This is an upper bound on the second term in the expression for the error that separates the part depending on the choice of the kernel evaluated at $(f(\cdot),y)$ from the part depending on the sample nodes.

1. Isotropic Density Kernels

Suppose that $k(z,y) = h^{-1}\tilde{k}((z-y)/h)$ where h is a bandwidth, and $K^y(y,z) = \widetilde{K}^y(y-z)$. Then the quantity measuring the quality of the density kernel in (4) may be written as

$$\begin{aligned}
&\left\| K^y(\cdot,y) - \int_{-\infty}^{\infty} k(z,y)\, K^y(z,\cdot)\,\mathrm{d}z \right\|_{K^y}^2 \\
&\quad = \widetilde{K}^y(0) - \frac{2}{h}\int_{-\infty}^{\infty} \tilde{k}((z-y)/h)\,\widetilde{K}^y(z-y)\,\mathrm{d}z \\
&\qquad\qquad + \frac{1}{h^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{k}((z-y)/h)\,\widetilde{K}^y(z-t)\,\tilde{k}((t-y)/h)\,\mathrm{d}z\,\mathrm{d}t \\
&\quad = \widetilde{K}^y(0) - 2\int_{-\infty}^{\infty} \tilde{k}(w)\,\widetilde{K}^y(hw)\,\mathrm{d}w \\
&\qquad\qquad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{k}(w)\,\widetilde{K}^y(h(w-v))\,\tilde{k}(v)\,\mathrm{d}w\,\mathrm{d}v.
\end{aligned}$$

via the variable transformations $w = (z-y)/h$ and $v = (t-y)/h$.

Let $\widetilde{K}^y(y) = \exp(-y^2/2)$ and $\tilde{k}(y) = \exp(-y^2/2)/\sqrt{2\pi}$.

References