# Asymptotics for density estimator with local divergence

Kanta Naito, Chiba University

We consider a density estimator composed by a localized optimization using Bregman divergence and a kernel function. Let $X_1, \ldots, X_n$ be a sample drawn from a $d$ variate density $f$. We aim at obtaining a density estimator with local adaptation. To do this we utilize a parametric $d$ variate density $g_\theta(x) = g(x, \theta)$, where $\theta \in \Theta \subset \mathbb{R}^p$ is a $p$-dimensional parameter vector. For a strictly convex function $U$, consider the Bregman divergence

$$D_{U^*}(g_\theta, f) = \int_{\mathbb{R}^d} \left[U^*(u(g(x,\theta))) - U^*(u(f(x))) - f(x)\left\{u(g(x,\theta)) - u(f(x))\right\}\right] dx,$$

where $U^*$ is the convex conjugate of $U$ and $u = U'$, the derivative of $U$.

A usual parametric density estimator can be composed by $\bar{f}(x) = g(x, \hat{\theta}_n)$, where

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \ell_n(\theta)$$

and

$$\ell_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} u(g(X_i, \theta)) + \int_{\mathbb{R}^d} U^*(u(g(x,\theta)))dx,$$

which is an empirical version of $D_{U^*}(g_\theta, f)$.

Our proposed density estimator $\hat{f}(t)$ at a target point $t$ is defined as

$$\hat{f}(t) = \frac{g(t, \hat{\theta}_n(t))}{\displaystyle\int_{\mathbb{R}^d} g(s, \hat{\theta}_n(s))ds},$$

where

$$\hat{\theta}_n(t) = \arg\min_{\theta \in \Theta} \ell_{n,t}(\theta)$$

and

$$\ell_{n,t}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i - t}{h}\right) u(g(X_i, \theta)) + \int_{\mathbb{R}^d} K\left(\frac{x - t}{h}\right) U^*(u(g(x,\theta)))dx.$$

Here $K(z)$ is kernel function which is a smooth unimodal integrable function symmetric around $z = 0$, and $h$ is the bandwidth. The $\ell_{n,t}(\theta)$ can be seen as a localized version of $\ell_n(\theta)$.

The risk of $\bar{f}$ and $\hat{f}$ are defined respectively as

$$\mathcal{D}(\bar{f}) = E_\chi\left[D_{U^*}(\bar{f}, f)\right] \quad \text{and} \quad \mathcal{D}(\hat{f}) = E_\chi\left[D_{U^*}(\hat{f}, f)\right]$$

where $E_\chi$ designates the expectation by the joint distribution of $X_1, ..., X_n$. We evaluate the risk difference $\mathcal{D}(\bar{f}) - \mathcal{D}(\hat{f})$ under the situation $n \to \infty$ and"$h \to \infty$", and we claim that $\hat{f}$ improves $\bar{f}$ in asymptotic sense.