MR3800980 62G07
Nagler, T. **[Nagler, Thomas]** (D-MUTU-DM)
Asymptotic analysis of the jittering kernel density estimator. **(English summary)**
Math. Methods Statist. **27** (2018), *no. 1,* 32–46.

This article is concerned with nonparametric multivariate density estimation for mixed data by a kernel method. Suppose that we are interested in estimating the unknown joint density f of $(\mathbf{Z},\mathbf{X})$ using iid observations $\{(\mathbf{Z}_i,\mathbf{X}_i)\}_{i=1}^n$, where $\mathbf{Z}\in\mathbb{Z}^p$ and $\mathbf{X}\in\mathbb{R}^q$ are discrete and continuous components, respectively. A typical approach for estimating f nonparametrically is to construct a product kernel from different univariate kernel functions for $\mathbf{Z}$ and $\mathbf{X}$ [see, e.g., Q. Li and J. S. Racine, J. Multivariate Anal. **86** (2003), no. 2, 266–292; MR1997765]; to put it in another way, discrete and continuous components are treated separately.

In contrast, the main focus of this article is on jittering—adding noises to discrete components to make them continuous. Let $\mathbf{E}\in\mathbb{R}^p$ be a random draw from a certain class of continuous distributions that is independent of $(\mathbf{Z},\mathbf{X})$. The jittering kernel density estimator is defined as

$$\tilde f(\mathbf{z},\mathbf{x})=\frac{1}{nh^pb^q}\sum_{i=1}^n\mathbb{K}\left(\frac{\mathbf{Z}_i+\mathbf{E}_i-\mathbf{z}}{h}\right)\mathbb{K}\left(\frac{\mathbf{X}_i-\mathbf{x}}{b}\right),$$

where $\mathbb{K}(\mathbf{u})=\prod_{j=1}^rK(u_j)$ is the product kernel constructed from a univariate continuous kernel $K(\cdot)$, and $(h,b)=(h_n,b_n)$ are bandwidths. Then, the following properties of $\tilde f(\mathbf{z},\mathbf{x})$ are explored:

(1) pointwise asymptotic normality;
(2) uniform strong consistency;
(3) relative asymptotic efficiency for the case of only one discrete variable (i.e., $\tilde f(\mathbf{z},\mathbf{x})=\tilde f(z)$) relative to the sample frequency estimator $f_n(z)=(1/n)\sum_{i=1}^n\mathbf{1}(Z_i=z)$; and
(4) minimax-optimal rates.

A few comparisons with convergence properties of the density estimator by Li and Racine [op. cit.], denoted by $\widehat f_{LR}(\mathbf{z},\mathbf{x})$, may be of interest. First, the leading bias term of $\tilde f(\mathbf{z},\mathbf{x})$ depends only on b; i.e., smoothing the discrete component generates no bias. This is a sharp contrast to the fact that the dominant bias term of $\widehat f_{LR}(\mathbf{z},\mathbf{x})$ includes both h and b (in the same notations). Second, $\mathrm{Var}\{\tilde f(\mathbf{z},\mathbf{x})\}=O\{(nh^pb^q)^{-1}\}$, which appears to be slower than $\mathrm{Var}\{\widehat f_{LR}(\mathbf{z},\mathbf{x})\}=O\{(nb^q)^{-1}\}$. However, the leading bias term of $\tilde f(\mathbf{z},\mathbf{x})$ is free of h, and thus it is optimal to choose h as large as possible so that $\mathrm{Var}\{\tilde f(\mathbf{z},\mathbf{x})\}=O\{(nb^q)^{-1}\}$ is the case. It can also be found that as long as a large value of h is taken, adding more discrete variables does not affect the convergence rate of $\tilde f(\mathbf{z},\mathbf{x})$.

Masayuki Hirukawa