Non-Standard Behavior of Density Estimators for Functions of Independent Observations

Wolfgang Wefelmeyer (University of Cologne)

based on joint work with
Anton Schick (Binghamton University)

mailto:wefelm@math.uni-koeln.de
http://www.mi.uni-koeln.de/∼wefelm/

Let $X_1, \ldots, X_n$ be real-valued and i.i.d. with density f.

We want to estimate the density p of a known function $q(X_1, \ldots, X_m)$ of $m \geq 2$ arguments at a point z.

Frees (1994) suggests a kernel estimator based on "observations" $q(X_{i_1}, \ldots, X_{i_m})$, i.e. a *local U-statistic*

$$\widehat{p}(z) = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \cdots < i_m \leq n} \frac{1}{b} k\Big(\frac{z - q(X_{i_1}, \ldots, X_{i_m})}{b}\Big).$$

This estimator does not behave like a usual kernel estimator. Frees shows that, *under appropriate assumptions*, $\widehat{p}(z)$ has the *parametric* rate $1/\sqrt{n}$. Giné and Mason (2007) prove a *functional* result for the process $z \mapsto \sqrt{n}(\widehat{p}(z) - p(z))$ in L_p for $1 \leq p \leq \infty$ (and uniformly in the bandwidth b).

We discuss, in special cases, when these results *fail* to hold.

Special case: density p of convolution of two (positive) powers,

$$q(X_1, X_2) = |X_1|^\nu + |X_2|^\nu, \quad \nu > 0.$$

The local U-statistic for p is

$$\widehat{p}(z) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \frac{1}{b} k\Big(\frac{z - |X_i|^\nu - |X_j|^\nu}{b}\Big).$$

If X has density f, then $|X|$ has density

$$h(y) = (f(y) + f(-y))\mathbf{1}[y > 0],$$

and $|X|^\nu$ has a density with a *peak* at 0:

$$g(y) = \frac{1}{\nu} y^{\frac{1}{\nu}-1} h(y^{\frac{1}{\nu}}).$$

The density p of $|X_1|^\nu + |X_2|^\nu$ has the convolution representation

$$p(z) = \int g(z-y)g(y)\,dy$$

and can also be estimated by a *plug-in* estimator, using X_j or $|X_j|^\nu$.

For $\nu < 2$, a Hoeffding decomposition of the local U-statistic gives

$$\widehat{p}(z) - p(z) = \frac{2}{n} \sum_{i=1}^{n} \Big(g(z - |X_i|^{\nu}) - p(z) \Big) + o_p(1/\sqrt{n}).$$

Theorem 1 *Let $\nu < 2$. Suppose h is of bounded variation and $h(0+) > 0$. Choose $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{n}(\widehat{p}(z) - p(z)) \Rightarrow N\Big(0, 4\mathit{Var}\, g(z - |X|^{\nu})\Big).$$

Note that the second moment of $g(z - |X|^{\nu})$ is

$$\int g^2(z-y) g(y)\, dy = \frac{1}{\nu^3} \int_0^1 (z-y)^{\frac{2}{\nu}-2} h^2\Big((z-y)^{\frac{1}{\nu}}\Big) y^{\frac{1}{\nu}-1} h\Big(y^{\frac{1}{\nu}}\Big).$$

This is *infinite* for $\nu \geq 2$ unless:
$h(z-) = 0$ (or $g(z-) = 0$) or $h(0+) = 0$.

A boundary case is $\nu = 2$, i.e. estimation of the density of $X_1^2 + X_2^2$. Then the variance of $g(z - X^2)$ is just barely infinite.

Theorem 2 *Let $\nu = 2$. Suppose h is of bounded variation and $h(0+)$ and $g(z-)$ are positive. Choose $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{\frac{n}{\log n}} \Big(\widehat{p}(z) - p(z)\Big) \Rightarrow N(0, h^2(0+)g(z-)).$$

The rate of the local U-statistic $\widehat{p}(z)$ is still close to $1/\sqrt{n}$, but its asymptotic variance now depends only on $h(0+)$ and $g(z-)$ (with h density of $|X|$ and g density of $|X|^\nu$). — One can still show efficiency, but a *functional* result for the process $z \mapsto \sqrt{n/\log n}(\widehat{p}(z) - p(z))$ is not possible.

(For $\nu < 2$, the rate of the local U-statistic $\widehat{p}(z)$ was $1/\sqrt{n}$, and its asymptotic variance was $4 \operatorname{Var} g(z - |X|^\nu)$.)

For $\nu > 2$, the density g of $|X|^\nu$ has an even more pronounced peak at 0. The local U-statistic $\widehat{p}(z)$ then converges more slowly than $1/\sqrt{n}$.

Theorem 3 *Let $\nu > 2$. Suppose h is of bounded variation and $h(0+)$ and $g(z-)$ are positive. Let $b \sim 1/n$. Then*

$$\widehat{p}(z) - p(z) = O_P(n^{-1/\nu}).$$

If $\nu \geq 1$ and g *vanishes* near z, then we still get the rate $1/\sqrt{n}$. This happens if g has compact support and z is outside it.

Theorem 4 *Let $\nu \geq 2$. Suppose h is of bounded variation, $h(0+)$ is positive, and g vanishes in a neighborhood of z. Let $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{n}(\widehat{p}(z) - p(z)) \Rightarrow N\Big(0, 4\mathit{Var}\, g(z - |X|^\nu)\Big).$$

The results translate to models with additional parameters and dependent observations.

Let $X_0, \dots, X_n$ be observations of a (uniformly ergodic) first-order nonlinear autoregressive process

$$X_j = r_\vartheta(X_{j-1}) + \varepsilon_j$$

with i.i.d. innovations ε_j with mean 0. Then the stationary density p of X_j at z can be estimated by the local U-statistic

$$\hat{p}(z) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} k_b(z - r_{\hat{\vartheta}}(X_i) - \hat{\varepsilon}_j)$$

with residuals $\hat{\varepsilon}_j = X_j - r_{\hat{\vartheta}}(X_{j-1})$ and $\hat{\vartheta}$ an estimator of ϑ.

The rate of the local U-statistic $\hat{p}(z)$ is $1/\sqrt{n}$ if the derivative of the autoregression function is bounded away from zero. This is in particular the case for *linear* autoregression.

For moving average: Saavedra and Cao (1999). For invertible linear processes: Schick and W. (2007). For nonlinear *regression*: Støve und Tjøstheim (2007), Müller (2009). For *nonparametric* regression: Jacho-Chávez and Escanciano (2009).

Suppose the autoregression function has derivative 0 at some point x. Then the rate of the local U-statistic $\hat{p}(z)$ depends on how flat r_ϑ is near x. Work in progress.

Analogous results hold for *products* (rather than *sums*) of independent random variables.

Let $(X_0, T_0), \dots, (X_n, T_n)$ be observations of a (uniformly ergodic) Markov renewal process. Assume that the inter-arrival times $T_j - T_{j-1}$ depend *multiplicatively* on the distance between the past and present states X_{j-1} and X_j of the embedded Markov chain,

$$T_j - T_{j-1} = |X_j - X_{j-1}|^\alpha W_j,$$

where $\alpha > 0$ is known and the W_j are positive, i.i.d., and independent of the embedded Markov chain. Then the inter-arrival density can be estimated by the local U-statistic

$$\widehat{p}(v) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} k_b(v - |X_i - X_{i-1}|^\alpha W_j).$$

Note that $W_j = |X_j - X_{j-1}|^{-\alpha}(T_j - T_{j-1})$ is observed. Schick and W. (2009) obtain the rate $1/\sqrt{n}$ for $\widehat{p}(v)$. — A *functional* result for the process $v \mapsto \sqrt{n}(\widehat{p}(v) - p(v))$ is not possible.