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

U. U. Müller[1]*, A. Schick[2]**, and W. Wefelmeyer[3]

[1] Department of Statistics, Texas A&M University
College Station, TX 77843-3143, USA
(email: uschi@stat.tamu.edu)

[2] Department of Mathematical Sciences, Binghamton University
Binghamton, NY 13902-6000, USA
(email: anton@math.binghamton.edu)

[3] Mathematical Institute, University of Cologne
50931 Cologne, Germany
(email: wefelm@math.uni-koeln.de)

Abstract. Densities of functions of two or more independent random variables can be estimated by local U-statistics. Frees (1994) gives conditions under which they converge pointwise at the parametric root-n rate. Giné and Mason (2007) give conditions under which this rate also holds in L_p-norms. We present several natural applications in which the parametric rate fails to hold in L_p or even pointwise.

1. The density estimator of a sum of squares of independent observations typically slows down by a logarithmic factor. For exponents greater than two, the estimator behaves like a classical density estimator.

2. The density estimator of a product of two independent observations typically has the root-n rate pointwise, but not in L_p-norms. An application is given to semi-Markov processes and estimation of an inter-arrival density that depends multiplicatively on the jump size.

3. The stationary density of a nonlinear or nonparametric autoregressive time series driven by independent innovations can be estimated by a local U-statistic (now based on dependent observations and involving additional parameters), but the root-n rate can fail if the derivative of the autoregression function vanishes at some point.

Keywords: Density estimator, Local U-statistic, Local von Mises statistic, Convergence rate, Autoregressive time series, Semi-Markov process.

1 Introduction

It is often of interest to estimate densities of known or unknown functions of independent observations. Consider for example a regression model $Y = r(X) + \varepsilon$ with independent error ε and covariate X. If we have independent observations (X_i, Y_i), $i = 1, \dots, n$, then the density of the response Y could be estimated by a kernel estimator based on $Y_1, \dots, Y_n$. However, a much

* Supported by NSF Grant DMS 0907014.

** Supported by NSF Grant DMS 0906551.

better estimator is obtained if we exploit the independence of ε and X and write Y as a sum $r(X)+\varepsilon$ of independent random variables. Then the density p of Y can be estimated by a local von Mises statistic

$$\hat{p}(z) = \frac{1}{n^2} \sum_{i,j=1}^{n} k_b(z - \hat{r}(X_i) - \hat{\varepsilon}_j).$$

Here $k_b(z) = k(z/b)/b$ with kernel k and bandwidth b, $\hat{r}$ is some estimator of the regression function r, and $\hat{\varepsilon}_j = Y_j - \hat{r}_j(X_j)$ are the corresponding residuals. Under appropriate conditions, the estimator $\hat{p}(z)$ converges at the parametric rate $n^{1/2}$; see Støve and Tjøstheim, 2011 [19], Escanciano and Jacho-Chávez, 2011 [1], and, for nonlinear regression and with responses missing at random, Müller, 2010 [5]. It is the purpose of this review to indicate why such rates are possible, and to illustrate when they fail.

The most straightforward version of the problem is the following. Let $X_1, \dots, X_n$ be independent real-valued observations with density f. We want to estimate the density p of some transformation $T(X_1, \dots, X_m)$ of m of these observations, with m at least 2. Frees, 1994 [2] proposed as an estimator of $p(z)$ the local U-statistic

$$\hat{p}(z) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \dots < i_m \le n} k_b(z - T(X_{i_1}, \dots, X_{i_m}))$$

with $k_b(x) = k(x/b)/b$ for a kernel k and a bandwidth b. He showed that this estimator can be pointwise $n^{1/2}$-consistent under some assumptions on f and T. Saavedra and Cao, 2000 [9] consider the function $T(X_1, X_2) = X_1 + \varphi X_2$. It is even possible to obtain $n^{1/2}$-consistency in various norms, together with functional central limit theorems in the corresponding spaces. Schick and Wefelmeyer, 2004 [11], 2007 [13] prove such results for transformations of the form $T(X_1, \dots, X_m) = T_1(X_1) + \dots + T_m(X_m)$ and $T(X_1, X_2) = X_1 + X_2$ in the sup-norm and in L_1-norms. Giné and Mason, 2007 [3] consider general transformations $T(X_1, \dots, X_m)$ and obtain such results in the L_p-norms. Their results hold locally uniformly in the bandwidth. More general results applicable here are in Nickl, 2007 [6] and Nickl, 2009 [7].

These results are less generally valid than appears at first sight. In Section 2 we restrict attention to $m = 2$ and to transformations of the special form $T(X_1, X_2) = T_1(X_1) + T_2(X_2)$ and explain under which conditions the local U-statistic $\hat{p}(z)$ is asymptotically linear, $n^{1/2}$-consistent, and asymptotically normal. The rate is typically slower when, say, $T_1(y) = T_1(x) + c(y - x)^\nu + o(|y - x|^\nu)$ for y to the left or right of some point x, with $\nu \ge 2$. Then the density of $T_1(X)$ has a strong peak. Specifically, we consider $T_1(x) = T_2(x) = x^\nu$ and describe the rates of the local U-statistic. Then we discuss the two-sample case and applications to regression, to time series driven by independent innovations, and to renewal processes with multiplicative waiting times.

2 Results and Applications

Let $X_1, \dots, X_n$ be independent real-valued observations with density f. An estimator for the density p of a transformation of the form $T(X_1, X_2) = T_1(X_1) + T_2(X_2)$ is the local U-statistic

$$\hat{p}(z) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} k_b(z - T_1(X_i) - T_2(X_j)),$$

where $k_b(z) = k(z/b)/b$ for a kernel k and a bandwidth b. Suppose that $T_1(X)$ and $T_2(X)$ have densities g_1 and g_2. The estimator $\hat{p}(z)$ has the Hoeffding decomposition

$$\hat{p}(z) = p * k_b(z) + \frac{1}{n} \sum_{i=1}^{n} \big(g_1 * k_b(z - T_2(X_i)) - p * k_b(z) \\ + g_2 * k_b(z - T_1(X_i)) - p * k_b(z) \big) + U(z),$$

where

$$U(z) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \big(k_b(z - T_1(X_i) - T_2(X_i)) - g_1 * k_b(z - T_2(X_i)) \\ - g_2 * k_b(z - T_1(X_i)) + p * k_b(z) \big)$$

is a degenerate local U-statistic. We have

$$n(n-1)E[U^2(z)] \le 2E[k_b^2(z - T_1(X_1) - T_2(X_2))] = 2p * k_b^2(z)$$

and

$$p * k_b^2(z) = \frac{1}{b} \int p(z - bu) k^2(u)\, du \le \frac{\|p\|_\infty}{b} \int k^2(u)\, du.$$

If p is bounded and $\int k^2(u)\, du$ is finite, we obtain $U(z) = O_P(1/(nb^{1/2}))$, which is of order $o_P(n^{-1/2})$ if $nb \to \infty$. The Hoeffding decomposition then says that the centered local U-statistic $\hat{p}(z) - p * k_b(z)$ is approximated by a sum of two centered and smoothed empirical "estimators" of $p(z)$ (that involve the unknown densities g_1 and g_2). Under mild assumptions one can remove the smoothing; see e.g. Schick and Wefelmeyer, 2004 [11]. If p is Hölder with exponent α, then the bias $p * k_b(z) - p(z)$ is of order $o(n^{-1/2})$ if $nb^{2\alpha} \to 0$. This implies that $\hat{p}(z)$ is asymptotically linear,

$$\hat{p}(z) = p(z) + \frac{1}{n} \sum_{i=1}^{n} \big(g_1(z - T_2(X_i)) + g_2(z - T_1(X_i)) - 2p(z) \big) + o_P(n^{-1/2}). \quad (1)$$

If $E[g_1^2(z - T_2(X_2))]$ and $E[g_2^2(z - T_1(X_1))]$ are finite, then $\hat{p}(z)$ is $n^{1/2}$-consistent and asymptotically normal.

Remark 1. (Convolution of density estimators.) The density p has the convolution representation

$$p(z) = \int g_2(z-y) g_1(y)\, dy.$$

Therefore, it can also be estimated by a convolution of density estimators

$$\hat{g}_{conv}(z) = \int \hat{g}_2(z-y)\hat{g}_1(y)\, dy$$

with kernel estimator for $g_1(y)$ based on $T_1(X_1), \dots, T_1(X_n)$,

$$\hat{g}_1(y) = \frac{1}{n}\sum_{i=1}^n k_b(y - T_1(X_i)),$$

and, correspondingly,

$$\hat{g}_2(y) = \frac{1}{n}\sum_{i=1}^n k_b(y - T_2(X_i)).$$

The estimator $\hat{g}_{conv}$ is asymptotically equivalent to $\hat{g}$. □

Remark 2. (Transform density estimator or transform observations.) Suppose that T_1, say, is strictly increasing and differentiable. Then the density of $T_1(X)$ at y is

$$g_1(y) = \frac{f(T_1^{-1}(y))}{T_1'(T_1^{-1}(y))}.$$

We obtain an alternative estimator of $g_1(y)$ by plugging in a kernel estimator for f,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^n k_b(x - X_i).$$

We expect that it depends on T_1 whether $\hat{g}_1(y)$ is better than

$$\tilde{g}_1(y) = \frac{\hat{f}(T_1^{-1}(y))}{T_1'(T_1^{-1}(y))}.$$

In the convolution representation $p(z) = \int g_2(z-y)g_1(y)\, dy$ we can use $\hat{g}_1$ or $\tilde{g}_1$. If T_2 is also strictly increasing and differentiable, we can combine $\hat{g}_1$ or $\tilde{g}_1$ with $\hat{g}_2$ or $\tilde{g}_2$. □

We now discuss cases in which $\hat{p}(z)$ is not $n^{1/2}$-consistent.

Remark 3. (Piecewise constant transformations.) The distribution of $T_1(X)$ does not always have a density. Suppose that T_1 is piecewise constant,

$$T_1(X) = \sum_{s=1}^t c_s \mathbf{1}[X \in I_s],$$

with $c_s \in \mathbb{R}$, and I_s, $s = 1, \dots, t$, a partition of $\mathbb{R}$. If $T_2(X)$ has a density g_2, then $T_1(X_1) + T_2(X_2)$ has a density p that is a finite mixture of shifts of g_2,

$$p(z) = \sum_{s=1}^{m} a_s g_2(z - c_s)$$

with weights $a_s = P(X \in I_s)$. As soon as each interval I_s contains at least one observation, the constants c_s can be observed, and $p(z)$ can be estimated by

$$\hat{p}(z) = \sum_{s=1}^{t} \hat{a}_s \hat{g}_2(z - c_s),$$

where $\hat{a}_s = (1/n) \sum_{i=1}^{n} \mathbf{1}[X_i = c_s]$. The rate of $\hat{p}(z)$ equals the pointwise rate of $\hat{g}_2$. □

Even if T_1 and T_2 are not constant on any interval, $\hat{p}(z)$ can fail to be $n^{1/2}$-consistent. In the following we describe a situation in which $T_1(X)$ and $T_2(X)$ have densities, but $g_1(z - T_2(X))$ does not necessarily have finite variance. For notational simplicity, assume that f is supported on $(0, \infty)$, and set $T_1(x) = T_2(x) = x^\nu$ for some $\nu > 0$. Then $g_1 = g_2 = g$ with

$$g(y) = \frac{1}{\nu} y^{1/\nu - 1} f(y^{1/\nu}),$$

and the stochastic expansion (1) of $\hat{p}(z)$ specializes to

$$\hat{p}(z) = p(z) + \frac{2}{n} \sum_{i=1}^{n} g(z - X_i^\nu) + o_P(n^{-1/2}). \tag{2}$$

In the theorems below, we take the kernel k to be continuously differentiable with support $[-1, 1]$. We also assume that f is bounded. First let $\nu < 2$. Then g is square-integrable, and $g(z - X^\nu)$ has finite variance. By the arguments of Schick and Wefelmeyer, 2004 [11] and 2009 [17] or Giné and Mason, 2007 [3], we have the following result.

Theorem 1. *Let $\nu < 2$. Suppose the density f is of bounded variation and $f(0+)$ is positive. Let $b \sim (\log n)^{1/2}/n$. Then $\hat{p}(z)$ has the stochastic expansion (2), and*

$$n^{1/2}\big(\hat{p}(z) - p(z)\big) \Rightarrow N\big(0, 4 \operatorname{Var}(g(z - X^\nu))\big).$$

For $\nu = 2$, square-integrability of g fails just barely, resulting in a rate for $\hat{p}(z)$ that is only slightly worse than $n^{-1/2}$. More precisely, Schick and Wefelmeyer, 2009 [16] prove the following result.

Theorem 2. *Let $\nu = 2$. Suppose f is of bounded variation, and $f(0+)$ and $g(z-)$ are positive. Let $b \sim (\log n)^{1/2}/n$. Then*

$$\Big(\frac{n}{\log n}\Big)^{1/2}\big(\hat{p}(z) - p(z)\big) \Rightarrow N(0, f^2(0+) g(z-)).$$

For $\nu > 2$, the rate of $\hat{p}(z)$ is of order $n^{-1/\nu}$ if f is of bounded variation and $f(0+)$ and $g(z-)$ are positive. Faster rates are possible under additional smoothness assumptions on p at z.

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

$$\hat{p}_b(z) - p(z) = O_P(n^{-\beta}).$$

Even in the case $\nu \geq 2$, the estimator $\hat{p}(z)$ can be $n^{1/2}$-consistent if $g(z-) = 0$ since this works against the peak of g at 0 in the representation $p(z) = g * g(z)$. For details we refer to Schick and Wefelmeyer, 2009 [16] and 2009 [17].

We will now briefly discuss possible applications of the above results.

Remark 4. (Several samples.) The above results carry over to m-sample cases. We restrict attention to $m = 2$. Suppose $X_1, \dots, X_n$ and $Z_1, \dots, Z_n$ are real-valued and independent with densities f_1 and f_2, respectively. An estimator for the density p of a transformation $T_1(X) + T_2(Z)$ is the local von Mises statistic

$$\hat{p}(z) = \frac{1}{n^2} \sum_{i,j=1}^{n} k_b(z - T_1(X_i) - T_2(Z_j)).$$

Let g_1 and g_2 denote the densities of $T_1(X)$ and $T_2(Z)$. As in the one-sample case (1) we obtain a stochastic expansion

$$\hat{p}(z) = p(z) + \frac{1}{n} \sum_{i=1}^{n} \big(g_1(z - T_2(Z_i)) + g_2(z - T_1(X_i)) - 2p(z)\big) + o_P(n^{-1/2}).$$

Appropriate versions of Theorems 1–3 continue to hold. □

Remark 5. (Regression.) Two-sample results can be applied to regression models $Y = r(X) + \varepsilon$ with ε independent of X. If we have independent observations (X_i, Y_i), $i = 1, \dots, n$, then the density p of Y can be estimated by the local von Mises statistic

$$\hat{p}(z) = \frac{1}{n^2} \sum_{i,j=1}^{n} k_b(z - \hat{r}(X_i) - \hat{\varepsilon}_j)$$

based on some estimator $\hat{r}$ of the regression function r, and on residuals $\hat{\varepsilon}_j = Y_j - \hat{r}(X_j)$. Note that the "pseudo-observations" $\hat{r}(X_i)$ and $\hat{\varepsilon}_j$ are only approximately independent, so we are close to the two-sample case with $Z = \varepsilon$, $T_1(X) = r(X)$, and $T_2(\varepsilon) = \varepsilon$. As seen above, we can expect a rate $n^{-1/2}$ for $\hat{p}(z)$ if r has a derivative that is bounded away from 0.

Suppose r is only piecewise monotone and continuously differentiable, and there are points x with

$$r(y) = c(y - x)^{\nu} + o(|y - x|^{\nu})$$

for y to the left or right of x. Then the convergence rate of $\hat{p}(z)$ will be determined by the largest such ν. □

Remark 6. (Time series.) Results for regression carry over to time series driven by independent innovations. Consider a first-order moving average process $X_i = \varepsilon_i + \varphi\varepsilon_{i-1}$, with independent innovations ε_i that have mean 0, finite variance, and density f. If $\varphi \neq 0$, the stationary density p can be estimated by a local von Mises statistic

$$\hat{p}(z) = \frac{1}{n^2} \sum_{i,j=1}^{n} k_b(z - \hat{\varepsilon}_i - \hat{\varphi}\hat{\varepsilon}_j)$$

with $\hat{\varphi}$ an estimator of φ. Saavedra and Cao, 1999 [8] obtain $n^{1/2}$-consistency; see also Schick and Wefelmeyer, 2004 [10]. Functional results for higher-order moving average processes and general linear processes are obtained in Schick and Wefelmeyer, 2004 [12], 2007 [14], 2008 [15] and 2009 [18]. Nonlinear and nonparametric time series can also be treated. □

Remark 7. (Renewal processes.) Here is a two-sample case where $T(X, Z)$ is a product rather than a sum of functions $T_1(X)$ and $T_2(Z)$. Let (X_i, T_i), $i = 0, \ldots, n$ be observations of a Markov renewal process with real state space. Assume that the embedded Markov chain is stationary. We make the structural assumption that the waiting times depend multiplicatively on some power of the distance between the previous and the present state of the embedded Markov chain,

$$T_i - T_{i-1} = |X_i - X_{i-1}|^{\nu} W_i,$$

where $\nu > 0$ and the W_i are independent with density g and independent of the embedded Markov chain. Note that W_i is observable as a function of the observations (X_{i-1}, T_{i-1}) and (X_i, T_i). We can estimate the waiting time density p of $T_i - T_{i-1}$ by the local von Mises statistic

$$\hat{p}(z) = \frac{1}{n^2} \sum_{i,j=1}^{n} k_b(z - |X_i - X_{i-1}|^{\nu} W_j).$$

Greenwood et al., 2011 [4] give conditions under which $\hat{p}(z)$ has rate $n^{-1/2}$ and is asymptotically linear and asymptotically normal. □

References

1. Escanciano, J. C. and Jacho-Chávez, D. T. "$\sqrt{n}$-Uniformly Consistent Density Estimation in Nonparametric Regression Models". To appear in *J. Econometrics* (2011).
2. Frees, E. W. "Estimating Densities of Functions of Observations". *J. Amer. Statist. Assoc.* 89:517–525 (1994).
3. Giné, E. and Mason, D. "On Local U-Statistic Processes and the Estimation of Densities of Functions of Several Sample Variables". *Ann. Statist.* 35:1105–1145 (2007).
4. Greenwood, P. E., Schick, A. and Wefelmeyer, W. "Estimating the Inter-Arrival Time Density of Semi-Markov Processes under Structural Assumptions on the Transition Distribution". *Statist. Probab. Lett.* 81:277–282 (2011).
5. Müller, U. U. "Estimating the Density of a Possibly Missing Response Variable in Nonlinear Regression". Submitted (2010).
6. Nickl, R. "Donsker-Type Theorems for Nonparametric Maximum Likelihood Estimators". *Probab. Theory Related Fields* 138:411–449 (2007). Erratum 141:331–332 (2008).
7. Nickl, R. "On Convergence and Convolutions of Random Signed Measures". *J. Theoret. Probab.* 22:38–56 (2009).
8. Saavedra, A. and Cao, R. "Rate of Convergence of a Convolution-Type Estimator of the Marginal Density of an MA(1) process". *Stochastic Process. Appl.* 80:129–155 (1999).
9. Saavedra, A. and Cao, R. "On the Estimation of the Marginal Density of a Moving Average Process". *Canad. J. Statist.* 28:799–815 (2000).
10. Schick, A. and Wefelmeyer, W. "Root n Consistent and Optimal Density Estimators for Moving Average Processes". *Scand. J. Statist.* 31:63–78 (2004).
11. Schick, A. and Wefelmeyer, W. "Root n Consistent Density Estimators for Sums of Independent Random Variables". *J. Nonparametr. Statist.* 16:925–935 (2004).
12. Schick, A. and Wefelmeyer, W. "Functional Convergence and Optimality of Plug-in Estimators for Stationary Densities of Moving Average Processes". *Bernoulli*, **10**, 889–917 (2004).
13. Schick, A. and Wefelmeyer, W. "Root-n Consistent Density Estimators of Convolutions in Weighted L_1-Norms". *J. Statist. Plann. Inference* 37:1765–1774 (2007).
14. Schick, A. and Wefelmeyer, W. "Uniformly Root-n Consistent Density Estimators for Weakly Dependent Invertible Linear Processes". *Ann. Statist.* 35:815–843 (2007).
15. Schick, A. and Wefelmeyer, W. "Root-n Consistency in Weighted L_1-Spaces for Density Estimators of Invertible Linear Processes". *Stat. Inference Stoch. Process.*, 11:281–310 (2008).
16. Schick, A. and Wefelmeyer, W. "Convergence Rates of Density Estimators for Sums of Powers of Observations". *Metrika* 69:249–264 (2009).
17. Schick, A. and Wefelmeyer, W. "Non-Standard Behavior of Density Estimators for Sums of Squared Observations". *Statist. Decisions* 27:55-73 (2009).
18. Schick, A. and Wefelmeyer, W. "Improved Density Estimators for Invertible Linear Processes". *Comm. Statist. Theory Methods* 38:3123-3147 (2009).
19. Støve, B. and Tjøstheim, D. "A Convolution Estimator for the Density of Nonlinear Regression Observations". To appear in *Scand. J. Statist.* (2011).