Root n consistent density estimators for sums of independent random variables

Anton Schick and Wolfgang Wefelmeyer

Abstract. The density of a sum of independent random variables can be estimated by the convolution of kernel estimators for the marginal densities. We show under mild conditions that the resulting estimator is $n^{1/2}$-consistent and converges in distribution in the spaces $C_0(\mathbb{R})$ and L_1 to a centered Gaussian process.

AMS 2000 subject classifications. 62G07, 62G20.

Key words and Phrases. Plug-in estimator, kernel density estimator, parametric convergence rate, von Mises statistic, functional central limit theorem.

1. Introduction

Smooth functionals of densities can often be estimated at the parametric rate $n^{-1/2}$, even though the density itself can be estimated only at the rate $n^{-\alpha/(2\alpha+1)}$, where α is the degree of smoothness of the density. A natural approach to estimating such functionals is to plug in a density estimator. Of particular interest in statistics are nonlinear integral functionals of the density f and of its derivatives $f^{(k)}$. For quadratic functionals $\int f^{(k)}(x)^2\,dx$ see Hall and Marron (1987) and Bickel and Ritov (1988). For generalizations $\int \phi(f(x),x)\,dx$ and $\int \phi(f(x),\dots,f^{(k)}(x),x)\,dx$ see Laurent (1996) and Birgé and Massart (1995). For the Shannon entropy $-\int f(x)\log f(x)\,dx$ see Dudewicz and van der Meulen (1981). For general results on plug-in estimators we refer to Donoho (1988), who also considers the Fisher information $\int f'(x)^2/f(x)\,dx$, and to Goldstein and Messer (1992) and Goldstein and Khas'minskii (1995). Almost sure i.i.d. representations are obtained by Eggermont and LaRiccia (1999, 2001) and Mason (2003). Abramson and Goldstein (1991) study the equidistribution functional $2\int f(x)g(x)/(f(x)+g(x))\,dx$ of *two* densities.

Frees (1994) shows that the density of a symmetric function $Y = h(X_1,\dots,X_m)$ of $m > 1$ independent and identically distributed random variables can be estimated at the

Anton Schick was partially supported by NSF Grant DMS 0072174.

parametric rate $n^{-1/2}$ if this density and the conditional density of Y given X_1 are sufficiently smooth. His result generalizes to non-identically distributed random variables. This covers in particular convolution densities $f * g(y) = \int f(y-x)g(x)\,dx$. A special case is the stationary density of a moving average process $Y_i = X_i - \vartheta X_{i-1}$ when ϑ is known. Saavedra and Cao (2000) have shown that the variance of an appropriate plug-in estimator of the stationary density at a fixed point y decreases as n^{-1}, implying the parametric rate. Saavedra and Cao (1999) obtain an analogous result for the more interesting case of unknown ϑ. Then ϑ and the innovations X_i must be replaced by estimators. Schick and Wefelmeyer (2004a) prove a stronger result, namely asymptotic normality (and efficiency). Schick and Wefelmeyer (2004b) generalize these results further. They treat higher-order moving average processes, and plug-in estimators of the stationary density as elements of function spaces. A natural space for densities is L_1. The key results of Schick and Wefelmeyer (2004b) are functional central limit theorems in the Banach spaces L_1 and $C_0(\mathbb{R})$, where $C_0(\mathbb{R})$ is the set of all continuous functions h from $\mathbb{R}$ to $\mathbb{R}$ that vanish at infinity in the sense that $\sup_{|y|>M} |h(y)| \to 0$ as $M \to \infty$. Endowed with the sup-norm, $C_0(\mathbb{R})$ becomes a separable Banach space. These results are the first non-local functional central limit theorems for density estimators.

In this note we obtain functional central limit theorems for plug-in estimators of convolution densities. The setting is the following. Suppose we observe independent and identically distributed random variables $X_1, \dots, X_n$. Let $u_1, \dots, u_m$ be known measurable functions such that $u_i(X_i)$ has a density f_i. We are interested in estimating the density g of the sum $u_1(X_1) + \dots + u_m(X_m)$. This density is the convolution $f_1 * \dots * f_m$ of $f_1, \dots, f_m$ and can be expressed as

$$g(y) = \int f_1(y - y_2 - \dots - y_m) f_2(y_2) \cdots f_m(y_m)\,dy_2 \cdots dy_m, \quad y \in \mathbb{R}.$$

An important special case is estimation of the density of a linear combination $a_1X_1 + \dots + a_mX_m$, where the X_i have common density and $a_1, \dots, a_m$ are known non-zero constants. If the constants are equal to one, then g is the m-fold convolution of the common density f of the X_i.

An obvious approach to estimating g is to estimate f_i by $\hat f_i$ and then g by the plug-in estimator $\hat g = \hat f_1 * \dots * \hat f_m$. We use the kernel estimators

$$\hat f_i(y) = \frac{1}{nb_i} \sum_{j=1}^n k_i\Big(\frac{y - u_i(X_j)}{b_i}\Big), \quad y \in \mathbb{R},$$

with kernels k_i and bandwidths b_i. In Section 2 we obtain the uniform i.i.d. representation

$$(1.1) \qquad \sup_{y\in\mathbb{R}} \Big| \hat g(y) - g(y) - \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n \big(g_i(y - u_i(X_j)) - g(y)\big)\Big| = o_p(n^{-1/2})$$

under smoothness of g and mild conditions on the kernels and the bandwidths. Here g_i is the density of $u_1(X_1)+\cdots+u_{i-1}(X_{i-1})+u_{i+1}(X_{i+1})+\cdots+u_m(X_m)$. We then use (1.1) to derive a functional limit theorem for $n^{1/2}(\hat{g}-g)$ in the Banach space $C_0(\mathbb{R})$. In Section 3 we prove an analogous result in the Banach space L_1. These two results imply corresponding results in L_p for $1<p<\infty$. In Section 4 we address computational aspects. We show that our estimator can be written as a von Mises statistic and is asymptotically equivalent to a U-statistic. We also address the choice of bandwidth. Simulations show that our estimator works well even for small sample sizes.

2. Convergence in $C_0(\mathbb{R})$

Let $(\Omega,\mathcal{F},P)$ be a probability space and $(S,\mathcal{B})$ a measurable space. A *random element* in S is a measurable mapping from Ω into S. Let $X,X_1,\dots,X_n$ be independent and identically distributed random elements in S. Let $u_1,\dots,u_m$ be measurable functions from S to the real line $\mathbb{R}$ for some integer $m>1$. Assume that $u_i(X)$ has a density f_i for $i=1,\dots,m$. We estimate f_i by the kernel estimator $\hat{f}_i$ introduced in Section 1.

Note that (1.1) can be written as

$$\|n^{1/2}(\hat{g}-g)-(\mathbb{H}_{n,1}+\cdots+\mathbb{H}_{n,m})\|_\infty = o_p(1), \tag{2.1}$$

where

$$\begin{aligned}\mathbb{H}_{n,i}(y) &= n^{-1/2}\sum_{j=1}^n \big(g_i(y-u_i(X_j))-g(y)\big)\\ &= n^{-1/2}\sum_{j=1}^n \big(g_i(y-u_i(X_j))-E(g_i(y-u_i(X_j))\big),\quad y\in\mathbb{R}.\end{aligned}$$

Let F_i denote the distribution function of f_i, let $\mathbb{F}_i$ denote the empirical distribution function based on $u_i(X_1),\dots,u_i(X_n)$, and let $\Delta_i=n^{1/2}(\mathbb{F}_i-F_i)$ denote the corresponding empirical process. Then we can write

$$\mathbb{H}_{n,i}(y)=\int g_i(y-u)\,d\Delta_i(u).$$

Suppose now that g_i is absolutely continuous with integrable almost everywhere derivative g_i'. Using integration by parts and a change of variables, this allows us to express

$$\mathbb{H}_{n,i}(y)=\int \Delta_i(u)g_i'(y-u)\,du=\int \Delta_i(y-u)g_i'(u)\,du=\Delta_i * g_i'(y).$$

We can use this representation to show that $\mathbb{H}_{n,i}$ has sample paths in $C_0(\mathbb{R})$. (Uniform) continuity of the sample paths follows from the bound

$$\sup_{|z-y|\le\delta}|\mathbb{H}_{n,i}(z)-\mathbb{H}_{n,i}(y)|\le\|\Delta_i\|_\infty \sup_{|z-y|\le\delta}\int |g_i'(z-t)-g_i'(y-t)|\,dt \tag{2.2}$$

and the fact that, by integrability of g_i',

$$\sup_{|z-y|\le\delta}\int |g_i'(z-t)-g_i'(y-t)|\,dt \to 0 \quad \text{as } \delta\downarrow 0;$$

see e.g. Theorem 9.5 in Rudin (1974) or Lemma 2 below. That the sample paths of $\mathbb{H}_{n,i}$ vanish at infinity follows from properties of the empirical process Δ_i and the bound

$$\sup_{|y|>2M}|\mathbb{H}_{n,i}(y)| \le \sup_{|t|>M}|\Delta_i(t)|\int_{|u|\le M}|g_i'(u)|\,du + \|\Delta_i\|_\infty \int_{|u|>M}|g_i'(u)|\,du, \quad M>0. \tag{2.3}$$

Let us now show that the above bounds also give tightness of the process $\mathbb{H}_{n,i}$. To this end let us first characterize compact subsets of $C_0(\mathbb{R})$.

Lemma 1. *A closed subset H of $C_0(\mathbb{R})$ is compact if and only if*

$$\lim_{\delta\downarrow 0}\sup_{h\in H}\sup_{|z-y|\le\delta}|h(z)-h(y)|=0$$

and

$$\lim_{K\to\infty}\sup_{h\in H}\sup_{|z|\ge K}|h(z)|=0.$$

Proof. The conditions are necessary by Dini's Theorem. To see that they are sufficient, choose $\varepsilon>0$. Now choose a K such that $\sup_{h\in H}\sup_{|z|\ge K}|h(z)|<\varepsilon$. In view of the Arzelà-Ascoli Theorem, see e.g. Billingsley (1968), this and the first condition show that the restrictions to $[-K,K]$ of the functions in H form a compact subset of $C([-K,K])$. Thus there are finitely many functions $h_1,\dots,h_m$ in H such that $\min_{1\le i\le m}\sup_{|z|\le K}|h(z)-h_i(z)|<\varepsilon$ for every $h\in H$. This shows that $\min_{1\le i\le m}\|h-h_i\|_\infty\le 3\varepsilon$. This in turn shows that H is totally bounded and hence compact. □

Now, by well-known properties of empirical processes, for every $\varepsilon>0$, there are finite constants N and M such that $P(\|\Delta_i\|_\infty>N)<\epsilon$ and $P(\sup_{|t|>M}|\Delta_i(t)|>\varepsilon)<\varepsilon$. These properties together with the bounds (2.2) and (2.3) give, in view of the above lemma, that the process $\mathbb{H}_{n,i}$ is tight. It is now easy to check that $\mathbb{H}_{n,1}+\cdots+\mathbb{H}_{n,m}$ converges in distribution in the space $C_0(\mathbb{R})$ to a centered Gaussian process. Thus, if we show (2.1), then we can conclude that $n^{1/2}(\hat g-g)$ converges in distribution to the same Gaussian process.

To obtain (2.1), we need smoothness of g, and appropriate kernels and bandwidths. For $r=1,2,\dots$ and $\alpha\in(0,1]$, let $\mathcal{G}_{r,\alpha}$ denote the set of all r-times differentiable functions whose r-th derivatives are Hölder of order α, and let $\mathcal{K}_{r,\alpha}$ denote the set of all uniformly continuous and integrable functions k such that $\int k(u)\,du=1$, $\int u^s k(u)\,du=0$ for $s=1,\dots,r$, and $\int |u|^{r+\alpha}|k(u)|\,du$ is finite. Let $\mathcal{A}$ be the set of integrable functions f for which there is an integrable function f' such that $f(y)=\int_{-\infty}^y f'(t)dt$, $t\in\mathbb{R}$. Such functions are absolutely continuous functions with integrable almost everywhere derivative.

THEOREM 1. *Let $g \in \mathcal{G}_{r,\alpha}$ for some r and α, and let $g_1, \dots, g_m \in \mathcal{A}$. Let $k_1, \dots, k_m \in \mathcal{K}_{r,\alpha}$. Assume the following rates on the bandwidths: $n(\min b_i)^2 \to \infty$ and $n(\max b_i)^{2(r+\alpha)} \to 0$. Then (2.1) holds. Moreover, the process $\mathbb{H}_{n,1}+\dots+\mathbb{H}_{n,m}$ and hence $n^{1/2}(\hat{g}-g)$ converges in distribution in the space $C_0(\mathbb{R})$ to a centered Gaussian process with covariance function*

$$\Gamma(s,t) = \operatorname{Cov}\Big(\sum_{i=1}^m g_i(s-u_i(X)), \sum_{i=1}^m g_i(t-u_i(X))\Big), \quad s,t \in \mathbb{R}. \tag{2.4}$$

PROOF. As a uniformly continuous integrable function, g belongs to $C_0(\mathbb{R})$. By assumption, the kernels k_i are uniformly continuous and integrable functions. Hence they also belong to $C_0(\mathbb{R})$. Thus the kernel estimators $\hat{f}_i$ are integrable elements of $C_0(\mathbb{R})$. Hence $\hat{g} = \hat{f}_1 * \dots * \hat{f}_m$ belongs to $C_0(\mathbb{R})$. Let $\bar{f}_i(y) = E\hat{f}_i(y)$. It is well-known that

$$E\|\hat{f}_i - \bar{f}_i\|_2^2 = E\int (\hat{f}_i(y) - \bar{f}_i(y))^2\,dy \le \frac{1}{nb_i}\int k_i^2(y)\,dy. \tag{2.5}$$

Note that k_i^2 is integrable since k_i is bounded and integrable.

Since $\bar{f}_i = f_i * k_{i,b_i}$ with $k_{i,b_i}(y) = k_i(y/b_i)/b_i$, we have $\bar{g} = \bar{f}_1 * \dots * \bar{f}_m = g * k_*$ with $k_* = k_{1,b_1} * \dots * k_{m,b_m}$ a member of $\mathcal{K}_{r,\alpha}$. Since $g \in \mathcal{G}_{r,\alpha}$, we obtain by a standard expansion that, with C the Hölder constant of the r-th derivative of g,

$$\begin{aligned}
\|\bar{g} - g\|_\infty &\le C\int |u|^{r+\alpha}|k_*(u)|\,du \\
&\le C\int\dots\int |u_1 + \dots + u_m|^{r+\alpha}|k_{1,b_1}(u_1)|\,du_1 \cdots |k_{m,b_m}(u_m)|\,du_m \\
&= O((\max b_i)^{r+\alpha}).
\end{aligned}$$

For a subset A of $\{1,\dots,m\}$, let $\gamma_A = (*_{i\in A}(\hat{f}_i - \bar{f}_i)) * (*_{i\notin A}\bar{f}_i)$, with the interpretation that $\gamma_\emptyset = \bar{g}$ and $\gamma_{\{1,\dots,m\}} = *_{i=1}^m(\hat{f}_i - \bar{f}_i)$. We have

$$\hat{g} = *_{i=1}^m \hat{f}_i = *_{i=1}^m(\bar{f}_i + (\hat{f}_i - \bar{f}_i)) = \sum_A \gamma_A = \sum_{r=0}^m \Gamma_r \tag{2.6}$$

with $\Gamma_r = \sum_{|A|=r}\gamma_A$. Note that $\Gamma_0 = \gamma_\emptyset = \bar{g}$ and $\Gamma_1 = \sum_{i=1}^m(\hat{f}_i - \bar{f}_i) * \bar{g}_i$ with $\bar{g}_i = *_{j\ne i}\bar{f}_j = g_i * (*_{j\ne i}k_{j,b_j})$. Thus we obtain

$$\Big\|\hat{g} - \bar{g} - \sum_{i=1}^m(\hat{f}_i - \bar{f}_i) * \bar{g}_i\Big\|_\infty \le \sum_{r=2}^m \|\Gamma_r\|_\infty.$$

We have $\|a*b\|_\infty \le \|a\|_2\|b\|_2$ and $\|a*b\|_\infty \le \|a\|_\infty\|b\|_1$. Hence for $|A| \ge 2$ and $i_1, i_2 \in A$,

$$\|\gamma_A\|_\infty \le \|\hat{f}_{i_1} - \bar{f}_{i_1}\|_2\|\hat{f}_{i_2} - \bar{f}_{i_2}\|_2 \prod_{i\in A\setminus\{i_1,i_2\}} \|\hat{f}_i - \bar{f}_i\|_1 \prod_{i\notin A}\|\bar{f}_i\|_1.$$

From this bound and (2.5) we obtain that

$$\sum_{r=2}^{m} \|\Gamma_r\|_\infty = O_p\Big(\frac{1}{n \min b_i}\Big).$$

Here we have also used that $\|\hat f_i - \bar f_i\|_1 \le \|\hat f_i\|_1 + \|\bar f_i\|_1$ and that $\|\hat f_i\|_1 \le \|k_{i,b_i}\|_1 = \|k_i\|_1$.

The desired result now follows if we show that

$$D_i = \|n^{1/2}(\hat f_i - \bar f_i) * \bar g_i - \mathbb{H}_{n,i}\|_\infty = o_p(1).$$

Let $\hat F_i$ and $\bar F_i$ denote the distribution functions of $\hat f_i$ and $\bar f_i$. Then $\hat F_i = \mathbb{F}_i * k_{i,b_i}$ and

$$n^{1/2}(\hat f_i - \bar f_i) * \bar g_i = n^{1/2}(\hat F_i - \bar F_i) * \bar g_i' = \Delta_i * k_{i,b_i} * \bar g_i' = \mathbb{H}_{n,i} * k_*.$$

Thus we obtain the bound $D_i \le \|\mathbb{H}_{n,i} * k_* - \mathbb{H}_{n,i}\|_\infty$. For every $y \in \mathbb{R}$, we have the bound

$$|\mathbb{H}_{n,i} * k_*(y) - \mathbb{H}_{n,i}(y)| \le \int |\mathbb{H}_{n,i}(y-u) - \mathbb{H}_{n,i}(y)||k_*(u)|\, du.$$

By distinguishing the cases $|u| \le \delta$ and $|u| > \delta$ we can bound the right-hand side by

$$\begin{aligned} &\sup_{|z-y|\le\delta} |\mathbb{H}_{n,i}(z) - \mathbb{H}_{n,i}(y)| \textstyle\int_{|u|\le\delta} |k_*(u)|\, du + 2\|\mathbb{H}_{n,i}\|_\infty \int_{|u|>\delta} |k_*(u)|\, du \\ &\le \sup_{|z-y|\le\delta} |\mathbb{H}_{n,i}(z) - \mathbb{H}_{n,i}(y)|\|k_*\|_1 + 2\|\mathbb{H}_{n,i}\|_\infty \delta^{-r-\alpha} \textstyle\int |u|^{r+\alpha}|k_*(u)|\, du. \end{aligned}$$

Since $\int |u|^{r+\alpha}|k_*(u)|\, du = O((\max b_i)^{r+\alpha})$, we obtain from this bound and the tightness of $\mathbb{H}_{n,i}$ that $\|\mathbb{H}_{n,i} * k_* - \mathbb{H}_{n,i}\|_\infty = o_p(1)$. This implies $D_i = o_p(1)$ and completes the proof. □

REMARK 1. The choice of bandwidths is possible only if the smoothness parameter $r+\alpha$ is greater than one. In this case we can choose $b_1 = \cdots = b_m \sim n^{-\beta}$ with $1/(2(r+\alpha)) < \beta < 1/2$.

REMARK 2. Let us now give sufficient conditions for the required properties of g in terms of the densities $f_1, \dots, f_m$. For this we will use that the convolution of a bounded function with an integrable function is bounded and uniformly continuous.

(1) If all the densities are in $\mathcal{A}$ and one of them has a bounded almost everywhere derivative, then $g \in \mathcal{G}_{m-1,1}$ and $g_1, \dots, g_m \in \mathcal{A}$, and we can choose $b_1 = \cdots = b_m \sim n^{-\beta}$ with $1/(2m) < \beta < 1/2$. In this case, the rate of the optimal bandwidth for a kernel estimator of a single f_i is $n^{-1/3}$. This means that for large m we can use a considerably over-smoothed estimator for f_i.

(2) If at least two of the densities belong to $\mathcal{A}$, and either one of their almost everywhere derivatives or one of the remaining densities is bounded, then $g_1, \dots, g_m \in \mathcal{A}$ and $g \in \mathcal{G}_{1,1}$. In this case we can choose $b_1 = \cdots = b_m \sim n^{-\beta}$ with $1/4 < \beta < 1/2$. Note that if $m > 2$ then some of the densities $f_1, \dots, f_m$ need not be smooth at all.

3. Convergence in L_1

We now give conditions under which $n^{1/2}(\hat{g}-g)$ converges in distribution in the space L_1 to a centered Gaussian process. We assume again that $g_1, \dots, g_m$ belong to $\mathcal{A}$. We proceed as in the previous section and show first that $\mathbb{H}_{n,1} + \cdots + \mathbb{H}_{n,m}$ converges in distribution in the space L_1 to a centered Gaussian process.

Since g_i' is integrable, $\mathbb{H}_{n,i}$ has integrable sample paths. Let us now show tightness of this process in L_1. For this we need the following characterization of compact sets in L_1, which is known as the Fréchet-Kolmogorov theorem, see Yosida (1980, p. 275).

LEMMA 2. *A closed subset H of L_1 is compact if and only if*

$$\sup_{h \in H} \|h\|_1 < \infty,$$

$$\lim_{\delta \downarrow 0} \sup_{|t|<\delta} \sup_{h \in H} \int |h(x-t) - h(x)|\, dx = 0,$$

$$\lim_{K \uparrow \infty} \sup_{h \in H} \int_{|x|>K} \int |h(x)|\, dx = 0.$$

Suppose that the functions $\psi_i = (1-F_i)^{1/2} F_i^{1/2}$ are integrable. Note that $\psi_i^2(z)$ is the second moment of $\Delta_i(z)$. Thus

$$E(\|\Delta_i\|_1) = \int E(|\Delta_i(z)|)\, dz \le \int \psi_i(z)\, dz < \infty. \tag{3.1}$$

This shows that Δ_i has almost surely integrable sample paths, and we may view Δ_i as an element of L_1. We also find that

$$E \int_{|z| \ge K} |\Delta_i(z)|\, dz \le \int_{|z|\ge K} \psi_i(z)\, dz \to 0 \quad \text{as } K \to \infty. \tag{3.2}$$

Moreover, for positive δ and finite K,

$$\|\mathbb{H}_{n,i}\|_1 \le \|g_i'\|_1 \|\Delta_i\|_1, \tag{3.3}$$

$$\sup_{|t|<\delta} \int |\mathbb{H}_{n,i}(z+t) - \mathbb{H}_{n,i}(z)|\, dz \le \sup_{|t|\le\delta} \int |g_i'(z+t) - g_i'(z)|\, dz\ \|\Delta_i\|_1, \tag{3.4}$$

$$\int_{|z|>2K} |\mathbb{H}_{n,i}(z)|\, dz \le \int_{|y|>K} |\Delta_i(y)|\, dy\ \|g_i'\|_1 + \int_{|z|>K} |g_i'(z)|\, dz\ \|\Delta_i\|_1. \tag{3.5}$$

By the integrability of g_i',

$$\sup_{|t|\le\delta} \int |g_i'(z+t) - g_i'(z)|\, dz \to 0 \quad \text{as } \delta \downarrow 0. \tag{3.6}$$

Applying (3.1), (3.2), and (3.6) to (3.3)–(3.5) and using Lemma 2, we see that $\mathbb{H}_{n,i}$ is tight in L_1. Consequently, $\mathbb{H}_{n,1}+\dots+\mathbb{H}_{n,m}$ converges in distribution in the space L_1 to a centered Gaussian process with covariance function (2.4).

Since we are now working in L_1, we need slightly different assumptions on g and $g_1,\dots,g_m$. For $r = 1,2,\dots,$ let $\mathcal{G}_r$ denote the set of all r-times differentiable functions whose r-th derivative belongs to $\mathcal{A}$.

THEOREM 2. *Let $(1-F_1)^{1/2}F_1^{1/2},\dots,(1-F_m)^{1/2}F_m^{1/2}$ be integrable. Let $g \in \mathcal{G}_r$ for some r, and let $g_1,\dots,g_m \in \mathcal{A}$. Let $k_1,\dots,k_m \in \mathcal{K}_{r,1}$. Assume the following rates on the bandwidths: $n(\min b_i)^2 \to \infty$ and $n(\max b_i)^{2(r+1)} \to 0$. Then*

$$\|n^{1/2}(\hat{g}-g) - (\mathbb{H}_{n,1}+\dots+\mathbb{H}_{n,m})\|_1 = o_p(1),$$

and $n^{1/2}(\hat{g}-g)$ converges in distribution in the space L_1 to a centered Gaussian process with covariance function given in (2.4).

PROOF. We begin by proving the following auxiliary result: If F is a distribution function such that $(1-F)^{1/2}F^{1/2}$ is integrable, then $\int |x|^{3/2}\,dF(x)$ is finite. Indeed, in this case,

$$\int_0^\infty x\,dF(x) = \int_0^\infty (1-F(x))\,dx \le \int_0^\infty (1-F(x))^{1/2}\,dx < \infty.$$

Thus $x(1-F(x)) \to 0$ as $x \to \infty$ and hence $x(1-F(x)) \le c$ for some for some $c > 0$ and all $x > 0$. Hence

$$\int_0^\infty x^{3/2}\,dF(x) = \frac{3}{2}\int_0^\infty x^{1/2}(1-F(x))\,dx \le \frac{3c^{1/2}}{2}\int_0^\infty (1-F(x))^{1/2}\,dx < \infty.$$

A similar argument yields $\int_{-\infty}^0 |x|^{3/2}\,dF(x) < \infty$.

Now we proceed as in the proof of Theorem 1. We use the same notation. By the above auxiliary result, f_i has finite moments of order 3/2. Thus, by Lemma 2 of Devroye (1992),

$$\|\hat{f}_i - \bar{f}_i\|_1 = O(n^{-1/2}b_i^{-1/2}).$$

Since $g \in \mathcal{G}_r$, we obtain by a standard expansion that

$$\|\bar{g}-g\|_1 \le \|g^{(r+1)}\|_1 \int |u|^{r+1}|k_*(u)|\,du = O((\max b_i)^{r+1}).$$

Using $\|a*b\|_1 \le \|a\|_1\|b\|_1$ we now get for subsets A of $\{1,\dots,m\}$ containing at least two elements i_1, i_2 that

$$\|\gamma_A\|_1 \le \|\hat{f}_{i_1}-\bar{f}_{i_1}\|_1\|\hat{f}_{i_2}-\bar{f}_{i_2}\|_1 \prod_{i\in A\setminus\{i_1,i_2\}} \|\hat{f}_i-\bar{f}_i\|_1 \prod_{i\notin A}\|\bar{f}_i\|_1.$$

From this bound and representation (2.6) we obtain

$$\Big\|\hat{g}-\bar{g}-\sum_{i=1}^m(\hat{f}_i-\bar{f}_i)*\bar{g}_i\Big\|_1 \le \sum_{r=2}^m \|\Gamma_r\|_1 = O_p\left(\frac{1}{n\min b_i}\right).$$

The desired result now follows if we show that

$$\|n^{1/2}(\hat{f}_i - \bar{f}_i) * g_i - \mathbb{H}_{n,i}\|_1 = \|\mathbb{H}_{n,i} * k_* - \mathbb{H}_{n,i}\|_1 = o_p(1). \tag{3.7}$$

For this we bound the left-hand side by

$$\begin{aligned}&\sup_{|t|\le\delta} \int |\mathbb{H}_{n,i}(y+t) - \mathbb{H}_{n,i}(y)|\, dy \int_{|u|\le\delta} |k_*(u)|\, du + 2\|\mathbb{H}_{n,i}\|_1 \int_{|u|>\delta} |k_*(u)|\, du \\ &\le \sup_{|t|\le\delta} \int |\mathbb{H}_{n,i}(y+t) - \mathbb{H}_{n,i}(y)|\, dy \|k_*\|_1 + 2\|\mathbb{H}_{n,i}\|_1 \delta^{r+1} \int |u|^{r+1} |k_*(u)|\, du.\end{aligned}$$

From this and tightness of $\mathbb{H}_{n,i}$ in L_1 it is easy to see that (3.7) holds. □

REMARK 3. A sufficient condition for integrability of $(1-F)^{1/2}F^{1/2}$ is that F has a finite moment of order greater than 2. This follows from the inequality

$$\int (1-F(x))^{1/2} F(x)^{1/2}\, dx \le \left(\int \frac{1}{(1+|x|^{1+\alpha})}\, dx \int (1+|x|^{1+\alpha})(1-F(x))F(x)\, dx \right)^{1/2}$$

and the fact that the last integral is finite for small enough $\alpha > 0$ by the moment assumption on F.

4. Computational Aspects

Our estimator can be written as

$$\hat{g}(y) = \frac{1}{n^m} \sum_{j_1=1}^{n} \cdots \sum_{j_m=1}^{n} k_*(y - u_1(X_{j_1}) - \cdots - u_m(X_{j_m})), \quad y \in \mathbb{R},$$

where k_* is the convolution of the kernels $k_{1,b_1}, \ldots, k_{m,b_m}$. In the special case that $b_1 = \cdots = b_m = b$, we have $k_*(x) = (k_1 * \cdots * k_m)(x/b)/b$. Replacing $k_1 * \cdots * k_m$ by a kernel K and setting $K_b(y) = K(y/b)/b$ we obtain the von Mises statistic

$$\hat{g}_v(y) = \frac{1}{n^m} \sum_{j_1=1}^{n} \cdots \sum_{j_m=1}^{n} K_b(y - u_1(X_{j_1}) - \cdots - u_m(X_{j_m})), \quad y \in \mathbb{R}.$$

A closely related estimator is the U-statistic

$$\hat{g}_u(y) = \frac{1}{(n)_m} \sum_{(j_1,\ldots,j_m)\in I_m^n} K_b(y - u_1(X_{j_1}) - \cdots - u_m(X_{j_m})), \quad y \in \mathbb{R},$$

with $I_m^n = \{(j_1, \ldots, j_m) \in \{1, \ldots, n\}^m : j_i \ne j_k \text{ if } i \ne k\}$, and $(n)_m = n!/(n-m)!$ the cardinality of I_m^n. This is a special case of the estimator considered by Frees (1994). If K is bounded and integrable, then

$$\int |\hat{g}_v(y) - \hat{g}_u(y)|\, dy \le 2\, \frac{n^m - (n)_m}{n^m} \int |K_b(y)|\, dy = O(n^{-1})$$

and

$$\sup_{y\in\mathbb{R}} |\hat{g}_v(y) - \hat{g}_u(y)| \le 2\, \frac{n^m - (n)_m}{n^m} \sup_{y\in\mathbb{R}} |K_b(y)| = O(b^{-1}n^{-1}).$$

Thus the estimators $\hat{g}_v$ and $\hat{g}_u$ are asymptotically equivalent for the two norms considered in the paper and for appropriate bandwidths. If K can be expressed as a convolution $K_1 * \cdots * K_m$, then $\hat{g}_v$ coincides with $\hat{g}$ upon taking $k_{i,b_i} = K_{i,b}$. Since $\mathcal{K}_{r,\alpha}$ is closed under convolutions, our results hold for these estimators $\hat{g}_v$ and $\hat{g}_u$. Since $\hat{g}_v$ and $\hat{g}_u$ are easy to compute, it is advantageous to work with these estimators and a *simple* kernel K that is expressible as a convolution of m kernels in the desired class $\mathcal{K}_{r,\alpha}$. For example, for $\mathcal{K}_{1,1}$ we can take K to be the standard normal density.

A possible way of selecting the bandwidth b is to minimize an estimator of the (scaled) integrated mean square error (IMSE)

$$\int nE(\hat{g}(y) - g(y))^2 \, dy.$$

The latter is approximately

$$\int \Big(nB^2(y) + \sum_{i=1}^m \operatorname{Var} g_i(y - u_i(X_i))\Big) \, dy,$$

where

$$B(y) = \int g(y - bt)K(t) \, dt - g(y).$$

We can estimate $B(y)$ by the von Mises statistic

$$\hat{B}(y) = \frac{1}{n^m} \sum_{j_1=1}^n \cdots \sum_{j_m=1}^n \tilde{K}_b(y - u_1(X_{j_1}) - \cdots - u_m(X_{j_m})),$$

where $\tilde{K}$ is a kernel with $\int \tilde{K}(t) \, dt = 0$ and $\int t^j \tilde{K}(t) \, dt = \int t^j K(t) \, dt$ for $j = 1, \ldots, s$ and some s greater than or equal to the order of K. This suggests estimating the IMSE by

$$\int \Big(n\hat{B}^2(y) + \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n (\hat{g}_i(y - u_i(X_j)) - \hat{g}(y))^2 \Big) \, dy,$$

where $\hat{g}_i$ is the von Mises estimator of g_i with kernel K_b, so that

$$\hat{g}_i(y - u_i(X_j)) = \frac{1}{n^{m-1}} \sum_{\mathbf{j}: j_i = j} K_b(y - u_1(X_{j_1}) - \cdots - u_m(X_{j_m})),$$

where the summation extends over indices $\mathbf{j} = (j_1, \ldots, j_m)$ such that $j_i = j$. To avoid calculating the above integral, we minimize instead

$$\sum_{l=1}^L \Big(n\hat{B}^2(y_l) + \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n (\hat{g}_i(y_l - u_i(X_j)) - \hat{g}(y_l))^2 \Big), \tag{4.1}$$

where $y_1 < \cdots < y_L$ are points in $\mathbb{R}$. This estimates the sum

$$\sum_{l=1}^L nE(\hat{g}(y_l) - g(y_l))^2. \tag{4.2}$$

We have performed a small simulation study to test this approach. We treat the case where the X_i are random variables with common density f, and take $m = 2$ and $u_1(x) = u_2(x) = x$. Then g is $f * f$, the convolution of f with itself. In the simulations we treat three densities f: the standard normal density: $f(x) = \phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$; a mixture of normal densities: $f(x) = (\phi(x) + \phi(x-1))/2$; and a Gamma density with shape parameter 2: $f(x) = x \exp(-x)\mathbf{1}_{(0,\infty)}(x)$.

We considered two kernels, one of order two and the other of order four. As order two kernel we took $K = \phi$ and matched it with $\tilde{K}$ given by $\tilde{K}(x) = (x^6 - 21x^4 + 105x^2 - 57)\phi(x)/48$ which shares the first six moments with $K = \phi$. In this case we selected the bandwidth as the minimizer of (4.1) over the grid 0.4, 0.6, 0.8, 1, 1.2. As order four kernel we chose $K(x) = (x^4 - 14x^2 + 27)\phi(x)/16$ and matched it with $\tilde{K}$ given by $\tilde{K}(x) = (-x^8 + 36x^6 - 354x^4 + 924x^2 - 297)\phi(x)/384$ which shares the first eight moments with K. In this case we selected the bandwidth as the minimizer of (4.1) over the grid 0.5, 0.75, 1, 1.25, 1.5. Both kernels can be expressed as convolutions $k * k$ of kernels k with the same order. For the order four kernel we can take $k(x) = (3 - x^2)\phi(x)/2$.

In the following table we give the asymptotic value of (4.2), which is $4 \sum_{l=1}^{L} \operatorname{Var} f(y_l - X_1)$, and the average of $\sum_{l=1}^{L} n(\hat{g}(y_l) - g(y_l))^2$ over 1000 simulated samples of sizes $n = 25$ and $n = 50$ for both kernels. For the standard normal density we took the y_l from the grid $-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2$; for the mixture density from the grid $-1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3$; and for the Gamma density from the grid 1, 1.75, 2.5, 3.25, 4, 4.75, 5.5, 6.25, 7.

		order two kernel		order four kernel	
density	asympt.	n=25	n= 50	n=25	n=50
normal	0.553	0.461	0.468	0.361	0.455
mixture	0.445	0.375	0.403	0.324	0.322
Gamma	0.471	0.332	0.386	0.292	0.329

The table shows that our choice of bandwidth works very well. Particularly for small sample sizes, the scaled IMSE is considerably smaller than the asymptotic value of (4.2). The order four kernel is noticeably better than the order two kernel. Simulations not presented here show that the higher order kernel has the additional advantage that the mean square error of the estimator is less sensitive to the choice of bandwidth in small sample sizes.

References

Abramson, I. and Goldstein, L. (1991). Efficient nonparametric testing by functional estimation. *J. Theoret. Probab.* **4,** 137–159.

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya Ser. A* **50,** 381–393.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23,** 11–29.

Devroye, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20,** 2037–2058.

Donoho, D. L. (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16,** 1390–1420.

Dudewicz, E. J. and van der Meulen, E. C. (1981). Entropy-based tests of uniformity. *J. Amer. Statist. Assoc.* **76,** 967–974.

Eggermont, P. P. B. and LaRiccia, V. N. (1999). Best asymptotic normality of the kernel density entropy estimator for smooth densities. *IEEE Trans. Inform. Theory* **45,** 1321–1326.

Eggermont, P. P. B. and LaRiccia, V. N. (2001). *Maximum Penalized Likelihood Estimation, Vol. I, Density Estimation*, Springer Series in Statistics, Springer-Verlag, New York.

Frees, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89,** 517–525.

Goldstein, L. and Khas'minskii, R. (1995). On efficient estimation of smooth functionals. *Theory Probab. Appl.* **40,** 151–156.

Goldstein, L. and Messer, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *Ann. Statist.* **20,** 1306–1328.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6,** 109–115.

Laurent, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24,** 659–681.

Mason, D. M. (2003). Representations for integral functionals of kernel density estimators. *Austrian J. Statist.* **32,** 131–142.

Rudin, W. (1974). *Real and Complex Analysis*, McGraw-Hill, New York.

Saavedra, A. and Cao, R. (1999). Rate of convergence of a convolution-type estimator of the marginal density of an MA(1) process. *Stochastic Process. Appl.* **80,** 129–155.

Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.* **28,** 799–815.

Schick, A. and Wefelmeyer, W. (2004a). Root n consistent and optimal density estimators for moving average processes. *Scand. J. Statist.* **31,** 63–78.

Schick, A. and Wefelmeyer, W. (2004b). Functional convergence and optimality of plug-in estimators for stationary densities of moving average processes. To appear in *Bernoulli.*

Yosida, K. (1980). *Functional Analysis*, Sixth Edition, Grundlehren der Mathematischen Wissenschaften 123, Springer-Verlag, Berlin.

Anton Schick, Binghamton University, Department of Mathematical Sciences, Binghamton, NY 13902-6000, USA

E-mail address: `anton@math.binghamton.edu`

Wolfgang Wefelmeyer, Universität zu Köln, Mathematisches Institut, Weyertal 86-90, 50931 Köln, Germany

E-mail address: `wefelm@math.uni-koeln.de`