On efficient estimation of densities for sums of squared observations

Anton Schick and Wolfgang Wefelmeyer

Abstract. Densities of functions of independent and identically distributed random observations can be estimated by a local U-statistic. Under an appropriate integrability condition, this estimator behaves asymptotically like an empirical estimator. In particular, it converges at the parametric rate. The integrability condition is rather restrictive. It fails for the sum of powers of two observations when the exponent is at least two. We have shown elsewhere that for exponent equal to two the rate of convergence slows down by a logarithmic factor on the support of the squared observation. Here we show that the estimator is efficient in the sense of Hájek and Le Cam. In particular, the convergence rate is optimal.

Key words: Local asymptotic normality, regular estimator, nonstandard convergence rates

1. Introduction

Suppose that $X_1, \dots, X_n$ are independent observations with density f. It is sometimes of interest to estimate the density p of a transformation $q(X_1, \dots, X_m)$ of m of these observations. Frees (1994) proposed as an estimator of $p(z)$ the local U-statistic

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

with $k_b(x) = k(x/b)/b$ for a kernel k and a bandwidth b. The $\sqrt{n}$-consistency of this estimator requires that the conditional density of $q(X_1, \dots, X_m)$ given X_i at z has a finite second moment for each $i = 1, \dots, m$. Under appropriate functional versions of these conditions, Schick and Wefelmeyer (2004, 2007) and Giné and Mason (2007) obtain even functional central limit theorems in various function spaces.

The research of Anton Schick was partially supported by NSF Grant DMS 0906551.

In many applications the integrability conditions do not hold. Consider transformations $q(X_1, X_2) = |X_1|^\nu + |X_2|^\nu$. Then $|X_1|$ has density

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

$|X_1|^\nu$ has density

$$g_\nu(y) = \beta y^{\beta-1} h(y^\beta)$$

with $\beta = 1/\nu$, and the conditional density of $|X_1|^\nu + |X_2|^\nu$ given $X_1 = x$ or $X_2 = x$ equals

$$g_\nu(y - |x|^\nu).$$

If h is bounded and $\nu < 2$, then $g_\nu(z - |X_1|^\nu)$ has a finite second moment. If $\nu \geq 2$, then $g_\nu(z - |X_1|^\nu)$ does not have a finite second moment if the one-sided limits $h(0+)$ and $g_\nu(z-)$ exist and are positive.

Schick and Wefelmeyer (2009a) obtained the following results for the local U-statistic

$$\hat{p}_b(z) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} k_b(z - |X_i|^\nu - |X_j|^\nu)$$

with continuously differentiable kernel k having support $[-1, 1]$ and a fixed positive z.

Theorem 1. *Let $\nu < 2$. Suppose the density h is of bounded variation and the right-hand limit $h(0+)$ at 0 is positive. Let $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{n}(\hat{p}_b(z) - p(z)) \xrightarrow{d} N\big(0, 4\operatorname{Var}(g_\nu(z - |X_1|^\nu))\big).$$

Theorem 2. *Let $\nu > 2$. Suppose h is of bounded variation and the right-hand limit $h(0+)$ and the left-hand limit $g_\nu(z-)$ are positive. Let $b \sim 1/n$. Then*

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

Theorem 3. *Let $\nu = 2$ and $g = g_2$. Suppose h is of bounded variation and the right-hand limit $h(0+)$ and the left-hand limit $g(z-)$ are positive. Let $b \sim \sqrt{\log n}/n$. Then*

$$\sqrt{\frac{n}{\log n}}\,\big(\hat{p}_b(z) - p(z)\big) \xrightarrow{d} N(0, h^2(0+)g(z-)).$$

In the situation of Theorem 1, it is easy to show we have local asymptotic normality and the estimator $\hat{p}_b$ is efficient at the parametric rate $\sqrt{n}$. In Section 2 we consider the situation of Theorem 3 and obtain efficiency of $\hat{p}_b(z)$ even though we now have the slower rate $\sqrt{n/\log n}$ of convergence. For this we exhibit a one-dimensional model that is locally asymptotically normal with a scale factor $1/\sqrt{n \log n}$ and is least favorable for estimating $p(z)$.

2. Result

In this section we recall the notion of local asymptotic normality for nonparametric models, and of regularity and efficiency for estimators of real-valued functionals; see Theorem 2 in Section 3.3 of Bickel et al. (1998). For each positive integer n, consider a family $\mathscr{P}_n = \{P_{n,f} : f \in \mathscr{F}\}$ of distributions on a measurable space $(\Omega_n, \mathscr{A}_n)$ with a common parameter set $\mathscr{F}$. Let ϕ be a function from $\mathscr{F}$ into $\mathbb{R}$. We are interested in efficient estimation of $\phi(f)$.

Let $f_{n,t}$, $t \in \mathbb{R}$, be a one-dimensional (local) submodel of $\mathscr{F}$, for which we have *local asymptotic normality* at $f_{n,0} = f$. This means the expansion

$$\Lambda_n(t) = \log \frac{dP_{n,f_{n,t}}}{dP_{n,f}} = tS_n - \frac{1}{2}t^2\sigma^2 + o_{P_{n,f}}(1) \tag{2.1}$$

holds for all $t \in \mathbb{R}$, some positive σ and random variables S_n which converge in distribution to σN, where N denotes a standard normal random variable. Let ϕ be differentiable at f in the sense that

$$a_n(\phi(f_{n,t}) - \phi(f)) \to t\gamma \tag{2.2}$$

for a sequence a_n tending to infinity and a non-zero real number γ.

We call an estimator $\hat{\phi}_n$ *regular* for ϕ at f with *limit* L, if L is a random variable such that

$$a_n(\hat{\phi}_n - \phi(f_{n,t})) \xrightarrow{d} L \quad \text{under } P_{n,f_{n,t}}$$

for each $t \in \mathbb{R}$. The convolution theorem says that L has the same distribution as $(\gamma/\sigma)N + M$ for some random variable M independent of N. Then $(\gamma/\sigma)N$ is more concentrated than L in symmetric intervals. This justifies calling the regular estimator $\hat{\phi}_n$ *efficient* for ϕ at f if L can be taken to be $(\gamma/\sigma)N$. An estimator $\hat{\phi}_n$ is regular and efficient if

$$a_n(\hat{\phi}_n - \phi(f)) = \frac{\gamma}{\sigma^2}S_n + o_{P_{n,f}}(1). \tag{2.3}$$

If such an estimator exists, we call $f_{n,t}$, $t \in \mathbb{R}$, *least favorable* for ϕ at f.

Here we have independent observations $X_1, \dots, X_n$ with density f under a measure P_f and take $P_{n,f}$ to be the restriction of P_f to the sigma-field generated by $X_1, \dots, X_n$. The parameter set $\mathscr{F}$ consists of all densities f which have bounded variation and for which $y \mapsto (f(y) + f(-y))\mathbf{1}[y > 0]$ has positive right-hand limit at 0 and positive left-hand limit at $\sqrt{z}$.

We write h for the density of $|X_1|$ and g for the density of X_1^2. As local model at f we take

$$f_{n,t}(x) = f(x)\big(1 + c_n t\mathbf{1}[|t| \le d_n]\chi_n(x)\big), \tag{2.4}$$

where $c_n = 1/\sqrt{n \log n}$, $d_n = (\log n)^{5/4}/(2\sup_{x\in\mathbb{R}} f(x))$, and

$$\chi_n(x) = \mathbf{1}_{(r_n,z]}(z - x^2)g(z - x^2) - \mu_n$$

with $r_n = (\log n)^{3/2}/n$ and

$$\mu_n = \int \mathbf{1}_{(r_n,z]}(z-u^2)g(z-u^2)f(u)\,du \to p(z).$$

We now check that $f_{n,t}$ defines a local model in $\mathscr{F}$. Since $\int \chi_n(x)f(x)\,dx$ equals zero and $c_n d_n|\chi_n(x)|$ is bounded by $1/2$, we see that $f_{n,t}$ is a density. It is easy to verify that $f_{n,t}$ has bounded variation. The limit requirements follow from those of f, the bound $c_n d_n|\chi_n(x)| \le 1/2$, and from

$$\chi_n(0+) = g(z-) - \mu_n = \frac{h(\sqrt{z}-)}{2\sqrt{z}} - \mu_n \quad \text{and} \quad \chi_n(\sqrt{z}-) = -\mu_n.$$

Thus $f_{n,t}$ belongs to $\mathscr{F}$ and defines a local model at $f_{n,0} = f$.

Next we verify local asymptotic normality. We have

$$\Lambda_n(t) = \sum_{j=1}^n \log(1 + t\mathbf{1}[|t| \le d_n]Z_{nj})$$

with

$$Z_{nj} = c_n\chi_n(X_j) = c_n\mathbf{1}_{(r_n,z]}(z-X_j^2)g(z-X_j^2) - c_n\mu_n.$$

We have

$$E[Z_{nj}] = 0, \quad j = 1,\dots,n,$$

$$\max_{1\le j\le n}|Z_{nj}| \le \frac{\sup_{x\in\mathbb{R}} f(x)}{(\log n)^{5/4}},$$

$$\begin{aligned}\sum_{j=1}^n E[Z_{nj}^2] &= \frac{1}{\log n}\int \chi_n^2(x)f(x)\,dx \\ &= \frac{1}{\log n}\left(\int_0^{z-r_n} g^2(z-y)g(y)\,dy - \mu_n^2\right) \\ &= \frac{1}{\log n}\left(\int_0^{z-r_n} \frac{h^2(\sqrt{z-y})g(y)}{4(z-y)}\,dy - \mu_n^2\right)\end{aligned}$$

and thus

$$\sum_{j=1}^n E[Z_{nj}^2] \to \sigma^2 = \frac{1}{4}h^2(0+)g(z-).$$

It follows that $\{Z_{nj} : j = 1,\dots,n\}$ is a Lindeberg array. Therefore we have

$$\sum_{j=1}^n Z_{nj} \xrightarrow{d} \sigma N$$

and

$$\sum_{j=1}^n Z_{nj}^2 = \sigma^2 + o_{P_{n,f}}(1).$$

In view of the inequality

$$\Big|\log(1+x) - x + \frac{x^2}{2}\Big| = \Big|\int_0^x \frac{y^2}{1+y}\,dy\Big| \le \frac{|x|^3}{3(1-|x|)}, \quad |x| < 1,$$

we obtain local asymptotic normality (2.1) as follows.

THEOREM 4. *For the local model (2.4) we have the stochastic expansion*

$$\log \frac{dP_{n,f_{n,t}}}{dP_{n,f}} = t\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\chi_n(X_j) - \frac{1}{2}t^2\sigma^2 + o_{P_{n,f}}(1)$$

with $\sigma^2 = \frac{1}{4}h^2(0+)g(z-)$.

If X has density $f_{n,t}$, then $|X|$ has density

$$h_{n,t}(y) = h(y)(1 + c_n t\mathbf{1}[|t| \le d_n]\chi_n(y)), \quad y \in \mathbb{R},$$

and X^2 has density

$$g_{n,t}(y) = g(y)(1 + c_n t\mathbf{1}[|t| \le d_n]\chi_n(\sqrt{y})), \quad y \in \mathbb{R}.$$

Here the functional is $\phi(f) = g * g(z) = p(z)$. We verify (2.2) with $a_n = \sqrt{n/\log n}$. The left-hand side of (2.2) is

$$\sqrt{\frac{n}{\log n}}\big(g_{n,t} * g_{n,t}(z) - g * g(z)\big)$$

and eventually equals

$$\frac{1}{\log n}\big(2t(\tilde{\chi}_n g) * g(z) + t^2 c_n(\tilde{\chi}_n g) * (\tilde{\chi}_n g)(z)\big),$$

where

$$\tilde{\chi}_n(y) = \chi_n(\sqrt{y}) = \mathbf{1}_{(r_n,z]}(z-y)g(z-y) - \mu_n.$$

It is easy to see that we have

$$\begin{aligned}(\tilde{\chi}_n g) * g(z) &= \int g(z-y)\tilde{\chi}_n(y)g(y)\,dy \\ &= \int_0^{z-r_n} g^2(z-y)g(y)\,dy - \mu_n p(z) \\ &= \log n\frac{1}{4}h^2(0+)g(z-) + o(\log n)\end{aligned}$$

and, with $\psi_n = \mathbf{1}_{(r_n,z)}g$,

$$\begin{aligned}(\tilde{\chi}_n g) * (\tilde{\chi}_n g)(z) &= \int g(z-y)(\psi_n(y) - \mu_n)g(y)(\psi_n(z-y) - \mu_n)\,dy \\ &= \psi_n^2 * \psi_n^2(z) - 2\mu_n\psi_n^2 * g(z) + p(z)\mu_n^2 \\ &= O\left(\int_{r_n}^{z-r_n}\frac{dy}{(z-y)y} + \int_{r_n}^{z}\frac{dy}{y\sqrt{z-y}}\right) \\ &= O(\log n).\end{aligned}$$

This shows that (2.2) holds with $\gamma = 2\sigma^2$.

Our estimator for $\phi(f) = p(z)$ is

$$\hat{p}_b(z) = \frac{2}{n(n-1)}\sum_{1\le i\le j\le n} k_b(z - X_i^2 - X_j^2)$$

with $b \sim \sqrt{\log n}/n$. Schick and Wefelmeyer (2009b) have shown the following stochastic expansion for $\hat{p}_b(z)$.

THEOREM 5.
$$\sqrt{\frac{n}{\log n}}\big(\hat{p}_b(z) - p(z)\big) = \frac{2}{\sqrt{n \log n}} \sum_{j=1}^{n} \chi_n(X_j) + o_{P_{n,f}}(1).$$

It follows from Theorem 5 and the characterization (2.3) with $a_n = \sqrt{n/\log n}$, $\gamma = 2\sigma^2$, $\sigma^2 = \frac{1}{4}h^2(0+)g(z-)$ and $S_n = \sum_{j=1}^{n} \chi_n(X_j)$ that $\hat{p}_b(z)$ is regular and efficient for $p(z)$.

References

[1] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

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

[3] Giné, E. and Mason, D. (2007). On local U-statistic processes and the estimation of densities of functions of several variables. *Ann. Statist.* **35**, 1104–1145.

[4] Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.* **16**, 925–935.

[5] Schick, A. and Wefelmeyer, W. (2007). Root-n consistent density estimators of convolutions in weighted L_1-norms. *J. Statist. Plann. Inference* **37**, 1765–1774.

[6] Schick, A. and Wefelmeyer, W. (2009a). Convergence rates of density estimators for sums of powers of observations. *Metrika* **69**, 249–264.

[7] Schick, A. and Wefelmeyer, W. (2009b). Non-standard behavior of density estimators for sums of squared observations. *Statist. Decisions* **27**, 55–73.

ANTON SCHICK, DEPARTMENT OF MATHEMATICAL SCIENCES, BINGHAMTON UNIVERSITY, BINGHAMTON, NY 13902-6000, USA

WOLFGANG WEFELMEYER, MATHEMATICAL INSTITUTE, UNIVERSITY OF COLOGNE, WEYERTAL 86–90, 50931 COLOGNE, GERMANY