Pointwise convergence rates and central limit theorems for kernel density estimators in linear processes

Anton Schick [*]
Binghamton University

Wolfgang Wefelmeyer
Universität zu Köln

Abstract

Convergence rates and central limit theorems for kernel estimators of the stationary density of a linear process have been obtained under the assumption that the innovation density is smooth (Lipschitz). We show that smoothness is not required. For example, it suffices that the innovation density has bounded variation.

Keywords. Lipschitz continuity, martingale approximation, smoothness of convolutions.

1 Introduction

Consider a linear process $X_t = \sum_{s=0}^{\infty} a_s \varepsilon_{t-s}$ with independent and identically distributed (i.i.d.) innovations ε_t that have mean zero, finite variance, and density f. We assume that $a_0 = 1$ and that the coefficients are summable, $\sum_{s=0}^{\infty} |a_s| < \infty$. Then X_t has a stationary density g. It can be estimated by the kernel estimator

$$\hat{g}(x) = \frac{1}{n} \sum_{j=1}^{n} k_b(x - X_j),$$

where $k_b(v) = k(v/b)/b$ with k a square-integrable kernel (i.e. a function such that $\int k(v)\, dv = 1$ and $\int k^2(v)\, dv < \infty$), and b is a bandwidth such that $b \to 0$ and $nb \to \infty$. Pointwise and uniform convergence rates have been studied by several authors, see for example Hall and Hart (1990), Tran (1992), Hallin and Tran (1996), Lu (2001), Wu and Mielniczuk (2002), and Bryk and Mielniczuk (2005).

Wu and Mielniczuk (2002) prove the following result: If f is Lipschitz, then $\sqrt{nb}(\hat{g}(x) - E[\hat{g}(x)])$ is asymptotically normal with mean zero and variance $\sigma^2(x) = g(x) \int k^2(v)\, dv$. We prove in Section 2 that this result can be obtained under weaker assumptions on f. In particular, it suffices that f is of bounded variation, except in the degenerate case with $a_s = 0$ for all $s \geq 1$. This exceptional case corresponds to the i.i.d. case, for which sufficient conditions are already known, see Parzen (1962). In the latter case moment assumptions are not needed. Instead Parzen requires continuity of $f = g$ at x and a kernel that is bounded and satisfies $|vk(v)| \to 0$ as $|v| \to \infty$.

---

2 Results

We consider observations $X_1, \ldots, X_n$ of a linear process with the following properties.

Assumption 1. *The process X_t, $t \in \mathbb{Z}$, has the representation $X_t = \sum_{s=0}^{\infty} a_s \varepsilon_{t-s}$ with i.i.d. innovations ε_t that have mean zero, finite variance, and density f. The coefficients a_s, $s \in \mathbb{N}$, are summable, $a_0 = 1$, and N is positive, where*

$$N = \sum_{s=1}^{\infty} \mathbf{1}(a_s \neq 0)$$

denotes the number of non-zero coefficients a_s, $s \geq 1$. The kernel k is square-integrable and the bandwidth satisfies $b \to 0$ and $nb \to \infty$.

We restrict ourselves to the case $N > 0$, since $N = 0$ corresponds to the i.i.d. case already treated by Parzen (1962).

Let g denote the stationary density and $\hat{g}$ the kernel estimator described above. First we obtain the above-mentioned conclusion of Theorem 1 in Wu and Mielniczuk (2002) under weaker assumptions. We do this by generalizing their approach. They express $\hat{g}(x) - E[\hat{g}(x)] = \hat{g}(x) - g * k_b(x)$ as a sum $D_0 + R_0$, where

$$D_0 = \frac{1}{n} \sum_{j=1}^{n} \left(k_b(x - X_j) - f * k_b(x - Y_{j,1}) \right)$$

is a martingale, and

$$R_0 = \frac{1}{n} \sum_{j=1}^{n} \left(f * k_b(x - Y_{j,1}) - g * k_b(x) \right).$$

Here we use the notation

$$Y_{j,i} = \sum_{s=i}^{\infty} a_s \varepsilon_{j-s}.$$

Then they show in their Lemma 2 that $\sqrt{nb} D_0 \Rightarrow N(0, \sigma^2(x))$ under the assumption that f is Lipschitz. In their proof, this assumption is used only to guarantee continuity of g at x, boundedness of f and hence g, and to prove

(2.1) $$\int k^2(v) |\bar{H}(bv) - \bar{H}(0)| \, dv = o_p(1),$$

where

$$\bar{H}(z) = \frac{1}{n} \sum_{j=1}^{n} \left(f(x - Y_{j,1} - z) - g(x - z) \right).$$

Assume now that f is bounded. This guarantees that g is bounded and uniformly continuous. Indeed, we can express g as the convolution of f and the density g_1 of $Y_{1,1}$ and obtain $g(x) \leq \|f\|_\infty \|g_1\|_1$ and

$$|g(x - t) - g(x)| \leq \|f\|_\infty \int |g_1(y - t) - g_1(y)| \, dy$$

2

for all $x \in \mathbb{R}$. Uniform continuity of g now follows from the continuity of translations in L_1, see e.g. Rudin (1974, Theorem 9.5). In view of these properties of g and the square-integrability of k, property (2.1) follows from

$$(2.2) \qquad \int k^2(v) \Big| \frac{1}{n} \sum_{j=1}^{n} \big(f(x - Y_{j,1} - bv) - f(x - Y_{j,1})\big) \Big| \, dv = o_p(1).$$

The expected value of the left-hand side of (2.2) is bounded by

$$\int k^2(v) \int |f(x - y - bv) - f(x - y)| g_1(y) \, dy \, dv$$
$$\leq \|g_1\|_\infty \int k^2(v) \int |f(t - bv) - f(t)| \, dt \, dv.$$

Since $\int |f(t - bv) - f(t)| \, dt$ is bounded by 2 and tends to zero by the continuity of translations in L_1, the Lebesgue dominated convergence theorem gives that

$$\int k^2(v) \int |f(t - bv) - f(t)| \, dt \, dv \to 0.$$

Thus the conclusion of their Lemma 2 remains true if f is bounded.

Wu and Mielniczuk (2002) also use Lipschitz continuity of f to conclude that $\sqrt{n} R_0 = O_p(1)$. To avoid this assumption, we express R_0 as

$$R_0 = \sum_{i-1}^{m} D_i + R_m,$$

where

$$D_i = \frac{1}{n} \sum_{j=1}^{n} \big(f_{i-1} * k_b(x - Y_{j,i}) - f_i * k_b(x - Y_{j,i+1})\big),$$

$$R_m = \frac{1}{n} \sum_{j=1}^{n} \big(f_m * k_b(x - Y_{j,m+1}) - g * k_b(x)\big).$$

Here f_i is the density of $\sum_{s=0}^{i} a_s \varepsilon_s$. Then $f_{i+1}(x) = \int f_i(x - a_{i+1}y) f(y) \, dy$ and $f_0 = f$. For $i = 1, 2, \ldots$, the sequence D_i is a square-integrable martingale with

$$nE[D_i^2] \leq E[(f_{i-1} * k_b)^2 (x - Y_{1,i})]$$
$$\leq \|f_{i-1} * k_b\|_\infty E[f_{i-1} * |k_b|(x - Y_{1,i})]$$
$$\leq \|f_{i-1}\|_\infty \int |k(v)| \, dv \, g * |k_b|(x)$$
$$\leq \Big(\|f\|_\infty \int |k(v)| \, dv\Big)^2.$$

Thus, if f is bounded, then $\sqrt{n} D_i = O_p(1)$.

Now consider R_m. If $a_s = 0$ for all $s > m$, then $Y_{j,m+1} = 0$ and thus $R_m = 0$. Now assume that there are infinitely many a_s that are not zero, and that f_m is Lipschitz. Then we can

3

proceed as in the proofs of Lemma 3 (with our f_m replacing their f_1) and Theorem 1 of Wu and Mielniczuk (2002) to conclude that $\sqrt{n}R_m = O_p(1)$. We arrive at the following result.

Theorem 1. *Suppose Assumption 1 holds and f is bounded. If $N = \infty$, assume also that f_m is Lipschitz for some m. Then*

$$\sqrt{nb}(\hat{g}(x) - g * k_b(x)) \Rightarrow N(0, \sigma^2(x)),$$

where $\sigma^2(x) = g(x) \int k^2(v)\, dv$.

Remark 1. A function h is L_1-*Lipschitz* if

$$\int |f(x - t) - f(x)|\, dx \le L|t|, \quad t \in \mathbb{R},$$

for some constant L. If $N = \infty$, the requirement that f_m is Lipschitz for some m can be replaced by the requirement that f_r is L_1-Lipschitz for some r. To see this, let $m = \inf\{s > r : a_s \ne 0\}$. Then $f_m(x) = \int f_r(x - a_m y)f(y)\, dy$. Since f is assumed bounded, we have

$$
\begin{aligned}
|f_m(x + t) - f_m(x)| &\le \int |f_r(x + t - a_m z) - f_r(x - a_m z)| f(z)\, dz \\
&\le \|f\|_\infty \int |f_r(x + t - a_m z) - f_r(x - a_m z)|\, dz \\
&\le \|f\|_\infty \frac{1}{|a_m|} \int |f_r(u + t) - f_r(u)|\, du \\
&\le \|f\|_\infty \frac{1}{|a_m|} L|t|.
\end{aligned}
$$

The requirement that a density is L_1-Lipschitz is rather mild: it is met by all densities of bounded variation. For details see Lemma 8 in Schick and Wefelmeyer (2005). Thus, if f is of bounded variation, we have the following result.

Corollary 1. *Suppose Assumption 1 holds and f is of bounded variation. Then*

$$\sqrt{nb}(\hat{g}(x) - g * k_b(x)) \Rightarrow N(0, \sigma^2(x)).$$

So far we have looked at the variance term $\hat{g}(x) - g * k_b(x)$. The bias term $g * k_b(x) - g(x)$ can be made small by choice of b and k if g is sufficiently smooth. Smoothness of g does not require smoothness of f if $N > 0$. For example, if f is of bounded variation and $N \ge m > 0$, then g has $m - 1$ bounded derivatives, and $g^{(m-1)}$ is Lipschitz. This follows by induction from Lemma 1 below. Thus we have the following theorem.

Theorem 2. *Suppose Assumption 1 holds. If f is of bounded variation, a kernel of order m is used, and $N \ge m$, then*

$$\hat{g}(x) - g(x) = O_p(n^{-1/2}b^{-1/2}) + O(b^m).$$

Remark 2. Suppose N is known to be infinite and f has bounded variation. Then we can control the rate $O(b^m)$ of the bias by choosing a kernel of high order m. A choice of bandwidth $b \sim n^{-1/(2m+1)}$ yields the rate $\hat{g}(x) - g(x) = O_p(n^{-m/(2m+1)})$. Thus we can achieve a rate close to the parametric rate $n^{-1/2}$.

3 An Auxiliary Result

Lemma 1. *Let p and q be integrable functions with q of bounded variation and p bounded and L_1-Lipschitz. Then $p * q$ is Lipschitz and has an almost everywhere derivative that is bounded, integrable, and L_1-Lipschitz.*

Proof. The proof is similar to the proof of Lemma 8 in Schick and Wefelmeyer (2005). Since q is integrable and of bounded variation, we can write q as difference of two bounded nondecreasing functions that vanish at $-\infty$. Without loss of generality we may assume that these functions are right-continuous, as this changes q only on a countable set and hence does not affect $p * q$. Thus we may assume that

$$q(x) = \mu_1(-\infty, x] - \mu_2(-\infty, x], \quad x \in \mathbb{R},$$

for two finite measures μ_1 and μ_2. Let $\mu = \mu_1 + \mu_2$ and set $h = h_1 - h_2$, where

$$h_i(x) = \int p(x - y)\mu_i(dy), \quad x \in \mathbb{R}, \quad i = 1, 2.$$

Note that h is integrable and bounded since

$$\|h_i\|_1 \leq \iint |p(x - y)|\mu_i(dy)\, dx \leq \|p\|_1 \mu_i(\mathbb{R}) \quad \text{and} \quad \|h_i\|_\infty \leq \|p\|_\infty \mu_i(\mathbb{R}), \quad i = 1, 2.$$

We have

$$\int_{-\infty}^{z} h_i(x)\, dx = \int \int_{x \leq z} p(x - y)\, dx\, \mu_i(dy)$$

$$= \int \int_{u \leq z - y} p(u)\, du\, \mu_i(dy)$$

$$= \int \mu_i(-\infty, z - u] p(u)\, du.$$

This shows that $p * q$ is absolutely continuous with almost everywhere derivative h. As h is bounded, $p * q$ is Lipschitz. In view of the inequality

$$\int |h(x + t) - h(x)|\, dx \leq \iint |p(x + t - y) - p(x - y)|\, dx\, \mu(dy)$$

$$\leq \mu(\mathbb{R}) \int |p(x + t) - p(x)|\, dx,$$

the function h inherits the L_1-Lipschitz property from p. $\qquad\square$

References

Bryk, A. and Mielniczuk, J. (2005), Asymptotic properties of density estimates for linear processes: application of projection method, *J. Nonparametr. Stat.* **17**, 121–133.

Hall, P. and Hart, J. D. (1990), Convergence rates in density estimation for data from infinite-order moving average processes, *Probab. Theory Related Fields* **87**, 253–274.

Hallin, M. and Tran, L. T. (1996), Kernel density estimation for linear processes: Asymptotic normality and optimal bandwidth derivation, *Ann. Inst. Statist. Math.* **48**, 429–449.

Lu, Z. (2001), Asymptotic normality of kernel density estimators under dependence, *Ann. Inst. Statist. Math.* **53**, 447–468.

Parzen, E. (1962), On estimation of a probability density function and mode, *Ann. Math. Statist.* **33**, 1065–1076.

Rudin, W. (1974), *Real and Complex Analysis* (McGraw-Hill, New York, 2nd ed.).

Schick, A. and Wefelmeyer, W. (2005), Root-n consistent density estimators of convolutions in weighted L_1-norms, Technical Report, Department of Mathematical Sciences, Binghamton University.

Tran, L. T. (1992), Kernel density estimation for linear processes, *Stochastic Process. Appl.* **41**, 281–296.

Wu, W. B. and Mielniczuk, J. (2002), Kernel density estimation for linear processes, *Ann. Statist.* **30**, 1441–1459.

ANTON SCHICK
DEPARTMENT OF MATHEMATICAL SCIENCES
BINGHAMTON UNIVERSITY
BINGHAMTON, NY 13902-6000
USA
ANTON@MATH.BINGHAMTON.EDU
HTTP://MATH.BINGHAMTON.EDU/ANTON/

WOLFGANG WEFELMEYER
MATHEMATISCHES INSTITUT
UNIVERSITÄT ZU KÖLN
50931 KÖLN
GERMANY
WEFELM@MATH.UNI-KOELN.DE
HTTP://WWW.MI.UNI-KOELN.DE/∼WEFELM/