Density estimators for invertible linear processes

Wolfgang Wefelmeyer
Mathematical Institute
University of Cologne

jointly with
Anton Schick (Binghamton University)

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

We present two (classes of) results on estimating the stationary density of linear time series:

1. Convergence rates of ordinary kernel density estimators for (possibly non-invertible) linear time series.

2. Parametric convergence rates of convolution estimators (or local U-statistics) for invertible linear time series.

Let $X_1, \ldots, X_n$ be observations of a linear process

$$X_j = \varepsilon_j + \sum_{s=1}^{\infty} \varphi_s \varepsilon_{j-s}.$$

Assume that the innovations ε_j have mean zero, finite variance, and density f, and that the coefficients φ_s are summable.

1. **Kernel density estimator.** An estimator for the density h of X_j at x is the kernel estimator

$$\hat{h}(x) = \frac{1}{n} \sum_{j=1}^{n} k_b(x - X_j) \quad \text{with} \quad k_b(x) = k(x/b)/b.$$

To obtain pointwise rates, Wu and Mielniczuk (2002) write $X_j = \varepsilon_j + Z_j$ and add and subtract $E(k_b(x - X_j)|Z_j) = k_b * f(Z_j)$:

$$\begin{aligned} \hat{h}(x) - h(x) = {} & \frac{1}{n} \sum_{j=1}^{n} \Big(k_b(x - X_j) - k_b * f(x - Z_j) \Big) \\ & + \frac{1}{n} \sum_{j=1}^{n} \Big(k_b * f(x - Z_j) - k_b * h(x) \Big) + k_b * h(x) - h(x). \end{aligned}$$

The first term is a martingale. The second term is a centered and *smoothed* kernel estimator. The third term is the ordinary bias term. — S/W (2006) refine the approach of Wu and Mielniczuk (2002), weaken their assumptions, and give conditions in terms of the innovation density only. S/W (2008) give L_1-rates.

We refine the approach of Wu and Mielniczuk (2002) as follows. The more coefficients φ_s are nonzero, the smoother is h. We exploit this by decomposing

$$X_j = Y_j + Z_j \quad \text{with} \quad Y_j = \varepsilon_j + \sum_{s=1}^{m-1} \varphi_s \varepsilon_{j-s}, \qquad Z_j = \sum_{s=m}^{\infty} \varphi_s \varepsilon_{j-s}.$$

With f_m denoting the density of Y_j, we have

$$\begin{aligned} \hat{h}(x) - h(x) =& \frac{1}{n} \sum_{j=1}^{n} \Big(k_b(x - X_j) - k_b * f_m(x - Z_j) \Big) \\ &+ \frac{1}{n} \sum_{j=1}^{n} \Big(k_b * f_m(x - Z_j) - k_b * h(x) \Big) + k_b * h(x) - h(x). \end{aligned}$$

For $m = 1$ this is the approach of Wu and Mielniczuk (2002).
If only finitely many φ_s are nonzero, the middle term *vanishes* for m large.

We show: If (the shift of) f is L_1-Lipschitz (e.g. if f has bounded variation), then f_m has an a.e. derivative of order $m-1$ which is L_1-Lipschitz (if $\varphi_1, \ldots, \varphi_{m-1}$ are nonzero.)

Result: *Assume that at least N coefficients φ_s are nonzero. Let f have finite variation. Choose a kernel of order $N+1$. Then*

$$\|\widehat{h} - h\|_1 = O_P(n^{-1/2} b^{-1/2}) + O(b^{N+1}).$$

To construct $\widehat{h}$:

Test for number of nonzero coefficients.
Then choose optimal bandwidth and order of kernel.
If many φ_s are nonzero, the rate of $\widehat{h}$ is close to $n^{-1/2}$.
(Not so good if the first few coefficients φ_s are zero.)

2. **Convolution estimator.** Suppose at least one φ_s is nonzero. Then $X_j = \varepsilon_j + Z_j$, so h has the convolution representation $f * g$, where h, f, g are the densities of X_j, ε_j, Z_j.

A better estimator for h than the ordinary kernel estimator would be given by a *convolution* of density estimators for f and g.
(The reason is that such a convolution is approximately the sum of two smoothed *empirical* estimators, as we will see. They converge at the $\sqrt{n}$ rate.)

But ε_j and Z_j are not observed and must be estimated. This is why we now need that the linear process is invertible.

To estimate ε_j, we assume that the linear process is *invertible*. This means that the innovations have a moving average representation in terms of the realizations of the process,

$$\varepsilon_j = X_j - \sum_{s=1}^{\infty} \varrho_s X_{j-s}.$$

Let $\hat{\varrho}_j$ be an estimator of ϱ_j. Let $p \to \infty$. Estimate ε_j and Z_j by

$$\hat{\varepsilon}_j = X_j - \sum_{s=1}^{p} \hat{\varrho}_s X_{j-s} \quad \text{and} \quad \hat{Z}_j = X_j - \hat{\varepsilon}_j.$$

Estimate the densities f of ε_j and g of Z_j by kernel estimators

$$\hat{f}(x) = \frac{1}{n-p} \sum_{j=p+1}^{n} k_b(x - \hat{\varepsilon}_j), \quad \hat{g}(x) = \frac{1}{n-p} \sum_{j=p+1}^{n} k_b(x - \hat{Z}_j).$$

Then estimate $h = f * g$ by the *convolution estimator* $\hat{h} = \hat{f} * \hat{g}$.

Result: *For appropriate choice of kernel and bandwidth, under (mild) conditions on the decay of φ_s and ϱ_s, if f has a moment > 3 and (essentially) finite variation, then $\hat{h} = \hat{f} * \hat{g}$ has the stochastic expansion*

$$\left\| \hat{h} - h - \mathbb{F} - \mathbb{G} + \sum_{s=1}^{p} (\hat{\varrho}_s - \varrho_s)\nu_s' \right\|_1 = o_P(n^{-1/2})$$

with

$$\mathbb{F}(x) = \frac{1}{n-p} \sum_{j=p+1}^{n} \Big(f(x - Z_j) - h(x) \Big),$$

$$\mathbb{G}(x) = \frac{1}{n-p} \sum_{j=p+1}^{n} \Big(g(x - \varepsilon_j) - h(x) \Big),$$

and $\nu_s(x) = E[X_0 f(x - Z_s)]$.

If $\hat{\varrho}_s$ are asymptotically linear (e.g. least squares estimators), then $n^{1/2}(\hat{h} - h)$ converges weakly in L_1 to a centered Gausian process.

The estimator $\widehat{h} = \widehat{f} * \widehat{g}$ is approximated by a sum of two smoothed empirical estimators:

$$\widehat{f} * \widehat{g} - f * g = f * (\widehat{g} - g) + g * (\widehat{f} - f) + (\widehat{f} - f) * (\widehat{g} - g).$$

This explains the rate $n^{-1/2}$ if the bandwidth is e.g. $(n \log n)^{-1/4}$.

The estimator $\widehat{f} * \widehat{g}(x)$ is equivalent to a local U-statistic:

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_b(x - \widehat{\varepsilon}_i - \widehat{Z}_j).$$

The estimator $\widehat{f} * \widehat{g}$ can be improved in several ways:
— Use efficient estimators for ϱ_s.
— Use empirical likelihood weights on $\widehat{f}$ and $\widehat{g}$ to exploit $E[\varepsilon_j] = 0$,
— Use the convolution representation of $Z_j = X_j - \varepsilon_j = \sum_{s=1}^{\infty} \varphi_s \varepsilon_{j-s}$ to estimate h by an increasing number $q \to \infty$ of convolutions,

$$\widehat{h}(x) = \int \cdots \int \widehat{f}\Big(x - \widehat{\varphi}_1 z_1 - \cdots - \widehat{\varphi}_q z_q\Big) \widehat{f}(z_1) \cdots \widehat{f}(z_q)\, dz_1 \cdots dz_q.$$