ADAPTIVE DENSITY ESTIMATION IN CONTINUOUS TIME

Delphine Blanke
L.S.T.A. (University Paris 6)
December 2002

- Estimation with known parameters : r_0, γ_0
- Simple adaptive procedure : r_0 known, γ_0 unknown
- Double adaptive procedure : r_0, γ_0 both unknown

DENSITY - DISCRETE TIME

$(X_t,\, t \in \mathbf{Z})$ $\mathbb{R}^d$-valued strongly mixing process where all X_t's have same density f. Observations : $X_1, \ldots, X_n$.

Regularity condition

$f \in H_r^d$, f defined on $\mathbb{R}^d$, r-times differentiable,

$$\left| \frac{\partial^{(r)} f}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}}(x) - \frac{\partial^{(r)} f}{\partial y_1^{j_1} \ldots \partial y_d^{j_d}}(y) \right| \leq \ell \|x - y\|^{\lambda}$$

with $j_1 + \cdots + j_d = r$.

Almost sure convergence

$$\limsup_{n \to \infty} \left(\frac{n}{\ln n} \right)^{\frac{r}{2r+d}} |f_n(x) - f(x)| < +\infty \quad a.s.$$

Kernel estimator

$$f_n(x) = \frac{1}{n h_n^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n} \right), \quad x \in \mathbb{R}^d$$

$h_n \sim \left(\ln n / n \right)^{\frac{1}{2r+d}}$, K kernel depending on r_0.

DENSITY - CONTINUOUS TIME

$\{X_t,\, t \in \mathbb{R}\}$: $\mathbb{R}^d$-valued measurable process such that all X_t's have same density f with sample paths displayed over $[0,\, T]$, $T \to +\infty$.

Some references

Castellana and Leadbetter (86), Bosq (97, 98), Bosq and Davydov (98), Bosq and al. (99), Kutoyants (95, 96, 97), Leblanc (95, 97), Van Zanten (02), Sköld and Hössjer (99), Sköld (01), Blanke and Bosq (97, 00), Blanke and Pumo (03)...

Estimator

$$f_T(x) = \frac{1}{Th_T^d} \int_0^T K\left(\frac{x - X_t}{h_T}\right) \, \mathrm{d}t, \quad x \in \mathbb{R}^d$$

Assumptions A

(i) $f \in H^d_{r_0}$, K as before,

(ii) f bounded and $f_{(X_s,X_t)} = f_{(X_0,X_{t-s})}$, $t > s$;

(iii) $f_{(X_0,X_u)}(y,z) \leq M(y,z)u^{-\gamma_0}$, $u \in]0, u_0[$, $u_0 > 0$, $\gamma_0 > 0$, $(y,z) \in \mathcal{V}^2$;

(iv) $\sup\limits_{(y,z)\in\mathcal{V}^2} \int_{u_0}^{+\infty} |g_u(y,z)|\, \mathrm{d}u < +\infty$

with $g_u := f_{(X_0,X_u)} - f \otimes f$.

(v) $\{X_t, \quad t \in \mathbb{R}\}$ GSM process.

Remark :

Set $Y_u := (\frac{X_u^{(1)} - X_0^{(1)}}{u^{\gamma_1}}, \ldots, \frac{X_u^{(d)} - X_0^{(d)}}{u^{\gamma_d}})$, with $0 < \gamma_i \leq 1$. If the joint density of (X_0, Y_u) is defined near $u = 0$ and bounded by $M(y,z)$, then

$$\sup_{(y,z)\in\mathcal{V}^2} f_{(X_0,X_u)}(y,z) \leq M(y,z)u^{-\sum_{i=1}^d \gamma_i}.$$

Hence for $\gamma_0 = \sum_{i=1}^d \gamma_i$, $A(iii)$ will be true (and $\gamma_0 \in]0, d]$).

'CL' : $\sup_{x,y} \int_0^{+\infty} |g_u(x,y)|\, \mathrm{d}u < +\infty$

Theorem 1 *Under Assumptions A, we obtain*

(1) if $\gamma_0 < 1$, $r_0 > d$, $h_{T_n} = O\Big(\big(\ln T_n / T_n\big)^{1/2r_0}\Big)$,

$$\limsup_{T_n \uparrow +\infty} \sqrt{\frac{T_n}{\ln T_n}}\, |f_{T_n}(x) - f(x)| < +\infty \ a.s.;$$

(2) if $\gamma_0 = 1$, $r_0 > d$, $h_{T_n} = O\Big(\big((\ln T_n)^2 / T_n\big)^{1/2r_0}\Big)$,

$$\limsup_{T_n \uparrow +\infty} \frac{\sqrt{T_n}}{\ln T_n}\, |f_{T_n}(x) - f(x)| < +\infty \ a.s.;$$

(3) if $\gamma_0 > 1$, $r_0 > d/\gamma_0$, $h_{T_n} \sim \left(\frac{\ln T_n}{T_n}\right)^{\frac{\gamma_0}{2r_0\gamma_0 + d(\gamma_0 - 1)}}$,

$$\limsup_{T_n \uparrow +\infty} \left(\frac{T_n}{\ln T_n}\right)^{\frac{r_0\gamma_0}{2r_0\gamma_0 + d(\gamma_0 - 1)}} |f_{T_n}(x) - f(x)| < +\infty \ a.s.$$

(4) if $r_0 \leq d/\max(1, \gamma_0)$, $h_{T_n} \sim \left(\frac{(\ln T_n)^2}{T_n}\right)^{\frac{1}{r_0 + d}}$

$$\limsup_{T_n \uparrow +\infty} \left(\frac{T_n}{(\ln T_n)^2}\right)^{\frac{r_0}{r_0 + d}} |f_{T_n}(x) - f(x)| < +\infty \ a.s.$$

Some extensions

• With additional conditions upon kernel K, T_n should be replaced by T in Theorem 1 ($\rightsquigarrow$ use of Borel-Cantelli lemma in continuous-time established in Bosq, 98).

• By strengthening Assumptions A, one may obtain uniform almost sure convergence of the kernel estimator (increasing compacts, whole of $\mathbb{R}^d$).

Some examples

• $d = 1$, $r_0 > 1$, $(X_t) =$ O.U. $\rightsquigarrow$ a.s. rate : $\sqrt{\frac{\ln T}{T}}$.

• $d = 1$, $r_0 > 1$, (X_t) mean-square differentiable

$d = 2$, $r_0 > 2$, $(X_t^{(1)}, X_t^{(2)})$ with $(X_t^{(i)})$ O.U.

$\rightsquigarrow$ a.s. rate : $\sqrt{\frac{(\ln T)^2}{T}}$.

• $d = 3$, $r_0 > 2$, $(X_t^{(i)})$ O.U. $\rightsquigarrow$ a.s. rate : $\left(\frac{\ln T}{T}\right)^{\frac{r_0}{2r_0+1}}$.

SIMPLE ADAPTIVE PROCEDURE

Goal : Perform an adaptive estimator with respect to parameter γ_0 ($\in]0, d]$) in the case of known r_0.

Some references

• Lepski (90) : i.i.d. case, Gaussian white noise model

• Efromovich (85) : i.i.d. case, density on [0,1], L_2 Sobolev classes.

• Golubev (92) : i.i.d. case, density on the real line, L_2 Sobolev classes.

• Butucea (00, 01) : i.i.d. case, exact adaptive density estimator, L_p Sobolev classes.

• Comte and Merlevède (02) : continuous time mixing processes, under C.L.'s conditions, relative to regularity of density in Besov spaces.

Grid for possible values of γ_0

$$\Gamma_{T_n} = \{\widetilde{\gamma}_0, 1, \gamma_{1,T_n}, \ldots, \gamma_{N_T,T_n}, d\}$$

where

$$\begin{cases} 0 < \widetilde{\gamma}_0 < 1, \ \gamma_{0,T_n} := 1, \\ \gamma_{j+1,T_n} - \gamma_{j,T_n} = \delta_{T_n}, \qquad j = 0, \ldots, N_{T_n} - 1, \\ d - \gamma_{N_T,T_n} \leq \delta_{T_n}. \end{cases}$$

Pointwise pseudo-estimator of γ_0: γ_0^*

$$\gamma_0^* = \min \{\gamma_1 \in \Gamma_{T_n} : \ \forall \gamma_2 \in \Gamma_{T_n} \ \gamma_2 \geq \gamma_1, \\ \Psi_{T_n}(\gamma_2) \left| \widehat{f}_{\gamma_1}(x) - \widehat{f}_{\gamma_2}(x) \right| \leq \eta, \ \}$$

where

$$\widehat{f}_\gamma(x) = \frac{1}{T_n h_{T_n}^d(\gamma)} \int_0^{T_n} K\left(\frac{x - X_t}{h_{T_n}(\gamma)}\right) \, \mathrm{d}t$$

with

$$h_{T_n}(\gamma) = \begin{cases} \left((\ln T_n)^4 / T_n\right)^{1/2d} & \text{if } \ \gamma < 1, \\ \left((\ln T_n)^2 / T_n\right)^{1/2r_0} & \text{if } \ \gamma = 1, \\ \left(\ln T_n / T_n\right)^{\frac{\gamma}{2r_0\gamma + d(\gamma-1)}} & \text{if } \ \gamma > 1. \end{cases}$$

and

$$\Psi_{T_n}(\gamma) = \begin{cases} (T_n/\ln T_n)^{1/2} & \text{if } \gamma < 1, \\ T_n^{1/2}/\ln T_n & \text{if } \gamma = 1, \\ (T_n/\ln T_n)^{\frac{r_0\gamma}{2r_0\gamma + d(\gamma-1)}} & \text{if } \gamma > 1. \end{cases}$$

Adaptive estimation for f at a fixed point $x \in \mathbb{R}^d$ follows with

$$\widehat{f}_{\gamma_0^*}(x) = \frac{1}{T_n h_{T_n}^d(\gamma_0^*)} \int_0^{T_n} K\left(\frac{x - X_t}{h_{T_n}(\gamma_0^*)}\right) \mathrm{d}t.$$

Remark: γ_0^* may not be a good estimator for γ_0 since :

$*$ $\widetilde{\gamma}_0$ stands for all values of γ_0 when $\gamma_0 \in]0, 1]$;

$*$ in some cases, γ_0^* 'undersmooth' the true regularity γ_0.

$\rightsquigarrow$ see e.g. Blanke (02) for related works upon estimation of holder index of continuous time processes.

Theorem 2 *Under Assumptions A, we obtain for* $\gamma_0 \in]0, d]$, $\eta > 2\,|b_{r_0}(x)|$ *and* $\delta_{T_n} \to 0$ *such that* $\delta_{T_n} \geq \delta_0 \frac{\ln \ln T_n}{\ln T_n}$ *with* $\delta_0 > 2r_0/d$

(1) if $r_0 > d/\max(1, \gamma_0)$,

$$\limsup_{T_n \uparrow +\infty} \Psi_{T_n}(\gamma_0) \left| \widehat{f}_{\gamma_0^*}(x) - f(x) \right| < +\infty \quad \text{a.s.}$$

(2) if $r_0 \leq d/\max(1, \gamma_0)$,

$$\limsup_{T_n \uparrow +\infty} \left(\frac{T_n}{(\ln T_n)^2} \right)^{\frac{r_0}{r_0 + d}} \left| \widehat{f}_{\gamma_0^*}(x) - f(x) \right| < +\infty \quad \text{a.s.}$$

Some extensions

• Uniform pseudo-estimator $\gamma^*_{0,\infty,a}$

$$\gamma^*_{0,\infty,a} = \min\Bigg\{\gamma_1 \in \Gamma_{T_n} : \forall \gamma_2 \in \Gamma_{T_n}, \gamma_2 \geq \gamma_1, \Psi_{T_n}(\gamma_2) \sup_{\|x\|_* \leq c_{(d)} T_n^a} \left|\widehat{f}_{\gamma_1}(x) - \widehat{f}_{\gamma_2}(x)\right| \leq \eta\Bigg\}$$

uniform results over compacts ($a = 0$), increasing compacts ($a > 0$) and whole of $\mathbb{R}^d$ ($a > 2$) for $\widehat{f}_{\gamma^*_{0,\infty,a}}(x)$ given by

$$\widehat{f}_{\gamma^*_{0,\infty,a}}(x) = \frac{1}{T_n h^d_{T_n}(\gamma^*_{0,\infty,a})} \int_0^{T_n} K\left(\frac{x - X_t}{h_{T_n}(\gamma^*_{0,\infty,a})}\right) \mathrm{d}t.$$

• Minimal threshold η depend on $|b_{r_0}(x)|$, an unknown quantity, relative to the asymptotic value of bias. Results are still true with η replaced by $\widehat{\eta} = (2 + \varepsilon)\left|\widehat{b}_{r_0}(x)\right|$, with $\varepsilon > 0$ and $\widehat{b}_{r_0}(x)$ is a 'rough' estimator of $b_{r_0}(x)$.

Sketch of proof - Case $r_0 > d/\max(1, \gamma_0)$

Decomposition:

$$\Psi_{T_n}(\gamma_0)\left|\widehat{f}_{\gamma_0^*}(x) - f(x)\right| \leq$$

$$\Psi_{T_n}(\gamma_0)\left|\widehat{f}_{\gamma_0^*}(x) - \widehat{f}_{\bar{\gamma}_0}(x)\right| + \Psi_{T_n}(\gamma_0)\left|\widehat{f}_{\bar{\gamma}_0}(x) - f(x)\right|$$

First term may be written as

$$\Psi_{T_n}(\gamma_0)\left|\widehat{f}_{\gamma_0^*}(x) - \widehat{f}_{\bar{\gamma}_0}(x)\right| \mathbb{I}_{\left\{\gamma_0^* \leq \bar{\gamma}_0\right\}}$$
$$+ \Psi_{T_n}(\gamma_0)\left|\widehat{f}_{\gamma_0^*}(x) - \widehat{f}_{\bar{\gamma}_0}(x)\right| \mathbb{I}_{\left\{\gamma_0^* > \bar{\gamma}_0\right\}}.$$

Definition of γ_0^* implies

$$\Psi_{T_n}(\gamma_0)\left|\widehat{f}_{\gamma_0^*}(x) - \widehat{f}_{\bar{\gamma}_0}(x)\right| \mathbb{I}_{\left\{\gamma_0^* \leq \bar{\gamma}_0\right\}} \leq \eta.$$

Finally,

$$\sum_n P\left(\Psi_{T_n}(\gamma_0)\left|\widehat{f}_{\gamma_0^*}(x)-\widehat{f}_{\bar{\gamma}_0}(x)\right|\mathbb{1}_{\{\gamma_0^*>\bar{\gamma}_0\}}>\varepsilon\right)$$

$$\leq \sum_n P\left(\gamma_0^*>\bar{\gamma}_0\right)$$

$$\leq \sum_n P\left(\bigcup_{j=j_0+1}^{\text{Card}(\Gamma_{T_n})}\Psi_{T_n}(\gamma_j)\left|\widehat{f}_{\bar{\gamma}_0}(x)-\widehat{f}_{\gamma_j}(x)\right|>\eta\right)$$

$$\leq \sum_n \sum_{j=j_0+1}^{\text{Card}(\Gamma_{T_n})} P\left(\Psi_{T_n}(\gamma_j)\left|\widehat{f}_{\bar{\gamma}_0}(x)-\widehat{f}_{\gamma_j}(x)\right|>\eta\right)$$

$\hookrightarrow$ result follows with exponential type inequalities for $\left|\widehat{f}_\gamma(x)-\mathrm{E}\,\widehat{f}_\gamma(x)\right|$ and η large enough.

DOUBLE ADAPTIVE PROCEDURE

Goal : Perform an adaptive estimator with respect to both parameters r_0 and γ_0.

2 grids : Γ_{T_n} and $\Delta_{T_n} = \{1, 2, \ldots, r_{T_n}\}$

Adaptive procedure in 2 steps

• Determine copulas $\big(r, \gamma(r)\big)$, $r \in \Delta_{T_n}$ and with $\gamma(r)$ given by

$$\gamma(r) = \min\Bigg\{\gamma : \ \forall \bar{\gamma} \in \Gamma_{T_n}, \ \bar{\gamma} \geq \gamma, \\ \Psi(r, \bar{\gamma}) \left| \widehat{f}_{(r,\bar{\gamma})}(x) - \widehat{f}_{(r,\gamma)}(x) \right| \leq \widehat{\eta_1}(r) \Bigg\}$$

where

$$\widehat{f}_{(r,\gamma)}(x) = \frac{1}{T_n h_{T_n}^d(r,\gamma)} \int_0^T K_{(r)}\left(\frac{x - X_t}{h_{T_n}(r,\gamma)}\right) \mathrm{d}t.$$

• Over previous copulas $(r, \gamma(r))$, compute again an adaptive procedure with respect to r:

$$r_0^* = \max\Big\{ r : \ \forall \bar{r} \in \Delta_{T_n}, \ \bar{r} \leq r,$$

$$\Psi(\bar{r}, \gamma(\bar{r})) \left| \widehat{f}_{(\bar{r},\gamma(\bar{r}))}(x) - \widehat{f}_{(r,\gamma(r))}(x) \right| \leq \widehat{\eta}_2(\bar{r}) \Big\}.$$

Proposed estimator :

$$\widehat{f}_{(r_0^*,\gamma(r_0^*))}(x) =$$

$$\frac{1}{T_n h_{T_n}^d (r_0^*, \gamma(r_0^*))} \int_0^{T_n} K\left(\frac{x - X_t}{h_{T_n}(r_0^*, \gamma(r_0^*))} \right) \mathrm{d}t$$