Least squares type estimation of the transition of a particular hidden Markov chain

Claire Lacour

MAP5 - University Paris 5

Model

$$Y_i = X_i + \varepsilon_i \qquad i = 1, \dots, n+1$$

with

- $(X_i)_{i \geq 0}$ homogeneous real-valued Markov chain, unknown
- ε_i i.i.d. noise, independent of (X_i). known

Transition kernel $P(x, A) = P(X_{i+1} \in A | X_i = x)$.
$\longrightarrow$ transition density $\pi(x, y)$.

Aim

Estimation of π from data $Y_1, Y_2, \dots, Y_{n+1}$.

deconvolution + regression

Assumptions

Let q_ε be the density of ε_i and q_ε^* its Fourier transform.

A1 $\forall x \in \mathbb{R}\ |q_\varepsilon^*(x)| \geq k_0(x^2+1)^{-\gamma/2}$ ordinary smooth noise

A2 (X_i) irreducible positive recurrent.

A3 (X_i) stationary with unknown density f.

A4 (X_i) geometrically β-mixing.

We estimate π on a compact set $A = A_1 \times A_2$.

A5 $\iint_{(x,y)\in A} |\pi(x,y)|^2 dxdy = \|\pi\|_A^2 < \infty$.

A6 $\forall x \in A_1 \quad 0 < f_0 \leq f(x) \leq \|f\|_\infty < \infty$.

Examples of such chains

- Diffusion processes : $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ we consider $(X_{i\Delta})_{1\le i\le n}$ under the assumptions of Pardoux and Veretennikov (2001).
- Nonlinear AR(1) processes : $X_n = \varphi(X_{n-1}) + \varepsilon_{X_{n-1},n}$ under the assumptions of Mokkadem (1987).
- ARX (1,1) models : $X_n = F(X_{n-1}, Z_n) + \xi_n$
- ARCH models : $X_{n+1} = F(X_n) + G(X_n)\varepsilon_{n+1}$

First idea

$$\pi(x,y) = \frac{F(x,y)}{f(x)} = \frac{\text{density of } (X_i, X_{i+1})}{\text{density of } X_i}$$

$\longrightarrow$ estimator $\hat{\pi} = \hat{F}/\hat{f}$.

Result in this framework

$$\mathbb{E}\|\pi - \hat{\pi}\|^2 \leq C_1 \mathbb{E}\|F - \hat{F}\|^2 + C_2 \mathbb{E}\|f - \hat{f}\|^2 + o(\frac{1}{n})$$

depends on the regularity of f and the regularity of F.

Problem

f can be less regular than π

$\longrightarrow$ non-optimal rate of convergence

Heuristic

- For all function G

$$G(X_{i+1}) = (\int \pi(., y)G(y)dy)(X_i) + \eta_i$$

where $\eta_i = G(X_{i+1}) - \mathbb{E}[G(X_{i+1})|X_i]$.

Heuristic

- For all function G

$$G(X_{i+1}) = (\int \pi(.,y)G(y)dy)(X_i) + \eta_i$$

where $\eta_i = G(X_{i+1}) - \mathbb{E}[G(X_{i+1})|X_i]$.

- Contrast to estimate $\int \pi(.,y)G(y)dy$:

$$\gamma_n(u) = \frac{1}{n}\sum_{i=1}^{n}[u^2(X_i) - 2u(X_i)G(X_{i+1})]$$

Let $t(x,y) = u(x)G(y)$. If $\int G^2 = 1$, $u = \int t(.,y)G(y)dy$ and the contrast becomes

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}[\int t^2(X_i,y)dy - 2t(X_i,X_{i+1})]$$

Problem

The X_i are not available, only the $Y_i = X_i + \varepsilon_i$ are known

Solution

$$\text{Find } \begin{cases} V_t \text{ such that } \mathbb{E}[V_t(Y_i, Y_{i+1})] = \mathbb{E}[t(X_i, X_{i+1})] \\ Q_t \text{ such that } \mathbb{E}[Q_t(Y_i)] = \mathbb{E}[\int t(X_i, y)dy] \end{cases}$$

We find $V_t^*(u, v) = \dfrac{t^*(u, v)}{q_\varepsilon^*(-u)q_\varepsilon^*(-v)}$ $\quad$ t^* Fourier transform of t

and $Q_t^*(u) = V_t^*(u, 0)$

Calculation of V_t

F_X density of (X_i, X_{i+1}), F_Y density of (Y_i, Y_{i+1})

$$F_Y = F_X * (q_\varepsilon \otimes q_\varepsilon) \Rightarrow F_Y^* = F_X^*(q_\varepsilon^* \otimes q_\varepsilon^*)$$

$q_\varepsilon \otimes q_\varepsilon : (x, y) \mapsto q_\varepsilon(x) q_\varepsilon(y)$

$$\mathbb{E}[t(X_i, X_{i+1})] = \iint t F_X = \frac{1}{2\pi} \iint t^* \overline{F_X^*} = \frac{1}{2\pi} \iint t^* \frac{\overline{F_Y^*}}{\overline{q_\varepsilon^*} \otimes \overline{q_\varepsilon^*}}$$
$$= \frac{1}{2\pi} \iint V_t^* \overline{F_Y^*} = \iint V_t F_Y = \mathbb{E}[V_t(Y_i, Y_{i+1})]$$

Then $V_t^* = \dfrac{t^*}{\overline{q_\varepsilon^*} \otimes \overline{q_\varepsilon^*}}$

A new contrast

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}[\int t^2(X_i, y)dy \quad -2 \quad t(X_i, X_{i+1})]$$

$$\int t^2(X_i, y)dy \downarrow Q_{t^2}(Y_i) \qquad t(X_i, X_{i+1}) \downarrow V_t(Y_i, Y_{i+1})$$

Definition

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}[Q_{t^2}(Y_i) - 2V_t(Y_i, Y_{i+1})].$$

- $\mathbb{E}[\gamma_n(t)] = \|t-\pi\|_f^2 - \|\pi\|_f^2$ where $\|t\|_f^2 = \int t^2(x,y)f(x)dxdy$.
- For S an approximation space, we want now to define

$$\hat{\pi} = \arg\min_{t \in S} \gamma_n(t).$$

Minimization of the contrast (1/2)

If $\hat{\pi}(x,y) = \sum \hat{a}_\lambda \psi_\lambda(x,y)$, $(\psi_\lambda)_\lambda$ basis of S

$$\gamma_n(\hat{\pi}) = \min_{t \in S} \gamma_n(t) \Leftrightarrow G\hat{A} = Z$$

with
$$\begin{cases} G = \left(\frac{1}{n} \sum_{i=1}^{n} Q_{\psi_\lambda \psi_\mu}(Y_i) \right)_{\lambda,\mu} \\ Z = \left(\frac{1}{n} \sum_{i=1}^{n} V_{\psi_\lambda}(Y_i, Y_{i+1}) \right)_{\lambda} \\ \hat{A} = (\hat{a}_\lambda)_\lambda . \end{cases}$$

Problem

G is not necessarily invertible.

Minimization of the contrast (2/2)

We introduce the set

$$\Gamma = \{\min \mathrm{Sp}(G) \geq (2/3) f_0\}$$

On Γ, G is invertible and $\gamma_n(t)$ can be minimized.

Definition

$$\hat{\pi}_m = \begin{cases} \arg\min_{t \in S_m} \gamma_n(t) & \textit{on } \Gamma \\ 0 & \textit{on } \Gamma^c \end{cases}$$

Remark : f_0 can be replaced by an estimator $\hat{f}_0$

Projection spaces

We take for S_m a wavelet space on $A_1 \times A_2$

Orthonormal wavelets with regularity $r > 2\gamma + 3/2$

γ regularity of the noise

The dimension of S_m is $D_m^2 \asymp 2^{2m}$

Risk

Let π_m be the orthogonal projection of π on S_m.

$$\mathbb{E}\|\pi - \hat{\pi}_m\|^2 = \underbrace{\|\pi - \pi_m\|^2}_{\text{bias}} + \underbrace{\mathbb{E}\|\pi_m - \hat{\pi}_m\|^2}_{\text{variance}}$$

Bias Depends on the regularity α of π

$$\|\pi - \pi_m\| \leq C D_m^{-\alpha}$$

Variance Depends on the regularity γ of the noise

$$\mathbb{E}\|\pi_m - \hat{\pi}_m\|^2 \leq C \frac{D_m^{4\gamma+2}}{n}$$

Model selection

Aim : to select automatically the m which performs the best bias-variance trade-off.

Method

$$\hat{m} = \arg\min_{m \in \mathcal{M}_n} \{\gamma_n(\hat{\pi}_m) + \text{pen}(m)\}$$

with pen a penalty function

and $\mathcal{M}_n = \{m \geq 1, \quad D_m^{4\gamma+2}/n \leq 1\}$

Estimator $\tilde{\pi} = \begin{cases} \hat{\pi}_{\hat{m}} & \text{if } \|\hat{\pi}_{\hat{m}}\| \leq \sqrt{n} \\ 0 & \text{else} \end{cases}$

Theorem

If $\gamma > 3/4$ *and* $\text{pen}(m) = K_0 f_0^{-1} \frac{D_m^{4\gamma+2}}{n}$

$$\mathbb{E}\|(\tilde{\pi} - \pi)\mathbb{1}_A\|^2 \leq C \inf_{m \in \mathcal{M}_n} \Big(\quad d^2(\pi, S_m) \quad + \text{pen}(m) \Big) + \frac{C'}{n}.$$

Non-asymptotic.

Remarks on the penalty :

- K_0 depends only on k_0, γ (noise) and r (approximation space)
- f_0 can be replaced by an estimator $\hat{f}_0$
- no mixing term

Theorem

If $\gamma > 3/4$ *and* $\text{pen}(m) = K_0 f_0^{-1} \frac{D_m^{4\gamma+2}}{n}$

$$\mathbb{E}\|(\tilde{\pi} - \pi)\mathbb{1}_A\|^2 \leq C \inf_{m \in \mathcal{M}_n} \Big(\underbrace{d^2(\pi, S_m)}_{D_m^{-2\alpha}} + \underbrace{\text{pen}(m)}_{D_m^{4\gamma+2}/n} \Big) + \frac{C'}{n}.$$

Non-asymptotic.

Remarks on the penalty :

- K_0 depends only on k_0, γ (noise) and r (approximation space)
- f_0 can be replaced by an estimator $\hat{f}_0$
- no mixing term

Corollary

We suppose that π belongs to the Besov space $B^{\alpha}_{2,\infty}$. Then,

$$\mathbb{E}\|(\tilde{\pi}-\pi)\mathbb{1}_A\|^2 = O(n^{-\frac{2\alpha}{2\alpha+4\gamma+2}}).$$

- adaptive estimator
- does not depend on the regularity of f
- no logarithmic loss