The coupling method for estimating β–mixing coefficients of Markov processes.

Oleg Butkovsky

(joint work with Alexander Yu. Veretennikov)

Department of Probability Theory
Moscow State University

11 April 2011

Two questions

Let X_n, $n \in \mathbb{Z}_+$ be a homogeneous Markov process. Suppose that X_n has a stationary distribution.

Two questions

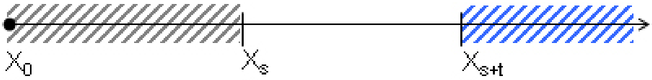
Let X_n, $n \in \mathbb{Z}_+$ be a homogeneous Markov process. Suppose that X_n has a stationary distribution.

- How fast is convergence to stationary regime?
- How fast is "mixing"?

Motivation

Mixing

- Sometimes it might be necessary to establish Central Limit Theorem (CLT) for dependent random variables.
- We can describe the dependence in terms of mixing coefficients.



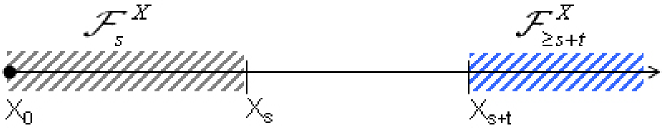
- Roughly speaking, if the dependence between events separated by large number of time steps is "weak", then CLT is satisfied.

Mixing

Let $\mathcal{F}_I^X$ be a σ-field generated by random variables $\{X_s,\ s \in I\}$.

$$\alpha(t) := \sup_{s \geqslant 0} \sup_{\substack{A \in \mathcal{F}^X_{\geqslant s+t} \\ B \in \mathcal{F}^X_s}} |\mathrm{P}(AB) - \mathrm{P}(A)\mathrm{P}(B)|,$$

$$\beta(t) := \sup_{s \geqslant 0} \mathsf{E} \sup_{A \in \mathcal{F}^X_{\geqslant s+t}} |\mathrm{P}(A|\mathcal{F}^X_s) - \mathrm{P}(A)|.$$



It is clear that $\alpha(t) = 0$ and $\beta(t) = 0$ for independent $X_0, X_1, \ldots$. In general case, if these coefficients are "sufficiently small", then under simple moment conditions the CLT holds.

Mixing and CLT

Theorem (I.A. Ibragimov). *Let* $(X_n)_{n\in\mathbb{Z}_+}$ *be a stationary process such that:*

1. $\mathsf{E}\,X_0 = 0$, $\mathsf{E}\,|X_0|^{2+\delta} < \infty$ *for some* $\delta > 0$;
2. $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$;
3. $\sigma^2 := \mathsf{E}\,X_0^2 + 2\sum_{k=1}^{\infty} \mathsf{E}\,X_0 X_k \neq 0$.

Then

$$\frac{X_0 + \cdots + X_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1), \quad n \to \infty.$$

Mixing and CLT

Theorem (I.A. Ibragimov). *Let $(X_n)_{n\in\mathbb{Z}_+}$ be a stationary process such that:*

1. $\mathsf{E}\, X_0 = 0$, $\mathsf{E}\,|X_0|^{2+\delta} < \infty$ *for some* $\delta > 0$;
2. $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$;
3. $\sigma^2 := \mathsf{E}\, X_0^2 + 2\sum_{k=1}^{\infty} \mathsf{E}\, X_0 X_k \neq 0$.

Then

$$\frac{X_0 + \cdots + X_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1), \quad n \to \infty.$$

- We prove that for certain class of Markov processes $\beta(n) = \underline{\underline{O}}(e^{-cn})$, $n \to \infty$.
- Since $\alpha(n) \leqslant \beta(n)$ this implies the CLT for these processes.

The coupling method

Total variation distance

- Let $(\Omega, \mathcal{F})$ be a measurable space.
- If Q and $\widetilde{Q}$ are two measures on $(\Omega, \mathcal{F})$, then **total variation distance** between Q and $\widetilde{Q}$ is defined by

$$d_{TV}(Q, \widetilde{Q}) := 2 \sup_{A \in \mathcal{F}} |Q(A) - \widetilde{Q}(A)| = \sup_f \left| \int_\Omega f(\omega)\, dQ - \int_\Omega f(\omega)\, d\widetilde{Q} \right|,$$

  where the second supremum is taken over measurable functions bounded by 1.
- Obviously, if Q_n are probability measures and $d_{TV}(Q_n, Q) \to 0,\ n \to \infty$, then $Q_n \xrightarrow{d} Q$.

Total variation distance

- Similarly, the total variation distance between two random variables X and Y is defined to be the total variation distance between their distributions.

$$d_{TV}(X,Y) := 2\sup_{A\in\mathcal{F}} |\mathrm{P}(X\in A) - \mathrm{P}(Y\in A)| = \sup_{f} |\mathsf{E}\, f(X) - \mathsf{E}\, f(Y)|.$$

- It follows from the definition, that

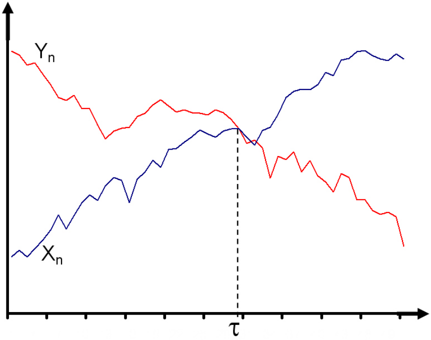
$$d_{TV}(X,Y) \leqslant 2\mathrm{P}(X\neq Y).$$

Setup

- Suppose $(X_n)_{n\in\mathbb{Z}_+}$ and $(Y_n)_{n\in\mathbb{Z}_+}$, are homogeneous Markov processes with the same transition functions.
- Our goal is to estimate $d_{TV}(X_n, Y_n)$.
- If μ is a stationary distribution of X_n and $\mathrm{Law}(Y_0) = \mu$, then $d_{TV}(X_n, Y_n) = d_{TV}(X_n, \mu)$

Coupling: a simple example

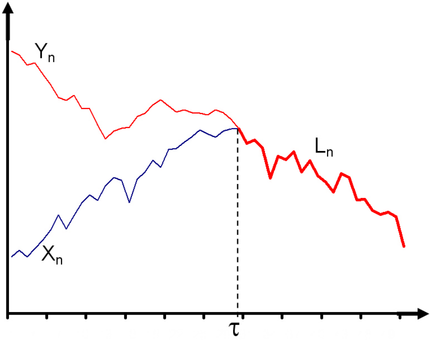
- To estimate $d_{TV}(X_n, Y_n)$ we use the coupling method.
- Let $\tau := \inf\{n : X_n = Y_n\}$ be the moment of the first meet of the processes.



- Then $L_n := X_n \mathrm{I}(n < \tau) + Y_n \mathrm{I}(n \geqslant \tau)$ is distributed as X_n.
- Hence
$d_{TV}(X_n, Y_n) = d_{TV}(L_n, Y_n) \leqslant 2\mathrm{P}(L_n \neq Y_n) \leqslant 2\mathrm{P}(\tau > n)$.

Coupling: a simple example

- To estimate $d_{TV}(X_n, Y_n)$ we use the coupling method.
- Let $\tau := \inf\{n : X_n = Y_n\}$ be the moment of the first meet of the processes.



- Then $L_n := X_n \,\mathrm{I}(n < \tau) + Y_n \,\mathrm{I}(n \geqslant \tau)$ is distributed as X_n.
- Hence
  $d_{TV}(X_n, Y_n) = d_{TV}(L_n, Y_n) \leqslant 2\mathrm{P}(L_n \neq Y_n) \leqslant 2\mathrm{P}(\tau > n)$.

Coupling: definition

- A coupling is a bivariate process $\left(\widetilde{X}_n, \widetilde{Y}_n\right)$, such that
  - $\widetilde{X}_n \stackrel{d}{=} X_n$ and $\widetilde{Y}_n \stackrel{d}{=} Y_n$, for all $n \in \mathbb{Z}_+$,
  - $\widetilde{X}_n(\omega) = \widetilde{Y}_n(\omega)$ for all $n > \tau(\omega)$.
- It is clear, that $d_{TV}(X_n, Y_n) = d_{TV}(\widetilde{X}_n, \widetilde{Y}_n) \leqslant 2\mathrm{P}(\tau > n)$.
- For instance, (L_n, Y_n) is a coupling. This construction was suggested by Doeblin.
- Can Doeblin's coupling be improved?

Doeblin's coupling vs Vaserstein's coupling

The Doeblin coupling

$(\widetilde{X}_n, \widetilde{Y}_n)$		$(\widetilde{X}_{n+1}, \widetilde{Y}_{n+1})$	$\tilde{p}_{..}$
(i,i)	$\rightarrow$	(k,k)	p_{ik}
(i,j)	$\rightarrow$	(k,k)	$p_{ik}p_{jk}$
(i,j)	$\rightarrow$	(k,l)	$p_{ik}p_{jl}$

The Vaserstein coupling

$(\widetilde{X}_n, \widetilde{Y}_n)$		$(\widetilde{X}_{n+1}, \widetilde{Y}_{n+1})$	$\tilde{p}_{..}$
(i,i)	$\rightarrow$	(k,k)	p_{ik}
(i,j)	$\rightarrow$	(k,k)	$p_{ik} \wedge p_{jk}$
(i,j)	$\rightarrow$	(k,l)	$c(i,j)(p_{ik} - p_{ik} \wedge p_{jk})(p_{jl} - p_{il} \wedge p_{jl})$

Suppose that processes X_n and Y_n have transition probability density $p(u, v)$.

Let us define

$$q(u, v) := \int_{-\infty}^{+\infty} p(u, t) \wedge p(v, t)\, dt.$$

Let $\eta_n = \left(\eta_n^1, \eta_n^2\right)$ be a Markov process with transition probability density $\varphi(x, y) := \varphi_1(x, y_1)\varphi_2(x, y_2)$, where $x = (x^1, x^2)$, $y = (y^1, y^2)$ and

$$\varphi_1(x, u) := (1 - q(x^1, x^2))^{-1} \left(p(x^1, u) - p(x^1, u) \wedge p(x^2, u)\right),$$
$$\varphi_2(x, u) := (1 - q(x^1, x^2))^{-1} \left(p(x^2, u) - p(x^1, u) \wedge p(x^2, u)\right).$$

We also set $\eta_0 = X_0$.

Lemma. *It is possible to construct a coupling* $\left(\widetilde{X}_n, \widetilde{Y}_n\right)$ *such that*

$$\mathrm{P}(\widetilde{X}_n \neq \widetilde{Y}_n) \leqslant \mathsf{E} \prod_{i=0}^{n-1} (1 - q(\eta_i^1, \eta_i^2)).$$

Idea of the proof

Let $\gamma_n = \mathrm{I}(\tau > n)$.

Consider the following decomposition

$$\widetilde{X}_n = \eta_n^1\, \mathrm{I}(\gamma_n = 1) + \zeta_n\, \mathrm{I}(\gamma_n = 0),$$
$$\widetilde{Y}_n = \eta_n^2\, \mathrm{I}(\gamma_n = 1) + \zeta_n\, \mathrm{I}(\gamma_n = 0).$$

Random processes η_n^1 and η_n^2 represent X_n and Y_n, correspondingly, under condition that coupling was not successful until time n.

On the other hand, ζ_n represents both X_n and Y_n if the coupling occurs before time n.

It is possible to show, that $(\eta_n^1, \eta_n^2, \zeta_n, \gamma_n)$ is a Markov process and

$$\mathrm{P}(\gamma_{n+1} = 1 | \gamma_n = 1, \eta_n^1 = x^1, \eta_n^2 = x^2) = 1 - q(x^1, x^2).$$

Estimates of convergence rate

Operator

Let us introduce operator $A: \ C_b \to C_b$

$$Af(x) := (1 - q(x))\, \mathsf{E}_x\, f(\eta_1).$$

Operator A has an interesting property:

$$\mathsf{E}\, A^n f(\eta_0) = \mathsf{E} \prod_{i=0}^{n-1} (1 - q(\eta_i)) f(\eta_n).$$

Therefore,

$$\frac{1}{2} d_{TV}(X_n, Y_n) \leqslant \mathsf{E}\, A^n 1(\eta_0) \leqslant \|A^n\|.$$

Theorem. *If operator A has a spectral radius* $r \neq 1$ *then for any* $\varepsilon > 0$ *and for sufficiently large* $n > n_0(\varepsilon)$

$$d_{TV}(X_n, Y_n) \leqslant 2e^{-n(|\ln r|-\varepsilon)}.$$

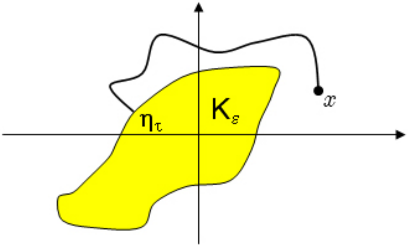
Alternative approach

- In some cases it might be difficult to check whether $r < 1$.
- We developed also an alternative approach.
- Clearly, $r \leqslant \|A\|$, however $\|A\| = 1 - \inf_x q(x) = 1$ in many cases.
- So we introduce operator B with $\|B\| \leqslant 1 - \varepsilon$.

Alternative approach

Let us consider a "good" set $K_\varepsilon := \{(x^1, x^2) : q(x^1, x^2) \geqslant \varepsilon\}$.

We denote by $\tau := \inf\{n > 0 : \ \eta_n \in K_\varepsilon\}$ a first hit time for the set K_ε.



Define operator $B : \ C_b \to C_b$ by the following formula

$$Bf(x) := \mathsf{E}_x\big((1 - q(\eta_1))(1 - q(\eta_2)) \cdot \ldots \cdot (1 - q(\eta_\tau))f(\eta_\tau)\big).$$

Operator B

$$
\begin{aligned}
Af(x) &:= (1 - q(x))\,\mathsf{E}_x\, f(\eta_1).\\
Bf(x) &:= \mathsf{E}_x\big((1 - q(\eta_1))(1 - q(\eta_2)) \cdot \ldots \cdot (1 - q(\eta_\tau))f(\eta_\tau)\big).
\end{aligned}
$$

We see, that since

$$
Bf(x) \leqslant \mathsf{E}_x\big((1 - q(\eta_\tau))f(\eta_\tau)\big) \leqslant (1 - \varepsilon)\,\mathsf{E}_x\, f(\eta_\tau),
$$

we have $\|B\| \leqslant 1 - \varepsilon$.

Operator B

Moreover,

$$\mathsf{E}(1-q(\eta_0)B^n f(\eta_0) = \mathsf{E}\prod_{i=0}^{\tau_n}\big(1-q(\eta_i)\big)f(\eta_{\tau_n}),$$

where τ_n is a $n-th$ hit time for the set K_ε.
On the other hand, it follows from Lemma that

$$\mathrm{P}(\widetilde{X}_{n+1} \neq \widetilde{Y}_{n+1}) \leqslant \mathsf{E}\prod_{i=0}^{n}(1-q(\eta_i)).$$

Operator B

Moreover,

$$\mathsf{E}(1 - q(\eta_0)B^n f(\eta_0) = \mathsf{E}\prod_{i=0}^{\tau_n}\big(1 - q(\eta_i)\big)f(\eta_{\tau_n}),$$

where τ_n is a $n - th$ hit time for the set K_ε.
On the other hand, it follows from Lemma that

$$\mathrm{P}(\widetilde{X}_{n+1} \neq \widetilde{Y}_{n+1}) \leqslant \mathsf{E}\prod_{i=0}^{n}(1 - q(\eta_i)).$$

Therefore if we introduce a "bad" set $\kappa_n := \{\omega : \tau_{\lfloor n/\Delta \rfloor} > n\}$ then

$$\mathrm{P}(\widetilde{X}_{n+1} \neq \widetilde{Y}_{n+1}) \leqslant \mathsf{E}\, B^{\lfloor n/\Delta \rfloor} 1(\eta_0) + \mathsf{E}\, \mathsf{I}(\kappa_n).$$

Theorem. *Assume that there exist* $\varepsilon > 0$, $\lambda > 0$, $M > 0$ *such that*
1. $\mathsf{E}\, e^{\lambda\tau} < \infty$.
2. For all $x \in K_\varepsilon$ *we have* $\mathsf{E}_x\, e^{\lambda\tau} < M$.
Then

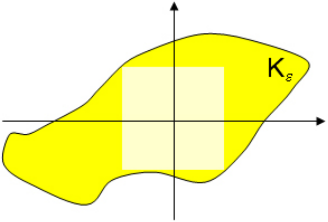
$$d_{TV}(X_n, Y_n) \leqslant Ce^{-n\theta},$$

where $C > 0$ *and*

$$\theta = \frac{|\ln(1-\varepsilon)|\lambda}{\ln M + |\ln(1-\varepsilon)|}.$$

- Recall that τ is a first hit time for process η.
- Let us reformulate conditions 1 and 2 in terms of a first hit time for process X_n.

Let us assume that $q(u, v) > \varepsilon$
for all u, v such that $|u| < K$, $|v| < K$.



Define $T := \inf\{n > 0 : \ |X_n| < K\}$.

Theorem. *Assume that there exist $\lambda > 0$, $K > 0$, $M > 0$, $\varkappa > 0$, such that*
1. $\mathsf{E}_u\, e^{\lambda T} < \infty$ for all u.
2. $\mathsf{E}_u\, e^{\lambda T} < M$ for all $|u| < K$.
3. $P_u(|X_1| < K) > \varkappa$ for all $|u| < K$.
Then

$$d_{TV}(X_n, Y_n) \leqslant Ce^{-n\theta_1},$$

wit some $C > 0$, $\theta_1 > 0$.

Estimates of convergence rate: Summary

- If $r < 1$ then $d_{TV}(X_n, Y_n) \leqslant 2e^{-n(|\ln r| - \varepsilon)}$.
- Under certain conditions on a first hit time for the set $[-K, K]$ we have $d_{TV}(X_n, Y_n) \leqslant Ce^{-n\theta}$.

Estimation of mixing coefficients

β-mixing

Theorem. *Assume that X_n has a stationary distribution. If $r \neq 1$ then for any $\varepsilon > 0$*

$$\beta(n) \leqslant Ce^{-n(|\ln r| - \varepsilon)},$$

for some $C > 0$ and sufficiently large $n > N(\varepsilon)$.

Central Limit Theorem

- Therefore, CLT for dependent random variables implies

Theorem. *Suppose that X_n has a stationary distribution and $r \neq 1$.*

Let X_n^μ, $n \in \mathbb{Z}_+$, be a stationary version of the process X_n.

Furthermore, assume that
1. $\mathrm{E}\,|X_1^\mu|^{2+\delta} < \infty$, *for some* $\delta > 0$,
2. $\sigma^2 := \mathrm{Var}\, X_1^\mu + 2 \sum_{k=1}^{\infty} \mathrm{cov}(X_1^\mu, X_{k+1}^\mu) \neq 0$.
Then the process X_n satisfies central limit theorem, i.e.

$$\frac{\sum_{i=1}^{n} X_i - n\,\mathrm{E}\,X_1^\mu}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{as } n \to \infty.$$