Mixing it up: A general framework for Markovian statistics beyond reversibility and the minimax paradigm

Mathematisches Seminar – Christian-Albrechts-Universität zu Kiel

Lukas Trottner (joint work with Niklas Dexheimer and Claudia Strauch)
Universität Mannheim

19.11.2020

Outline

1. Status Quo in nonparametric Markovian statistics
2. The β-mixing framework
3. Moment bounds for Markovian integral functionals
4. Statistical Applications

Status Quo in nonparametric Markovian statistics

Continuous-time Markov processes

- Let $\mathbf{X} = (X_t)_{t \geqslant 0}$ be a continous-time $\mathcal{X}$-valued Markov process with probabilities $(\mathbb{P}^x)_{x \in \mathcal{X}}$ and filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geqslant 0}$, i.e. $\mathbb{P}^x$-a.s. for all $s, t \geqslant 0$

$$\mathbb{E}^x[f(X_{t+s})|\mathcal{F}_t] = \mathbb{E}^{X_t}[f(X_s)] = P_s f(X_t), \quad f \in \mathcal{B}_b(\mathcal{X}) \qquad \text{(Markov property)}$$

- Stationarity is a minimum requirement for statistical analysis of general Markov processes, i.e. $\exists$(unique) probability measure μ s.t.

$$\forall t \geqslant 0, B \in \mathcal{B}(\mathcal{X}) : \mathbb{P}^\mu(X_t \in B) = \int_{\mathcal{X}} \mathbb{P}^x(X_t \in B)\, \mu(\mathrm{d}x) = \mu(B), \quad \text{(stat. dist.)}$$

- For non-stationary specific classes of Markov processes we may resort to either transforming the process to obtain stationarity or focus on other properties (e.g. martingale structures)

Nonparametric statistics for Markov processes (so far)

Up to now, no treatment of Markov processes in a structurally general framework, but clear focus on specific models such as

- (Multivariate) diffusions,

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad t \geqslant 0$$

- Ornstein–Uhlenbeck processes driven by a Lévy process Z,

$$\mathrm{d}X_t = -BX_t\,\mathrm{d}t + \mathrm{d}Z_t, \quad t \geqslant 0$$

- (More recently) Lévy driven jump diffusions

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t + \gamma(X_{t-})\,\mathrm{d}Z_t, \quad t \geqslant 0$$

Nonparametric statistics for Markov processes (so far)

All these investigations have mostly two things in common

1. Knowledge of the structural dynamics of the processes allows investigation of estimator performance for characterisitics in the minimax sense.
   E.g. for estimating the drift b of a diffusion with unit diffusion matrix, choose some functional class Σ and determine upper/lower bounds for

$$\inf_{\widehat{b}_T} \sup_{b \in \Sigma} \mathcal{R}(\widehat{b}_T, b),$$

   where the infimum is taken over all $\sigma(X_t, 0 \leqslant t \leqslant T)$-measurable ($\rightsquigarrow$ continuous observations) functions $\widehat{b}_T$ (estimators) and $\mathcal{R}(\widehat{b}_T, b)$ is some risk quantifier (pointwise L^2, MISE, sup-norm etc.)

Nonparametric statistics for Markov processes (so far)

2. The analysis relies heavily on imposing assumptions on coeffcients that guarantee some form of strong stability of the process. For this, there are two approaches
   - Make the right assumptions so that $\mathbf{X}$ is reversible, i.e. $P_t = P_t^*$ on $L^2(\mu)$,

     $$\int P_t f \, g \, \mathrm{d}\mu = \int f \, P_t g \, \mathrm{d}\mu, \quad f, g \in L^2(\mu) \qquad \text{(symmetry)}$$

     and moreover a Poincaré inequality holds,

     $$\int f^2 \, \mathrm{d}\mu - \Big(\int f \, \mathrm{d}\mu \Big)^2 \leqslant -c_P \int f \mathcal{A} f \, \mathrm{d}\mu \qquad \text{(PI)}$$

   - Make sure that process can be included in Meyn and Tweedie framework for stability of Markov processes.

     Ultimate goal: Show that $\mathbf{X}$ is exponentially ergodic, i.e. $\exists$ function $V \geqslant 1$ s.t.

     $$\| P_t(x, \cdot) - \mu \|_{\mathrm{TV}} \coloneqq \sup_{|f| \leqslant 1} |P_t f(x) - \mu(f)| \leqslant C V(x) \mathrm{e}^{-\kappa t}, \quad t \geqslant 0, x \in \mathcal{X} \quad \text{(exp. ergod.)}$$

The β-mixing framework

Motivation

- From a classical nonparametric point of view the statistical analysis of reversible diffusion processes is quite satisfactory
- However, the price to be paid for a "full picture" are heavy structural constraints that exclude many models of applied and theoretical interest
- Moreover the minimax paradigm is not undisputed due to its pessimistic nature
- Instead of choosing a Markov process with specified dynamics up to a choice of characteristics, we raise the question:
  What are the essential stability properties needed to build a robust statistical theory upon for general Markov processes?

What are the essential stability conditions needed?

Any answer to this question should be

1. broad enough to include a wide range of processes
2. comprehensive enough to yield theoretical results with which rates for statistical procedures can be obtained that match (or at least come close) to the ones known for well-behaved Markov models

We show that a framework based on the exponential β-mixing property for Markov processes fulfills both criteria.

The β-mixing property

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{A}, \mathcal{B}$ be sub-σ-algebras of $\mathcal{F}$.
- Denote by $\mathbb{P}|_{\mathcal{A}}, \mathbb{P}|_{\mathcal{B}}$ the restrictions of $\mathbb{P}$ to $\mathcal{A}, \mathcal{B}$ and let $\mathbb{P}|_{\mathcal{A}\otimes\mathcal{B}}$ be the restriction of the image measure of $\mathbb{P}$ under the injection map $\iota(\omega) = (\omega, \omega)$ to $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{B})$

$$\rightsquigarrow \text{for } A \times B \in \mathcal{A} \otimes \mathcal{B}: \quad \mathbb{P}|_{\mathcal{A}\otimes\mathcal{B}}(A \times B) = \mathbb{P}(A \cap B).$$

- β-mixing coefficient $\beta(\mathcal{A}, \mathcal{B})$ defined by

$$\beta(\mathcal{A}, \mathcal{B}) \coloneqq \sup_{C \in \mathcal{A}\otimes\mathcal{B}} \left| \mathbb{P}|_{\mathcal{A}\otimes\mathcal{B}}(C) - \mathbb{P}|_{\mathcal{A}} \otimes \mathbb{P}|_{\mathcal{B}}(C) \right|$$

- Note that

$$\beta(\mathcal{A}, \mathcal{B}) \geqslant \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \eqqcolon \alpha(\mathcal{A}, \mathcal{B})$$

The β-mixing property

- Berbee's coupling lemma: If $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough, then for any pair of random variables X, Y we can construct $Y^* \perp\!\!\!\perp X$ such that $Y^* \overset{\mathrm{d}}{=} Y$ and $\mathbb{P}(Y \neq Y^*) = \beta(X, Y)$
- Define for the Markov process $\mathbf{X}$ started in some distribution η the β-mixing coefficient

$$\beta_\eta(t) = \sup_{s \geqslant 0} \beta_\eta(\mathcal{F}_s, \overline{\mathcal{F}}_{s+t}), \quad \mathcal{F}_s \coloneqq \sigma(X_u, u \leqslant s), \overline{\mathcal{F}}_s \coloneqq \sigma(X_u, u \geqslant s).$$

If $\mathbf{X}$ is stationary this reduces to

$$\beta(t) \coloneqq \beta_\mu(t) = \int_{\mathcal{X}} \|P_t(x, \cdot) - \mu\|_{\mathrm{TV}} \, \mu(\mathrm{d}x).$$

The β-mixing framework

Central Assumption $(\mathscr{A}\beta)$

$\mathbf{X}$ is Borel right, stationary and exponentially β-mixing, i.e. $\exists c_\kappa, \kappa > 0$ s.t. $\beta(t) \leqslant c_\kappa \mathrm{e}^{-\kappa t}$.

- Exponential β-mixing is intimately connected with exponential ergodicity. If $\mathbf{X}$ is exponentially ergodic, then
$$\beta(t) = \int_{\mathcal{X}} \|P_t(x,\cdot) - \mu\|_{\mathrm{TV}}\, \mu(\mathrm{d}x) \leqslant c_\kappa \mathrm{e}^{-\kappa t} \int_{\mathcal{X}} V(x)\, \mu(\mathrm{d}x)$$
and hence $\mathbf{X}$ is exponentially β-mixing if $\mu(V) < \infty$.
- Criteria for exponential ergodicity if $\mathbf{X}$ is aperiodic stated as Lyapunov drift criteria in terms of generator $\mathcal{A}$ and resolvent $(U_\lambda)_{\lambda>0}$ of $\mathbf{X}$:
$$\exists V \geqslant 1, b, c > 0, C \text{ petite set}: \quad \mathcal{A}V \leqslant -cV + b\mathbf{1}_C;$$
$$\exists V_\lambda \geqslant 1, \lambda > 0, b < \infty, \beta \in (0,1), C \text{ petite}: \quad \lambda U_\lambda V_\lambda \leqslant \beta V_\lambda + b\mathbf{1}_C.$$

Moment bounds for Markovian integral functionals

Main result

Let $\mathcal{G}$ be a countable class of bounded functions g with $\mu(g) = 0$ and define for $T > 0$

$$\mathbb{G}_T(g) \coloneqq \frac{1}{\sqrt{T}} \int_0^T g(X_s)\,\mathrm{d}s, \quad d_{\mathbb{G},T}(f,g) \coloneqq \sqrt{\mathrm{Var}(\mathbb{G}_T(f-g))}.$$

Theorem (Dexheimer, Strauch and T. (2020))

Given $(\mathscr{A}\beta)$ and $m_T \in (0, T/4]$, there exists $\tau \in [m_T, 2m_T]$ such that for any $1 \leqslant p < \infty$

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}} |\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1 \int_0^\infty \log \mathcal{N}\left(u, \mathcal{G}, \tfrac{2m_T}{\sqrt{T}} d_\infty\right) \mathrm{d}u + \widetilde{C}_2 \int_0^\infty \sqrt{\log \mathcal{N}(u, \mathcal{G}, d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+ 4 \sup_{g\in\mathcal{G}} \left(\frac{2m_T}{\sqrt{T}} \|g\|_\infty \widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau} \widetilde{c}_2 \sqrt{p} + \frac{1}{2}\|g\|_\infty c_\kappa \sqrt{T} \mathrm{e}^{-\frac{\kappa m_T}{p}}\right),$$

with some constants $\widetilde{c}_1, \widetilde{c}_2, \widetilde{C}_1, \widetilde{C}_2 > 0$.

Breaking the moment bound down

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1\int_0^\infty \log\mathcal{N}\left(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\right)\mathrm{d}u + \widetilde{C}_2\int_0^\infty\sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+\,4\sup_{g\in\mathcal{G}}\left(\frac{2m_T}{\sqrt{T}}\|g\|_\infty\widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau}\widetilde{c}_2\sqrt{p} + \frac{1}{2}\|g\|_\infty c_\kappa\sqrt{T}\mathrm{e}^{-\frac{\kappa m_T}{p}}\right),$$

- Break up integral $\int_0^T g(X_s)\,\mathrm{d}s$ into $2\times n_T$ parts of length m_T:

$$I_g(\mathbf{X}^{j,1}) \coloneqq \int_{2(j-1)m_T}^{(2j-1)m_T} g(X_s)\,\mathrm{d}s,\quad I_g(\mathbf{X}^{j,2}) \coloneqq \int_{(2j-1)m_T}^{2jm_T} g(X_s)\,\mathrm{d}s,\quad j=1,\ldots,n_T$$

Breaking the moment bound down

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\big)\,\mathrm{d}u + \widetilde{C}_2\int_0^\infty \sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+ 4\sup_{g\in\mathcal{G}}\left(\frac{2m_T}{\sqrt{T}}\|g\|_\infty\widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau}\widetilde{c}_2\sqrt{p} + \frac{1}{2}\|g\|_\infty c_\kappa\sqrt{T}\mathrm{e}^{-\frac{\kappa m_T}{p}}\right),$$

- Break up integral $\int_0^T g(X_s)\,\mathrm{d}s$ into $2\times n_T$ parts of length m_T:

$$I_g(\mathbf{X}^{j,1}) := \int_{2(j-1)m_T}^{(2j-1)m_T} g(X_s)\,\mathrm{d}s, \quad I_g(\mathbf{X}^{j,2}) := \int_{(2j-1)m_T}^{2jm_T} g(X_s)\,\mathrm{d}s, \quad j=1,\dots,n_T$$

- Use Berbee's coupling lemma to obtain i.i.d. vectors $(\widehat{\mathbf{X}}^{j,1})_{j=1,\dots,n_T}$ and $(\widehat{\mathbf{X}}^{j,2})_{j=1,\dots,n_T}$ s.t. $\widehat{\mathbf{X}}^{j,k} \overset{\mathrm{d}}{=} \mathbf{X}^{j,k}$ and $\mathbb{P}(\widehat{\mathbf{X}}^{j,k}\neq\mathbf{X}^{j,k}) \leqslant c_\kappa\mathrm{e}^{-\kappa m_T}$ for $k=1,2$

Breaking the moment bound down

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\big)\,\mathrm{d}u + \widetilde{C}_2\int_0^\infty\sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+4\sup_{g\in\mathcal{G}}\left(\frac{2m_T}{\sqrt{T}}\|g\|_\infty\widetilde{c}_1 p+\|g\|_{\mathbb{G},\tau}\widetilde{c}_2\sqrt{p}+\frac{1}{2}\|g\|_\infty c_\kappa\sqrt{T}\mathrm{e}^{-\frac{\kappa m_T}{p}}\right),$$

- Use Berbee's coupling lemma to obtain i.i.d. vectors $(\widehat{\mathbf{X}}^{j,1})_{j=1,\ldots,n_T}$ and $(\widehat{\mathbf{X}}^{j,2})_{j=1,\ldots,n_T}$ s.t. $\widehat{\mathbf{X}}^{j,k}\stackrel{\mathrm{d}}{=}\mathbf{X}^{j,k}$ and $\mathbb{P}(\widehat{\mathbf{X}}^{j,k}\neq\mathbf{X}^{j,k})\leqslant c_\kappa\mathrm{e}^{-\kappa m_T}$ for $k=1,2$
- Use this to obtain a bound for coupling error

$$\left(\mathbb{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{1}{\sqrt{T}}\int_0^T(g(X_s)-g(\widehat{X}_s))\,\mathrm{d}s\Big|^p\Big]\right)^{1/p}=\left(\mathbb{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{1}{\sqrt{T}}\sum_{k=1}^2\sum_{j=1}^{n_T}(I_g(\mathbf{X}^{j,k})-I_g(\widehat{\mathbf{X}}^{j,k}))\Big|^p\Big]\right)^{1/p}$$

Breaking the moment bound down

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1\int_0^\infty \log\mathcal{N}\Big(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\Big)\,\mathrm{d}u + \widetilde{C}_2\int_0^\infty \sqrt{\log\mathcal{N}\big(u,\mathcal{G},d_{\mathbb{G},\tau}\big)}\,\mathrm{d}u$$
$$+4\sup_{g\in\mathcal{G}}\left(\frac{2m_T}{\sqrt{T}}\|g\|_\infty\widetilde{c}_1 p+\|g\|_{\mathbb{G},\tau}\widetilde{c}_2\sqrt{p}+\frac{1}{2}\|g\|_\infty c_\kappa\sqrt{T}\mathrm{e}^{-\frac{\kappa m_T}{p}}\right),$$

- It remains to analyze

$$\left(\mathbb{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{n_T} I_g(\widehat{\mathbf{X}}^{j,k})\Big|^p\Big]\right)^{1/p}$$

Breaking the moment bound down

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\big)\,\mathrm{d}u + \widetilde{C}_2\int_0^\infty \sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$

$$+4\sup_{g\in\mathcal{G}}\Big(\frac{2m_T}{\sqrt{T}}\|g\|_\infty\widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau}\widetilde{c}_2\sqrt{p} + \frac{1}{2}\|g\|_\infty c_\kappa\sqrt{T}\mathrm{e}^{-\frac{\kappa m_T}{p}}\Big),$$

- It remains to analyze

$$\left(\mathbb{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{n_T} I_g(\widehat{\mathbb{X}}^{j,k})\Big|^p\Big]\right)^{1/p}$$

- Bernstein's inequality for sums of i.i.d. random variables yields

$$\mathbb{P}\left(\Big|\frac{1}{\sqrt{T}}\sum_{j=1}^{n_T} I_{g-f}(\widehat{\mathbb{X}}^{j,k})\Big| > d_{\mathbb{G},m_T}(f,g)\sqrt{u} + \frac{m_T}{\sqrt{T}}d_\infty(f,g)u\right) \leqslant \mathrm{e}^{-u}$$

$\rightsquigarrow$ $(T^{-1/2}\sum_{j=1}^{n_T} I_g(\widehat{\mathbb{X}}^{j,k}))_{g\in\mathcal{G}}$ has mixed tail w.r.t. $d_{\mathbb{G},m_T}$ and $\frac{m_T}{\sqrt{T}}d_\infty$

Breaking the moment bound down

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\big)\,\mathrm{d}u + \widetilde{C}_2\int_0^\infty \sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+4\sup_{g\in\mathcal{G}}\left(\frac{2m_T}{\sqrt{T}}\|g\|_\infty\widetilde{c}_1 p+\|g\|_{\mathbb{G},\tau}\widetilde{c}_2\sqrt{p}+\frac{1}{2}\|g\|_\infty c_\kappa\sqrt{T}\mathrm{e}^{-\frac{\kappa m_T}{p}}\right),$$

- Using generic chaining for empirical processes we obtain

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{n_T} I_g(\widehat{X}^{j,k})\right|^p\right]\right)^{1/p} \leqslant \frac{\widetilde{C}_1}{2}\int_0^\infty \log\mathcal{N}\big(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\big)\,\mathrm{d}u + \frac{\widetilde{C}_2}{2}\int_0^\infty \sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+2\sup_{g\in\mathcal{G}}\left(\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{n_T} I_g(\widehat{X}^{j,k})\right|^p\right]\right)^{1/p}.$$

Breaking the moment bound down

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}|\mathbb{G}_T(g)|^p\right]\right)^{1/p} \leqslant \widetilde{C}_1\int_0^\infty \log\mathcal{N}\left(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\right)\mathrm{d}u + \widetilde{C}_2\int_0^\infty\sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+\,4\sup_{g\in\mathcal{G}}\left(\frac{2m_T}{\sqrt{T}}\|g\|_\infty\widetilde{c}_1 p + \|g\|_{\mathbb{G},\tau}\widetilde{c}_2\sqrt{p} + \frac{1}{2}\|g\|_\infty c_\kappa\sqrt{T}\mathrm{e}^{-\frac{\kappa m_T}{p}}\right),$$

- Using generic chaining for empirical processes + $1:1$-correspondence between moment bounds and deviation inequalities we obtain

$$\left(\mathbb{E}\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{n_T} I_g(\widehat{\mathbb{X}}^{j,k})\right|^p\right]\right)^{1/p} \leqslant \frac{\widetilde{C}_1}{2}\int_0^\infty \log\mathcal{N}\left(u,\mathcal{G},\tfrac{2m_T}{\sqrt{T}}d_\infty\right)\mathrm{d}u + \frac{\widetilde{C}_2}{2}\int_0^\infty\sqrt{\log\mathcal{N}(u,\mathcal{G},d_{\mathbb{G},\tau})}\,\mathrm{d}u$$
$$+\,2\sup_{g\in\mathcal{G}}\left(\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\sum_{j=1}^{n_T} I_g(\widehat{\mathbb{X}}^{j,k})\right|^p\right]\right)^{1/p}.$$

Implication for the ergodic theorem

- Recall that for a Markov process $\mathbf{X}$ with invariant distribution μ it holds for any $g \in L^p(\mathbb{P}^\mu)$

$$\frac{1}{T}\int_0^T g(X_s)\,\mathrm{d}s \underset{T\to\infty}{\longrightarrow} \mathbb{E}^\mu[g(X_0)|\mathcal{I}], \quad \text{in } L^p(\mathbb{P}^\mu),$$

  where $\mathcal{I} = \{\Lambda \in \mathcal{F}_\infty : \Lambda = \theta_t^{-1}\Lambda,\ \forall t \geqslant 0\}$
- exponential β-mixing implies that $\mathcal{I}$ is $\mathbb{P}^\mu$-trivial and hence

$$\frac{1}{T}\int_0^T g(X_s)\,\mathrm{d}s \underset{T\to\infty}{\longrightarrow} \mu(g), \quad \text{in } L^p(\mathbb{P}^\mu),$$

- Moreover, for bounded g, independently of $t > 0$,

$$\|g\|_{\mathbb{G},t}^2 = \frac{1}{t}\mathrm{Var}\Big(\int_0^t g(X_s)\,\mathrm{d}s\Big) \leqslant 2\|g\|_\infty^2 \frac{c_\kappa}{\kappa}$$

Implication for the ergodic theorem

Theorem (Dexheimer, Strauch, T. (2020))

Under $(\mathscr{A}\beta)$ it holds for any bounded g and $T > 0$ that

$$\left\| \frac{1}{T} \int_0^T g(X_t)\,\mathrm{d}t - \mu(g) \right\|_{L^p(\mathbb{P})} \lesssim p \|g\|_\infty \frac{1}{\sqrt{T}}.$$

Statistical Applications

Estimation strategy based on moment bounds

- Suppose that we aim to estimate a characteristic $f\colon \mathcal{X} \to \mathbb{R}$ on the state space based on observations $(X_t)_{0 \leqslant t \leqslant T}$ by finding an appropriate mean estimator $\widehat{f}_T$ of the form
$$\widehat{f}_T(x) = \frac{1}{T}\int_0^T g(X_t; x)\,\mathrm{d}t,$$
with a careful choice of g
- The sup-moment bound is perfectly suited for establishing upper bounds on the sup-norm risk
$$\mathcal{R}_\infty^{(p)}(\widehat{f}_T, f; D) \coloneqq \mathbb{E}\Big[\sup_{x \in D} |f(x) - \widehat{f}_T(x)|^p\Big]^{1/p}, \quad D \text{ bounded}$$
- The additional analytical effort needed is dependent on whether $\widehat{f}_T(x)$ is unbiased or not $\rightsquigarrow$ determines tightness of variance bound needed for $\widehat{f}_T$

Example 1: Kernel invariant density estimation

- Classical problem in statistics for $\mathbb{R}^d$-valued Markov processes $\mathbf{X}$: estimation of the invariant density ρ (if it exists)
- Based on a continuous record $(X_s)_{0 \leqslant s \leqslant T}$, a natural estimator is given by

$$\widehat{\rho}_{h,T}(x) = \frac{1}{Th^d} \int_0^T K\Big(\frac{x - X_t}{h}\Big) \,\mathrm{d}t, \quad x \in \mathbb{R}^d,$$

  where h is a bandwidth and K is some kernel function
- Fixing some bounded, open domain D, we want to obtain tight upper bounds for the risk

$$\mathcal{R}_\infty^{(p)}(\widehat{\rho}_{h,T}, \rho; D) \coloneqq \Big(\mathbb{E}\Big[\sup_{x \in D} |\widehat{\rho}_{h,T}(x) - \rho(x)|^p\Big]\Big)^{1/p}, \quad p \geqslant 1$$

  for $\mathbf{X}$ belonging to a class of Markov processes as general as possible

Risk decomposition

- Triangle inequality yields the following decomposition into bias and volatility of the estimator

$$\Big(\mathbb{E}\Big[\sup_{x\in D}\big|\widehat{\rho}_{h,T}(x)-\rho(x)\big|^p\Big]\Big)^{1/p}$$
$$=\sup_{x\in D}\big|\mathbb{E}[\widehat{\rho}_{h,T}(x)]-\rho(x)\big|+\Big(\mathbb{E}\Big[\sup_{x\in D}\big|\widehat{\rho}_{h,T}(x)-\mathbb{E}[\widehat{\rho}_{h,T}(x)]\big|^p\Big]\Big)^{1/p}=:\mathbf{B}+\mathbf{V}$$

- The bias is taken care of by making the right assumptions on the invariant density ρ: assume that $\rho|_D\in\mathcal{H}_D(\beta,L)$ for some $\beta>0$, then $\mathbf{B}\lesssim h^{\beta}$
- Assuming $(\mathcal{A}\beta)$, we can rely on the moment bound for Markovian integral functionals to bound the the stochastic term $\mathbf{V}$, since

$$\mathbf{V}=\frac{1}{\sqrt{T}h^d}\Big(\mathbb{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{1}{\sqrt{T}}\int_0^T g(X_t)\,\mathrm{d}t\Big|^p\Big]\Big)^{1/p},\quad \mathcal{G}=\Big\{K\big(\tfrac{x-\cdot}{h}\big)-\mu\big(K\big(\tfrac{x-\cdot}{h}\big)\big):x\in D\cap\mathbb{Q}^d\Big\}$$

Variance bounds for Markovian integral functionals

- Tight upper bounds for $\mathbf{V}$ require careful control of the variance of $\int_0^T g(X_t)\,\mathrm{d}t = \int_0^T K((x - X_t)/h)\,\mathrm{d}t$ in terms of h
- β-mixing is not quite sufficient: we also need control of the small time transition probabilities
- We restrict to $\mathbb{R}^d$-valued $\mathbf{X}$ having transition densities $(p_t)_{t \geqslant 0}$ with dimension $d \geqslant 2$

Heat kernel Assumption $(\mathscr{A}1)$

There exists $c > 0$ such that

$$\forall t \in (0,1] : \sup_{x,y \in \mathbb{R}^d} p_t(x,y) \leqslant c t^{-d/2}.$$

Variance bounds for Markovian integral functionals

Proposition (Dexheimer, Strauch and T. (2020))

Grant $(\mathscr{A}\beta)$ and $(\mathscr{A}1)$, and let f be a bounded function with compact support $\mathcal{S}$ fulfilling $\boldsymbol{\lambda}(\mathcal{S}) < 1$. Then, there exists a constant $C > 0$ not depending on f such that, for any $T > 0$,

$$\operatorname{Var}\left(\int_0^T f(X_t)\,\mathrm{d}t\right) \leqslant CT\|f\|_\infty^2\|\rho\|_\infty\boldsymbol{\lambda}^2(\mathcal{S})\psi_d^2(\boldsymbol{\lambda}(\mathcal{S}))$$

with

$$\psi_d(x) \coloneqq \begin{cases} \sqrt{1+\log(1/x)}, & d = 2, \\ x^{\frac{1}{d}-\frac{1}{2}}, & d \geqslant 3. \end{cases}$$

Rates for sup-norm invariant kernel density estimation

Theorem (Dexheimer, Strauch and T. (2020))

Grant $(\mathscr{A}\beta)$ and $(\mathscr{A}1)$. Let $D \subset \mathbb{R}^d$ be open and bounded. Suppose that $\rho|_D \in \mathcal{H}_D(\beta, \mathrm{L})$ with $\beta > 2$ and let K be a smooth, Lipschitz-continuous kernel of order $\lfloor \beta \rfloor$. Then, for any $p \geqslant 1$,

$$\left(\mathbb{E}\Big[\sup_{x \in D}|\widehat{\rho}_{h,T}(x) - \rho(x)|^p\Big]\right)^{1/p} \in \begin{cases} \mathrm{O}(\log T/\sqrt{T}), & d = 2, \\ \mathrm{O}\big((\log T/T)^{\beta/(2\beta+d-2)}\big), & d \geqslant 3, \end{cases}$$

if

$$h = h(T) \sim \begin{cases} \log T/T^{1/4}, & d = 2, \\ (\log T/T)^{1/(2\beta+d-2)}, & d \geqslant 3. \end{cases}$$

Incorporating Lévy driven SDEs into our framework

- Let
$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t + \gamma(X_{t-})\,\mathrm{d}Z_t \qquad \text{(SDE)}$$
be a d-dimensional SDE ($d \geqslant 2$), where Z is a pure jump Lévy process with Lévy measure ν, independent of Brownian Motion W
- We want to uniformly estimate the invariant density ρ (if it exists) of the Markov process solving (SDE)
- To this end, it is sufficient to introduce assumptions on the coefficients that guarantee
  (a) existence of a strong solution of (SDE) such that the corresponding Markov process $\mathbf{X}$ possesses transition densities and a stationary distribution
  (b) exponential β-mixing
  (c) heat kernel bound

Making the right assumptions

The following set of assumptions on the coefficients is sufficient:

($\mathfrak{J}1$) b, γ, σ are globally Lipschitz and $\exists c \geqslant 1$ s.t.

$$c^{-1}\mathbb{I}_{d\times d} \leqslant \sigma(x)\sigma(x)^T \leqslant c\mathbb{I}_{d\times d}, \quad \forall x \in \mathbb{R}^d.$$

Additionally, b and γ are bounded.

($\mathfrak{J}2$) $\nu \ll \text{Leb}$ and for an $\alpha \in (0,2)$, the function $\|\gamma(x)z\|^{d+\alpha}\nu(z)$ is bounded and measurable. If $\alpha = 1$,

$$\int_{r<\|\gamma(x)z\|\leqslant R} \gamma(x)z\,\nu(\mathrm{d}z) = 0$$

holds for any $0 < r < R < \infty, x \in \mathbb{R}^d$.

($\mathfrak{J}3$) $\exists c_1, c_2 > 0$ s.t.

$$\langle x, b(x)\rangle \leqslant -c_1\|x\|, \quad \forall x : \|x\| \geqslant c_2,$$

and $\eta_0 > 0$ such that

$$\int_{\mathbb{R}^d} \|z\|^2 e^{\eta_0\|z\|}\nu(\mathrm{d}z) < \infty.$$

Example 2: Estimation of generator functionals of Lévy processes

- Let ξ be a subordinator with drift d and Lévy measure Π, which is not compound Poisson and such that $\mathbb{E}^0[\xi_1] < \infty$
- The extended generator $\mathcal{A}$ acts on "nice" test functions γ via

$$\begin{aligned}\mathcal{A}\gamma(x) &= d\gamma'(x) + \int_0^\infty (\gamma(x+y) - \gamma(x))\,\Pi(\mathrm{d}y)\\ &= d\gamma'(x) + \int_0^\infty \gamma'(x+y)\Pi((y,\infty))\,\mathrm{d}y\end{aligned}$$

- We are interested in estimating $\mathcal{A}\gamma$ uniformly for γ' with bounded domain
- We cannot work directly with ξ in our framework because Lévy processes do not exhibit ergodic behavior $\rightsquigarrow$ need to use some space/time transformation to obtain ergodic process

Ergodicity of overshoots of subordinators

Define $T_t \coloneqq \inf\{s \geqslant 0 : \xi_s > t\}$ and let

$$\mathcal{O}_t = \xi_{T_t} - t, \quad t \geqslant 0 \qquad \text{(overshoot)}$$

Theorem (Döring and T. (2020+))

If either $d > 0$ or there exists $(a, b) \subset (0, \infty)$ such that $\mathrm{Leb}|_{(a,b)} \ll \Pi|_{(a,b)}$, then $(\mathcal{O}_t)_{t \geqslant 0}$ is ergodic with invariant distribution

$$\mu(\mathrm{d}x) = \mathbb{E}^0[\xi_1]^{-1}\big(d\delta_0(\mathrm{d}x) + \Pi((x, \infty))\,\mathrm{d}x\big), \quad x \geqslant 0,$$

If moreover, there exists $\lambda > 0$ such that $\mathbb{E}^0[\exp(\lambda\xi_1)] < \infty$, then $(\mathcal{O}_t)_{t \geqslant 0}$ is exponentially ergodic and exponentially β-mixing.

Relation of stationary overshoots and generator

- Observe that

$$\begin{aligned}\mathcal{A}\gamma(x) &= d_H\gamma'(x) + \int_0^\infty \gamma'(x+y)\Pi((y,\infty))\,\mathrm{d}y \\ &= \mathbb{E}^0[\xi_1]\int_0^\infty \gamma'(x+y)\,\mu(\mathrm{d}y)\end{aligned}$$

- It is therefore natural to consider as an estimator of $f(x) = \mathcal{A}\gamma(x)$,

$$\widehat{f}_S(x) = \frac{1}{S}\int_0^S \eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t,$$

given an overshoot sample $(\mathcal{O}_t)_{0\leqslant t\leqslant S}$ and $\eta = \mathbb{E}^0[\xi_1]$

- The moment bound theorem for Markovian integral functionals yields

$$\mathbb{E}^\mu\big[\|\widehat{f}_S - f\|_\infty\big] \lesssim \frac{1}{\sqrt{S}}.$$

Uniform risk bound for estimation of $\mathcal{A}\gamma$

- To obtain an estimator $\widetilde{f}_T$ for given data $(\xi_t)_{0\leqslant t\leqslant T}$ instead of $(\xi_{T_s})_{0\leqslant s\leqslant S}$ we consider

$$\widetilde{f}_T(x) \coloneqq \frac{1}{\xi_T}\int_0^{\xi_T} \eta\gamma'(x+\mathcal{O}_t)\,\mathrm{d}t, \quad x \geqslant 0$$

- Note that indeed, up to the level $S = \xi_T$, we have observed all overshoots $(\mathcal{O}_t)_{0\leqslant s<S}$

Theorem (Christensen, Strauch and T. (2020+))

Let $\gamma \in \mathcal{D}(\mathcal{A}) \cap \mathcal{C}^2(\mathbb{R}_+)$ such that γ' has bounded support D, and suppose that the conditions for exponential ergodicity of overshoots holds. Then, for any $\varepsilon > 0$,

$$\mathbb{E}^{\mu}\left[\|\widetilde{f}_T - f\|_\infty\right] \in \mathrm{O}\Big(T^{-\frac{1}{2+\varepsilon}}\Big).$$

References

Amorino, C. and A. Gloter (2020). *Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes*. arXiv: 2001.07422.

Christensen, S., C. Strauch, and L. Trottner (2020+). "Learn how to reflect – data-driven optimal strategies in singular control problems".

Dalalyan, A. and M. Reiß (2007). "Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case". In: *Probab. Theory Related Fields* 137.1-2, pp. 25–47. ISSN: 0178-8051.

Dexheimer, N., C. Strauch, and L. Trottner (2020). *Mixing it up: A general framework for Markovian statistics beyond reversibility and the minimax paradigm*. arXiv: 2011.00308.

Döring, L. and L. Trottner (2020+). "Stability of Overshoots of Markov Additive Processes".

Down, D., S. P. Meyn, and R. L. Tweedie (1995). "Exponential and uniform ergodicity of Markov processes". In: *Ann. Probab.* 23.4, pp. 1671–1691. ISSN: 0091-1798.

Masuda, H. (2007). "Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps". In: *Stochastic Processes and their Applications* 117.1, pp. 35 –56. ISSN: 0304-4149.

Meyn, S. P. and R. L. Tweedie (1993). "Stability of Markovian processes. II. Continuous-time processes and sampled chains". In: *Adv. in Appl. Probab.* 25.3, pp. 487–517. ISSN: 0001-8678.

Thank you!