Statistical inference for Bifurcating Markov Chains

Marc Hoffmann

Université Paris-Dauphine

Luminy, February 17, 2016

Joint work with V. Bitseki-Penda and A. Olivier

Content

- Some technical remarks about variance estimates for geometrically ergodic Markov chains
- Bifurcating Markov Chains (BMC)
- Deviation bounds for uniformly geometrically ergodic BMC
- Estimation in BMC: invariant density, mean transition and transition densities
- Application in a transport-fragmentation model

MC setting

- $(X_n)_{n\geq 0}$ homogeneous MC with state space $(\mathcal{S}, \mathfrak{S})$ and transition $\mathcal{P}$.
- **Assumption (UE)** there exists a unique invariant probability ν $(\nu\mathcal{P} = \nu)$ such that for all $m \geq 0$:

$$\left|\mathcal{P}^m\phi(x) - \nu(\phi)\right| \leq R|\phi|_\infty \rho^m \text{ for all } x \in \mathcal{S}$$

  with $0 < \rho < 1$ and $R > 0$.
- Actually a *very strong* assumption: uniform geometric ergodicity v.s. geometric ergodicity

Objective

- Let ϕ with $\nu(\phi) = 0$ and $\mathcal{E}_n(\phi) := n^{-1} \sum_{i=1}^{n} \phi(X_i)$.
- Objective: control of $V_n(\phi) := \mathbb{E}_\nu \left[\mathcal{E}_n(\phi)^2 \right]$.
- In the IID case,
$$V_n(\phi) = n^{-1} |\phi|^2_{L^2(\nu)}.$$
- In the general Markov case, we have an additional covariance term

$$\begin{aligned} 2n^{-2} \sum_{i<j} \mathbb{E}_\nu \left[\phi(X_i)\phi(X_j) \right] &= 2n^{-2} \sum_{i<j} \mathbb{E}_\nu \left[\phi(X_i) \mathcal{P}^{j-i} \phi(X_i) \right] \\ &= 2n^{-2} \sum_{i<j} \nu(\phi \mathcal{P}^{j-i} \phi) \end{aligned}$$

Covariance control in the Markov case

- We have by **(UE)**

$$|\nu(\phi\mathcal{P}^{j-i}\phi)| \lesssim |\phi|_\infty |\phi|_{L^1(\nu)}\rho^{j-i}.$$

- It follows that

$$\begin{aligned} 2n^{-2}\sum_{i<j}\mathbb{E}_\nu\big[\phi(X_i)\phi(X_j)\big] &\lesssim n^{-1}|\phi|_\infty|\phi|_{L^1(\nu)}n^{-1}\sum_{i\neq j}\rho^{j-i} \\ &\lesssim n^{-1}|\phi|_\infty|\phi|_{L^1(\nu)} \end{aligned}$$

- Finally

$$V_n(\phi) := \mathbb{E}_\nu\big[\mathcal{E}_n(\phi)^2\big] \lesssim n^{-1}\big(|\phi|^2_{L^2(\nu)} + |\phi|_\infty|\phi|_{L^1(\nu)}\big)$$

- Nonparametrically-wise, we are happy.

The Markov case on a binary tree

- Binary tree

$$\mathbb{T}_n = \bigcup_{m=0}^{n} \mathbb{G}_m, \quad \mathbb{G}_m = \{0,1\}^m.$$

- Notation $\mathbb{T} = \bigcup_{m=0}^{\infty} \mathbb{G}_m$ and $|u| = m$ if $u \in \mathbb{T}$ and $u \in \mathbb{G}_m$.
- $|\mathbb{T}_m| = 2^{m+1} - 1 \lesssim 2^m$ and $|\mathbb{G}_m| = 2^m$.
- Let $(X_u)_{u\in\mathbb{T}}$ be a Markov chain indexed by $\mathbb{T}$ under **(UE)**.
- With ϕ such that $\nu(\phi) = 0$, how does

$$V_n(\phi) = \mathbb{E}_\nu\Big[\Big(|\mathbb{T}_n|^{-1} \sum_{u\in\mathbb{T}_n} \phi(X_u)\Big)^2\Big]$$

behave?

First variance estimate

- Decompose the sum along generations. Set

$$Y_m(\phi) = \sum_{u \in \mathbb{G}_m} \phi(X_u) \text{ so that } \sum_{u \in \mathbb{T}_n} \phi(X_u) = \sum_{m=0}^{n} Y_m(\phi).$$

- We have (triangle inequality)

$$V_n(\phi) = |\mathbb{T}_n|^{-2}\mathbb{E}_\nu\Big[\Big(\sum_{m=0}^{n} Y_m(\phi)\Big)^2\Big] \leq |\mathbb{T}_n|^{-2}\Big(\sum_{m=0}^{n} \mathbb{E}_\nu\big[Y_m(\phi)^2\big]^{1/2}\Big)^2$$

- Next step: control of $\mathbb{E}_\nu\big[Y_m(\phi)^2\big]$.

Control of $\mathbb{E}_\nu\left[Y_m(\phi)^2\right]$

- $\mathbb{E}_\nu\left[Y_m(\phi)^2\right] = I + II$, with $I = |\mathbb{G}_m|\nu(\phi^2)$ and

$$
\begin{aligned}
II &= \sum_{u \neq v, u,v \in \mathbb{G}_m} \mathbb{E}_\nu\left[\phi(X_u)\phi(X_v)\right] \\
&= \sum_{u \neq v, u,v \in \mathbb{G}_m} \mathbb{E}_\nu\Big[\mathbb{E}_\nu[\phi(X_u)|\mathcal{F}_{|u\wedge v|}]\mathbb{E}_\nu[\phi(X_v)|\mathcal{F}_{|u\wedge v|}]\Big] \\
&= \sum_{u \neq v, u,v \in \mathbb{G}_m} \mathbb{E}_\nu\Big[\mathcal{P}^{|u|-|u\wedge v|}\phi(X_{u\wedge v})\mathcal{P}^{|v|-|u\wedge v|}\phi(X_{u\wedge v})\Big]
\end{aligned}
$$

- $u \wedge v =$ most recent common ancestor of u and v.

Control of $\mathbb{E}_\nu\left[Y_m(\phi)^2\right]$ (cont.)

- Re-index the sum on $u \neq v$, $|u| = |v| = m$ as a functional of $X_{u\wedge v}$:

$$\{(u,v) \in \mathbb{G}_m^2\} = \bigcup_{\ell=1}^{m} \{(u,v) \in \mathbb{G}_m^2, u \wedge v = m - \ell\}$$

- $|\{(u,v) \in \mathbb{G}_m^2, u \wedge v = m - \ell\}| \lesssim 2^{2\ell} |\mathbb{G}_{m-\ell}|$.
- We obtain

$$\begin{aligned} II &= \sum_{u \neq v, u,v \in \mathbb{G}_m} \mathbb{E}_\nu\left[\mathcal{P}^{|u|-|u\wedge v|}\phi(X_{u\wedge v})\mathcal{P}^{|u|-|u\wedge v|}\phi(X_{u\wedge v})\right] \\ &= \sum_{\ell=1}^{m} 2^{2\ell-1}\mathbb{E}_\nu\big[\sum_{w\in\mathbb{G}_{m-\ell}} \big(\mathcal{P}^\ell\phi\big)(X_w)^2\big] \\ &\lesssim 2^m \sum_{\ell=1}^{m} 2^\ell \nu \mathcal{P}^{m-\ell}\big((\mathcal{P}^\ell\phi)^2\big). \end{aligned}$$

Too quick estimates

- Now, we have $(\mathcal{P}^\ell \phi)^2 \lesssim |\phi|_\infty^2 \rho^{2\ell}$ and

$$II \lesssim 2^m \sum_{\ell=1}^{m} (2\rho^2)^\ell |\phi|_\infty^2 \lesssim 2^m |\phi|_\infty^2 \quad \text{(provided } \rho < \tfrac{\sqrt{2}}{2}\text{)}$$

- Putting together $I + II$ plus taking $\sqrt{\ }$, we obtain

$$\mathbb{E}_\nu \left[Y_m(\phi)^2 \right]^{1/2} \lesssim 2^{m/2} \sqrt{|\phi|_{L^2(\nu)}^2 + |\phi|_\infty^2},$$

so $\left(\sum_{m=0}^{n} 2^{m/2} \sqrt{\ } \right)^2 \lesssim |\mathbb{T}_n| (|\phi|_{L^2(\nu)}^2 + |\phi|_\infty^2)$ and finally

$$V_n(\phi) \lesssim |\mathbb{T}_n|^{-1} (|\phi|_{L^2(\nu)}^2 + |\phi|_\infty^2)$$

- Nonparametrically-wise, we are NOT happy...

Reversible (L^2) vs nonreversible (L^∞) theory

- Where do we lose? We have $(\mathcal{P}^\ell \phi)^2 \lesssim |\phi|_\infty^2 \rho^{2\ell}$ because of **(UE)**.
- If we had a reversible process, we could hope for

$$(\mathcal{P}^\ell \phi)^2 \lesssim |\phi|_{L^2}^2 \rho^{2\ell}.$$

- Solution: we can sacrifice *some of* the geometric ergodicity if we have some *regularization* of $\mathcal{P}$: for $\ell \geq 1$

$$|\mathcal{P}^\ell \phi|_\infty \leq |\mathcal{P}\phi|_\infty$$

and for a nice state space (hereafter $\mathcal{S} = \mathbb{R}$)

$$|\mathcal{P}\phi(x)| = |\int_{\mathcal{S}} \phi(y)\mathcal{P}(x, dy)| \stackrel{!}{=} |\int_{\mathcal{R}} \phi(y)\mathcal{P}(x, y)dy| \leq \sup_{x\in\mathcal{S}, y\in \mathrm{supp}\phi} \mathcal{P}(x, y)|\phi|_{L^1(\mathrm{Leb})}$$

Toward a compromise between the action of $\mathcal{P}$ and **(UE)**

- Under this additional regularity property on $\mathcal{P}$

$$|\mathcal{P}^\ell \phi(x)| \lesssim |\phi|_{L^1(\mathrm{Leb})} \wedge |\phi|_\infty \rho^\ell.$$

- Trade-off in the control of the covariance term II of $\mathbb{E}_\nu\big[Y_m(\phi)^2\big]$:

$$\begin{aligned} II &\lesssim 2^m \sum_{\ell=1}^{\ell^\star} 2^\ell |\phi|^2_{L^1(\mathrm{Leb})} + 2^m \sum_{\ell=\ell^\star+1}^{m} (2\rho^2)^2 |\phi|^2_\infty \\ &\lesssim 2^m \inf_{\ell^\star} \big(2^{\ell^\star} |\phi|^2_{L^1(\mathrm{Leb})} + (2\rho^2)^{\ell^\star} |\phi|^2_\infty\big) \end{aligned}$$

and optimise in $\ell^\star$...

- Nonparametrically-wise, we are happy again: if $\phi(x) = h^{-1}\psi(h^{-1}x)$, we obtain the refinement of the bad constant $|\phi|^2_\infty$ into a constant that behaves like $|\phi|_{L^1(\mathrm{Leb})}|\phi|_\infty$.

BMC

Definition

A bifurcating Markov chain is a family $(X_u)_{u\in\mathbb{T}}$ of random variables with value in $(\mathcal{S},\mathfrak{S})$ such that X_u is $\mathcal{F}_{|u|}$-measurable for every $u \in \mathbb{T}$ and

$$\mathbb{E}\Big[\prod_{u\in\mathbb{G}_m} g_u(X_u, X_{u0}, X_{u1})\Big|\mathcal{F}_m\Big] = \prod_{u\in\mathbb{G}_m} \mathcal{P}g_u(X_u)$$

for every $m \geq 0$ and $(g_u)_{u\in\mathbb{G}_m}$, where
$\mathcal{P}g(x) = \int_{\mathcal{S}\times\mathcal{S}} g(x,y,z)\mathcal{P}(x, dy\,dz)$

Essential object: the mean transition

- The tagged-branch chain $(Y_m)_{m\geq 0}$: $Y_0 = X_\emptyset$ and for $m \geq 1$,

$$Y_m = X_{\emptyset\epsilon_1\cdots\epsilon_m},$$

  $(\epsilon_m)_{m\geq 1}$ IID Bernoulli with parameter $1/2$, independent of $(X_u)_{u\in\mathbb{T}}$.
- Transition (mean transition)

$$\mathcal{Q} = (\mathcal{P}_0 + \mathcal{P}_1)\,/2,$$

  obtained from the marginals $\mathcal{P}_0(x, dy) = \int_{z\in\mathcal{S}} \mathcal{P}(x, dy\, dz)$ and $\mathcal{P}_1(x, dz) = \int_{y\in\mathcal{S}} \mathcal{P}(x, dy\, dz)$.

Digest

- Guyon (2007) proves that if $(Y_m)_{m\geq 0}$ is ergodic with invariant measure ν, then

$$\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(X_u) \to \int_{\mathcal{S}} g(x)\nu(dx)$$

  holds almost-surely as $n\to\infty$.
- We also have

$$\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} g(X_u, X_{u0}, X_{u1}) \to \int_{\mathcal{S}} \mathcal{P}g(x)\nu(dx)$$

  almost-surely as $n\to\infty$.
- These results are appended with central limit theorems.

Toward statistical inference

- $\mathcal{D} \subseteq \mathcal{S}$ that will be later needed for statistical purposes.
- Mean transition $\mathcal{Q} = \frac{1}{2}(\mathcal{P}_0 + \mathcal{P}_1)$.

Assumptions

- **Assumption (D)** The family $\{\mathcal{Q}(x, dy), x \in \mathcal{S}\}$ is dominated:

$$\mathcal{Q}(x, dy) = \mathcal{Q}(x, y)\mathfrak{n}(dy) \text{ for every } x \in \mathcal{S},$$

for some $\mathcal{Q} : \mathcal{S}^2 \to [0, \infty)$ such that

$$|\mathcal{Q}|_{\mathcal{D}} = \sup_{x \in \mathcal{S}, y \in \mathcal{D}} \mathcal{Q}(x, y) < \infty.$$

- **Assumption (UE)** $\mathcal{Q}$ admits a unique invariant probability measure ν and there exist $R > 0$ and $0 < \rho < 1/2$ such that

$$\left|\mathcal{Q}^m g(x) - \nu(g)\right| \leq R|g|_\infty \rho^m, \quad x \in \mathcal{S}, \quad m \geq 0,$$

Variance definitions

- For $g : \mathcal{S}^d \to \mathbb{R}$, define $\Sigma_{1,1}(g) = |g|_2^2$ and for $n \geq 2$,

$$\Sigma_{1,n}(g) = |g|_2^2 + \min_{1 \leq \ell \leq n-1} \left(|g|_1^2 2^{\ell} + |g|_\infty^2 2^{-\ell} \right). \tag{1}$$

- Define also $\Sigma_{2,1}(g) = |\mathcal{P} g^2|_1$ and for $n \geq 2$,

$$\Sigma_{2,n}(g) = |\mathcal{P} g^2|_1 + \min_{1 \leq \ell \leq n-1} \left(|\mathcal{P} g|_1^2 2^{\ell} + |\mathcal{P} g|_\infty^2 2^{-\ell} \right). \tag{2}$$

One-step deviations

Theorem

Under **(D)** *and* **(UE)**, *for every* $n \geq 1$*:*

(i) *For any* $\delta > 0$ *such that* $\delta \geq 4R|g|_\infty|\mathbb{G}_n|^{-1}$, *we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(X_u) - \nu(g) \geq \delta\Big) \leq \exp\Big(\frac{-|\mathbb{G}_n|\delta^2}{\kappa_1\Sigma_{1,n}(g) + \kappa_2|g|_\infty\delta}\Big).$$

(ii) *For any* $\delta > 0$ *such that* $\delta \geq 4R(1-2\rho)^{-1}|g|_\infty|\mathbb{T}_n|^{-1}$, *we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} g(X_u) - \nu(g) \geq \delta\Big) \leq \exp\Big(\frac{-|\mathbb{T}_n|\delta^2}{\kappa_3\Sigma_{1,n}(g) + \kappa_4|g|_\infty\delta}\Big).$$

Two-steps deviations

Theorem

Under **(D)** *and* **(UE)**, *for every* $n \geq 2$*:*

(i) *For any* $\delta > 0$ *such that* $\delta \geq 4R|\mathcal{P}g|_\infty |\mathbb{G}_n|^{-1}$, *we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(X_u, X_{u0}, X_{u1}) - \nu(\mathcal{P}g) \geq \delta\Big) \leq \exp\Big(\frac{-|\mathbb{G}_n|\delta^2}{\kappa_1\Sigma_{2,n}(g) + \kappa_2|g|_\infty\delta}\Big)$$

(ii) *For any* $\delta > 0$ *such that* $\delta \geq 4(nR|\mathcal{P}g|_\infty + |g|_\infty)|\mathbb{T}_{n-1}|^{-1}$, *we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{u\in\mathbb{T}_{n-1}} g(X_u, X_{u0}, X_{u1}) - \nu(\mathcal{P}g) \geq \delta\Big)$$

$$\leq \exp\Big(\frac{-n^{-1}|\mathbb{T}_{n-1}|\delta^2}{\kappa_1\Sigma_{2,n-1}(g) + \kappa_2|g|_\infty\delta}\Big).$$

Statistical inference

- From now on $(\mathcal{S}, \mathfrak{S}) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ and $\mathcal{D} \subset \mathcal{S}$ compact interval
- **Assumption (S)** The family $\{\mathcal{P}(x, dy\,dz), x \in \mathcal{S}\}$ is dominated w.r.t. the Lebesgue measure:

$$\mathcal{P}(x, dy\,dz) = \mathcal{P}(x, y, z) dy\,dz \;\; \text{for every} \;\; x \in \mathcal{S}$$

  for some $\mathcal{P} : \mathcal{S}^3 \to [0, \infty)$ such that

$$|\mathcal{P}|_{\mathcal{D}} = \sup_{(x,y,z) \in \mathcal{D}^3} |\mathcal{P}(x, y, z)| < \infty.$$

Statistical inference (cont.)

- For some $n \geq 1$, we observe $(X_u)_{u \in \mathbb{T}_n}$
- Under **(D), (S)**, with $\mathfrak{n}(dy) = dy$, we have
  - $\mathcal{P}(x, dy\, dz) = \mathcal{P}(x, y, z) dy\, dz$
  - $\mathcal{Q}(x, dy) = \mathcal{Q}(x, y) dy$
  - $\nu(dx) = \nu(x) dx$
- Goal: estimate nonparametrically $x \rightsquigarrow \nu(x)$, $(x, y) \rightsquigarrow \mathcal{Q}(x, y)$ and $(x, y, z) \rightsquigarrow \mathcal{P}(x, y, z)$ for $x, y, z \in \mathcal{D}$.

Nonparametric estimation of $\nu(x)$

- For a σ-regular wavelet basis, we approximate the representation

$$\nu(x) = \sum_{\lambda \in \Lambda} \nu_\lambda \psi^1_\lambda(x), \quad \nu_\lambda = \langle \nu, \psi^1_\lambda \rangle$$

by

$$\widehat{\nu}_n(x) = \sum_{|\lambda| \leq J} \widehat{\nu}_{\lambda,n} \psi^1_\lambda(x),$$

with

$$\widehat{\nu}_{\lambda,n} = \mathcal{T}_{\lambda,\eta}\Big(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \psi^1_\lambda(X_u)\Big).$$

- $\mathcal{T}_{\lambda,\eta}(x) = x \mathbf{1}_{|x| \geq \eta}$ threshold operator (with $\mathcal{T}_{\lambda,\eta}(x) = x$ for the low frequency part.
- $\widehat{\nu}_n$ is specified by the maximal resolution level J and the threshold η.

Theorem

Under **(D)** *and* **(UE)** *with* $\mathfrak{n}(dx) = dx$*, specify* $\widehat{\nu}_n$ *with*

$$J = \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \quad \text{and} \quad \eta = c\sqrt{\log |\mathbb{T}_n| / |\mathbb{T}_n|}$$

for some $c > 0$*. For every* $\pi \in (0, \infty]$*,* $s \in (1/\pi, \sigma]$ *and* $p \geq 1$*, for large enough* n *and* c*, the following estimate holds*

$$\Big(\mathbb{E}\big[\|\widehat{\nu}_n - \nu\|^p_{L^p(\mathcal{D})}\big]\Big)^{1/p} \lesssim \Big(\frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_1(s,p,\pi)},$$

with $\alpha_1(s,p,\pi) = \min\big\{\frac{s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi}\big\}$*, up to a constant that depends on* $s, p, \pi, \|\nu\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})}$*,* ρ*,* R *and* $|\mathcal{Q}|_{\mathcal{D}}$ *and that is continuous in its arguments.*

- The estimator $\widehat{\nu}_n$ is *smooth-adaptive* in the following sense: for every $s_0 > 0$, $0 < \rho_0 < 1/2$, $R_0 > 0$ and $\mathcal{Q}_0 > 0$, define the sets $\mathcal{A}(s_0) = \{(s, \pi), s \geq s_0, s_0 \geq 1/\pi\}$ and

$$\mathcal{Q}(\rho_0, R_0, \mathcal{Q}_0) = \{\mathcal{Q} \text{ such that } \rho \leq \rho_0, R \leq R_0, |\mathcal{Q}|_{\mathcal{D}}, \leq \mathcal{Q}_0\},$$

where $\mathcal{Q}$ is taken among mean transitions for which **(UE)** holds. Then, for every $C > 0$, there exists $c^\star = c^\star(\mathcal{D}, p, s_0, \rho_0, R_0, \mathcal{Q}_0, C)$ such that $\widehat{\nu}_n$ specified with $c^\star$ satisfies

$$\sup_n \sup_{(s,\pi)\in\mathcal{A}(s_0)} \sup_{\nu,\mathcal{Q}} \Big(\frac{|\mathbb{T}_n|}{\log|\mathbb{T}_n|}\Big)^{p\alpha_1(s,p,\pi)} \mathbb{E}\big[\|\widehat{\nu}_n - \nu\|^p_{L^p(\mathcal{D})}\big] < \infty$$

where the supremum is taken among $(\nu, \mathcal{Q})$ such that $\nu\mathcal{Q} = \nu$ with $\mathcal{Q} \in \mathcal{Q}(\rho_0, R_0, \mathcal{Q}_0)$ and $\|\nu\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})} \leq C$.

Nonparametric estimation of the mean transition $\mathcal{Q}(x, y)$

- First estimate
$$f_{\mathcal{Q}}(x, y) = \nu(x)\mathcal{Q}(x, y)$$
of the distribution of (X_{u^-}, X_u) (when $\mathcal{L}(X_\emptyset) = \nu$) by
$$\widehat{f}_n(x, y) = \sum_{|\lambda| \leq J} \widehat{f}_{\lambda,n}\psi_\lambda^2(x, y),$$
with
$$\widehat{f}_{\lambda,n} = \mathcal{T}_{\lambda,\eta}\Big(\frac{1}{|\mathbb{T}_n^\star|}\sum_{u \in \mathbb{T}_n^\star}\psi_\lambda^2(X_{u^-}, X_u)\Big),$$
($\mathbb{T}_n^\star = \mathbb{T}_n \setminus \mathbb{G}_0$.)
- Estimate $\mathcal{Q}(x, y)$ via
$$\widehat{\mathcal{Q}}_n(x, y) = \frac{\widehat{f}_n(x, y)}{\max\{\widehat{\nu}_n(x), \varpi\}} \qquad (3)$$
for some $\varpi > 0$.
- Thus $\widehat{\mathcal{Q}}_n$ is specified by J, η and ϖ.

Theorem

Under **(D)** *and* **(UE)** *with* $\mathfrak{n}(dx) = dx$, *specify* $\widehat{\mathcal{Q}}_n$ *with*

$$J = \tfrac{1}{2} \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \quad \text{and} \quad \eta = c\sqrt{(\log |\mathbb{T}_n|)^2 / |\mathbb{T}_n|}$$

for some $c > 0$ *and* $\varpi > 0$. *For every* $\pi \in [1, \infty]$, $s \in (2/\pi, \sigma]$ *and* $p \geq 1$, *for large enough* n *and* c *and small enough* ϖ, *the following estimate holds*

$$\Big(\mathbb{E}\big[\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|^p_{L^p(\mathcal{D}^2)}\big]\Big)^{1/p} \lesssim \Big(\frac{(\log |\mathbb{T}_n|)^2}{|\mathbb{T}_n|}\Big)^{\alpha_2(s,p,\pi)}, \qquad (4)$$

with $\alpha_2(s, p, \pi) = \min\left\{\frac{s}{2s+2}, \frac{s/2+1/p-1/\pi}{s+1-2/\pi}\right\}$, *provided* $m(\nu) = \inf_{x \in \mathcal{D}} \nu(x) \geq \varpi > 0$ *and up to a constant that depends on* $s, p, \pi, \|\mathcal{Q}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^2)}$, $m(\nu)$ *and that is continuous in its arguments.*

- This rate is moreover (nearly) optimal: define $\varepsilon_2 = s\pi - (p - \pi)$. We have

$$\inf_{\widehat{\mathcal{Q}}_n} \sup_{\mathcal{Q}} \Big(\mathbb{E}\big[\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|^p_{L^p(\mathcal{D}^2)}\big]\Big)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_2(s,p,\pi)} & \text{if} \quad \varepsilon_2 > 0 \\ \Big(\dfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_2(s,p,\pi)} & \text{if} \quad \varepsilon_2 \leq 0. \end{cases}$$

  where the infimum is taken among all estimators of $\mathcal{Q}$ based on $(X_u)_{u\in\mathbb{T}_n}$ and the supremum is taken among all $\mathcal{Q}$ such that $\|\mathcal{Q}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^2)} \leq C$ and $m(\nu) \geq C'$ for some $C, C' > 0$.
- The calibration of the threshold ϖ needed to define $\widehat{\mathcal{Q}}_n$ requires an *a priori* bound on $m(\nu)$.
- The $(\log|\mathbb{T}_n|)^2$ comes from the slow term in the deviations inequality and from the wavelet thresholding procedure.

Nonparametric estimation of the transition $\mathcal{P}(x, y, z)$

- First estimate the density

$$f_{\mathcal{P}}(x, y, z) = \nu(x)\mathcal{P}(x, y, z)$$

of the distribution of (X_u, X_{u0}, X_{u1}) (when $\mathcal{L}(X_\emptyset) = \nu$) by

$$\widehat{f}_n(x, y, z) = \sum_{|\lambda| \leq J} \widehat{f}_{\lambda,n}\psi_\lambda^3(x, y, z),$$

with

$$\widehat{f}_{\lambda,n} = \mathcal{T}_{\lambda,\eta}\Big(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{u \in \mathbb{T}_{n-1}} \psi_\lambda^3(X_u, X_{u0}, X_{u1})\Big),$$

- Next estimate the density $\mathcal{P}$ by

$$\widehat{\mathcal{P}}_n(x, y, z) = \frac{\widehat{f}_n(x, y, z)}{\max\{\widehat{\nu}_n(x), \varpi\}} \qquad (5)$$

for some threshold $\varpi > 0$.

- Thus the estimator $\widehat{\mathcal{P}}_n$ is specified by J, η and ϖ.

Theorem

Under **(D), (UE), (S)**. *Specify* $\widehat{\mathcal{P}}_n$ *with*

$$J = \tfrac{1}{3}\log_2 \frac{|\mathbb{T}_n|}{\log|\mathbb{T}_n|} \quad \text{and} \quad \eta = c\sqrt{(\log|\mathbb{T}_n|)^2/|\mathbb{T}_n|}$$

for some $c > 0$ *and* $\varpi > 0$*. For every* $\pi \in [1,\infty]$, $s \in (3/\pi, \sigma]$ *and* $p \geq 1$*, for large enough* n *and* c *and small enough* ϖ*, the following estimate holds*

$$\left(\mathbb{E}\big[\|\widehat{\mathcal{P}}_n - \mathcal{P}\|^p_{L^p(\mathcal{D}^3)}\big]\right)^{1/p} \lesssim \left(\frac{(\log|\mathbb{T}_n|)^2}{|\mathbb{T}_n|}\right)^{\alpha_3(s,p,\pi)}, \qquad (6)$$

with $\alpha_3(s,p,\pi) = \min\left\{\frac{s}{2s+3}, \frac{s/3+1/p-1/\pi}{2s/3+1-2/\pi}\right\}$*, provided* $m(\nu) \geq \varpi > 0$ *and up to a constant that depends on* $s, p, \pi, \|\mathcal{P}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^3)}$ *and* $m(\nu)$ *and that is continuous in its arguments.*

- This rate is moreover (nearly) optimal: define $\varepsilon_3 = \frac{s\pi}{3} - \frac{p-\pi}{2}$. We have

$$\inf_{\widehat{\mathcal{P}}_n} \sup_{\mathcal{P}} \Big(\mathbb{E}\big[\|\widehat{\mathcal{P}}_n - \mathcal{P}\|_{L^p(\mathcal{D}^3)}^p\big]\Big)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_3(s,p,\pi)} & \text{if} \quad \varepsilon_3 > 0 \\ \Big(\dfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_3(s,p,\pi)} & \text{if} \quad \varepsilon_3 \leq 0, \end{cases}$$

where the infimum is taken among all estimators of $\mathcal{P}$ based on $(X_u)_{u\in\mathbb{T}_n}$ and the supremum is taken among all $\mathcal{P}$ such that $\|\mathcal{P}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^3)} \leq C$ and $m(\nu) \geq C'$ for some $C, C' > 0$.

Application: cell division by growth

- To each node (or cell) $u \in \mathbb{T}$, we associate as trait $X_u \in \mathcal{S} \subset (0, \infty)$ the size at birth of the cell u.
- Each cell grows exponentially with a common rate $\tau > 0$.
- A cell of size x splits into two newborn cells of size $x/2$ each (thus $X_{u0} = X_{u1}$ here), with a size-dependent splitting rate $B(x)$ for some $B : \mathcal{S} \to [0, \infty)$.
- Two newborn cells start a new life independently of each other.

- If ζ_u denotes the lifetime of the cell u, we thus have

$$\mathbb{P}\big(\zeta_u \in [t, t+dt)\big|\zeta_u \geq t, X_u = x\big) = B\big(x \exp(\tau t)\big)dt$$

  and

$$X_u = \tfrac{1}{2} X_{u^-} \exp(\tau \zeta_{u^-})$$

  that entirely determine the evolution of the population.
- Goal: estimate $x \rightsquigarrow B(x)$ for $x \in \mathcal{D}$ where $\mathcal{D} \subset \mathcal{S}$ is a given compact interval.

- The process $(X_u)_{u\in\mathbb{T}}$ is a bifurcating Markov chain with state space $\mathcal{S}$ and $\mathbb{T}$-transition

$$\mathcal{P}_B(x, dy\, dz) = \mathbb{P}\big(X_{u0} \in dy, X_{u1} \in dz \,|X_{u^-} = x\big).$$

- it is not difficult to check that

$$\mathcal{P}_B(x, dy\, dz) = Q_B(x, dy) \otimes \delta_y(dz)$$

and

$$Q_B(x, dy) = \frac{B(2y)}{\tau y} \exp\Big(- \int_{x/2}^{y} \frac{B(2s)}{\tau s} ds\Big) \mathbf{1}_{\{y\geq x/2\}} dy.$$

- $x \rightsquigarrow B(x)$ is continuous implies **(D)** with $\mathcal{Q} = Q_B$ and $\mathfrak{n}(dx) = dx$.
- Let $\mathcal{S} = (0, C]$. Pick $r \in \mathcal{S}$ and $L > 0$ and let

$$\mathcal{C}(r, L) = \Big\{ B, \int^{\sup \mathcal{S}} \frac{B(x)}{x} dx = \infty, \ \int_0^r \frac{B(x)}{x} dx \leq L \Big\}.$$

- We comply with **(UE)** for $\mathcal{Q} = Q_B$ with $0 < \rho < 1/2$ if $r > \sup \mathcal{S}/2$ and $0 < L < \tau \log 2$.
- We know by Proposition 2 in Doumic *et al.* (2015) that

$$B(x) = \frac{\tau x}{2} \frac{\nu_B(x/2)}{\int_{x/2}^{x} \nu_B(z) dz},$$

where ν_B denotes the unique invariant probability of the transition $\mathcal{Q} = Q_B$.

- For a given compact interval $\mathcal{D} \subset \mathcal{S}$, define

$$\widehat{B}_n(x) = \frac{\tau x}{2} \frac{\widehat{\nu}_n(x/2)}{\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \mathbf{1}_{\{x/2 \le X_u < x\}}\right) \vee \varpi}, \tag{7}$$

where $\widehat{\nu}_n$ is the wavelet thresholding estimator specified by a maximal resolution level J and a threshold η and $\varpi > 0$

Theorem

Specify $\widehat{B}_n$ *with*

$$J = \tfrac{1}{2}\log_2 \frac{|\mathbb{T}_n|}{\log|\mathbb{T}_n|} \quad \text{and} \quad \eta = c\sqrt{\log|\mathbb{T}_n|/|\mathbb{T}_n|}$$

for some $c > 0$. *For every* $B \in \mathcal{C}(r, L)$, *for every* $\pi \in (0, \infty]$, $s \in (1/\pi, \sigma]$ *and* $p \geq 1$, *for large enough* n *and* c *and small enough* ϖ, *the following estimate holds*

$$\Big(\mathbb{E}\big[\|\widehat{B}_n - B\|^p_{L^p(\mathcal{D})}\big]\Big)^{1/p} \lesssim \Big(\frac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_1(s,p,\pi)},$$

provided that $\inf \mathcal{D} \leq r/2$, *with* $\alpha_1(s, p, \pi) = \min\big\{\frac{2s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi}\big\}$, *up to a constant that depends on* $s, p, \pi, \|B\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})}$, r *and* L *and that is continuous in its arguments.*

- This rate is moreover (nearly) optimal: define $\varepsilon_1 = s\pi - \frac{1}{2}(p-\pi)$. We have

$$\inf_{\widehat{B}_n} \sup_{B} \left(\mathbb{E}\left[\|\widehat{B}_n - B\|_{L^p(\mathcal{D})}^p\right]\right)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_1(s,p,\pi)} & \text{if} \quad \varepsilon_1 > 0 \\ \left(\dfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\right)^{\alpha_1(s,p,\pi)} & \text{if} \quad \varepsilon_1 \leq 0, \end{cases}$$

  where the infimum is taken among all estimators of B based on $(X_u)_{u\in\mathbb{T}_n}$ and the supremum is taken among all $B \in \mathcal{C}(r,L)$ such that $\|B\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})} \leq C$.
- We improve on the results of Doumic *et al.* in two directions: smoothness adaptation + minimax results
- Quite stringent restriction that $\mathcal{S}$ is bounded

- We consider a perturbation of the baseline splitting rate $\widetilde{B}(x) = x/(5-x)$ over the range $x \in \mathcal{S} = (0,5)$ of the form

$$B(x) = \widetilde{B}(x) + \mathfrak{c}\, T\big(2^j(x - \tfrac{7}{2})\big)$$

  with $(\mathfrak{c}, j) = (3,1)$ or $(\mathfrak{c}, j) = (9,4)$, and where $T(x) = (1+x)\mathbf{1}_{\{-1 \leq x < 0\}} + (1-x)\mathbf{1}_{\{0 \leq x \leq 1\}}$ is a tent shaped function.
- the trial splitting rate with parameter $(\mathfrak{c}, j) = (9,4)$ is more localized around $7/2$ and higher than the one associated with parameter $(\mathfrak{c}, j) = (3,1)$.
- For a given B, we simulate $M = 100$ Monte Carlo trees up to the generation $n = 15$ with $\tau = 2$.

Numerical illustration

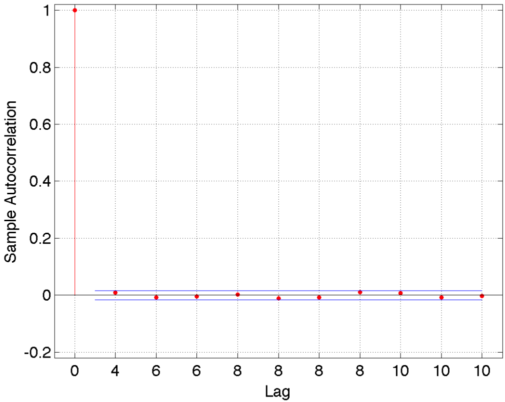

Figure : *Sample autocorrelation of ordered* $(X_{u0}, u \in \mathbb{G}_{n-1})$ *for* $n = 15$. *Note: due to the binary tree structure the lags are* $\{4, 6, 6, \ldots\}$. *As expected, we observe a fast decorrelation.*

- We implement the estimator $\widehat{B}_n$ defined by (7) using the Matlab wavelet toolbox.
- We use compactly supported Daubechies wavelets of order 8. $J := \frac{1}{2}\log_2(|\mathbb{T}_n|/\log|\mathbb{T}_n|)$ and we threshold the coefficients $\widehat{\nu}_{\lambda,n}$ which are too small by hard thresholding.
- We choose the threshold proportional to $\sqrt{\log|\mathbb{T}_n|/|\mathbb{T}_n|}$ (and we calibrate the constant to 10 or 15 for respectively the two trial splitting rates, mainly by visual inspection).
- We evaluate $\widehat{B}_n$ on a regular grid of $\mathcal{D} = [1.5, 4.8]$ with mesh $\Delta x = (|\mathbb{T}_n|)^{-1/2}$. For each sample we compute the empirical error

$$e_i = \frac{\|\widehat{B}_n^{(i)} - B\|_{\Delta x}}{\|B\|_{\Delta x}}, \quad i = 1, \dots, M,$$

where $\|\cdot\|_{\Delta x}$ denotes the discrete L^2-norm over the numerical sampling and sum up the results through the mean-empirical error $\bar{e} = M^{-1}\sum_{i=1}^{M} e_i$, together with the empirical standard deviation $\big(M^{-1}\sum_{i=1}^{M}(e_i - \bar{e})^2\big)^{1/2}$.

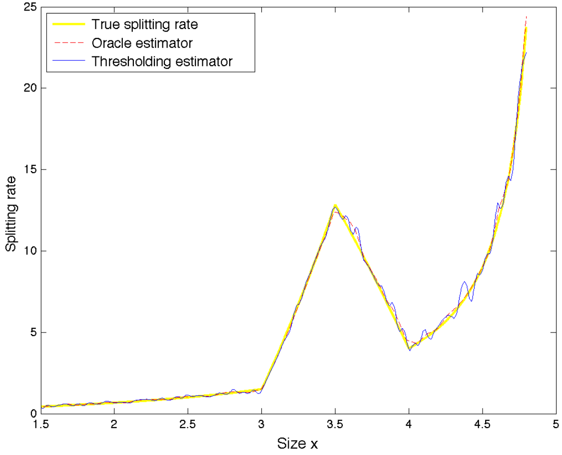

Figure : *Large spike: reconstruction of the trial splitting rate B specified by $(\mathfrak{c}, j) = (3, 1)$ over $\mathcal{D} = [1.5, 4.8]$ based on one sample $(X_u, u \in \mathbb{T}_n)$ for $n = 15$ (i.e. $\frac{1}{2}|\mathbb{T}_n| = 32\ 767$).*

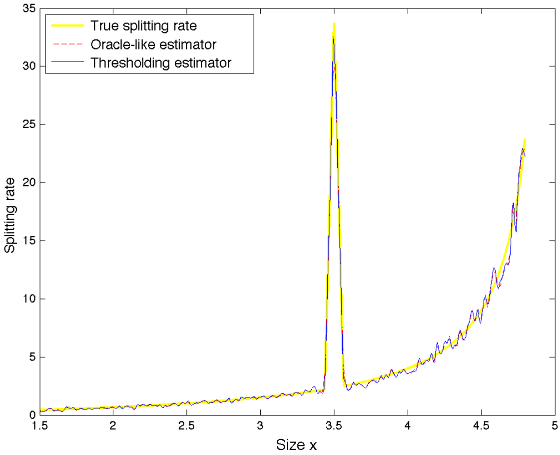

Figure : *High spike: reconstruction of the trial splitting rate* B *specified by* $(\mathfrak{c}, j) = (9, 4)$ *over* $\mathcal{D} = [1.5, 4.8]$ *based on one sample* $(X_u, u \in \mathbb{T}_n)$ *for* $n = 15$ *(i.e.* $\frac{1}{2}|\mathbb{T}_n| = 32\ 767$*).*

THANK YOU FOR YOUR ATTENTION!