Sharp Bernstein and Hoeffding type inequalities for regenerative Markov chains

Gabriela Ciołek

LTCI, Télécom ParisTech, Université Paris-Saclay,
46 Rue Barrault, 75013 Paris, France
gabrielaciolek@gmail.com

May 11, 2017

Outline

Atomic Markov chains

- Let $X = (X_n)_{n \in \mathbb{N}}$ be a homogeneous Markov chain on a countably generated state space $(E, \mathcal{E})$ with transition probability Π and initial probability ν.
- Chain X is assumed to be ψ -irreducible and aperiodic.

Regenerative Markov chain

We say that the chain X is regenerative, when there exists a measurable set A such that $\mu(A) > 0$ and $\Pi(x, .) = \Pi(y, .)$ for all $(x, y) \in A^2$

Atomic Markov chains

- Define the sequence of regeneration times $(\tau_A(j))_{j\geq 1}$.
- Let

$$\tau_A = \tau_A(1) = \inf\{n \geq 1 : X_n \in A\}$$

be the first time when the chain hits the regeneration set A and

$$\tau_A(j) = \inf\{n > \tau_A(j-1), X_n \in A\} \text{ for } j \geq 2.$$

- The segments of data are of the form:

$$\mathcal{B}_j = (X_{1+\tau_A(j)}, \cdots, X_{\tau_A(j+1)}),\ j \geq 1$$

and take values in the torus $\cup_{k=1}^{\infty} E^k$.
- By the strong Markov property blocks corresponding to the consecutive visitis of the chain to atom A are i.i.d.

Harris recurrent Markov chains

Harris recurrence

Assume that X is a ψ-irreducible Markov chain. We say that X is *Harris recurrent* iff, starting from any point $x \in E$ and any set such that $\psi(A) > 0$, we have

$$\mathbb{P}_x(\tau_A < +\infty) = 1.$$

We construct an artificial regeneration set via Nummelin technique.

Small set

We say that a set $S \in \mathcal{E}$ is small if there exists a parameter $\delta > 0$, a positive probability measure Φ supported by S and an integer $m \in N^*$ such that

$$\forall x \in S,\ A \in \mathcal{E} \ \ \Pi^m(x, A) \geq \delta\, \Phi(A), \tag{1}$$

where Π^m denotes the m-th iterate of the transition probability Π.

Nummelin's splitting technique

- Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of independent r.v.'s with parameter δ.
- We construct the bivriate chain $X^{\mathcal{M}} = (X_n, Y_n)_{n\in\mathbb{N}}$ with a joint distribution $\mathbb{P}_{\nu,\mathcal{M}}$.
- The construction relies on the mixture representation of Π on S, namely $\Pi(x, A) = \delta\Phi(A) + (1-\delta)\frac{\Pi(x,A)-\delta\Phi(A)}{1-\delta}$. It can be retrieved by the following randomization of the transition probability Π each time the chain X visits the set S. If $X_n \in S$ and
    - if $Y_n = 1$ (which happens with probability $\delta \in]0,1[$), then X_{n+1} is distributed according to the probability measure Φ,
    - if $Y_n = 0$ (that happens with probability $1-\delta$), then X_{n+1} is distributed according to the probability measure $(1-\delta)^{-1}(\Pi(X_n,\cdot) - \delta\Phi(\cdot))$.

  $\hat{S} = S \times \{1\}$ is an atom for the split chain.

Nummelin's splitting technique

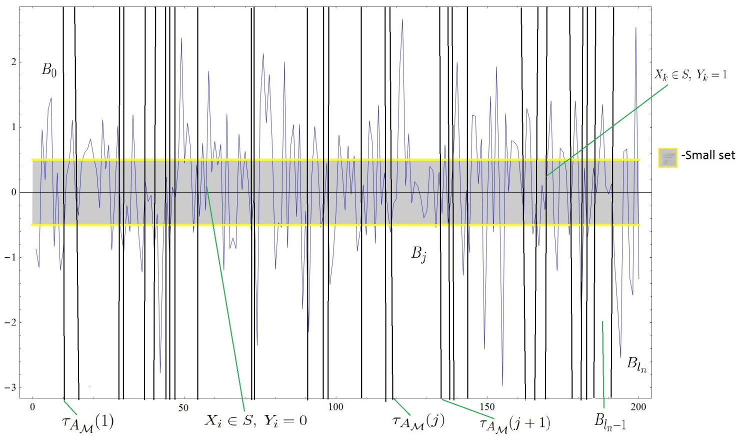

Figure: Regeneration block construction for AR(1) model.

We introduce the following notation for partial sums of the regeneration cycles $f(B_i) = \sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i)$. In the following, we assume that the mean inter-renewal time $\alpha = \mathbb{E}_A[\tau_A] < \infty$ and write $l_n = \sum_{i=1}^{n} \mathbb{I}\{X_i \in A\}$ for the total number of consecutive visits of the chain to the atom A. The regenerative approach is based on the following decomposition of the sum $\sum_{i=1}^{n} f(X_i)$:

$$\sum_{i=1}^{n} f(X_i) = \sum_{i=1}^{\lfloor \frac{n}{\alpha} \rfloor} f(B_i) + \Delta_n,$$

where

$$\Delta_n = \frac{1}{n}\sum_{i=1}^{\tau_A} f(X_i) + \frac{1}{n}\sum_{i=l_{n_1}}^{l_{n_2}} f(B_i) + \frac{1}{n}\sum_{i=\tau_A(l_n-1)}^{n} f(X_i),$$

where $l_{n_1} = \min\left(\lfloor \frac{n}{\alpha} \rfloor - 1, l_n - 1\right)$, $l_{n_2} = \max\left(\lfloor \frac{n}{\alpha} \rfloor - 1, l_n - 1\right)$ and

$$\sigma^2(f) = \frac{1}{\mathbb{E}_A(\tau_A)}\mathbb{E}_A\left(\sum_{i=1}^{\tau_A}\{f(X_i) - \mu(f)\}^2\right)$$

is the asymptotic variance.

In empirical processes theory for processes indexed by class of functions, it is important to assess the complexity of considered classes. The information about entropy of $\mathcal{F}$ helps us to inspect how large our class is.

Covering and uniform entropy number

The *covering number* $N_p(\epsilon, Q, \mathcal{F})$ is the minimal number of balls $\{g : \|g - f\|_{L^p(Q)} < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{F}$. The centers of the balls need not to belong to $\mathcal{F}$, but they should have finite norms. The *entropy* (without bracketing) is the logarithm of the covering number. We define uniform entropy number as $N_p(\epsilon, \mathcal{F}) = \sup_Q N_p(\epsilon, Q, \mathcal{F})$, where the supremum is taken over all discrete probability measures Q.

We impose the following conditions on the chain:

A1. (Bernstein's block moment condition) There exists a positive constant M such that for any $p \geq 2$ and for every $f \in \mathcal{F}$

$$\mathbb{E}_A |f(B_1)|^p \leq \frac{1}{2} p! \sigma^2(f) M^{p-2}. \tag{2}$$

A2. (Block length moment assumption) There exists a positive constant N such that for any $p \geq 2$

$$\mathbb{E}_A |\tau_A|^p \leq N^p. \tag{3}$$

A3. (Non-regenerative block exponential moment assumption) There exists $\lambda_0 > 0$ such that for every $f \in \mathcal{F}$ we have $\mathbb{E}_\nu \left[\exp\left[\lambda_0 \left|\sum_{i=1}^{\tau_A} f(X_i)\right|\right]\right] < \infty$.

A4. (Exponential block moment assumption) There exists $\lambda_1 > 0$ such that for every $f \in \mathcal{F}$ we have $\mathbb{E}_A [\exp[\lambda_1 |f(B_1)|]] < \infty$.

A5. (uniform entropy number condition) $\mathcal{N}_2(\epsilon, \mathcal{F}) < \infty$.

Bernstein type maximal inequality for regenerative Markov chains

Assume that $X = (X_n)_{n\in\mathbb{N}}$ is a regenerative positive recurrent Markov chain. Then, under assumptions A1-A5 and for $\epsilon < x$, we have for some positive explicit constants $L,\ R,\ > 0$ and any $q_1, q_2 > 1$, and n large enough

$$\mathbb{P}_\nu\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\left|\sum_{i=1}^{n} f(X_i) - \mu(f)\right| \geq x\right] \leq N_2(\epsilon,\mathcal{F})\left\{2\exp\left[-\frac{(x-2\epsilon)^2 n}{8\left(\frac{\sigma_m^2}{\alpha}+\frac{M(x-2\epsilon)}{n}\right)}\right]\right.$$
$$+\, C_1\exp\left[-\frac{(x-2\epsilon)n}{6}\right] + C_2\exp\left[-\frac{(x-2\epsilon)n}{6}\right]$$
$$\left.+\exp\left[\frac{1}{q_1(2q_1-2)}-\frac{(x-2\epsilon)n^{1/2}}{6Lq_1}\right]+\exp\left[\frac{1}{q_2(2q_2-2)}-\frac{(x-2\epsilon)n^{1/2}}{6Rq_2}\right]\right\}, \qquad (4)$$

where $C_1, C_2,\ L,\ R$ can be explicitly computed. F is an envelope function for $\mathcal{F}$.

Remark

Notice that our bound is a deviation bound in that it holds only for n large enough. This is due to the control of the covering functions (under $\mathbb{P}_n$) by a control under $\mathbb{P}$. However, by making additional assumptions on the regularity of the class of functions and by choosing the adequate norm, it is possible to obtain by the same arguments an exponential inequality valid for any n. Indeed, if $\mathcal{F}$ belongs to a ball of a Hölder space $\mathbb{C}^P(E')$ on a compact set E' of an Euclidean space endowed with the norm

$$||f||_{\mathbb{C}^P(E')} = \sup_{x \in E'} |f(x)| + \sup_{x_1 \in E', x_2 \in E'} \left(\frac{f(x_1) - f(x_2)}{d(x_1, x_2)^p} \right),$$

then one can obtain concentration maximal inequality.

Sketch of the proof

For the simplicity's sake we introduce one piece of notation $\bar{f}(x) = f(x) - \mu(f)$. Notice that as $n \to \infty$ we have with $\mathbb{P}_\nu$-probability one that $l_n \sim \left\lfloor \frac{n}{\alpha} \right\rfloor$. Thus, we consider the sum of random variables of the following form

$$Z_n(\bar{f}) = \frac{1}{n} \sum_{i=1}^{\lfloor \frac{n}{\alpha} \rfloor} \bar{f}(B_j). \tag{5}$$

Furthermore, we have that $S_n(\bar{f}) = Z_n(\bar{f}) + \Delta_n(\bar{f})$.

Note that

$$\mathbb{P}_A\left[\frac{1}{n}\left|\sum_{i=1}^{\lfloor \frac{n}{\alpha} \rfloor} \bar{f}(B_i)\right| \geq x\right] \leq 2\exp\left[-\frac{x^2 n}{8\left(\frac{\sigma^2(f)}{\alpha}+\frac{Mx}{n}\right)}\right] \tag{6}$$

since $\bar{f}(B_i),\ i=1,\cdots,\lfloor \frac{n}{\alpha} \rfloor$ are independent and identically distributed sub-exponential random variables.

Control the remainder term Δ_n is challenging. We want to bound the tail probabilities:

$$\begin{aligned}
&\mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=1}^{\tau_A}\bar{f}(X_i)+\frac{1}{n}\sum_{i=l_{n_1}}^{l_{n_2}}\bar{f}(B_i)+\frac{1}{n}\sum_{i=\tau_A(l_n-1)}^{n}\bar{f}(X_i)\right|\geq x\right\}\\
&\leq \mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=1}^{\tau_A}\bar{f}(X_i)\right|\geq\frac{x}{6}\right\}+\mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=l_{n_1}}^{l_{n_2}}\bar{f}(B_i)\right|\geq\frac{x}{6}\right\}\\
&+\mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=\tau_A(l_n-1)}^{n}\bar{f}(X_i)\right|\geq\frac{x}{6}\right\}. \qquad (7)
\end{aligned}$$

First and the last terms on the right hand side of (7) can be easily controlled by the Markov's inequality. In order to control the middle term, firstly note that $\mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=l_{n_1}}^{l_{n_2}}\bar{f}(B_i)\right|\geq\frac{x}{6}\right\}$ can be written as

$$\mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=l_{n_1}}^{l_{n_2}}\bar{f}(B_i)\right|\geq\frac{x}{6}\right\}=\mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=l_n-1}^{\lfloor\frac{n}{\alpha}\rfloor-1}\bar{f}(B_i)\mathbb{I}_{\{l_n<\lfloor\frac{n}{\alpha}\rfloor\}}\right|\geq\frac{x}{6}\right\}$$
$$+\,\mathbb{P}_\nu\left\{\left|\frac{1}{n}\sum_{i=\lfloor\frac{n}{\alpha}\rfloor-1}^{l_n-1}\bar{f}(B_i)\mathbb{I}_{\{l_n>\lfloor\frac{n}{\alpha}\rfloor\}}\right|\geq\frac{x}{6}\right\}$$

The control of the middle term comes down to control of the moment generating functions of the processes (technical, see the proof of Ciołek and Bertail (2017) for details)

$$\frac{1}{n}\sum_{i=l_n-1}^{\lfloor \frac{n}{\alpha} \rfloor -1} \bar{f}(B_i)\mathbb{I}_{\{l_n<\lfloor \frac{n}{\alpha} \rfloor\}}$$

and

$$\frac{1}{n}\sum_{i=\lfloor \frac{n}{\alpha} \rfloor -1}^{l_n-1} \bar{f}(B_i)\mathbb{I}_{\{l_n<\lfloor \frac{n}{\alpha} \rfloor\}}$$

We obtain the maximal inequality by applying similar arguments like in Pollard (1984) and Kosorok (2008). We obtain that

$$\mathbb{P}_\nu\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}(f(X_i)-\mu(f))\right|\geq x\right]$$
$$\leq N_2(\epsilon,\mathcal{F})\max_{j\in N_2(\epsilon,\mathcal{F})}\mathbb{P}_\nu\left\{\frac{1}{n}\sum_{i=1}^{n}|h_j(X_i)-\mu(h_j)|\geq x-2\epsilon\right\}$$

where $h_1, h_2, \cdots, h_W$ are functions such that $W = N_2(\epsilon,\mathcal{F})$.

We can obtain even sharper upper bound when class $\mathcal{F}$ is uniformly bounded. In the following, we will show that it is possible to get a Hoeffding's type inequality and have a stronger control of moments of the sum $S_n(f)$ which is a natural consequence of uniform boundedness assumption imposed on $\mathcal{F}$.

A6. Class of functions $\mathcal{F}$ is uniformly bounded.

Hoeffding type maximal inequality for regenerative Markov chains

Assume that $X = (X_n)_{n\in\mathbb{N}}$ is a regenerative positive recurrent Markov chain. Then, under assumptions A1-A6 and for $\epsilon < x$, we have for some positive explicit constants $L,\ R,\ D > 0$ and any $q_1, q_2 > 1$

$$\mathbb{P}_\nu\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\left|\sum_{i=1}^{n}\frac{f(X_i)-\mu(f)}{\sigma(f)}\right|\geq x\right]\leq N_2\left(\epsilon,\mathcal{F}\right)\left\{2\exp\left[-\frac{(x-2\epsilon)^2n^2}{8\left\lfloor\frac{\sigma_m^2}{\alpha}\right\rfloor D^2}\right]\right.$$
$$+\,C_1\exp\left[-\frac{(x-2\epsilon)n}{6}\right]+C_2\exp\left[-\frac{(x-2\epsilon)n}{6}\right]$$
$$\left.+\exp\left[\frac{1}{q_1(2q_1-2)}-\frac{(x-2\epsilon)n^{1/2}}{6Lq_1}\right]+\exp\left[\frac{1}{q_2(2q_2-2)}-\frac{(x-2\epsilon)n^{1/2}}{6Rq_2}\right]\right\},$$

where C_1 and C_2 are constants that can be explicitly computed. F is an envelope function for $\mathcal{F}$.

It is noteworthy that presented theorems are also valid in Harris recurrent case under slightly modified assumptions. It is well known that it is possible to retrieve all regeneration techniques also in Harris case via the Nummelin splitting technique which allows to extend the probabilistic structure of any chain in order to artificially construct a regeneration set.

We will formulate Bernstein type maximal inequality for unbounded classes of functions in Harris recurrent case. We impose the following conditions:

AH1. (Bernstein's block moment condition) There exists a positive constant M such that for any $p \geq 2$ and for every $f \in \mathcal{F}$

$$\sup_{y \in S} \mathbb{E}_y \left| f(B_1) \right|^p \leq \frac{1}{2} p! \sigma^2(f) M^{p-2}. \tag{8}$$

AH2. (Block length moment assumption) There exists a positive constant N such that for any $p \geq 2$

$$\sup_{y \in S} \mathbb{E}_y \left| \tau_S \right|^p \leq N^p. \tag{9}$$

AH3. (Non-regenerative block exponential moment assumption) There exists a constant $\lambda_0 > 0$ such that for every $f \in \mathcal{F}$ we have $\mathbb{E}_\nu \left[\exp \left| \sum_{i=1}^{\tau_S} f(X_i) \right| \right] < \infty$.

AH4. (Exponential block moment assumption) There exists a constant $\lambda_1 > 0$ such that for every $f \in \mathcal{F}$ we have $\sup_{y \in S} \mathbb{E}_y \left[\exp \left| f(B_1) \right| \right] < \infty$.

Let $\sup_{y \in S} \mathbb{E}_y \left| \tau_S \right| = \alpha_{\mathcal{M}} < \infty$.

Bernstein type inequality for Harris recurrent Markov chains

Assume that $X_{\mathcal{M}}$ is a Harris recurrent, strongly aperiodic Markov chain. Suppose also that $\mathcal{N}_2(\epsilon, \mathcal{F}) < \infty$.Then, under assumptions AH1-AH4, we have for some positive explicit constants $L,\ R\ > 0$ and any $q_1,\ q_2 > 1$

$$\mathbb{P}_\nu\left[\frac{1}{n}\left|\sum_{i=1}^{n} f(X_i) - \mu(f)\right| \geq x\right] \leq \mathcal{N}_2(\epsilon, \mathcal{F})\left\{2\exp\left[-\frac{(x-2\epsilon)^2 n}{8\left(\frac{\sigma^2(f)}{\alpha_{\mathcal{M}}} + \frac{M(x-2\epsilon)}{n}\right)}\right]\right.$$
$$+ C_1\exp\left[-\frac{(x-2\epsilon)n}{6}\right] + C_2\exp\left[-\frac{(x-2\epsilon)n}{6}\right]$$
$$\left.+\exp\left\{\frac{1}{q_1(2q_1-2)} - \frac{(x-2\epsilon)n^{1/2}}{6Lq_1}\right\} + \exp\left\{\frac{1}{q_2(2q_2-2)} - \frac{(x-2\epsilon)n^{1/2}}{6Rq_2}\right\}\right\},$$

where $C_1,\ C_2$ and $L,\ R$ are constants that can be explicitly computed.

References I

- Bertail, P., Ciołek, G. (2017). Sharp Bernstein and Hoeffding type inequalities for regenerative Markov chains. Submitted. hal-01440167
- Bertail, P. , Clémençon, S., (2010). Sharp bounds for the tail of functionals of Markov chains. *Probability Theory and its applications.*. *54*, 3, 505-515.
- Kosorok, M.R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.
- Nummelin, E. (1978). A splitting technique for Harris recurrent chains. Z. Wahrsch. Verw. Gebiete, **43**, 309-318.
- Petrov, V. (1975). *Sums of Independent Random Variables.* Springer-Verlag Berlin Heidelberg.
- Pollard D. (1984). Convergence of Stochastic Processes. Springer, New York.