Bernstein-like inequality for Markov chains

Michał Lemańczyk (Warsaw Uniwersity)
CIRM (Marseille), 11.07.2017

Classical Bernstein's inequality

If $(\xi_i)_i$ is a sequence of i.i.d. (independent, identically distributed) centered random variables such that $\sup_i \|\xi_i\|_\infty \leqslant M$ and $\sigma^2 := \mathbb{E}\xi_i^2$ then:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \xi_i\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}Mt}\right).$$

Bernstein-like inequality for Markov chains

Main Theorem (Bernstein-like inequality for Markov chains)

Let X be a geometrically ergodic, irreducible Markov chain with state space $\mathcal{X}$ and let π be its unique stationary measure. Moreover, let f be a bounded, Borel function ($\|f\|_\infty < \infty$) defined on $\mathcal{X}$ such that $\mathbb{E}_\pi f = 0$. Then, for all $x \in \mathcal{X}$, there exists a constant $\tau = \tau(X, x)$ such that for all $t > 0$, we have:

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \leqslant$$

$$K \exp\left(-\frac{t^2}{100n\sigma^2_{Mrv} + 2100\tau^2\|f\|_\infty \log(n)}\right),$$

where $\sigma^2_{Mrv} := Var_\pi(f(X_0)) + 2\sum_{i=1}^\infty Cov_\pi(f(X_0), f(X_i))$ denotes the asymptotic variance of Markov chain X.

Idea of the proof of the theorem

We have:

$$\left|\sum_{i=0}^{n-1} f(X_i)\right| \leqslant U_n(f) + V_n(f) + W_n(f),$$

where

$$U_n(f) := \left|\sum_{k=0}^{\sigma_0} f(X_k)\right|,$$

$$V_n(f) := \left|\sum_{i=1}^{N} s_{i-1}(f)\right|,$$

$$W_n(f) := \left|\sum_{k=n}^{\sigma_N} f(X_k)\right|,$$

Idea of the proof of the theorem

where

$$s_i := s_i(f) := \sum_{j=\sigma_i+1}^{\sigma_{i+1}} f(X_j)$$

$$N := \inf\{i \geqslant 0 |\ \sigma_i \geqslant n-1\}.$$

Thus we must estimate the following probabilities:

$$\mathbb{P}(U_n(f) \geqslant t),\ \mathbb{P}(V_n(f) \geqslant t),\ \mathbb{P}(W_n(f) \geqslant t).$$

The crucial part lies in estimating:

$$\mathbb{P}(V_n(f) \geqslant t) = \mathbb{P}(\left|\sum_{i=0}^{N} s_{i-1}(f)\right| > t).$$

It is known that the sequence $(s_i)_{i\geqslant 1}$ is 1-dependent and stationary. Moreover, it turns out that the problem of dealing with $\mathbb{P}(V_n(f) \geqslant t)$ reduces (up to technicalities) to the following lemma:

Idea of the proof of the theorem

where

$$s_i := s_i(f) := \sum_{j=\sigma_i+1}^{\sigma_{i+1}} f(X_j)$$

$$N := \inf\{i \geqslant 0 |\ \sigma_i \geqslant n-1\}.$$

Thus we must estimate the following probabilities:

$$\mathbb{P}(U_n(f) \geqslant t),\ \mathbb{P}(V_n(f) \geqslant t),\ \mathbb{P}(W_n(f) \geqslant t).$$

The crucial part lies in estimating:

$$\mathbb{P}(V_n(f) \geqslant t) = \mathbb{P}(\left|\sum_{i=0}^{N} s_{i-1}(f)\right| > t).$$

It is known that the sequence $(s_i)_{i \geqslant 1}$ is 1-dependent and stationary. Moreover, it turns out that the problem of dealing with $\mathbb{P}(V_n(f) \geqslant t)$ reduces (up to technicalities) to the following lemma:

Idea of the proof of the theorem

where

$$s_i := s_i(f) := \sum_{j=\sigma_i+1}^{\sigma_{i+1}} f(X_j)$$

$$N := \inf\{i \geqslant 0 |\ \sigma_i \geqslant n-1\}.$$

Thus we must estimate the following probabilities:

$$\mathbb{P}(U_n(f) \geqslant t),\ \mathbb{P}(V_n(f) \geqslant t),\ \mathbb{P}(W_n(f) \geqslant t).$$

The crucial part lies in estimating:

$$\mathbb{P}(V_n(f) \geqslant t) = \mathbb{P}(\left|\sum_{i=0}^{N} s_{i-1}(f)\right| > t).$$

It is known that the sequence $(s_i)_{i\geqslant 1}$ is 1-dependent and stationary. Moreover, it turns out that the problem of dealing with $\mathbb{P}(V_n(f) \geqslant t)$ reduces (up to technicalities) to the following lemma:

Idea of the proof of the theorem

where

$$s_i := s_i(f) := \sum_{j=\sigma_i+1}^{\sigma_{i+1}} f(X_j)$$

$$N := \inf\{i \geqslant 0 | \; \sigma_i \geqslant n-1\}.$$

Thus we must estimate the following probabilities:

$$\mathbb{P}(U_n(f) \geqslant t), \; \mathbb{P}(V_n(f) \geqslant t), \; \mathbb{P}(W_n(f) \geqslant t).$$

The crucial part lies in estimating:

$$\mathbb{P}(V_n(f) \geqslant t) = \mathbb{P}(\left|\sum_{i=0}^{N} s_{i-1}(f)\right| > t).$$

It is known that the sequence $(s_i)_{i \geqslant 1}$ is 1-dependent and stationary. Moreover, it turns out that the problem of dealing with $\mathbb{P}(V_n(f) \geqslant t)$ reduces (up to technicalities) to the following lemma:

Idea of the proof of the theorem

Lemma

If $(\xi_i)_{i\in\mathbb{N}}$ is a stationary, 1-dependent Markov chain such that $\xi_i \in \mathbb{R}$, $\mathbb{E}\xi_i = 0$, $\sup_i \|\xi_i\|_\infty \leqslant M$ and $\sigma_\infty^2 := \lim_{n\to\infty} \frac{1}{n} Var(\sum_{i=1}^n \xi_i) = \mathbb{E}\xi_1^2 + 2\mathbb{E}\xi_1\xi_2$, then:

$$\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i\right| \geqslant t\right) \leqslant 6\exp\left(-\frac{t^2}{64n\sigma_\infty^2 + 12Mt}\right).$$

Idea of the proof of the theorem. Proof of the lemma

- Consider the natural filtration $\mathcal{F}_i := \sigma\{\xi_j \mid j \leqslant i\}$ and $Z_i := \xi_i + \mathbb{E}(\xi_{i+1}|\mathcal{F}_i) - \mathbb{E}(\xi_i|\mathcal{F}_{i-1})$.
  1. (Z_i) is 2-dependent
  2. $\mathbb{E}Z_i = 0$
  3. $\|Z_i\|_\infty \leqslant 3M$
  4. $\mathbb{E}Z_i^2 = \sigma_\infty^2$
  5. $\sum_{i=1}^n Z_i = \left(\sum_{i=1}^n \xi_i\right) + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) - \mathbb{E}(\xi_1|\mathcal{F}_0)$
- Now, we proceed as follows. Firstly, we establish Bernstein's inequality for Z_i:

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| \geqslant t\right) \leqslant \mathbb{P}\left(|Z_1 + Z_4 + Z_7 + \cdots| \geqslant t/3\right) +$$

$$\mathbb{P}\left(|Z_2 + Z_5 + Z_8 + \cdots| \geqslant t/3\right) + \mathbb{P}\left(|Z_3 + Z_6 + Z_9 + \cdots| \geqslant t/3\right)$$

$$\leqslant 6 \exp\left(-\frac{t^2}{18n\sigma_\infty^2 + 6Mt}\right).$$

- Consider the natural filtration $\mathcal{F}_i := \sigma\{\xi_j \mid j \leqslant i\}$ and $Z_i := \xi_i + \mathbb{E}(\xi_{i+1}|\mathcal{F}_i) - \mathbb{E}(\xi_i|\mathcal{F}_{i-1})$.
  1. (Z_i) is 2-dependent
  2. $\mathbb{E}Z_i = 0$
  3. $\|Z_i\|_\infty \leqslant 3M$
  4. $\mathbb{E}Z_i^2 = \sigma_\infty^2$
  5. $\sum_{i=1}^n Z_i = \left(\sum_{i=1}^n \xi_i\right) + \mathbb{E}(\xi_{n+1}|\mathcal{F}_n) - \mathbb{E}(\xi_1|\mathcal{F}_0)$
- Now, we proceed as follows. Firstly, we establish Bernstein's inequality for Z_i:

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| \geqslant t\right) \leqslant \mathbb{P}\left(|Z_1 + Z_4 + Z_7 + \cdots| \geqslant t/3\right) +$$

$$\mathbb{P}\left(|Z_2 + Z_5 + Z_8 + \cdots| \geqslant t/3\right) + \mathbb{P}\left(|Z_3 + Z_6 + Z_9 + \cdots| \geqslant t/3\right)$$

$$\leqslant 6 \exp\left(-\frac{t^2}{18n\sigma_\infty^2 + 6Mt}\right).$$

Finally, without loss of generality $t \geqslant 6M$ and:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \xi_i\right| \geqslant t\right) \leqslant \mathbb{P}\left(\left|\sum_{i=1}^{n} Z_i\right| \geqslant t/2\right) +$$

$$\mathbb{P}(\,|\mathbb{E}(\xi_{n+1}|\mathcal{F}_n) - \mathbb{E}(\xi_1|\mathcal{F}_0)| \geqslant t/2) =$$

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Z_i\right| \geqslant t/2\right) \leqslant 6\exp\left(-\frac{t^2}{64n\sigma_\infty^2 + 12Mt}\right).$$