Hoeffding's inequalities for Markov chains on general state space

Błażej Miasojedow

University of Warsaw

MCQMC 2010, Warsaw

Introduction

- Let $\mu := \int_{\mathcal{X}} f(x)\pi(dx)$
- In MCMC methods we construct a Markov chain $X_1, X_2, \ldots$ with stationary distribution π and we approximate μ by $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} f(X_i)$.
- Our goal is to derive bounds of probabilities of estimation errors

$$\mathbf{P}\left(|\bar{X} - \mu| > \varepsilon\right) \leq ? \tag{1}$$

- We will present a Hoeffding's type inequalities for bounded functions f.
- Using these inequalities we will construct sub-exponential bounds of (1) for functions f with exponential tails.

Introduction

- Let $\mu := \int_{\mathcal{X}} f(x)\pi(dx)$
- In MCMC methods we construct a Markov chain $X_1, X_2, \ldots$ with stationary distribution π and we approximate μ by $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} f(X_i)$.
- Our goal is to derive bounds of probabilities of estimation errors

$$\mathbf{P}\left(|\bar{X} - \mu| > \varepsilon\right) \leq ? \tag{1}$$

- We will present a Hoeffding's type inequalities for bounded functions f.
- Using these inequalities we will construct sub-exponential bounds of (1) for functions f with exponential tails.

Introduction

- Let $\mu := \int_{\mathcal{X}} f(x)\pi(dx)$
- In MCMC methods we construct a Markov chain $X_1, X_2, \dots$ with stationary distribution π and we approximate μ by $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} f(X_i)$.
- Our goal is to derive bounds of probabilities of estimation errors

$$\mathbf{P}\left(|\bar{X} - \mu| > \varepsilon\right) \leq ? \tag{1}$$

- We will present a Hoeffding's type inequalities for bounded functions f.
- Using these inequalities we will construct sub-exponential bounds of (1) for functions f with exponential tails.

Introduction

- Let $\mu := \int_{\mathcal{X}} f(x)\pi(dx)$
- In MCMC methods we construct a Markov chain $X_1, X_2, \ldots$ with stationary distribution π and we approximate μ by $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} f(X_i)$.
- Our goal is to derive bounds of probabilities of estimation errors

$$\mathbf{P}\left(|\bar{X} - \mu| > \varepsilon\right) \leq ? \tag{1}$$

- We will present a Hoeffding's type inequalities for bounded functions f.
- Using these inequalities we will construct sub-exponential bounds of (1) for functions f with exponential tails.

Introduction

- Let $\mu := \int_{\mathcal{X}} f(x)\pi(dx)$
- In MCMC methods we construct a Markov chain $X_1, X_2, \ldots$ with stationary distribution π and we approximate μ by $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} f(X_i)$.
- Our goal is to derive bounds of probabilities of estimation errors

$$\mathbf{P}\left(|\bar{X} - \mu| > \varepsilon\right) \leq ? \tag{1}$$

- We will present a Hoeffding's type inequalities for bounded functions f.
- Using these inequalities we will construct sub-exponential bounds of (1) for functions f with exponential tails.

Classic Hoeffding's inequalities

Theorem (Hoeffding 1963)

If $X_1, X_2, \ldots$ *is a seqeunce of i.i.d. random variables with* $0 \le X_1 \le 1$ *then for all* n *and all* $0 < \varepsilon < 1 - \mu$ *we have*

$$P(\bar{X} - \mu > \varepsilon) \le \left\{ \left(\frac{\mu}{\mu + \varepsilon} \right)^{\mu+\varepsilon} \left(\frac{1-\mu}{1-\mu-\varepsilon} \right)^{1-\mu-\varepsilon} \right\}^n \le e^{-2n\varepsilon^2},$$

where $\mu := \mathbf{E}X_1$ *and* $\bar{X} := \frac{\sum_{i=1}^n X_i}{n}$.

León & Perron

- Consider $X_1, X_2, \ldots$ a **reversible** Markov chain on finite state space $\mathcal{X}$ with transition matrix M and stationary distribution π.
- Let ρ_1 be the second largest eigenvalue of matrix M and let $\rho := \max\{0, \rho_1\}$.

León & Perron

- Consider $X_1, X_2, \ldots$ a **reversible** Markov chain on finite state space $\mathcal{X}$ with transition matrix M and stationary distribution π.
- Let ρ_1 be the second largest eigenvalue of matrix M and let $\rho := \max\{0, \rho_1\}$.

León & Perron

Theorem (León & Perron 2004)

Let $f : \mathcal{X} \to [0,1]$ *with stationary mean* $\mu := \pi f$ *then for all* $0 < \varepsilon - \mu$ *and all* n

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq \left[\frac{\mu+\bar{\mu}\rho}{1-2(\bar{\mu}-\varepsilon)/(1+\sqrt{\Delta})}\right]^{n(\mu+\varepsilon)} \left[\frac{\bar{\mu}+\mu\rho}{1-2(\mu+\varepsilon)/(1+\sqrt{\Delta})}\right]^{n(\bar{\mu}-\varepsilon)}$$

$$\leq \exp\left\{-2\frac{1-\rho}{1+\rho}\varepsilon^2 n\right\},$$

where $S_n := \sum_{i=1}^n f(X_i)$, $\bar{\mu} = 1-\mu$ *and* $\Delta = 1 + \frac{4\rho(\mu+\varepsilon)(\bar{\mu}-\varepsilon)}{\mu\bar{\mu}(1-\rho)^2}$.

Main result

- We generelize this result in two direction:
- We assume only that state space $\mathcal{X}$ is Polish space with Borel σ-field $\mathcal{B}(\mathcal{X})$.
- We do not assume reversibility of chain $X_1, X_2, \dots$.

Main result

- We generelize this result in two direction:
- We assume only that state space $\mathcal{X}$ is Polish space with Borel σ-field $\mathcal{B}(\mathcal{X})$.
- We do not assume reversibility of chain $X_1, X_2, \ldots$.

Main result

- We generelize this result in two direction:
- We assume only that state space $\mathcal{X}$ is Polish space with Borel σ-field $\mathcal{B}(\mathcal{X})$.
- We do not assume reversibility of chain $X_1, X_2, \ldots$.

Main result

- So $X_1, X_2, \ldots$ is a Markov chain on $\mathcal{X}$ with transition kernel $P(x, \cdot)$ and stationary distribution π.
- Let L^2_π be a Hilbert space of function $g : \mathcal{X} \to \mathbb{R}$ with norm $\|\cdot\|_2$ defined by inner product

$$\langle f, g \rangle := \int_{\mathcal{X}} f(x)g(x)\pi(dx).$$

Main result

- So $X_1, X_2, \ldots$ is a Markov chain on $\mathcal{X}$ with transition kernel $P(x, \cdot)$ and stationary distribution π.
- Let L^2_π be a Hilbert space of function $g : \mathcal{X} \to \mathbb{R}$ with norm $\|\cdot\|_2$ defined by inner product

$$\langle f, g \rangle := \int_{\mathcal{X}} f(x)g(x)\pi(dx).$$

Main result

- So $X_1, X_2, \ldots$ is a Markov chain on $\mathcal{X}$ with transition kernel $P(x, \cdot)$ and stationary distribution π.
- Let L^2_π be a Hilbert space of function $g : \mathcal{X} \to \mathbb{R}$ with norm $\|\cdot\|_2$ defined by inner product

$$\langle f, g \rangle := \int_{\mathcal{X}} f(x) g(x) \pi(dx).$$

Main result

- **Assumption (A1)**

$$\|P - \Pi\|_2 \le \rho < 1$$

Theorem (Main result)

If Markov chain $X_1, X_2, \ldots$ *satisfies (A1). Then for all functions* $f : \mathcal{X} \to [0,1]$ *with stationary mean* $\mu := \pi f$, *all* n *and all* $0 < \varepsilon < 1 - \mu$ *we have:*

$$\begin{aligned} \mathbf{P}_\pi\left(S_n \ge n(\mu+\varepsilon)\right) &\le \left[\frac{\mu + \bar{\mu}\rho}{1 - 2(\bar{\mu} - \varepsilon)/(1+\sqrt{\Delta})}\right]^{n(\mu+\varepsilon)} \left[\frac{\bar{\mu} + \mu\rho}{1 - 2(\mu+\varepsilon)/(1+\sqrt{\Delta})}\right]^{n(\bar{\mu}-\varepsilon)} \\ &\le \exp\left\{-2\frac{1-\rho}{1+\rho}\varepsilon^2 n\right\}, \end{aligned}$$

where $S_n := \sum_{i=1}^n f(X_i)$, $\bar{\mu} = 1 - \mu$ *and* $\Delta = 1 + \frac{4\rho(\mu+\varepsilon)(\bar{\mu}-\varepsilon)}{\mu\bar{\mu}(1-\rho)^2}$.

Main result

- **Assumption (A1)**

$$\|P - \Pi\|_2 \leq \rho < 1$$

Theorem (Main result)

If Markov chain $X_1, X_2, \ldots$ *satisfies (A1). Then for all functions* $f : \mathcal{X} \to [0,1]$ *with stationary mean* $\mu := \pi f$*, all* n *and all* $0 < \varepsilon < 1 - \mu$ *we have:*

$$\begin{aligned} \mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) &\leq \left[\frac{\mu + \bar{\mu}\rho}{1 - 2(\bar{\mu} - \varepsilon)/(1+\sqrt{\Delta})}\right]^{n(\mu+\varepsilon)} \left[\frac{\bar{\mu} + \mu\rho}{1 - 2(\mu+\varepsilon)/(1+\sqrt{\Delta})}\right]^{n(\bar{\mu}-\varepsilon)} \\ &\leq \exp\left\{-2\frac{1-\rho}{1+\rho}\varepsilon^2 n\right\}, \end{aligned}$$

where $S_n := \sum_{i=1}^n f(X_i)$, $\bar{\mu} = 1 - \mu$ *and* $\Delta = 1 + \frac{4\rho(\mu+\varepsilon)(\bar{\mu}-\varepsilon)}{\mu\bar{\mu}(1-\rho)^2}$.

Reversible case

- In **reversible** case assumption **(A1)** is equivalent to geometric ergodicity. Reversible Markov chain, for all n, satisfies

$$\|P^n(x,\cdot) - \pi\|_{tv} \leq M_x \rho^n$$

iff holds **(A1)** with the same ρ

Non-reversible case

- In non-reversible case there exist simple examples of geometric ergodic chain such that for all n norm

$$\|P^n - \Pi\|_2 = 1$$

- We define an adjoint transition kernel $P^*(x, \cdot)$ by

$$\pi(dx)P(x, dy) = \pi(dy)P^*(y, dx)$$

Lemma

*Markov chain with transition kernel P satisfies **(A1)** $\Longleftrightarrow$ the Markov chain with transition kernel $PP^*(x, \cdot)$ is geometrically ergodic.*

Non-reversible case

- In non-reversible case there exist simple examples of geometric ergodic chain such that for all n norm
$$\|P^n - \Pi\|_2 = 1$$
- We define an adjoint transition kernel $P^*(x, \cdot)$ by
$$\pi(dx)P(x, dy) = \pi(dy)P^*(y, dx)$$

Lemma

*Markov chain with transition kernel P satisfies **(A1)** $\iff$ the Markov chain with transition kernel $PP^*(x, \cdot)$ is geometrically ergodic.*

Non-reversible case

- In non-reversible case there exist simple examples of geometric ergodic chain such that for all n norm
$$\|P^n - \Pi\|_2 = 1$$
- We define an adjoint transition kernel $P^*(x, \cdot)$ by
$$\pi(dx)P(x, dy) = \pi(dy)P^*(y, dx)$$

Lemma

*Markov chain with transition kernel P satisfies **(A1)** $\iff$ the Markov chain with transition kernel $PP^*(x, \cdot)$ is geometrically ergodic.*

Drift condition

- We construct a drift condition which implies **(A1)**.
- Let function $V : \mathcal{X} \to [1, \infty)$ satisfies following properties
  1. The level sets $C_r := \{x \in \mathcal{X} \ : \ V(x) \le r\}$ are small sets for all $r \in \mathcal{X}$ i.e.

     $$P(x, dy) \ge \beta_r \mathbb{I}(x \in V_r)\nu_r(dy),$$

     where $\beta_r > 0$ and ν_r is a probability measure.
  2. There exist λ_1 and λ_2 with $\lambda_1\lambda_2 < 1$ and constant K such that

     $$\begin{aligned} PV(x) &\le \lambda_1 V(x) + K \\ P^*V(x) &\le \lambda_2 V(x) + K \end{aligned}$$

Drift condition

- We construct a drift condition which implies **(A1)**.
- Let function $V : \mathcal{X} \to [1, \infty)$ satisfies following properties
  1. The level sets $C_r := \{x \in \mathcal{X} \ : \ V(x) \leq r\}$ are small sets for all $r \in \mathcal{X}$ i.e.
  $$P(x, dy) \geq \beta_r \mathbb{I}(x \in V_r)\nu_r(dy),$$
  where $\beta_r > 0$ and ν_r is a probability measure.
  2. There exist λ_1 and λ_2 with $\lambda_1 \lambda_2 < 1$ and constant K such that
  $$\begin{aligned} PV(x) &\leq \lambda_1 V(x) + K \\ P^* V(x) &\leq \lambda_2 V(x) + K \end{aligned}$$

Drift condition

- We construct a drift condition which implies **(A1)**.
- Let function $V : \mathcal{X} \to [1, \infty)$ satisfies following properties
  1. The level sets $C_r := \{x \in \mathcal{X} \ : \ V(x) \leq r\}$ are small sets for all $r \in \mathcal{X}$ i.e.

$$P(x, dy) \geq \beta_r \mathbb{I}(x \in V_r) \nu_r(dy),$$

     where $\beta_r > 0$ and ν_r is a probability measure.
  2. There exist λ_1 and λ_2 with $\lambda_1 \lambda_2 < 1$ and constant K such that

$$\begin{aligned} PV(x) &\leq \lambda_1 V(x) + K \\ P^* V(x) &\leq \lambda_2 V(x) + K \end{aligned}$$

Drift condition

- We construct a drift condition which implies **(A1)**.
- Let function $V : \mathcal{X} \to [1, \infty)$ satisfies following properties
  1. The level sets $C_r := \{x \in \mathcal{X} \ : \ V(x) \leq r\}$ are small sets for all $r \in \mathcal{X}$ i.e.

     $$P(x, dy) \geq \beta_r \mathbb{I}(x \in V_r)\nu_r(dy),$$

     where $\beta_r > 0$ and ν_r is a probability measure.
  2. There exist λ_1 and λ_2 with $\lambda_1\lambda_2 < 1$ and constant K such that

     $$\begin{aligned} PV(x) &\leq \lambda_1 V(x) + K \\ P^* V(x) &\leq \lambda_2 V(x) + K \end{aligned}$$

Method of proof

- **Notation**:
  - Let $X_1', X_2', \ldots$ be a Markov chain on $\mathcal{X}$ with transition kernel $Q(x, A) := (1-\rho)\pi(A) + \rho\mathbb{I}(x \in A)$.
  - Let $\widehat{Q}_{t,f}$ be a linear self-adjoint operator on space L^2_π: for all $t > 0$

$$\widehat{Q}_{t,f}(g)(x) := e^{\frac{t}{2}f(x)} Q\left(g(x) e^{\frac{t}{2}f(x)}\right)$$

  - Let $Y_1, Y_2, \ldots$ be a Markov chain on $\{0, 1\}$ with transition matrix

$$M := (1-\rho)\mathbf{1}\mu^T + \rho\mathbb{I},$$

    where $\mu := [1-\mu, \mu]^T$
  - Let $\widehat{M}_t$ be a matrix

$$\widehat{M}_t = D_t M D_t,$$

    where $D_t := \mathrm{diag}[1, e^{\frac{t}{2}}]$
  - Let θ_t be the largest eigenvalue of $\widehat{M}_t$

Method of proof

- **Notation**:
  - Let $X_1', X_2', \dots$ be a Markov chain on $\mathcal{X}$ with transition kernel $Q(x, A) := (1-\rho)\pi(A) + \rho\mathbb{I}(x \in A)$.
  - Let $\widehat{Q}_{t,f}$ be a linear self-adjoint operator on space L^2_π: for all $t > 0$

$$\widehat{Q}_{t,f}(g)(x) := e^{\frac{t}{2}f(x)} Q\left(g(x) e^{\frac{t}{2}f(x)}\right)$$

  - Let $Y_1, Y_2, \dots$ be a Markov chain on $\{0, 1\}$ with transition matrix

$$M := (1-\rho)\mathbf{1}\mu^T + \rho\mathbb{I},$$

    where $\mu := [1-\mu, \mu]^T$
  - Let $\widehat{M}_t$ be a matrix

$$\widehat{M}_t = D_t M D_t,$$

    where $D_t :=$ diag$[1, e^{\frac{t}{2}}]$
  - Let θ_t be the largest eigenvalue of $\widehat{M}_t$

Method of proof

- Idea of proof is the same as in León & Perron. We reduce problem to two-state space case:
  1. We show that
  $$\mathbf{E}_\pi \exp(tS_n) \leq \left\| \widehat{Q}_{t,f} \right\|_2^n$$
  2. We obtain
  $$\left\| \widehat{Q}_{t,f} \right\|_2 \leq \theta_t,$$
  3. Finally we apply León & Perron theorem for two-state case.

Method of proof

- Idea of proof is the same as in León & Perron. We reduce problem to two-state space case:
  1. We show that
  $$\mathbf{E}_\pi \exp(tS_n) \leq \left\| \widehat{Q}_{t,f} \right\|_2^n$$
  2. We obtain
  $$\left\| \widehat{Q}_{t,f} \right\|_2 \leq \theta_t,$$
  3. Finally we apply León & Perron theorem for two-state case.

Method of proof

- Idea of proof is the same as in León & Perron. We reduce problem to two-state space case:
  1. We show that
  $$\mathbf{E}_\pi \exp(tS_n) \leq \left\| \widehat{Q}_{t,f} \right\|_2^n$$
  2. We obtain
  $$\left\| \widehat{Q}_{t,f} \right\|_2 \leq \theta_t,$$
  3. Finally we apply León & Perron theorem for two-state case.

Method of proof

- Idea of proof is the same as in León & Perron. We reduce problem to two-state space case:
  1. We show that
  $$\mathbf{E}_\pi \exp(tS_n) \le \left\| \widehat{Q}_{t,f} \right\|_2^n$$
  2. We obtain
  $$\left\| \widehat{Q}_{t,f} \right\|_2 \le \theta_t,$$
  3. Finally we apply León & Perron theorem for two-state case.

New elements

- Under assumption (A1) we obtain

$$\begin{aligned}
\left\|e^{\frac{t}{2}f}Pe^{\frac{t}{2}f}\right\|_2 &= \sup_{\{g,h\,:\,\|g\|_2=\|h\|_2=1\}} \left\langle e^{\frac{t}{2}f}h, Pe^{\frac{t}{2}f}g\right\rangle \\
&= \sup_{\{g,h\,:\,\|g\|_2=\|h\|_2=1\}} \left\{\pi(e^{\frac{t}{2}f}h)\pi(e^{\frac{t}{2}f}g) + \left\langle (e^{\frac{t}{2}f}h)_C, P(e^{\frac{t}{2}f}g)_C\right\rangle\right\} \\
&\le \sup_{\{g,h\,:\,\|g\|_2=\|h\|_2=1\}} \left\{\pi(e^{\frac{t}{2}f}h)\pi(e^{\frac{t}{2}f}g) + \rho\left\|(e^{\frac{t}{2}f}h)_C\right\|_2\left\|(e^{\frac{t}{2}f}g)_C\right\|_2\right\} \\
&\le \sup_{\{g\,:\,\|g\|_2=1\}} \left\{\left[\pi(e^{\frac{t}{2}f}g)\right]^2 + \rho\left\|(e^{\frac{t}{2}f}g)_C\right\|_\pi^2\right\} \\
&= \sup_{\{g\,:\,\|g\|_2=1\}} \left\langle e^{\frac{t}{2}f}g, \pi(e^{\frac{t}{2}f}g) + \rho(e^{\frac{t}{2}f}g)_C\right\rangle \\
&= \left\|\widehat{Q}_{t,f}\right\|_2,
\end{aligned}$$

where $g_C := g - \pi g$.

New elements

- Above observation gives us, that without reversiblity we have

$$\mathbf{E}_\pi \exp(tS_n) \leq \left\| \widehat{Q}_{t,f} \right\|_2^n$$

- As in León & Perron we obtain

$$\mathbf{E}_\pi \exp\left(t \sum_{i=1}^{n} f(X_i') \right) \leq \mathbf{E}_\mu \exp\left(t \sum_{i=1}^{n} Y_i \right)$$

- In finite space case it's obvious that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t \sum_{i=1}^{n} f(X_i') \right) = \log\left(\left\| \widehat{Q}_{t,f} \right\|_2 \right),$$

in general space case we need to prove this.

New elements

- Above observation gives us, that without reversiblity we have

$$\mathbf{E}_\pi \exp(tS_n) \leq \left\| \widehat{Q}_{t,f} \right\|_2^n$$

- As in León & Perron we obtain

$$\mathbf{E}_\pi \exp\left(t\sum_{i=1}^n f(X_i')\right) \leq \mathbf{E}_\mu \exp\left(t\sum_{i=1}^n Y_i\right)$$

- In finite space case it's obvious that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t\sum_{i=1}^n f(X_i')\right) = \log\left(\left\| \widehat{Q}_{t,f} \right\|_2\right),$$

in general space case we need to prove this.

New elements

- Above observation gives us, that without reversiblity we have

$$\mathbf{E}_\pi \exp(tS_n) \leq \left\| \widehat{Q}_{t,f} \right\|_2^n$$

- As in León & Perron we obtain

$$\mathbf{E}_\pi \exp\left(t \sum_{i=1}^n f(X_i')\right) \leq \mathbf{E}_\mu \exp\left(t \sum_{i=1}^n Y_i\right)$$

- In finite space case it's obvious that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t \sum_{i=1}^n f(X_i')\right) = \log\left(\left\| \widehat{Q}_{t,f} \right\|_2\right),$$

in general space case we need to prove this.

New elements

- Let $g_k \searrow f$ and $h_k \nearrow f$ be a sequences of step functions. We obtain that

$$\left\|\widehat{Q}_{t,g_k}\right\|_2 = u_k$$

and

$$\left\|\widehat{Q}_{t,h_k}\right\|_2 = l_k,$$

where u_k and l_k are eigenvalues of $\widehat{Q}_{t,g_k}$ and $\widehat{Q}_{t,h_k}$ respectively.

- Moreover

$$u_k \searrow \left\|\widehat{Q}_{t,f}\right\|_2, \qquad l_k \nearrow \left\|\widehat{Q}_{t,f}\right\|_2$$

New elements

- Let $g_k \searrow f$ and $h_k \nearrow f$ be a sequences of step functions. We obtain that

$$\left\|\widehat{Q}_{t,g_k}\right\|_2 = u_k$$

and

$$\left\|\widehat{Q}_{t,h_k}\right\|_2 = l_k,$$

where u_k and l_k are eigenvalues of $\widehat{Q}_{t,g_k}$ and $\widehat{Q}_{t,h_k}$ respectively.
- Moreover

$$u_k \searrow \left\|\widehat{Q}_{t,f}\right\|_2, \qquad l_k \nearrow \left\|\widehat{Q}_{t,f}\right\|_2$$

New elements

- Since u_k are eigenvalues then it's easy to show

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbf{E}_\pi \exp\left(t\sum_{i=1}^{n} g_k(X_i')\right) = \log(u_k),$$

  the same we have for pair l_k and h_k.
- Hence for all k

$$\log(l_k) \le \lim_{n\to\infty}\frac{1}{n}\log \mathbf{E}_\pi \exp\left(t\sum_{i=1}^{n} f_k(X_i')\right) \le \log(u_k)$$

- So if $k\to\infty$ we have

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbf{E}_\pi \exp\left(t\sum_{i=1}^{n} f_k(X_i')\right) = \log\left(\left\|\widehat{Q}_{t,f}\right\|_2\right)$$

New elements

- Since u_k are eigenvalues then it's easy to show

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t \sum_{i=1}^{n} g_k(X_i') \right) = \log(u_k),$$

  the same we have for pair l_k and h_k.

- Hence for all k

$$\log(l_k) \leq \lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t \sum_{i=1}^{n} f_k(X_i') \right) \leq \log(u_k)$$

- So if $k \to \infty$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t \sum_{i=1}^{n} f_k(X_i') \right) = \log\left(\left\| \widehat{Q}_{t,f} \right\|_2 \right)$$

New elements

- Since u_k are eigenvalues then it's easy to show

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t\sum_{i=1}^n g_k(X_i')\right) = \log(u_k),$$

  the same we have for pair l_k and h_k.
- Hence for all k

$$\log(l_k) \leq \lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t\sum_{i=1}^n f_k(X_i')\right) \leq \log(u_k)$$

- So if $k \to \infty$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{E}_\pi \exp\left(t\sum_{i=1}^n f_k(X_i')\right) = \log\left(\left\|\widehat{Q}_{t,f}\right\|_2\right)$$

Sub-exponential bounds

- Assumptions:

A1

$$\|P - \Pi\|_2 \leq \rho < 1$$

A2 Let $f : \mathcal{X} \to \mathbb{R}$ satisfies: for all $M > M_0$

$$\mathbf{P}_\pi\left(|f(X_1)| > \frac{M}{2}\right) \leq ae^{bM}$$

A3 Let n be sufficently large and ε sufficently small :

$$\log(an) > bM_0 \quad \text{and} \quad \varepsilon < \frac{\log(an)}{2b} - \mu$$

or

$$\left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}} > M_0, \quad \text{and} \quad \varepsilon + \mu < \frac{1}{2}\left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}$$

Sub-exponential bounds

- Assumptions:

A1

$$\|P - \Pi\|_2 \le \rho < 1$$

A2 Let $f : \mathcal{X} \to \mathbb{R}$ satisfies: for all $M > M_0$

$$\mathbf{P}_\pi \left(|f(X_1)| > \frac{M}{2} \right) \le ae^{bM}$$

A3 Let n be sufficently large and ε sufficently small :

$$\log(an) > bM_0 \quad \text{and} \quad \varepsilon < \frac{\log(an)}{2b} - \mu$$

or

$$\left(2\frac{1-\rho}{b(1+\rho)} n\varepsilon^2 \right)^{\frac{1}{3}} > M_0, \quad \text{and} \quad \varepsilon + \mu < \frac{1}{2} \left(2\frac{1-\rho}{b(1+\rho)} n\varepsilon^2 \right)^{\frac{1}{3}}$$

Sub-exponential bounds

- Assumptions:

A1

$$\|P - \Pi\|_2 \leq \rho < 1$$

A2 Let $f : \mathcal{X} \to \mathbb{R}$ satisfies: for all $M > M_0$

$$\mathbf{P}_\pi \left(|f(X_1)| > \frac{M}{2} \right) \leq ae^{bM}$$

A3 Let n be sufficently large and ε sufficently small :

$$\log(an) > bM_0 \quad \text{and} \quad \varepsilon < \frac{\log(an)}{2b} - \mu$$

or

$$\left(2\frac{1-\rho}{b(1+\rho)} n\varepsilon^2 \right)^{\frac{1}{3}} > M_0, \quad \text{and} \quad \varepsilon + \mu < \frac{1}{2} \left(2\frac{1-\rho}{b(1+\rho)} n\varepsilon^2 \right)^{\frac{1}{3}}$$

Sub-exponential bounds

Theorem

Under assumptions A1-A3. For all initial distribution ν with density $\frac{d\nu}{d\pi}$ w.r.t. π and all $1 \le p, q \le \infty$, with $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\mathbf{P}_\nu \left(S_n \ge n(\mu + \varepsilon) \right) \le \left\| \frac{d\nu}{d\pi} \right\|_p 2 \exp \left\{ -\frac{2}{q} \frac{1-\rho}{1+\rho} \varepsilon^2 \frac{n}{{M^*}^2} \right\},$$

where M^* is a solution of

$$M^3 - M^2 \frac{\log(an)}{b} - 2 \frac{1-\rho}{b(1+\rho)} n\varepsilon^2 = 0.$$

Moreover

$$\left(2 \frac{1-\rho}{b(1+\rho)} n\varepsilon^2 \right)^{\frac{1}{3}} \le M^* \le 1.465571 \max \left\{ \left(2 \frac{1-\rho}{b(1+\rho)} n\varepsilon^2 \right)^{\frac{1}{3}}, \frac{\log(an)}{b} \right\}.$$

Sub-exponential bounds

- If Markov chain satisfies dirft condtion with $V, \lambda_1, \lambda_2, K$ then for all functions $|f| \leq \log(V)$ we have

$$\mathbf{P}_\pi\left(|f(X_1)| > \frac{M}{2}\right) \leq \frac{K}{1-\lambda} e^{\frac{M}{2}},$$

where $\lambda = \min\{\lambda_1, \lambda_2\}$.

Sketch of proof

- Under these assumptions we can decompose

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq \mathbf{P}_\pi\left(S_{n,M} \geq n(\mu_M+\varepsilon)\right) + n\mathbf{P}_\pi\left(|f(X_1)| > \frac{M}{2}\right),$$

  where $S_{n,M} := \sum_{i=1}^n f(X_i)\mathbb{I}(|X_i| \leq \frac{M}{2})$ and $\mu_M := \pi(f\mathbb{I}(|X| \leq \frac{M}{2}))$.
- To first term we use Hoeffding's ineqaulity to second term we use A2 and we have

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq \exp\left\{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{M^2}\right\} + \exp\left\{-bM+\log(an)\right\}.$$

- Now we can optimize by M and obtain

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq 2\exp\left\{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{{M^*}^2}\right\}$$

  where M^* ia a solution of equation $M^3 - M^2\frac{\log(an)}{b} - 2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2 = 0$.

Sketch of proof

- Under these assumptions we can decompose

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq \mathbf{P}_\pi\left(S_{n,M} \geq n(\mu_M+\varepsilon)\right) + n\mathbf{P}_\pi\left(|f(X_1)| > \frac{M}{2}\right),$$

  where $S_{n,M} := \sum_{i=1}^n f(X_i)\mathbb{I}(|X_i| \leq \frac{M}{2})$ and $\mu_M := \pi(f\mathbb{I}(|X| \leq \frac{M}{2}))$.
- To first term we use Hoeffding's ineqaulity to second term we use A2 and we have

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq \exp\left\{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{M^2}\right\} + \exp\left\{-bM+\log(an)\right\}.$$

- Now we can optimize by M and obtain

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq 2\exp\left\{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{{M^*}^2}\right\}$$

  where M^* ia a solution of equation $M^3 - M^2\frac{\log(an)}{b} - 2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2 = 0$.

Sketch of proof

- Under these assumptions we can decompose

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq \mathbf{P}_\pi\left(S_{n,M} \geq n(\mu_M+\varepsilon)\right) + n\mathbf{P}_\pi\left(|f(X_1)| > \frac{M}{2}\right),$$

  where $S_{n,M} := \sum_{i=1}^n f(X_i)\mathbb{I}(|X_i| \leq \frac{M}{2})$ and $\mu_M := \pi(f\mathbb{I}(|X| \leq \frac{M}{2}))$.
- To first term we use Hoeffding's ineqaulity to second term we use A2 and we have

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq \exp\left\{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{M^2}\right\} + \exp\left\{-bM + \log(an)\right\}.$$

- Now we can optimize by M and obtain

$$\mathbf{P}_\pi\left(S_n \geq n(\mu+\varepsilon)\right) \leq 2\exp\left\{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{{M^*}^2}\right\}$$

  where M^* ia a solution of equation $M^3 - M^2\frac{\log(an)}{b} - 2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2 = 0$.

Some References

- León, C. A. and Perron, F. (2004). Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.* Volume 14, Number 2, 958-970.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58 13-30.