Mathematical foundations of Infinite-Dim Statistical models
chap.3 Empirical Processes (3.1.1 $\sim$ 3.1.2)

Presented by Jiin Seo

September 21, 2018

Outline

Outline

3.1.1 Definitions and Overview

- This chapter develops empirical process theory with an emphasis on finite sample sizes.

3.1.1 Definitions and Overview

Let $(S, \mathcal{S}, P)$ be a prob. space $X_i, i \in \mathbf{N}$, be the coordinate ftns of the infinite product prob. space $(\boldsymbol{\Omega}, \boldsymbol{\Sigma}, Pr) := (S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, P^{\mathbb{N}}), X_i : S^{\mathbb{N}} \mapsto \mathcal{S}$ which are i.i.d. S-valued r.v.s with law P.

Def. (Empirical Measure) The empirical measure corresponding to the 'observations' $X_1, \cdots, X_n$, $\forall n \in \mathbb{N}$, is defined as the random discrete probability measure

$$P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}, \tag{3.1}$$

where δ_x is Dirac measure at x.

3.1.1 Definitions and Overview

Notations

- $\mathcal{Q}f = \mathcal{Q}(f) = \int_{\Omega} f d\mathcal{Q}$: integral of f w.r.t $\mathcal{Q}$

- Empirical process indexed by $\mathcal{F}$:
  Let $\mathcal{F}$ be a collection of P-integrable ftns $f : S \mapsto \mathbb{R}$, usually infinite.
  For any such class of ftns $\mathcal{F}$, the empirical measure defines a stochastic process

$$f \mapsto P_n f, f \in \mathcal{F}. \tag{3.2}$$

- Empirical process for the centred and normalised process :

$$f \mapsto \nu_n(f) := \sqrt{n}(P_n f - Pf), f \in \mathcal{F} \tag{3.3}$$

,

3.1.1 Definitions and Overview

Goal of empirical process theory is

- To study the properties of the approximation of Pf by $P_n f$, uniformly in $\mathcal{F}$
- To obtain both probability estimates for the random quantities

$$||P_n - P||_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |P_n f - Pf|$$

and probabilistic limit thms for the processes $\{(P_n - P)(f) : f \in \mathcal{F}\}$

3.1.1 Definitions and Overview

Main Questions

- 1) Concentration of $||P_n - P||_{\mathcal{F}}$ about its mean $\Rightarrow$ Talagrand's inequality
- 2) Estimation for $E||P_n - P||_{\mathcal{F}} \Rightarrow$ bracketing , Vapnik-Cervonenkis classes of functions
- 3) Limit theorems: L.L.N. and C.L.T.
- Inequalities

  - Exponential Inequalities for sums of centred bdd. indep. random variables and and the associated Maximal Inequalities.

  - Levy's inequality and Hoffmann-Jorgensen's Inequality

  - Randomisation/Symmetrisation Inequalities

Outline

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

- $\xi_{n,i}, i \in \mathbb{N}$: indep. centred real random variables.
- Tail probabilities of $S_n = \sum_{i=1}^{n} \xi_{n,i}$ are similar to those of Poisson r.v. and normal variable.
- Chebyshev's inequality

$$Pr\{|S_n| \geq t\} \leq \frac{\sum_{i=1}^{n} E\xi_{n,i}^2}{t^2}, t > 0$$

- Construction exponential inequalities for S_n
  - m.g.f. $\rightarrow$ applying Markov's inequality to $e^{\lambda S_n}$
- Types of inequalities when the variables $\xi_{n,i}$ are bounded :
  - the range of the variable
  - variance into account

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Lemma 3.1.1 Let X be a centred r.v. taking values in $[a, b]$ for some $-\infty < a < 0 \leq b < \infty$. Then, $\forall \lambda > 0$, setting $L(\lambda) := logEe^{\lambda X}$, we have

$$L(0) = L'(0) = 0, L''(\lambda) \leq (b-a)^2/4 \qquad (3.6)$$

and hence

$$Ee^{\lambda X} \leq e^{\lambda^2 (b-a)^2/8} \qquad (3.7)$$

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Theorem 3.1.2 (Hoeffding's inequality) Let X_i be a indep. centred r.v. taking values taking values, respectively, in $[a_i, b_i]$ for some $-\infty < a_i < 0 \leq b_i < \infty, i = 1, \cdots, n, \forall n \in \mathbb{N}$, and let $S_n = \sum_{i=1}^{n} X_i$. Then, $\forall \lambda > 0$,

$$Ee^{\lambda S_n} \leq e^{\lambda^2 \sum_{i=1}^{n}(b_i - a_i)^2/8}, \tag{3.8}$$

and $\forall t \geq 0$,

$$\begin{aligned} Pr\{S_n \geq t\} &\leq exp(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}), \\ Pr\{S_n \leq -t\} &\leq exp(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}). \end{aligned} \tag{3.9}$$

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Proof of Theorem 3.1.2 By Lemma 3.1.1 and independence,

$$Ee^{\lambda S_n} = \prod_{i=1}^{n} Ee^{\lambda X_i} \leq e^{\lambda^2 \sum_{i=1}^{n}(b_i - a_i)^2/8}.$$

We then have, by Markov's inequality,

$$Pr\{S_n \geq t\} = Pr\{e^{\lambda S_n} \geq e^{\lambda t}\} \leq Ee^{\lambda S_n}/e^{\lambda t} \leq exp(\lambda^2 \sum_{i=1}^{n}(b_i - a_i)^2/8 - \lambda t).$$

This bound is smallest for $\lambda = 4t/\sum_{i=1}^{n}(b_i - a_i)^2$, which gives the first inequality in (3.9).The second inequality follows by applying the first to $-X_i$.

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Theorem 3.1.5 Let X be a centred r.v. taking value s.t $|X| \le c$ a.s., for some $c < \infty$, and $EX^2 = \sigma^2$. Then $\forall \lambda > 0$,

$$Ee^{\lambda X} \le exp(-\frac{\sigma^2}{c^2}(e^{\lambda c} - 1 - \lambda c)), \qquad (3.11)$$

As a consequence, if $X_i, 1 \le i \le n < \infty$, are centred, indep. and a.s. bdd by $c < \infty$ in absolute value, then setting

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} EX_i^2 \qquad (3.12)$$

and $S_n = \sum_{i=1}^{n} X_i$, we have , $\forall \lambda > 0$.

$$Ee^{\lambda S_n} \le exp(\frac{n\sigma^2}{c^2}(e^{\lambda c} - 1 - \lambda c)), \qquad (3.13)$$

and the same inequality holds for $-S_n$.

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Prrof of Theorem 3.1.5 Since $EX = 0$, expansion of the exponential gives

$$Ee^{\lambda X} = 1 + \sum_{k=2}^{\infty} \frac{\lambda^k EX^k}{k!} \leq exp(\sum_{k=2}^{\infty} \frac{\lambda^k EX^k}{k!}),$$

whereas, since $|EX^k| \leq c^{k-2}\sigma^2, \forall k \geq 2$, this exponent can be bounded by

$$|\sum_{k=2}^{\infty} \frac{\lambda^k EX^k}{k!}| \leq \lambda^2\sigma^2 \sum_{k=2}^{\infty} \frac{(\lambda c)^{k-2}}{k!} = \frac{\sigma^2}{c^2} \sum_{k=2}^{\infty} \frac{(\lambda c)^k}{k!} = \frac{\sigma^2}{c^2}(e^{\lambda c} - 1 - \lambda c).$$

This gives inequality (3.11). Inequality (3.13) follows from (3.11) by independence. The foregoing also applies to $Y_i = -X_i$.

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

$$\phi(x) := e^{-x} - 1 + x, \textit{ for } x \in \mathbb{R}, \text{ and}$$
$$h_1(x) := (1+x)log(1-x) - x, \textit{ for } x \geq 0. \quad (3.14)$$

Proposition 3.1.6 Let Z be a r.v whose m.g.f satisfies the bound

$$Ee^{\lambda Z} \leq exp(\nu(e^{\lambda} - 1 - \lambda)), \lambda > 0,$$

for some $\nu > 0$. Then, $\forall t \geq 0$,

$$Pr\{Z \geq t\} \leq exp(-\nu h_1(t/\nu)) \leq exp(-\frac{3t}{4} log(1+\frac{2t}{3\nu})) \leq exp(-\frac{t^2}{2\nu + 2t/3})$$

and

$$Pr\{Z \geq \sqrt{2\nu x} + x/3\} \leq e^{-x}, x \geq 0$$

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Theorem 3.1.7(Inequalities of Bennett, Prokhorov and Bernstein) Let $X_i, 1 \leq i \leq n$, be indep. centred r.v. a.s. bdd by $c < \infty$ in absolute value. Set $\sigma^2 = 1/n \sum_{i=1}^n EX_i^2$ and $S_n = \sum_{i=1}^n X_i$. Then, $\forall u \geq 0$,

$$\begin{aligned} Pr\{S_n \geq u\} &\leq exp(-\frac{n\sigma^2}{c^2} h_1(\frac{uc}{n\sigma^2})) \\ &\leq exp(-\frac{3u}{4c} log(1 + \frac{2uc}{3n\sigma^2})) \\ &\leq exp(-\frac{u^2}{2n\sigma^2 + 2cu/3}) \end{aligned} \tag{3.23}$$

and

$$Pr\{S_n \geq \sqrt{2n\sigma^2 u} + \frac{cu}{3}\} \leq e^{-u} \tag{3.24}$$

where h_1 is as defined in (3.14), and the same inequalities hold for $Pr\{S_n < -u\}$.

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Proposition 3.1.8 (Bernstein's inequality) Let $X_i, 1 \leq i \leq n$, be centred indep. random variables s.t., $\forall k \geq 2$ and all $1 \leq i \leq n$,

$$E|X_i|^k \leq \frac{k!}{2}\sigma_i^2 c^{k-1}, \qquad (3.25)$$

and set $\sigma^2 = \sum_{i=1}^n \sigma_i^2, S_n = \sum_{i=1}^n X_i$. Then

$$Pr\{S_n \geq t\} \leq exp(-\frac{t^2}{2\sigma^2 + 2ct}), t \geq 0 \qquad (3.26)$$

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Proof of Proposition 3.1.8 Assuming that $c|\lambda| < 1$, the moment-growth hypothesis implies that, for $1 \leq k \leq n$,

$$Ee^{\lambda X_k} \leq 1 + \frac{\sigma_k^2}{2} \sum_{i=2}^{\infty} |\lambda|^k c^{k-2} = 1 + \frac{\lambda^2 \sigma_k^2}{2(1 - |\lambda| c)} \leq e^{\lambda^2 \sigma_k^2/(2-2c|\lambda|)},$$

which, by independence and the exponential Chebyshev's inequality, implies that

$$Pr\{S_n \geq t\} \leq \frac{Ee^{\lambda S_n}}{e^{\lambda t}} \leq exp(\frac{\lambda^2 \sigma^2}{2 - 2c|\lambda|} - \lambda t).$$

The result obtains by taking $\lambda = t/(\sigma^2 + ct)$.

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

- Inequalities of Hoeffding, Bennet, Bernstein and Prohorov also hold for the maximum of the partial-sums $\max_{k\leq n} S_k$ by virtue of Doob's submartingale inequality.

- Given a symmetric matrix A with eigenvalues λ_i, its Hilbert-Schmidt norm $||A||_{HS}$ is defined as $||A||_{HS}^2 = \sum \lambda_i^2$.
  $||A||$: maximum of its eigenvalues

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Theorem 3.1.9(Hanson-Wright's inequality) Let $A = (a_{ij})_{i,j=1}^{n}$ be a symm. matrix with all its diagonal terms a_{ii} equal to zero, let $g_i, i = 1, \cdots, n$, be indep. standard normal variables and set

$$X = \sum_{i,j} a_{ij} g_i g_j = 2 \sum_{i<j} a_i j g_i g_j.$$

Alternatively, let A be a diagonal matrix with eigenvalues τ_i, and set

$$X = \sum_{i} \tau_i (g_i^2 - 1), .$$

g_i indep. $N(0, 1)$, as earlier. Then both random variables satisfy

$$Ee^{\lambda X} \leq e^{||A||_{HS}^2 \lambda^2/(1-2\lambda||A||)} = e\varphi^{2||A||_{HS}^2 2||A||}(\lambda), for 0 < \lambda < 1/2||A||. \quad (3.27)$$

Consequently, for $t \geq 0$,

$$Pr\{X > t\} \leq e^{-t^2/4(||A||_{HS}^2+||A||t)} \text{or} Pr\{X \geq \sqrt{4||A||_{HS}^2 t} + 2||A||t\} \leq e^{-t}, \quad (3.28)$$

3.1.2 Exponential and Maximal Inequalities for Sums of Independent Centred and Bounded Real Random Variables

Finally, we see that control of m.g.f. of a collection of r.v. translates into control of the expected value of their maximum:

Theorem 3.1.10) (a) Let $X_i, i = 1, \cdots, N$, be random variables s.t. $Ee^{\lambda X_i} \leq e^{\lambda^2 \sigma_i^2/2}$, for $0 \leq \sigma_i < \infty \forall \lambda > 0$ and $i \leq N$. Then

$$E \max_{i \leq N} X_i \leq \sqrt{2 logN} \max_i \sigma_i. \tag{3.31}$$

(b) Let X_i be random variables s.t $Ee^{\lambda X_i} \leq e^{\varphi_{\nu_i, c}(\lambda)}$ for $0 < \lambda \leq 1/c$ and $i = 1, \cdots, N$, where $\nu_i, c > 0$ and $\varphi_{\nu_i, c}$ is defined in (3.21). In particular, by (3.22), this holds with $c = 1/3$ if $Ee^{\lambda X_i} \leq exp(\nu_i(e^{\lambda} - 1 - \lambda))$. Then

$$E \max_{i \leq N} X_i \leq \sqrt{2\nu logN} + clogN, \tag{3.32}$$

where $\nu = \max_{i \leq N} \nu_i$.