Mathematical Foundations of Infinite-Dimensional Statistical Models
Chapter 3.1.3 $\sim$ 3.1.4

Evarist Giné and Richard Nickl

Presenter: Yongchan Kwon

Department of Statistics, Seoul National University, Seoul, Korea

ykwon0407@snu.ac.kr

TL;DR

- Groundworks for showing (i) almost sure convergence and (ii) L^p convergence of independent sample bounded processes are given.
- Advanced machineries, symmetrization and randomization techniques, are introduced. They allow general results and weaken assumptions.

Reviews

- In section 3.1.1, we simply set up some notations and define empirical processes.
- In section 3.1.2, we looked into two main types of inequalities: (i) concentration inequalities for sums of independent random variables and (ii) entropy bounds type inequality for the expected value of maximum of random variables.
- The representative examples are the celebrated inequalities of Bennet, Prokhorov, and Bernstein in Theorem 3.1.7 and the basic maximal inequalities in Theorem 3.1.10.
- In the section 3.1.3 and 3.1.4, we cover some inequalities for (sample bounded) stochastic processes and some advanced machinery.

Contents

3.1.3 The Lévy and Hoffmann-Jørgensen Inequalities

Notations

- T: a countable index set. (quite strong)
- $l_\infty(T)$: a set of real bounded functions defined on T with the supremum norm

$$\|x\|_T = \sup_{t \in T} |x(t)|,$$

which is measurable.
- $X(t)$: a stochastic process with index set T.
- We say that a process is $SBC(T)$ if almost all its sample paths are bounded and the set T is countable.
- A symmetric process if $Pr\{Y \in A\} = Pr\{-Y \in A\}$ for all A in the cylindrical σ-algebra.
- Given a sequence of independent sample bounded processes Y_i, $i = 1, \ldots, n$, indexed by T, we set

$$S_k = \sum_{i=1}^{k} Y_i, i = 1, \ldots, n, \quad \text{and} \quad Y_n^* = \max_{1 \le i \le n} \|Y_i\|_T .$$

Lévy's inequalities

Theorem (Theorem 3.1.11)

Let Y_i, $i = 1, \ldots, n$, be independent symmetric SBC(T) processes. Then, for every $t > 0$,

$$Pr\left\{\max_{1\leq k\leq n} \|S_k\|_T > t\right\} \leq 2Pr\left\{\|S_n\|_T > t\right\},$$

and

$$Pr\left\{Y_n^* > t\right\} \leq 2Pr\left\{\|S_n\|_T > t\right\}.$$

In particular,

$$E\left(\max_{1\leq k\leq n} \|S_k\|_T\right)^p \leq 2E\,\|S_n\|_T^p, \quad \text{and} \quad E\left(Y_n^*\right)^p \leq 2E\,\|S_n\|_T^p,$$

for all $p > 0$.

Lévy-Ottaviani's inequalities

Theorem (Theorem 3.1.12)

Let Y_i, $i = 1, \ldots, n$, be independent SBC(T) processes. Then, for every $u, v > 0$,

$$Pr\left\{\max_{1\le k\le n} \|S_k\|_T > u + v\right\} \le \frac{1}{1 - \max_{k\le n} Pr\{\|S_n - S_k\|_T > v\}} Pr\left\{\|S_n\|_T > u\right\},$$

and for all $t \ge 0$,

$$Pr\left\{\max_{1\le k\le n} \|S_k\|_T > t\right\} \le 3 \max_{k\le n} Pr\left\{\|S_k\|_T > \frac{t}{3}\right\}.$$

The authors drop the subindex T from the norms but I keep it for clearer notations.

Two inequalities

- Since techniques used in proofs are rather simple, I will skip all the proofs.
- Take-home messages: The two inequalities are useful to derive almost surely convergence from convergence in probability.
- What about L^p convergence?

Hoffmann-Jørgensen's inequality

Theorem (Theorem 3.1.15)

For each $p > 0$, if Y_i, $i = 1, \dots, n$, are independent symmetric SBC(T) processes, and if $t_0 \geq 0$ is defined as

$$t_0 = \inf\{t > 0 : Pr\{\|S_n\|_T > t\} \leq 1/8\},$$

then

$$\|\|S_n\|_T\|_p \leq 2^{(p+2)/p}(p+1)^{(p+1)/p}[4^{1/p}t_0 + \|Y_n^*\|_p].$$

- the L^p-norm of a sum of independent symmetric processes is dominated by L^p-norm of the maximum of their norms plus a quantile of the sum.
- I intentionally use the notation $\|\|S_n\|_T\|_p$, instead $\|S_n\|_p$ to keep clear notations.
- The proof is very long... though not difficult.

3.1.4 Symmetrisation, Randomisation, Contraction

Preliminaries

- Though the results in previous section lay theoretical grounds on convergence of random variables, the assumption *symmetry* is often strong.
- In this section, by invoking symmetrization and randomization, authors weaken the assumption: from symmetric to centered.

A contraction principle

Theorem (Theorem 3.1.17)

For $n \in \mathbb{N}$, let Y_i, $i = 1, \dots, n$, be independent SBC(T) processes, let $a_i, i \leq n$, be real numbers and let F be a nonnegative, nondecreasing convex function on $[0, \infty)$. Then, if either
(C1) $0 \leq a_i \leq 1$ and the process Y_i are centered, meaning $E(\|Y_i\|_T) < \infty$ and $E(Y_i) = 0$, or
(C2) $|a_i| \leq 1$ and the processes Y_i are symmetric,
we have

$$E\left\{F\left(\left\|\sum_{i=1}^{n} a_i Y_i\right\|_T\right)\right\} \leq E\left\{F\left(\left\|\sum_{i=1}^{n} Y_i\right\|_T\right)\right\}.$$

A contraction principle (cont'd)

Corollary (Corollary 3.1.18)

For $n \in \mathbb{N}$, let Y_i, $i = 1, \ldots, n$, be independent SBC(T) processes, let $|a_i| \leq 1, i \leq n$, be real numbers. Then, for all $p \geq 1$,

$$E\left\{\left\|\sum_{i=1}^{n} a_i Y_i\right\|_T^p\right\} \leq 2^p E\left\{\left\|\sum_{i=1}^{n} Y_i\right\|_T^p\right\}.$$

- Use $|f+g|^p \leq 2^{p-1}\{|f|+|g|\}$ and apply the previous theorem.
- The contraction inequalities are applied to random a_i and the most famous random sequence is Rademacher sequences $\{a_i\}_{i=1}^n$: $Pr\{a_i = 1\} = Pr\{a_i = -1\} = 1/2$ and a_i's are independent. [Definition 3.1.19].

A contraction principle - random version(cont'd)

Corollary (Corollary 3.1.20)

Let $\{\epsilon_i\}_{i=1}^n$ be a Rademacher sequence independent of a sequence $\{Z_i\}_{i=1}^n$ consisting of independent SBC(T) processes. Let $C_i \subset l_\infty(T)$ be such that the variable $\|\sum_{i=1}^n \tau_i Z_i I_{Z_i \in C_i}\|_T$ is measurable for all choices of $\tau_i = \pm 1$. Then, for all $p \geq 1$,

$$E\left\{\left\|\sum_{i=1}^n \epsilon_i Z_i I_{Z_i \in C_i}\right\|_T^p\right\} \leq E\left\{\left\|\sum_{i=1}^n \epsilon_i Z_i\right\|_T^p\right\}.$$

- Note that combining random variables (Rademacher) ϵ_i makes random variables $\epsilon_i Z_i$ centered.

An extension of the Lévy inequality (Thm. 3.1.11)

Theorem (Theorem 3.1.21)

Let $Y_i, i \leq n < \infty$, be independent centered SBC(T) processes with supremum norms in L^p for some $p \geq 1$, and let $\{\epsilon_i\}_{i=1}^n$ be a Rademacher sequence independent of a sequence $\{Y_i\}_{i=1}^n$. Then,

$$2^{-p} E \left\{ \left\| \sum_{i=1}^n \epsilon_i Y_i \right\|_T^p \right\} \leq E \left\{ \left\| \sum_{i=1}^n Y_i \right\|_T^p \right\} \leq 2^p E \left\{ \left\| \sum_{i=1}^n \epsilon_i (Y_i - c_i) \right\|_T^p \right\},$$

for any functions $c_i = c_i(t)$ defined on T, and

$$E \left\{ \max_{k \leq n} \left\| \sum_{i=1}^n Y_i \right\|_T^p \right\} \leq 2^{p+1} E \left\{ \left\| \sum_{i=1}^n \epsilon_i Y_i \right\|_T^p \right\}.$$

Key idea of the proof

Key idea: Symmetrization by ghost samples.

Let $\{Y_i'\}$ be a copy (ghost samples) of the sequence $\{Y_i\}$ independent of $\{Y_i\}$ and of the Rademacher sequence, and let E' denote integration with respect to these variables only. Then,

$$\begin{aligned} E\left\{\left\|\sum_{i=1}^n Y_i\right\|_T^p\right\} &= E\left\{\left\|\sum_{i=1}^n Y_i - E'\left(\sum_{i=1}^n Y_i'\right)\right\|_T^p\right\} = E\left\{\left\|E'\left(\sum_{i=1}^n Y_i - \sum_{i=1}^n Y_i'\right)\right\|_T^p\right\} \\ &\leq E\left\{\left\|\sum_{i=1}^n Y_i - \sum_{i=1}^n Y_i'\right\|_T^p\right\} = E\left\{\left\|\sum_{i=1}^n (Y_i - c_i) - \sum_{i=1}^n (Y_i' - c_i)\right\|_T^p\right\} \\ &= E\left\{\left\|\sum_{i=1}^n \epsilon_i\{(Y_i - c_i) - (Y_i' - c_i)\}\right\|_T^p\right\} \\ &\leq 2^p E\left\{\left\|\sum_{i=1}^n \epsilon_i(Y_i - c_i)\right\|_T^p\right\}. \end{aligned}$$

□

Extension to centered independent SBC(T) processes

- The left inequality can be proved by using Corollary 3.1.18.
- Theorem 3.1.22 provides an extension of the Hoffmann-Jørgensen's inequality.
- Next questions:
  (i) Can we develop tail probabilities for centered independent SBC(T) processes as well? (Proposition 3.1.23 and Proposition 3.1.24)
  (ii) Can we develop the same thing with other symmetric random variables, not Rademacher? (Proposition 3.1.25)

Tail probabilities

Theorem (Proposition 3.1.23)

Let $Y_i, i \leq n < \infty$, be independent centered SBC(T) processes, and let $|a_i| \leq 1$. Then, for all $t > 0$,

$$Pr\left\{\left\|\sum_{i=1}^{n} a_i Y_i\right\|_T > t\right\} \leq 3 \max_{j \leq n} Pr\{\|S_j\|_T > t/9\}.$$

- Note that the inequality is in a bit different form with the Proposition 3.1.12.

Tail probabilities of a SBC(T) process

Theorem (Proposition 3.1.24)

Let $Y(t), Y'(t), t \in T$, *be two SBC(T) processes defined on the factors of* $(\Omega \times \Omega', \Sigma \otimes \Sigma', Pr = P \times P')$; *that is,* $Y(t, \omega, \omega') = Y(t, \omega)$ *and* $Y'(t, \omega, \omega') = Y'(t, \omega')$, *for* $t \in T, \omega \in \Omega, \omega' \in \Omega'$. *Then, for all* $s > 0$ *and* $0 < u \leq s$ *such that* $\sup_{t \in T} Pr\{|Y'(t)| \geq u\} < 1$, *we have*

$$
\begin{aligned}
&Pr\{\|Y\|_T > s\} \\
&\leq \frac{1}{1 - \sup_{t \in T} Pr\{|Y'(t)| \geq u\}} Pr\{\|Y - Y'\|_T > s - u\}.
\end{aligned}
$$

Further, if $\theta \geq \sup_{t \in T} E[\{Y'(t)\}^2]$, *then for any* $s \geq (2\theta)^{1/2}$,

$$
Pr\{\|Y\|_T > s\} \leq 2Pr\{\|Y - Y'\|_T > s - (2\theta)^{1/2}\}.
$$

Tail probabilities of a cummulative sum

Corollary (Corollary 3.1.25)

Let $\{Y_i'\}$ be a copy (ghost samples) of the sequence $\{Y_i\}$ independent of $\{Y_i\}$ and of the Rademacher sequence. Let $\sigma^2 = \sup_{t \in T} E\{Y_i^2(t)\} < \infty$. Then for all $s \geq \sqrt{2n\sigma^2}$ and for any real numbers a_i,

$$Pr\left\{\left\|\sum_{i=1}^{n} Y_i\right\|_T > s\right\} \leq 4Pr\left\{\left\|\sum_{i=1}^{n} \epsilon_i(Y_i - a_i)\right\|_T > (s - \sqrt{2n\sigma^2})/2\right\}.$$