**Lecture No 4** March 18, 2022 (Fri)

Recall the definition of conditional expectation and conditional probability.

- $(\Omega, \mathcal{F}, P)$: Probability space
- $\mathcal{G}$: sub $\sigma$-field of $\mathcal{F}$ i.e., $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{G}$ is a $\sigma$-field
- $X = X(\omega)$: real-valued integrable r.v., i.e., $E[|X|] < \infty$
- Recall $E[X, A] := \int_A X dP$ for $A \in \mathcal{F}$.

(1) $Y = Y(\omega)$: conditional expectation of $X$ under $\mathcal{G}$, if it satisfies the following two conditions:

$$(CE) \qquad \begin{cases} ① \quad A \in \mathcal{G} \quad \Longrightarrow \quad E[X, A] = E[Y, A] \\ ② \quad Y \text{ is } \mathcal{G}\text{–measurable real-valued r.v.} \end{cases}$$

We denote $Y$ by $E[X|\mathcal{G}](\omega)$ or $E[X|\mathcal{G}]$.

(2) $P(A|\mathcal{G}) = E[1_A|\mathcal{G}](\omega)$: conditional probability of $A \in \mathcal{F}$ under $\mathcal{G}$. □

## 6.3 Properties of conditional expectation

We summarize properties of $E[X|\mathcal{G}]$ which will be useful later.
Here we denote $a.s.$ for $P$-$a.s.$

[Proposition 6.3] $X, Y$: real-valued integrable r.v.'s
(1) For $^{\forall}a, b \in \mathbb{R}$,

$$E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}] \quad a.s.$$

(Note: the exceptional set depends on $a, b$)
(2) $X \geq 0$ a.s. $\Longrightarrow E[X|\mathcal{G}] \geq 0$ a.s.
(3) $X$: $\mathcal{G}$–measurable and $Y$:bounded
(this condition can be relaxed as $E[|XY|] < \infty$)

$$\Longrightarrow E[XY|\mathcal{G}] = XE[Y|\mathcal{G}] \quad a.s.$$

In particular, $E[X|\mathcal{G}] = X$ a.s.
(4) (Tower property) $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ and $\mathcal{H}$ is also $\sigma$-field

$$\Longrightarrow E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}] \quad a.s.$$

(5) $X \perp\!\!\!\perp \mathcal{G}$ (i.e. $\sigma(X) \perp\!\!\!\perp \mathcal{G}$) $\Longrightarrow E[X|\mathcal{G}] = E[X]$ a.s. □

[Proof] Proofs are all given by checking conditions ①, ②. We omit (1), (2) (shown by contradiction) and show (3)–(5) only.

(3) Since RHS is $\mathcal{G}$–measurable, it's enough to show

$$E[XY, B] = E[XE[Y|\mathcal{G}], B]$$

for $^\forall B \in \mathcal{G}$. First, the case $X = 1_A, A \in \mathcal{G}$ is OK, since

$$E[1_A E[Y|\mathcal{G}], B] = E[E[Y|\mathcal{G}], A \cap B] = E[Y, A \cap B] = E[1_A Y, B]$$

$\therefore$ OK also for $\mathcal{G}$-measurable simple functions. For general $X$, as usual, we may approximate $X$ by simple functions.

(4) Since $E[X|\mathcal{H}]$ is obviously $\mathcal{H}$–measurable, it is enough to show

$$E[E[X|\mathcal{G}], B] = E[E[X|\mathcal{H}], B], \qquad {}^\forall B \in \mathcal{H}$$

However, both sides coincide with $E[X, B]$.

(5) Since $X \perp\!\!\!\perp \mathcal{G}$, we have

$$E[X, B] = E[X] \cdot P(B) = E[E[X], B]$$

for $^\forall B \in \mathcal{G}$. $\therefore$ $E[X|\mathcal{G}] = E[X]$ a.s. □

Jensen's inequality can be extended to the conditional expectation as follows.

[Proposition 6.4] Let $\psi : \mathbb{R} \to \mathbb{R}$ be convex and assume that a real-valued r.v. $X$ satisfies $E[|X|] < \infty, E[|\psi(X)|] < \infty$. Then,

$$\psi(E[X|\mathcal{G}]) \leq E[\psi(X)|\mathcal{G}] \quad a.s.$$

In particular, if $X \in L^p$ (i.e., $E[|X|^p] < \infty$) for $p \geq 1$, then

$$|E[X|\mathcal{G}]|^p \leq E[|X|^p|\mathcal{G}] \quad a.s. \qquad \square$$

⚇ For $^\forall a \in \mathbb{R}$, there exists $c = c(a) \in \mathbb{R}$ such that

$$\psi(x) \geq \psi(a) + c(a) \cdot (x - a),\ x \in \mathbb{R}.$$

Take $x = X, a = E[X|\mathcal{G}]$, and then take the conditional expectation $E[\cdot|\mathcal{G}]$ of both sides. For the second assertion, we may note that $\psi(x) = |x|^p, p \geq 1$ is convex. $\square$

§7 Law of large numbers

## 7.1 Weak law

- $(\Omega, \mathcal{F}, P)$: Probability space
- $\{X_n\}_{n=1}^{\infty}$: a sequence of real-valued r.v.'s s.t. $E[|X_n|] < \infty$

[Definition 7.1] (1) We say $\{X_n\}_{n=1}^{\infty}$ satisfies the weak law of large numbers, if

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - E[X_k]) \longrightarrow 0, \quad \text{in probability.}$$

(2) We say $\{X_n\}_{n=1}^{\infty}$ satisfies the strong law of large numbers, if this convergence holds in a.s.-sense (i.e. a.s.-convergence).

□

[Remark] • We call $\frac{1}{n}\sum_{k=1}^{n} X_k$ a sample mean of $\{X_n\}_{n=1}^{\infty}$.

• If $\{X_n\}_{n=1}^{\infty}$ have same distribution, $m = E[X_k]$ does not depend on $k$, and the weak/strong law of large numbers is equivalent to $\frac{1}{n}\sum_{k=1}^{n} X_k \to m$, in probability or a.s.

• In other words, the sample mean converges to expectation (called ensemble average in statistical physics).

[Theorem 7.1] If $(X_n)_{n=1}^{\infty}$ is pairwise independent (i.e., $X_i \perp\!\!\!\perp X_j$ for $^{\forall} i \neq j$) and $\sup_n \mathrm{Var}(X_n) < \infty$, then the weak law of large numbers holds. □

⚇ Set $m_k := E[X_k]$. Then, for ${}^\forall\varepsilon > 0$, we have

$$P\Big(\Big|\frac{1}{n}\sum_{k=1}^{n}(X_k - m_k)\Big| > \varepsilon\Big) \underset{\text{Chebyshev}}{\leq} \frac{1}{\varepsilon^2 n^2} E\Big[\Big|\sum_{k=1}^{n}(X_k - m_k)\Big|^2\Big]$$

$$= \frac{1}{\varepsilon^2 n^2}\sum_{j,k=1}^{n} E[(X_j - m_j)(X_k - m_k)]$$

$$\underset{\perp\!\!\!\perp}{=} \frac{1}{\varepsilon^2 n^2}\sum_{k=1}^{n} E[(X_k - m_k)^2] \leq \frac{1}{\varepsilon^2 n}\sup_n \mathrm{Var}(X_n) \quad \underset{n\to\infty}{\longrightarrow} 0 \qquad \square$$

## 7.2 Strong law

- **Kolmogorov's inequality**

The following inequality is fundamental to prove strong LLN and also extended to Doob's inequality for martingales.

[Lemma 7.2] $\{X_n\}_{n=1}^{\infty}$: a sequence of real-valued r.v.'s, $\perp\!\!\!\perp$, $E[X_n] = 0$, $V_n = \mathrm{Var}(X_n) < \infty$

$$\Longrightarrow P\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k} X_i\right| \ge a\right) \le \frac{1}{a^2}\sum_{i=1}^{n} V_i, \quad {}^{\forall} a > 0 \qquad \square$$

[Remark] $\max_{1\le k\le n} P\left(\left|\sum_{i=1}^{k} X_i\right| \ge a\right) \le$ (RHS) is obvious by applying Chebyshev's inequality for each probability. It is essential that max is inside of the probability. We need to observe the maximum of $\left|\sum_{i=1}^{k} X_i\right|$ when the time $k$ varies from 1 to $n$. $\square$

[Proof] Set $Z_k := \sum_{i=1}^{k} X_i, \quad A := \{\max_{1\le k\le n} |Z_k| \ge a\}$
$A_k := \left\{|Z_i| < a \text{ for } 1 \le^{\forall} i \le k-1 \text{ and } |Z_k| \ge a\right\}$
(i.e. the event that the time $|Z|$ first exceeds $a$ is $k$)

Then, the event $A$ of which we want to estimate probability is decomposed as

$$A = \bigcup_{k=1}^{n} A_k \qquad \text{(disjoint)}.$$

Therefore,

$$P(A) = \sum_{k=1}^{n} P(A_k) \underset{\text{Chebyshev}}{\le} \sum_{k=1}^{n} \frac{1}{a^2} E\left[Z_k^2, A_k\right]$$

However, we have

$$E\left[Z_k^2, A_k\right] \le E\left[Z_n^2, A_k\right]$$

$\because$ Noting that $Z_n^2 - Z_k^2 = (Z_n - Z_k)^2 + 2(Z_n - Z_k)Z_k$, we have

$$\begin{aligned} E\left[Z_n^2 - Z_k^2, A_k\right] &\ge 2E\left[Z_k(Z_n - Z_k), A_k\right] \\ &= 2E\left[1_{A_k} Z_k (Z_n - Z_k)\right] \end{aligned}$$

However, $1_{A_k} Z_k$ is $\sigma(\{X_i\}_{i=1}^k)$–measurable and $Z_n - Z_k = \sum_{i=k+1}^n X_i$ is $\sigma(\{X_i\}_{i=k+1}^n)$–measurable so that these are independent. Thus,

$$\text{The last expectation} = 2E\left[1_{A_k} Z_k\right] E\left[Z_n - Z_k\right] = 0 \qquad \square$$

We now return to the estimate of $P(A)$, and obtain

$$\begin{aligned} P(A) &\leq \sum_{k=1}^n \frac{1}{a^2} E\left[{Z_n}^2, A_k\right] = \frac{1}{a^2} E\left[Z_n^2, A\right] \\ &\leq \frac{1}{a^2} E[Z_n^2] = \frac{1}{a^2} E\left[\left(\sum_{i=1}^n X_i\right)^2\right] \\ &= \frac{1}{a^2} \sum_{i=1}^n V_i. \end{aligned}$$

For the last identity, we again used the independence of $\{X_i\}_{i=1}^n$.

$\square$

- **Strong law of large numbers**

[Theorem 7.3] (Kolmogorov's 1st theorem) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real-valued r.v.'s, $\perp\!\!\!\perp$, $\sum_{n=1}^{\infty} \frac{1}{n^2} \mathrm{Var}(X_n) < \infty$. Then, the strong law of large numbers holds, i.e.

$$\frac{1}{n} \sum_{k=1}^{n} (X_k - E[X_k]) \to 0 \quad \text{a.s.} \qquad \square$$

[Proof] Considering $X_k - E[X_k]$ instead of $X_k$, we may assume $E[X_k] = 0, {}^{\forall} k$. Set $Y_n := \frac{1}{n} \sum_{k=1}^{n} X_k$.

[Step 1] Once $P(A(\varepsilon)) = 1, {}^{\forall} \varepsilon > 0$ is shown, the theorem is proved, where

$$A(\varepsilon) := \varliminf_{n\to\infty} \{|Y_n| < \varepsilon\} \equiv \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|Y_n| < \varepsilon\}$$

(→See next page)

⚆ Set $A := \bigcap_{k=1}^{\infty} A(\frac{1}{k})$. Then, from $P(A(\frac{1}{k})) = 1$, we have $P(A) = 1$. However, "$\omega \in A = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|Y_n| < \frac{1}{k}\}$ $\Longrightarrow$ for $^\forall k$, $^\exists N = N(\omega, k)$ s.t. $|Y_n| < \frac{1}{k}$ for $^\forall n \geq N$".
Thus, $\omega \in A$ implies $\lim_{n\to\infty} Y_n(\omega) = 0$, which shows the conclusion of the theorem. □

[Step 2] Our goal is to show $P(A(\varepsilon)^c) = 0$. To this end, set

$$B_m(\varepsilon) := \bigcup_{n=2^{m-1}}^{2^m-1} \{|Y_n| \geq \varepsilon\} = \left\{ \max_{2^{m-1} \leq n < 2^m} |Y_n| \geq \varepsilon \right\}.$$

Then, since

$$A(\varepsilon)^c = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|Y_n| \geq \varepsilon\} \subset \bigcup_{m=\ell}^{\infty} B_m(\varepsilon)$$

for $^\forall \ell = 1, 2, \ldots$, we have

$$P(A(\varepsilon)^c) \leq \sum_{m=\ell}^{\infty} P(B_m(\varepsilon)) \underset{\ell\to\infty}{\longrightarrow} 0, \quad \text{if} \quad \sum_{m=1}^{\infty} P(B_m(\varepsilon)) < \infty.$$

[Step 3] Finally, we may show $\sum_{m=1}^{\infty} P(B_m(\varepsilon)) < \infty$.

Setting $Z_n = \sum_{k=1}^{n} X_k \, (= nY_n)$,

$$
\begin{aligned}
P(B_m(\varepsilon)) &= P\left( \max_{2^{m-1} \le n < 2^m} \frac{1}{n} |Z_n| \ge \varepsilon \right) \\
&\le P\left( \max_{2^{m-1} \le n < 2^m} |Z_n| \ge \varepsilon 2^{m-1} \right) \quad (\because \tfrac{1}{n} \le \tfrac{1}{2^{m-1}}) \\
&\le P\left( \max_{1 \le n \le 2^m} |Z_n| \ge \varepsilon 2^{m-1} \right) \\
&\le \frac{1}{\varepsilon^2 2^{2m-2}} \sum_{k=1}^{2^m} \mathrm{Var}(X_k) \quad (\because \text{Kolmogorov's inequality})
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\sum_{m=1}^{\infty} P(B_m(\varepsilon)) &\leq \frac{4}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{1}{2^{2m}} \sum_{k=1}^{2^m} \mathrm{Var}(X_k) \\
&= \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \mathrm{Var}(X_k) \sum_{m'=m_k}^{\infty} \frac{1}{2^{2m'}} \\
&\leq \frac{16}{3\varepsilon^2} \sum_{k=1}^{\infty} \mathrm{Var}(X_k) \cdot \frac{1}{k^2} < \infty
\end{aligned}$$

Here, for the second line, we first determine $m_k \in \mathbb{N}$ by $2^{m_k-1} < k \leq 2^{m_k}$ and then interchange the orders of two sums. The third line follows by $\sum_{m'=m_k}^{\infty} 2^{-2m'} = \frac{4}{3} 2^{-2m_k} \leq \frac{4}{3k^2}$. The last sum converges by our assumption. Thus, the theorem is shown. □

[Kolmogorov's 2nd theorem]

- In Theorem 7.3 (Kolmogorov's 1st theorem), we assumed the $L^2$ property of r.v.'s: $E[X_n^2] < \infty$. However, only expectation appears in the statement of LLN.
- In fact, if the distribution of $\{X_n\}$ is the same, $L^2$-integrability is unnecessary:

[Theorem 7.4] (Kolmogorov's 2nd theorem) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real-valued r.v.'s, *i.i.d.* (=independent and identically distributed) such that $E[|X_1|]\,(=E[|X_n|]) < \infty$. Then, the strong law of large numbers holds, i.e.

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - E[X_k]) \to 0 \quad \text{a.s.} \qquad \square$$

[Proof] • As before, we may assume $E[X_n] = 0$.

[Step 1] First, we introduce a cutoff to $X_k$:

$$Z_k := X_k 1_{[-k,k]}(X_k) - m_k, \quad m_k = E[X_1 1_{[-k,k]}(X_1)]$$

Then $\{Z_k\}$ satisfies the assumption of Theorem 7.3, i.e. $\sum_{k=1}^{\infty} \frac{1}{k^2} \mathrm{Var}(Z_k) < \infty$ holds (this requires some computation, but we omit the details. We use $E[|X_1|] < \infty$ only).

Thus, $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Z_k = 0 \quad a.s.$

[Step 2] Since we have bounds $|X_1 1_{[-k,k]}(X_1)| \le |X_1|$ by $k$-independent integrable function, by Lebesgue's convergence theorem, we see $m_k \underset{k\to\infty}{\to} E[X_1] = 0$. Thus, $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} m_k = 0$ for the Cesàro mean. Therefore, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k 1_{[-k,k]}(X_k) = 0 \text{ a.s.}$$

[Step 3] Finally we show $|X_k| \leq k$ e.v. Once this is shown, since $X_k = X_k 1_{[-k,k]}(X_k)$ holds except finitely many $k$'s in a.s.-sense, we obtain the conclusion.

To show $|X_k| \leq k$ e.v., by Borel-Cantelli's lemma, it is enough to show $\sum_{k=1}^{\infty} P\big(|X_k| > k\big) < \infty$. However,

$$\begin{aligned}\sum_{k=1}^{\infty} P\big(|X_k| > k\big) &= \sum_{k=1}^{\infty}\sum_{j=k}^{\infty} P\big(j < |X_1| \leq j+1\big) \\ &= \sum_{j=1}^{\infty} jP\big(j < |X_1| \leq j+1\big) \\ &\leq E[|X_1|] < \infty. \qquad \square\end{aligned}$$

[Remark] If *i.i.d.* $\{X_n\}$ satisfies $E[X_n^4] < \infty$, the proof of the strong law of large numbers is relatively easy.

∵ We may assume $E[X_n] = 0$. Then a simple computation shows $E[(\sum_{k=1}^{n} X_k)^4] \leq Cn^2$. Thus, $\sum_{n=1}^{\infty} P(|\frac{1}{n}\sum_{k=1}^{n} X_k| > \varepsilon) < \infty$ for ${}^\forall\varepsilon > 0$. Then, an argument based on Borel-Cantelli's lemma leads to the conclusion. We omit the details; see for example, Durrett, Chapter 1, (6.3) Theorem (or (6.5) Theorem in the 2nd edition). $\square$

## §8 Characteristic function and central limit theorem

Strong LLN: For $\{X_n\}$ *i.i.d.*, $m = E[X_1] < \infty$, we have

$$Y_n := \frac{1}{n}\sum_{k=1}^{n} X_k \to m \quad \text{a.s.}$$

Central limit theorem (CLT): Consider a fluctuation of $Y_n$ around the limit $m$:

$$Z_n := \sqrt{n}(Y_n - m) \equiv \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} X_k - m\right)$$

- We will show that $Z_n \Longrightarrow N(0,1)$, that is, the convergence of $Z_n$ in law sense to the Gaussian (normal) distribution with mean 0 and variance 1.
- The convergence in law is the weakest in several concepts of convergence for r.v.'s.
- We first summarize results about the convergence in law.

## 8.1 Convergence in law and weak convergence of measures

[Definition 8.1] (1) Let $\mu_n, n = 1, 2, \dots$ and $\mu$ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, we define

$\mu_n \Longrightarrow \mu$ (weak convergence) (weak*–convergence)
$\underset{\text{def}}{\Longleftrightarrow}$ For $^\forall f \in C_b(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R}, \text{ bounded continuous}\}$,

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\, \mu_n(dx) = \int_{\mathbb{R}} f(x)\, \mu(dx)$$

(2) Let $X_n, n = 1, 2, \dots$ and $X$ be real-valued r.v.'s.
We call $X_n$ converges to $X$ in law sense ($X_n \to X$ in law)
$\underset{\text{def}}{\Longleftrightarrow}$ Corresponding sequence of distributions converges weakly, i.e., $P_{X_n} \Longrightarrow P_X$
$\Longleftrightarrow \lim_{n\to\infty} E[f(X_n)] = E[f(X)], \quad {}^\forall f \in C_b(\mathbb{R})$ □

[Rem] Recall the formula of change of variables for integrals:

$$\int_{\mathbb{R}} f(x)\, P_X(dx) = E[f(X)]$$ □

[Theorem 8.1] Let $\mu_n, n = 1, 2, \dots$ and $\mu$ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, the following five conditions are mutually equivalent.

(1) $\mu_n \Longrightarrow \mu$ (weak convergence)
(2) For $^\forall G$: open set of $\mathbb{R}$, $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$
(3) For $^\forall C$: closed set of $\mathbb{R}$, $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$
(4) For $^\forall A \in \mathcal{B}(\mathbb{R})$ s.t. $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$,
where $\partial A = \bar{A} \setminus A^\circ$ is the boundary of $A$.
(5) Let $F_n$, $F$ be distributions of $\mu_n$, $\mu$, respectively, i.e. $F(x) = \mu((-\infty, x])$ etc. Then, at $^\forall$continuity point $x$ of $F$,

$$\lim_{n\to\infty} F_n(x) = F(x),$$

where $x$ is called a continuity point of $F$, if $F(x) = F(x-)\Big(:= \lim_{y\uparrow x} F(y)\Big)$ holds. □

[Proof] • (2) $\Longleftrightarrow$ (3): Mutually take complements of $G$, $C$.

• (2)+(3) $\Longrightarrow$ (4)

$\because$

$$\mu(A^\circ) \underset{(2)}{\leq} \liminf_{n\to\infty} \mu_n(A^\circ) \leq \limsup_{n\to\infty} \mu_n(\bar{A}) \underset{(3)}{\leq} \mu(\bar{A})$$

Thus, if $\mu(\partial A) = 0$, we have $\mu(A^\circ) = \mu(\bar{A}) = \mu(A)$, which implies (4). □

• (4) $\Longrightarrow$ (5)

$\because$ Note "$x$ is continuity point of $F \Longleftrightarrow \mu(\{x\}) = 0$". Take $A = (-\infty, x]$ in (4) and, if $x$ is a continuity point of $F$, then $\mu(\partial A) = 0$ so that (5) is shown. □

Next time, we will show (1) $\Longrightarrow$ (3) and (5) $\Longrightarrow$ (1).