Lecture No 2 March 11, 2022 (Fri)

1.4 Expectation and variance – continuation

In Lecture No 1, we discussed

> *Probability space, σ-fields, Kolmogorov's σ-field (on infinite product space), Random variables, Probability distributions.*

Then, for a real-valued r.v. X defined on a probability space $(\Omega, \mathcal{F}, P)$, we defined Expectation (or Mean) as an integral with respect to the measure P in Lebesgue's sense:

$$E[X] := \int_{\Omega} X(\omega)\, P(d\omega)$$

Fundamental inequalities

(1) Chebyshev's inequality: For $p > 0, \varepsilon > 0$,

$$P(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} E[|X|^p]$$

∵ $P(|X| \geq \varepsilon) = E[1_{\{|X| \geq \varepsilon\}}] \leq E\left[\left(\frac{|X|}{\varepsilon}\right)^p\right]$ □

(2) Jensen's inequality: Assume $\psi : \mathbb{R} \to \mathbb{R}$ is convex, i.e., for ${}^\forall \lambda \in (0,1), {}^\forall x, y \in \mathbb{R}$

$$\psi(\lambda x + (1-\lambda)y) \leq \lambda\psi(x) + (1-\lambda)\psi(y)$$

hold. Then, if $E[|X|] < \infty$, $E[|\psi(X)|] < \infty$, we have

$$\psi(E[X]) \leq E[\psi(X)]$$

∵ By the convexity of ψ, for ${}^\forall a \in \mathbb{R}$, there exists ${}^\exists c = c(a) \in \mathbb{R}$ s.t. $\psi(a) + c(a)(x-a) \leq \psi(x), x \in \mathbb{R}$. Take $x = X, a = E[X]$ and then take the expectations of both sides. □

(3) Hölder's inequality: For $1 < p, q < \infty$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$|E[XY]| \leq E[|X|^p]^{\frac{1}{p}} E[|Y|^q]^{\frac{1}{q}}$$

In particular, when $p = q = 2$, it is called Schwarz's inequality.

Interchange of the expectation and limits

Let $(X_n)_{n=1,2,\ldots}, X$ be real-valued r.v.'s defined on a probability space $(\Omega, \mathcal{F}, P)$.

(1) Lebesgue's convergence theorem: If $X_n \to X$ $(a.s.)$ (a.s.-convergence, i.e. $P({}^{\exists}\lim_{n\to\infty} X_n = X) = 1$) and ${}^{\exists}$non-negative integrable r.v. Y s.t. $|X_n| \le Y$ for ${}^{\forall}n$, then

$$\lim_{n\to\infty} E[X_n] = E[X]$$

(2) Monotone convergence theorem: If $0 \le X_1 \le X_2 \le \cdots$ and $X_n \to X$ $(a.s.)$, then

$$\lim_{n\to\infty} E[X_n] = E[X]$$

(3) Fatou's lemma: If $X_n \ge 0$,

$$E\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty} E[X_n]$$

P: Give an example that "$<$" holds in (3).

Variance, covariance and moment

For real-valued r.v.'s X, X_1, X_2, we define

$$\mathrm{Var}(X) := E[(X - E[X])^2] \quad \text{variance of } X$$

$$\mathrm{Cov}(X_1, X_2) := E\left[(X_1 - E[X_1])(X_2 - E[X_2])\right]$$

$$\text{covariance of } X_1, X_2$$

$$E[X^n] \;: n\,\text{th moment of } X$$

These exist if $E[X^2] < \infty$; $E[X_1^2] < \infty$, $E[X_2^2] < \infty$; $E[|X|^n] < \infty$, respectively.

[Example] (Higher dimensional Gaussian (normal) distribution)
For $m = (m_i) \in \mathbb{R}^d$ and $d \times d$ positive definite real symmetric matrix $V = (V_{ij})$, the measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$

$$\mu_{m,V}(dx) = \frac{1}{(2\pi)^{d/2}(\det V)^{1/2}} \exp\left\{-\frac{1}{2}(x-m)\cdot V^{-1}(x-m)\right\} dx$$

is called a Gaussian distribution with mean m and covariance matrix V. Here, dx is the Lebesgue measure on $\mathbb{R}^d$, V^{-1} denotes the inverse matrix of V and $\cdot$ denotes the inner product in $\mathbb{R}^d$. Moreover, an $\mathbb{R}^d$–valued r.v. $X = (X_1, X_2, \ldots, X_d)$ having distribution $\mu_{m,V}$ is called a Gaussian random variable. It holds

$$E[X] \equiv \big(E[X_i]\big)_{i=1,2,\ldots,d} = m,$$
$$\mathrm{Cov}(X) \equiv \big(\,\mathrm{Cov}(X_i, X_j)\big)_{i,j=1,2,\ldots,d} = V$$

□

§2 Dynkin's π–λ theorem

- This theorem is useful when we want to extend "some fact" (or "some property") to all sets in a σ-field.
- For example, this is used to show the uniqueness of measure in Theorem 2.5 below.
- We first introduce families of sets called π-system and λ-system.
- "λ-system" represents the sets for which we want to show the fact, while "π-system" represents the sets for which we already know the fact holds.

Dynkin, 2003 (from Wikipedia)

[Definition 2.1] (1) A family $\mathcal{P}$ of subsets of Ω is called π–system, if the followings are satisfied

1. $\Omega \in \mathcal{P}$
2. $A, B \in \mathcal{P} \Longrightarrow A \cap B \in \mathcal{P}$

(2) A family $\mathcal{L}$ of subsets of Ω is called λ–system (Dynkin's family), if the followings are satisfied

1. $\Omega \in \mathcal{L}$
2. $A, B \in \mathcal{L}$ and $A \subset B \Longrightarrow B \setminus A \in \mathcal{L}$
3. $A_n \in \mathcal{L} \nearrow A$ (i.e. $A_1 \subset A_2 \subset \cdots$, $\bigcup_{n=1}^{\infty} A_n = A$) $\Longrightarrow A \in \mathcal{L}$

□

[Theorem 2.1] (Dynkin's π–λ theorem) (1) Let $\mathcal{P}$ be a π–system and $\mathcal{L}$ be a λ–system such that $\mathcal{P} \subset \mathcal{L}$. Then, $\sigma(\mathcal{P}) \subset \mathcal{L}$ holds.
(2) Set $\mathcal{L}(\mathcal{P})$ the smallest λ–system which includes $\mathcal{P}$. Then, if $\mathcal{P}$ is a π–system, $\mathcal{L}(\mathcal{P}) = \sigma(\mathcal{P})$ holds. □

(2) follows from (1). Indeed, (1) shows $\sigma(\mathcal{P}) \subset \mathcal{L}(\mathcal{P})$ and the converse inclusion is obvious, since a σ–field is a λ–system. Therefore, it is enough to show (1) only.

[Proof of Theorem 2.1 (1)] We omit the details, but the followings are shown; see [1] (Williams' book), Chapter A1.

(a) $\mathcal{P}$: π–system $\Longrightarrow \mathcal{L}(\mathcal{P})$: π-system

(b) $\mathcal{A}$: π–system and λ–system $\Longrightarrow \mathcal{A}$: σ–field
($\Longleftarrow$ is obvious)

From (a), $\mathcal{L}(\mathcal{P})$ is a π-system so that, by (b), we see that $\mathcal{L}(\mathcal{P})$ is a σ-field. $\therefore$ We know $\sigma(\mathcal{P}) \subset \mathcal{L}(\mathcal{P})$. However, from the assumption "$\mathcal{P} \subset \mathcal{L}$ and $\mathcal{L}$: λ–system", $\mathcal{L}(\mathcal{P}) \subset \mathcal{L}$. Thus, we obtain $\sigma(\mathcal{P}) \subset \mathcal{L}$. $\square$

[Important examples of π–system] Families of semi-open intervals, rectangles with sides parallel to the axes.

(1) $\Omega = \mathbb{R}$, $\mathcal{P} = \{(y, x]; -\infty \le y \le x \le \infty\}$
where $(x, x] := \emptyset, (y, \infty] := (y, \infty)$

(2) $\Omega = \mathbb{R}^2$, $\mathcal{P} = \{(a, b] \times (c, d]\}$,
where $-\infty \le a \le b \le \infty, -\infty \le c \le d \le \infty$

(3) $\Omega = S^{\mathbb{N}}$, $\mathcal{P} = \{C : \text{cylinder sets}\}$ $\square$

[Example of application]

[Theorem 2.5] (Uniqueness of probability measure) Let $\mathcal{P}$ be a π–system and two probability measures P_1, P_2 on $\sigma(\mathcal{P})$ be given. If P_1 and P_2 coincide on $\mathcal{P}$, then they coincide also on $\sigma(\mathcal{P})$, that is, $P_1 = P_2$ holds.

☺ • Set $\mathcal{L} = \{A \in \sigma(\mathcal{P});\ P_1(A) = P_2(A)\}$. If one can show that $\mathcal{L}$ is a λ–system, since the assumption implies $\mathcal{P} \subset \mathcal{L}$, by π–λ theorem, we obtain $\sigma(\mathcal{P}) \subset \mathcal{L}$. This shows the conclusion.

• However, it is easy to show that $\mathcal{L}$ is a λ–system. Indeed, $\Omega \in \mathcal{L}$ is obvious. The axiom 2 of λ–system is also obvious. The axiom 3 follows from the continuity of measures: For $A_n \in \mathcal{L} \nearrow A$,

$$P_1(A) = \lim_{n\to\infty} P_1(A_n) = \lim_{n\to\infty} P_2(A_n) = P_2(A),$$

which implies $A \in \mathcal{L}$. □

P: If $\mathcal{P}$ is not a π–system, even if two measures coincide on $\mathcal{P}$, they may not coincide on $\sigma(\mathcal{P})$. Give an example.

In particular, for two measures μ, ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, one can naturally define a product measure $\mu \times \nu$ on rectangles. Then, by Theorem 2.5, its extension to $\mathcal{B}(\mathbb{R}^2)$ (if exists) is unique.

3.1 Definition

Let a sequence of real-valued r.v.'s $(X_n)_{n=1,2,\dots}$ and a real-valued r.v. X be given on a probability space $(\Omega, \mathcal{F}, P)$.

[Definition 3.1] (1) (a.s.-convergence) $X_n \to X$ a.s.

$$\underset{\text{def}}{\iff} P\Big(\lim_{n\to\infty} X_n = X\Big) = 1 \quad \Big(\text{i.e. } P\Big(\Big\{\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}\Big) = 1\Big)$$

(2) (Convergence in probability) $X_n \to X$ in prob.

$$\underset{\text{def}}{\iff} \text{For } {}^\forall\varepsilon > 0,\ \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

(3) (L^p-convergence, Convergence in mean of p th order)

For $p \geq 1$, $X_n \to X$ in L^p

$$\underset{\text{def}}{\iff} \lim_{n\to\infty} E[|X_n - X|^p] = 0$$

(4) (Convergence in law) $X_n \to X$ in law

$$\underset{\text{def}}{\iff} \text{For } {}^\forall f \in C_b(\mathbb{R}),\ \lim_{n\to\infty} E[f(X_n)] = E[f(X)]$$

where $C_b(\mathbb{R})$ stands for the set of all real-valued bounded continuous functions on $\mathbb{R}$. □

[Remark] • Convergence in law can be defined for X_n, X defined on different probability spaces.
• (2), (3), (4) determine metrizable topologies.
• It is known that metric equivalent to the convergence (1) does not exists and it is not defined through any topology. See Durrett, "Probability: Theory and Examples", Chapter 1, Ex 6.2.

3.2 Relation between convergences

[Theorem 3.1] Let $(X_n)_{n=1,2,\dots}, X$ be real-valued r.v.'s.
(1) a.s.-convergence $\Longrightarrow$ convergence in probability
(2) L^p-convergence$\Longrightarrow$convergence in probability
(3) convergence in probability $\Longrightarrow$ convergence in law □

⚆ (1) follows by
$P(|X_n - X| > \varepsilon) = E\left[1_{(\varepsilon,\infty)}(|X_n - X|)\right] \to 0$, for which one can use Lebesgue's convergence theorem.
(2) is an easy consequence of Chebyshev's inequality:

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} E\left[|X_n - X|^p\right] \to 0.$$

To show (3), we prepare a lemma.

[Lemma 3.2] If $X_n \to X$ in prob., one can choose a subsequence X_{n_k} such that $X_{n_k} \to X$ a.s. holds.
(Note that n_k is independent of ω) $\square$

∵ By the convergence in probability, one can choose an increasing sequence $n_k \in \mathbb{N}$ such that
$P(|X_{n_k} - X| \geq \frac{1}{k}) \leq 2^{-k}$ holds.
Since the sum in RHS in k is finite, by Borel-Cantelli's lemma (stated below) , we see $P(|X_{n_k} - X| < \frac{1}{k}, \text{e.v.}) = 1$, and this shows the conclusion. $\square$

[Lemma 3.3] (Borel-Cantelli's lemma) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then we have

$$P\left(\overline{\lim_{n\to\infty}} A_n\right) = 0, i.e., P\left(\underline{\lim_{n\to\infty}} A_n^c\right) = 1,$$

where

$$\omega \in \overline{\lim_{n\to\infty}} A_n := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \iff \omega \text{ belongs infinitely many } A_n$$

$$\omega \in \underline{\lim_{n\to\infty}} A_n^c := \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^c \iff {}^{\exists}k \text{ s.t. } \omega \in A_n^c \text{ for } {}^{\forall}n \geq k$$

We write the former as "A_n i.o. (infinitely often)", the latter as "A_n^c e.v. (eventually)". □

∵

$$P\left(\overline{\lim_{n\to\infty}} A_n\right) \underset{({}^{\forall}k)}{\leq} P\left(\bigcup_{n=k}^{\infty} A_n\right) \leq \sum_{n=k}^{\infty} P(A_n) \to 0 \ \ (k \to \infty) \qquad \square$$

We return to the proof of (3): For $X_n \to X$ in prob., take its any subsequence $(X_{n'})$. Then, by Lemma 3.2, one can find its further subsequence $(X_{n''})$ such that $X_{n''} \to X$ a.s. Then, by Lebesgue's convergence theorem, for ${}^\forall f \in C_b(\mathbb{R})$, we obtain $E[f(X_{n''})] \to E[f(X)]$. However, this means that, without taking subsequence, $E[f(X_n)] \to E[f(X)]$ holds. □

[Recall] Let (a_n) be a real sequence. Assume that its any subsequence $(a_{n'})$ contains a further converging subsequence $(a_{n''})$ such that $a_{n''} \to a$. If the limit a is common, we see that (a_n) itself converges to a without taking subsequence.

P: Show above by a simple ε-δ (or ε-N) argument.

We summarize the relation for convergences of r.v.'s.

$$\begin{array}{lcl} \text{a.s.-conv.} & \searrow & \\ & & \text{conv. in prob.} \quad \Longrightarrow \quad \text{conv. in law} \\ L^p\text{-conv.} & \nearrow & \end{array}$$

§4 Independence, sub σ-field as an amount of information

Independence is a unique notion in probability theory.
Reflecting amount of information, we consider several types of sub σ-fields of $\mathcal{F}$.

4.1 Independence of events

Probability space $(\Omega, \mathcal{F}, P)$ is always given.

[Definition 4.1] A family of events $\{A_\lambda \in \mathcal{F}\}_{\lambda \in \Lambda}$ parametrized by Λ is independent
$\underset{\text{def}}{\iff}$ For ${}^\forall\{\lambda_1, \ldots, \lambda_n\} \subset \Lambda$ (finite set of different points),

$$P\left(\bigcap_{k=1}^{n} A_{\lambda_k}\right) = \prod_{k=1}^{n} P\left(A_{\lambda_k}\right)$$

holds. □

[Remark] • In particular, if Λ is a set of two points

$$A, B \in \mathcal{F} \text{ is independent} \Longleftrightarrow P(A \cap B) = P(A)P(B)$$

This is a well-known definition.
• Even if $\{A_1, \ldots, A_n\} \subset \mathcal{F}$ satisfy

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{k=1}^{n} P(A_k),$$

they need not be independent. This identity should be satisfied by any finite number of sets taken from $\{A_1, \ldots, A_n\}$. □

P: Construct a counterexample.

4.2 Independence of σ–fields

[Definition 4.2] A family $\{\mathcal{F}_\lambda\}_{\lambda\in\Lambda}$ of sub σ-fields of $\mathcal{F}$ parametrized by Λ is independent

$\underset{\text{def}}{\iff}$ For $^\forall A_\lambda \in \mathcal{F}_\lambda$, $\{A_\lambda\}_{\lambda\in\Lambda}$ is independent. □

4.3 Independence of random variables

Random variables may take values in different spaces, but they are defined on a common probability space $(\Omega, \mathcal{F}, P)$.

[Definition 4.3] A family of $(S_\lambda, \mathcal{S}_\lambda)$-valued r.v.'s $X_\lambda, \lambda \in \Lambda$ parametrized by Λ is independent

$\underset{\text{def}}{\iff}$ $\{\sigma(X_\lambda)\}_{\lambda\in\Lambda}$ is independent □

[Recall] We set $\sigma(X) = \{X^{-1}(A); A \in \mathcal{S}\}$, which is also denoted by $\mathcal{F}_X$.

- Independence of $(X_\lambda)_{\lambda \in \Lambda}$ can be rephrased as

$$P(X_{\lambda_k} \in A_k, k = 1, 2, \ldots, n) = \prod_{k=1}^{n} P(X_{\lambda_k} \in A_k)$$

  holds for any finite subset $\lambda_1, \ldots, \lambda_n \in \Lambda$ of Λ and $A_k \in \mathcal{S}_{\lambda_k}$ $(1 \leq k \leq n)$.
- We denote Independence as $A \perp\!\!\!\perp B$, $\{A_\lambda\} \perp\!\!\!\perp$, $\{\mathcal{F}_\lambda\} \perp\!\!\!\perp$, $\{X_\lambda\} \perp\!\!\!\perp$.