Lecture No 3 March 15, 2022 (Tue)

4.3 Independence of random variables – continuation

In the last part of Lecture No 3, we defined

> *the notion of Independence of events* $\{A_\lambda \in \mathcal{F}\}_{\lambda\in\Lambda}$*,*
> *sub* σ*-fields* $\{\mathcal{F}_\lambda\}_{\lambda\in\Lambda}$ *of* $\mathcal{F}$ *and r.v.'s* $\{X_\lambda\}_{\lambda\in\Lambda}$

Independence of r.v.'s is invariant under the composition with measurable functions.

[Lemma 4.1] Assume that a family of S_λ–valued r.v.'s $(X_\lambda)_{\lambda\in\Lambda}$ is independent and $g_\lambda : S_\lambda \to S'_\lambda$ are measurable. Here, $(S'_\lambda, \mathcal{S}'_\lambda)$ are other measurable spaces. Then, $\{Y_\lambda := g_\lambda(X_\lambda)\}_{\lambda\in\Lambda}$ is also independent. □

⚇ This follows by showing $\sigma(Y_\lambda) \subset \sigma(X_\lambda)$: For ${}^\forall A \in \mathcal{S}'_\lambda$, we have $\{Y_\lambda \in A\} = \{X_\lambda \in g_\lambda^{-1}(A)\} \in \sigma(X_\lambda)$. □

P: (Exercise) Let $(X_k)_{k=1,2,\ldots,n}$ be an independent family of real-valued r.v.'s and let $g = g(x_1, x_2, \ldots, x_i) : \mathbb{R}^i \to \mathbb{R}$ be Borel measurable with $i < n$. Set $Y = g(X_1, X_2, \ldots, X_i)$. Then, $Y, X_{i+1}, \ldots, X_n$ is a family of independent r.v.'s. □

[Hint] Show the independence of

$$\{X = (X_1, X_2, \ldots, X_i), X_{i+1}, \ldots, X_n\}$$

by π–λ theorem, and then use Lemma 4.1. Note that X is an $\mathbb{R}^i$-valued r.v. obtaining by gathering $\{X_1, X_2, \ldots, X_i\}$. □

[Lemma 4.2] Let $(X_k)_{k=1,2,\dots,n}$ be an independent family of integrable r.v.'s. Then, the product $X = X_1 X_2 \cdots X_n$ of these r.v.'s is also integrable (i.e. $E[|X|] < \infty$) and

$$E[X] = E[X_1]E[X_2]\cdots E[X_n]$$

holds. □

☺ • It's enough to show the conclusion when each X_k is simple. Indeed, first, by decomposing $X_k = X_k^+ - X_k^-$, we may assume $X_k \geq 0$. Then, since such X_k can be approximated by an increasing sequence of non-negative simple functions, one may apply the monotone convergence theorem.

• A simple function is written as

$$X_k(\omega) = \sum_{i \in I_k} a_i^k 1_{A_i^k}(\omega)$$

Here, we may assume $a_i^k \neq a_{i'}^k (i \neq i', i, i' \in I_k)$ and $\{A_i^k\}_{i \in I_k}$ are mutually disjoint. Then, the independence follows by showing $A_i^k \in \sigma(X_k)$ and, based on this, one can compute $E[X]$. □

[Lemma 4.3] If a sequence of real-valued r.v.'s $(X_k)_{k=1,2,\dots,n}$ is pairwise independent (i.e. any two of them is independent) and $\mathrm{Var}(X_k) < \infty \, (k = 1, 2, \dots, n)$, then

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k)$$

holds. □

☺ This is easily shown by the definition of variance. After expanding square, cross terms vanish by independence. □

§5 Product of probability spaces

To construct a sequence of independent r.v.'s, we construct product of probability spaces.

5.1 Product of finitely many probability spaces

First, we consider the product of 2 probability spaces.

- $(S_1, \mathcal{S}_1, \mu_1)$, $(S_2, \mathcal{S}_2, \mu_2)$: 2 probability spaces
- $\Omega := S_1 \times S_2$: product space
- $\mathcal{P} := \{A_1 \times A_2 \subset \Omega; A_1 \in \mathcal{S}_1, A_2 \in \mathcal{S}_2\}$: a family of rectangles
- $\mathcal{S}_1 \times \mathcal{S}_2 := \sigma\{\mathcal{P}\}$: we call product σ-field
- Define a function P on $\mathcal{P}$ as

$$P(A_1 \times A_2) := \mu_1(A_1)\mu_2(A_2), \qquad A_1 \times A_2 \in \mathcal{P}$$

Then,

[Proposition5.1] P can be uniquely extended as a probability measure on $\mathcal{F} := \mathcal{S}_1 \times \mathcal{S}_2$. We denote $P =: \mu_1 \times \mu_2$. □

∵ • This was shown in the course of Lebesgue measure theory.

• $^\exists P$ on $\mathcal{F}$ is shown by Carathéodory-Hopf extension theorem:

> *Let $\mathcal{C}(\subset \mathcal{P}(\Omega))$ be a finitely additive family and assume that P is σ-additive on $(\Omega, \mathcal{C}) \Longrightarrow P$ can be uniquely extended to $(\Omega, \sigma(\mathcal{C}))$ to be σ-additive.*

• The uniqueness follows by π-λ theorem noting that $\mathcal{P}$ is a π-system. □


Carathéodory (from Wikipedia)

Let n probability spaces $(S_k, \mathcal{S}_k, \mu_k), k = 1, 2, \ldots, n$ be given

- $\Omega = \prod_{k=1}^{n} S_k$
- $\mathcal{P} := \{A_1 \times \cdots \times A_n \subset \Omega; A_k \in \mathcal{S}_k\}$: all of rectangular parallelepipeds
- $P(A_1 \times \cdots \times A_n) := \mu_1(A_1) \cdots \mu_n(A_n)$ on $\mathcal{P}$
- $\mathcal{F} \equiv \mathcal{S}_1 \times \cdots \times \mathcal{S}_n := \sigma(\mathcal{P})$
- $P = \mu_1 \times \cdots \times \mu_n$: product measure on $\mathcal{F}$ is uniquely constructed.

⚇ Construction: Once product measure of 2 probability spaces is constructed, one can define P inductively in n

$$P = (\mu_1 \times \cdots \times \mu_{n-1}) \times \mu_n.$$

Uniqueness: Apply π-λ theorem. □

Based on product of probability spaces, one can construct a sequence of finitely many independent r.v.'s.

[Proposition 5.2] Let real-valued r.v.'s $X^{(k)} : S_k \to \mathbb{R}$ be given on probability spaces $(S_k, \mathcal{S}_k, \mu_k)$, $k = 1, 2, \ldots, n$.
For $\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega = \prod_{k=1}^n S_k$, set

$$X_k(\omega) = X^{(k)}(\omega_k).$$

Then, $(X_k)_{k=1,2,\ldots,n}$ is an independent sequence of r.v.'s defined on the probability space $(\Omega, \mathcal{F}, P)$ and each X_k has a same distribution as $X^{(k)}$, where $\mathcal{F} = \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n$ and $P = \mu_1 \times \mu_2 \times \cdots \times \mu_n$. □

[Proof] • X_k and $X^{(k)}$ has same distribution: For $^\forall B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} P(X_k \in B) &= \big(\mu_1 \times \mu_2 \times \cdots \times \mu_n\big)(X^{(k)}(\omega_k) \in B) \\ &= \mu_k(X^{(k)} \in B) \end{aligned}$$

• Independence: For $^\forall B_1, B_2, \dots, B_n \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} &P(X_1 \in B_1, \dots, X_n \in B_n) \\ &\quad = \big(\mu_1 \times \mu_2 \times \cdots \times \mu_n\big)(X^{(1)} \in B_1, \dots, X^{(n)} \in B_n) \\ &\quad = \prod_{k=1}^{n} \mu_k(X^{(k)} \in B_k) = \prod_{k=1}^{n} P(X_k \in B_k). \end{aligned}$$

This implies the independence of $(X_k)_{k=1,2,\dots,n}$. □

5.2 Product of infinitely many probability spaces

Let a sequence of infinitely many probability spaces $\{(S_n, \mathcal{S}_n, \mu_n)\}_{n=1,2,\dots}$ be given.

- $\Omega := \prod_{n=1}^{\infty} S_n$: infinite product space
- $\mathcal{C} := \{C :$ cylinder sets of $\Omega\}$, where

$$C \equiv C_A^{(n)} = \{\omega = (\omega_1, \omega_2, \dots) \in \Omega\,;\, (\omega_1, \dots, \omega_n) \in A\},$$
$$n \in \mathbb{N}, A \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$$

- $\mathcal{F} := \sigma(\mathcal{C})$: Kolmogorov's σ–field
- $P(C_A^{(n)}) := (\mu_1 \times \mu_2 \times \cdots \times \mu_n)(A)$ on $\mathcal{C}$

  (Remark: well-defined on $\mathcal{C}$ independently of choice of n, A)

Then,

- Probability measure P on $(\Omega, \mathcal{F})$ can be uniquely constructed as an extension of P on $(\Omega, \mathcal{C})$. For the proof, we again use Carathéodory-Hopf extension theorem.
- We denote $P =: \prod_{n=1}^{\infty} \mu_n$ and call it infinite product measure.
- Not necessarily product measure, but in more general setting, Kolmogorov's extension theorem is known. $\rightarrow$ We state Kolmogorov's theorem later for the construction of Brownian motion. This is actually useful for the construction of general Markov processes.

Based on infinite product of probability spaces, one can construct an infinite sequence of independent r.v.'s. Assuming that r.v.'s are real-valued, we formulate in a different manner from Proposition 5.2.

[Proposition 5.3] Let $\mu_1, \mu_2, \ldots$ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, on some probability space $(\Omega, \mathcal{F}, P)$, one can construct real-valued r.v.'s X_n such that

(1) $\{X_n\}_{n=1,2,\ldots}$ $\perp\!\!\!\perp$

(2) Each X_n has μ_n as its distribution. □

⨀ Take $\Omega = \mathbb{R}^{\mathbb{N}}, \mathcal{F} =$ Kolmogorov's σ–field, $P = \prod_{k=1}^{\infty} \mu_k$ and $X_n(\omega) := \omega_n, \omega = (\omega_1, \omega_2, \ldots) \in \Omega$. Then, the proof is similar to that of Proposition 5.2. □

§6 Conditional probability and Conditional expectation

These notions are fundamental in probability theory, and especially, necessary to develop martingale theory, which is a base of stochastic analysis, since stochastic integrals (defined later) determine martingales.

6.1 Radon-Nikodým's theorem

We apply Radon-Nikodým's theorem known in measure theory to define conditional probability and conditional expectation.

Radon

Nikodým–Banach (from Wikipedia)

[Definition 6.1] (1) Let $(S, \mathcal{S})$ be a measurable space. A set function $\nu : \mathcal{S} \to \mathbb{R}$ is called a finite signed measure, if it is σ-additive:

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \nu(A_n) \text{ for disjoint } A_n \in \mathcal{S}, n = 1, 2, \ldots.$$

(2) Let μ be a measure on $(S, \mathcal{S})$ and let ν be a finite signed measure on $(S, \mathcal{S})$. We say ν is absolutely continuous with respect to μ (denoted by $\nu \ll \mu$) if and only if $\nu(A) = 0$ holds for $A \in \mathcal{S}$ such that $\mu(A) = 0$. □

[Remark for (1)] It is important that ν does not take infinite values $\pm\infty$. From this fact, through Hahn decomposition (decomposition of ν to positive and negative parts), we see that $\sum_{n=1}^{\infty} \nu(A_n)$ always converges absolutely.

[Example] Let a measure space $(S, \mathcal{S}, \mu)$ and a real-valued integrable function f on S (i.e. $\int_S |f(x)|\,\mu(dx) < \infty$) be given and set

$$\nu(A) = \int_A f(x)\,\mu(dx), \quad A \in \mathcal{S}.$$

Then, ν is a finite signed measure which is absolutely continuous with respect to μ. □

∵ Integrability of f shows the finiteness of ν and σ-additivity of Lebesgue integral in A implies that ν is a signed measure. If $\mu(A) = 0$, then $1_A(x)f(x) = 0$, μ-a.e. so that $\nu(A) = \int_S 1_A(x)f(x)\mu(dx) = 0$. This shows the absolute continuity of ν with respect to μ. □

The converse of [Example] is also true. This is Radon-Nikodým's theorem.

[Theorem 6.1] (Radon-Nikodým's theorem) Let μ be a finite measure on $(S, \mathcal{S})$ (i.e. $\mu(S) < \infty$) and let ν be a finite signed measure on $(S, \mathcal{S})$, which is absolutely continuous with respect to μ. Then,

(1) There exists a real-valued $\mathcal{S}$–measurable function f on S, which is μ-integrable, such that ν can be represented (as above [Example]) as

$$\nu(A) = \int_A f(x)\,\mu(dx), \quad A \in \mathcal{S}.$$

(2) (Uniqueness of f) If another function $\tilde{f}$ exists and represents ν as above, then we have
$f(x) = \tilde{f}(x), \quad \mu$-a.e.$x$ □

We call f a density function of ν with respect to μ and denote

$$\frac{d\nu}{d\mu}(x).$$

- The proof of Radon-Nikodým's theorem can be found in a usual textbook of Lebesgue measure theory.
- The uniqueness of f (the assertion (2)) is easy. Indeed, from (1), for $^\forall A \in \mathcal{S}$

$$\int_A \{f(x) - \tilde{f}(x)\}\, \mu(dx) = 0$$

holds. In particular, taking $A = \{f - \tilde{f} > 1/n\}$, we see $\mu(\{f - \tilde{f} > 1/n\}) = 0, n = 1, 2, \ldots$. Thus, by the σ–additivity of μ, we obtain $\mu(\{f - \tilde{f} > 0\}) = 0$. Similarly, we obtain $\mu(\{f - \tilde{f} < 0\}) = 0$ and therefore $\mu(\{f - \tilde{f} \neq 0\}) = 0$. This shows $f = \tilde{f}, \mu$-$a.e.\ x$.

6.2 Definition of conditional probability and conditional expectation

Setting: We first give an abstract definition.

- $(\Omega, \mathcal{F}, P)$: Probability space
- $\mathcal{G}$: sub σ-field of $\mathcal{F}$ i.e., $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{G}$ is a σ-field
- $X = X(\omega)$: real-valued integrable r.v., i.e., $E[|X|] < \infty$
- For $A \in \mathcal{F}$, we denote the integral of X on A by

$$E[X, A] := \int_A X dP \left(= \int_A X(\omega) P(d\omega) \right).$$

(For example, $E[X, \Omega] = E[X]$)

[Definition 6.2] (1) We call $Y = Y(\omega)$ a conditional expectation of X under $\mathcal{G}$, if it satisfies the following two conditions:

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

Y is unique in P-$a.s.$ sense and we denote $Y = Y(\omega)$ by $E[X|\mathcal{G}](\omega)$ or $E[X|\mathcal{G}]$.

(2) For $A \in \mathcal{F}$,

$$P(A|\mathcal{G}) = E[1_A|\mathcal{G}](\omega)$$

is called a conditional probability of A under $\mathcal{G}$. □

- Existence and uniqueness of Y

  $\because$ Set $\nu(A) := E[X, A]$ $(= \int_A X(\omega)P(d\omega))$, $A \in \mathcal{G}$ and $\mu(A) := P(A)$, $A \in \mathcal{G}$ (restriction of P to $\mathcal{G}$). Then, since $\nu \ll \mu$ as measures on $\mathcal{G}$, by Radon-Nikodým's theorem, there exists a $\mathcal{G}$-measurable function $Y(\omega)$ such that

$$\nu(A) = \int_A Y(\omega)P(d\omega) = E[Y, A]$$

  holds for ${}^{\forall}A \in \mathcal{G}$. Such Y is unique in P-a.s. sense. □

- When $\mathcal{G} = \mathcal{F}$, $E[X|\mathcal{G}] = X$ (P-a.s.)

  $\because$ $Y = X$ satisfies ①, ② in the condition (CE). □

- When $\mathcal{G} = \{\emptyset, \Omega\}$, $E[X|\mathcal{G}] = E[X]$ (P-a.s.)

  $\because$ $Y = E[X]$ (constant function) satisfies ①, ② in the condition (CE). □

Let's see the abstract definition of conditional probability and conditional expectation coincides with a simple definition in case that $\mathcal{G}$ is determined from a finite division of Ω.

- For given $B \in \mathcal{F}$ s.t. $P(B) > 0$, we simply define conditional probability of A by

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

(i.e. Once we know B happens, the probability that A happens is given by this)

- Corresponding conditional expectation (i.e. expectation under the measure $P(\cdot|B)$) is defined by

$$E[X|B] := \frac{E[X, B]}{P(B)}$$

In fact, when $X = 1_A$, this coincides with $P(A|B)$. For general X, we first approximate it by simple functions and then take the limit.

- This can be extended to the case of finite division $\{B_i\}_{i=1}^n$ of Ω (i.e., $B_i \in \mathcal{F}$, disjoint, $\Omega = \bigcup_{i=1}^n B_i$, $0 < P(B_i) < 1$).
- Namely, we define the conditional probability of A under the division $\{B_i\}_{i=1}^n$ by

$$P(A|\{B_i\}_{i=1}^n)(\omega) := \sum_{i=1}^n P(A|B_i) 1_{B_i}(\omega)$$

- Corresponding conditional expectation is defined by

$$E[X|\{B_i\}_{i=1}^n](\omega) = \sum_{i=1}^n E[X|B_i] 1_{B_i}(\omega)$$

- In other words, on B_i, we define these concepts by $P(A|B_i)$ and $E[X|B_i]$.

Abstract and simple definitions coincide:

[Proposition 6.2] When $\mathcal{G} = \sigma(\{B_i\}_{i=1}^n)$, we have $E[X|\{B_i\}_{i=1}^n] = E[X|\mathcal{G}]$, P-a.s. □

∵ Set $Y(\omega) = \sum_{i=1}^n E[X|B_i] 1_{B_i}(\omega)$. Then Y is $\mathcal{G}$–measurable so that it satisfies ② in the condition (CE). Thus it's enough to show ① in the condition (CE). To this end, for $^\forall A \in \mathcal{G}$,

$$E[Y, A] = E\left[\sum_{i=1}^n E[X|B_i] 1_{B_i}, A\right] = \sum_{i=1}^n \frac{E[X, B_i]}{P(B_i)} P(B_i \cap A)$$

However, since $A \in \mathcal{G}$, A is a union of some of $\{B_i\}$ and

$$P(B_i \cap A) = \begin{cases} P(B_i), & \text{if } B_i \subset A \\ 0 \quad , & \text{if } B_i \cap A = \emptyset \end{cases}$$

Therefore,

$$E[Y, A] = \sum_{i: B_i \subset A} E[X, B_i] = E[X, A] = \text{LHS of ①} \qquad \square$$