LECTURE NOTES IN STOCHASTIC PROCESSES

GUAN-YU CHEN

1. Conditional probability and expectation

1.1. **Definitions and properties.** Recall that if E is an event with $\mathbb{P}(E) > 0$, then the conditional probability of A given E is defined by

$$\mathbb{P}(A|E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)},$$

or equivalently

$$\mathbb{P}(A \cap E) = \mathbb{P}(A|E)\mathbb{P}(E). \tag{1.1}$$

Formally, when $\mathbb{P}(E) = 0$, (1.1) becomes

$$0 = \mathbb{P}(A \cap E) = \mathbb{P}(A|E)\mathbb{P}(E) = 0 \times \mathbb{P}(A|E).$$

From the above identity, it seems that the value of $\mathbb{P}(A|E)$ can be any real number (or $\pm\infty$ if $0 \times \infty := 0$) when $\mathbb{P}(E) = 0$. Such a concept actually applies for discrete probabilities but not for general cases. Note that if E is a mutually disjoint union of events with positive probabilities, say $(E_n)_{n=1}^{\infty}$, then

$$\mathbb{P}(A \cap E) = \sum_{n=1}^{\infty} \mathbb{P}(A|E_n)\mathbb{P}(E_n). \tag{1.2}$$

This gives an idea of defining the conditional probability in a more general setting.

For convenience, let's consider the case that E is given by a random variable. Let X be a random variable and $E = \{a < X \leq b\}$. For $n \in \mathbb{N}$ and $1 \leq i \leq n$, set $x_i = a + i(b-a)/n$ and $E_i = \{x_{i-1} < X \leq x_i\}$. Assuming $\mathbb{P}(x_{i-1} < X \leq x_i) > 0$ gives

$$\mathbb{P}(A \cap E) = \sum_{i=1}^{n} \mathbb{P}(A|E_i)\mathbb{P}(E_i).$$

In a formal computation, if the following limit exists (in any proper sense)

$$\mathbb{P}(A|X = x) := \lim_{h \to 0} \frac{\mathbb{P}(A \cap \{x - h \leq X \leq x + h\})}{\mathbb{P}(x - h \leq X \leq x + h)},$$

then

$$\mathbb{P}(A \cap \{a < X \leq b\}) = \int_{(a,b]} \mathbb{P}(A|X = x)F_X(dx).$$

The last equality was given by the Radon-Nikodym theorem.

Theorem 1.1. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ, ν be measures on $(\Omega, \mathcal{F})$, where ν is non-negative and σ-finite and μ is a signed measure. If μ is absolutely continuous w.r.t. ν, that is, $\nu(B) = 0$ implies $|\mu|(B) = 0$, then there exists a $\mathcal{F}$-measurable function f such that*

$$\mu(B) = \int_B f(x)\nu(dx), \quad \forall |\mu|(B) < \infty.$$

Back to the conditional probability, let $A \in \mathcal{F}$ and set, for $B \in \mathcal{B}(\mathbb{R})$,

$$\nu(B) = \mathbb{P}(X \in B), \quad \mu(B) = \mathbb{P}(A \cap \{X \in B\}).$$

Since μ is absolutely continuous w.r.t. ν, we may define the conditional probability by the Radon-Nykodym theorem.

Definition 1.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$ be a random element. For $A \in \mathcal{F}$, the conditional probability of A given $X = x$ is denoted by $\mathbb{P}(A|X = x)$ and defined to be a $\mathcal{B}$-measurable function such that

$$\mathbb{P}(A \cap \{X \in B\}) = \int_B \mathbb{P}(A|X = x)\mathbb{P}_X(dx), \quad \forall B \in \mathcal{B}.$$

In a similar way, one may consider the case that $\nu(E) = \mathbb{P}(E)$ and $\mu(E) = \mathbb{P}(A \cap E)$ for all $E \in \mathcal{F}(X)$.

Definition 1.2. For any random element X and any event $A \in \mathcal{F}$, the conditional probability of A given $X(\omega)$ is denoted by $\mathbb{P}(A|X(\omega))$ or briefly $\mathbb{P}(A|X)$ and defined to be a $\mathcal{F}(X)$-measurable function such that

$$\mathbb{P}(A \cap E) = \int_E \mathbb{P}(A|X(\omega))\mathbb{P}(d\omega), \quad \forall E \in \mathcal{F}(X).$$

Remark 1.1. Note that $\mathbb{P}(A|X = x)$ (resp. $\mathbb{P}(A|X)$) is almost surely non-negative. Furthermore, if φ and φ' are versions of $\mathbb{P}(A|X = x)$ (resp. $\mathbb{P}(A|X)$), then $\varphi = \varphi'$ $\mathbb{P}_X$-a.s. (resp. $\mathbb{P}$-a.s.)

Proposition 1.2. *Let $A \in \mathcal{F}$ and $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$ and $Y : (\Omega, \mathcal{F}) \to (R', \mathcal{B}')$ be random elements.*

(1) *If $\varphi(x) = \mathbb{P}(A|X = x)$ $\mathbb{P}_X$-a.s., then $\varphi(X) = \mathbb{P}(A|X)$ $\mathbb{P}$-a.s.*
(2) *If $\mathbb{P}(A|X) = \psi(X)$ $\mathbb{P}$-a.s. for some $\mathcal{B}$-measurable function ψ, then $\psi(x) = \mathbb{P}(A|X = x)$ $\mathbb{P}_X$-a.s..*
(3) *If $\mathcal{F}(X) = \mathcal{F}(Y)$, then $\mathbb{P}(A|X) = \mathbb{P}(A|Y)$ $\mathbb{P}$-a.s..*
(4) *If $\sigma(\{A\})$ and $\mathcal{F}(X)$ are independent, then*

$$\mathbb{P}(A|X = x) = \mathbb{P}(A) \ \mathbb{P}_X\text{-a.s.}, \quad \mathbb{P}(A|X) = \mathbb{P}(A) \ \mathbb{P}\text{-a.s.}$$

Proof. For (1), let $E \in \mathcal{F}(X)$ and choose $B \in \mathcal{B}$ such that $E = \{X \in B\}$. Then,

$$\int_E \varphi(X)d\mathbb{P} = \int_\Omega \varphi(X)\mathbf{1}_B(X)d\mathbb{P} = \int_R \varphi(x)\mathbf{1}_B(x)\mathbb{P}_X(dx) = \mathbb{P}(A \cap \{X \in B\}) = \mathbb{P}(A \cap E).$$

This implies $\varphi(X) = \mathbb{P}(A|X)$ $\mathbb{P}$-a.s.

To see (2), note that, for any version of $\mathbb{P}(A|X)$, there is always a $\mathcal{B}$-measure function ψ such that $\mathbb{P}(A|X) = \psi(X)$. For $B \in \mathcal{B}$ and $E = \{X \in B\}$, one has

$$\mathbb{P}(A \cap \{X \in B\}) = \mathbb{P}(A \cap E) = \int_E \psi(X)d\mathbb{P} = \int_B \psi(x)\mathbb{P}_X(dx).$$

This implies, $\psi(x) = \mathbb{P}(A|X = x)$ $\mathbb{P}_X$-a.s. (3) and (4) are immediate from the definition of $\mathbb{P}(A|X)$ and $\mathbb{P}(A|X = x)$. ☐

Next, we introduce the conditional expectation. Recall that if $\mathbb{P}(E) > 0$ and $\mathbb{E}|Y| < \infty$, then $\mathbb{E}(Y|E) := \int_\Omega Y(\omega)\mathbb{P}(d\omega|E)$, where $\mathbb{P}(A|E) = \mathbb{P}(A \cap E)/\mathbb{P}(E)$ for all $A \in \mathcal{F}$. Note that

$$\int_\Omega Y(\omega)\mathbb{P}(d\omega|E) = \int_E Y(\omega)\mathbb{P}(d\omega) \times \frac{1}{\mathbb{P}(E)}.$$

This implies $\int_E Y d\mathbb{P} = \mathbb{E}(Y|E)\mathbb{P}(E)$. As before, if E is a mutually disjoint union of $(E_n)_{n=1}^\infty$ and $\mathbb{P}(E_n) > 0$, then

$$\int_E Y(\omega)\mathbb{P}(d\omega) = \sum_{n=1}^\infty \mathbb{E}(Y|E_n)\mathbb{P}(E_n),$$

which is another identity similar to (1.2). We may then extend the definition of conditional probabilities to that of conditional expectations. Let $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$ be a random element and let Y be a random variable satisfying $\mathbb{E}|Y| < \infty$. For $B \in \mathcal{B}$, set

$$\nu(B) = \mathbb{P}(X \in B), \quad \mu(B) = \int_{\{X\in B\}} Y d\mathbb{P},$$

and, for $E \in \mathcal{F}(X)$, set

$$\nu'(E) = \mathbb{P}(E), \quad \mu'(E) = \int_E Y d\mathbb{P}.$$

Obviously, μ and μ' are absolutely continuous w.r.t. ν and ν'. Again, one may apply the Radon-Nykodym theorem to achieve the following definitions.

Definition 1.3. Let Y be a random variable satisfying $\mathbb{E}|Y| < \infty$. For any random element $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$, the conditional expectation of Y given $X = x$ (resp. $X(\omega)$) is denoted by $\mathbb{E}(Y|X = x)$ (resp. $\mathbb{E}(Y|X)$) and is defined to be a $\mathcal{B}$-measurable (resp. $\mathcal{F}(X)$-measurable) function satisfying

$$\int_{\{X\in B\}} Y d\mathbb{P} = \int_B \mathbb{E}(Y|X = x)\mathbb{P}_X(dx), \quad \forall B \in \mathcal{B}.$$

$$\left(\text{resp. } \int_E Y d\mathbb{P} = \int_E \mathbb{E}(Y|X)d\mathbb{P}, \quad \forall E \in \mathcal{F}(X).\right)$$

In the above definition, it is clear that, for $A \in \mathcal{F}$,

$$\mathbb{P}(A|X = x) = \mathbb{E}(\mathbf{1}_A|X = x) \ \mathbb{P}_X\text{-a.s.}, \quad \mathbb{P}(A|X) = \mathbb{E}(\mathbf{1}_A|X) \ \mathbb{P}\text{-a.s.}$$

Proposition 1.3. *Let $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$ be a random element and $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variables satisfying $\mathbb{E}|Y| < \infty$.*

(1) *If $\varphi(x) = \mathbb{E}(Y|X = x)$ $\mathbb{P}_X$-a.s., then $\varphi(X) = \mathbb{E}(Y|X)$ $\mathbb{P}$-a.s.*

(2) *If $\mathbb{E}(Y|X) = \psi(X)$ $\mathbb{P}$-a.s. for some $\mathcal{B}$-measurable function ψ, then $\psi(x) = \mathbb{E}(Y|X = x)$ $\mathbb{P}_X$-a.s.*

(3) *If X_1, X_2 are random elements satisfying $\mathcal{F}(X_1) = \mathcal{F}(X_2)$, then $\mathbb{E}(Y|X_1) = \mathbb{E}(Y|X_2)$ $\mathbb{P}$-a.s..*

(4) *If X, Y are independent, then*

$$\mathbb{E}(Y|X = x) = \mathbb{E}Y \ \mathbb{P}_X\text{-a.s.}, \quad \mathbb{E}(Y|X) = \mathbb{E}Y \ \mathbb{P}\text{-a.s.}$$

(5) *If $\mathcal{F}(Y) \subset \mathcal{F}(X)$ and ψ is a $\mathcal{B}$-measurable random variable satisfying $Y = \psi(X)$ $\mathbb{P}$-a.s., then*

$$\mathbb{E}(Y|X = x) = \psi(x) \ \mathbb{P}_X\text{-a.s.}, \quad \mathbb{E}(Y|X) = Y \ \mathbb{P}\text{-a.s.}.$$

Proof. The proof for (1)-(4) is similar to the proof of Proposition 1.2. For (5), note that, for $E \in \mathcal{F}(X)$,

$$\int_E \mathbb{E}(Y|X)d\mathbb{P} = \int_E Y d\mathbb{P}.$$

This implies $\mathbb{E}(Y|X) = Y$ $\mathbb{P}$-a.s., which proves the second identity. The first one is obvious from (2) since $\mathbb{E}(Y|X) = Y = \psi(X)$ $\mathbb{P}$-a.s. and ψ is $\mathcal{B}$-measurable. □

Exercise 1.1. Prove by following the definition that if $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$ is a random element taking values on $\{x_i : i = 1, 2, ...\}$ and $A \in \mathcal{F}$, then

$$\mathbb{P}(A|X) = \sum_{i=1}^{\infty} \delta_{x_i}(X)\mathbb{P}(A|X = x_i), \quad \mathbb{P}\text{-a.s.},$$

where $\mathbb{P}(A|X = x) := \mathbb{P}(A \cap \{X = x\})/\mathbb{P}(X = x)$ if $\mathbb{P}(X = x) > 0$ and $\mathbb{P}(A|X = x) := 0$ otherwise.

Exercise 1.2. Let $\Omega = [-1, 1]$, $\mathcal{F} = \mathcal{B}(\Omega)$ and $\mathbb{P}(d\omega) = \frac{1}{2}\mu(d\omega)$, where μ is the Lebesgue measure on $[-1, 1]$.

(1) Let X be a random variable defined by $X(\omega) = |\omega|$. Find a version of $\mathbb{P}(A|X)$ and $\mathbb{P}(A|X = x)$ and do the same problem for the case $X(\omega) = \omega^2$.
(2) Assume that $\mathbb{E}|Y| < \infty$. Find a version for $\mathbb{E}(Y|X)$ and $\mathbb{E}(Y|X = x)$ for respective X in (1).

Exercise 1.3. Let X and Y be random variables with joint density f and set $f_X(x) = \int_{\mathbb{R}} f(x, y)dy$.

(1) Prove that, for $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(Y \in B|X = x) = \int_B \frac{f(x, y)}{f_X(x)} dy, \quad \mathbb{P}_X\text{-a.s.}$$

(2) Assume that $\mathbb{E}|Y| < \infty$. Prove that

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y\frac{f(x, y)}{f_X(x)} dy, \quad \mathbb{P}_X\text{-a.s.}$$

1.2. **A general definition.**

Definition 1.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability and $\mathcal{C} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$. Let $A \in \mathcal{F}$ and Y be a random variable satisfying $\mathbb{E}|Y| < \infty$.

(1) The conditional probability of A given $\mathcal{C}$ is a $\mathcal{C}$-measurable function satisfying

$$\int_C \mathbb{P}(A|\mathcal{C})d\mathbb{P} = \mathbb{P}(A \cap C), \quad \forall C \in \mathcal{C}.$$

(2) The conditional expectation of Y given $\mathcal{C}$ is a $\mathcal{C}$-measurable function satisfying

$$\int_C \mathbb{E}(Y|\mathcal{C})d\mathbb{P} = \int_C Y d\mathbb{P} \quad \forall C \in \mathcal{C}.$$

In particular, if $\mathcal{C} = \mathcal{F}(X_1, X_2, ...)$, we also write $\mathbb{P}(A|X_1, X_2, ...)$ and $\mathbb{E}(Y|X_1, X_2, ...)$ instead.

Remark 1.2. Note that $\mathbb{P}(A|\mathcal{C}) = \mathbb{E}(\mathbf{1}_A|\mathcal{C})$ $\mathbb{P}$-a.s..

Remark 1.3. For $\mathbb{E}|Y| < \infty$, $\mathbb{E}(Y|X) = \mathbb{E}(Y|\mathcal{F}(X))$.

Proposition 1.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability and $\mathcal{C} \subset \mathcal{F}$. Assume that Y, Y_1, Y_2 are random variables with finite mean.*

(1) *For any constants a, b,*

$$\mathbb{E}(aY_1 + bY_2|\mathcal{C}) = a\mathbb{E}(Y_1|\mathcal{C}) + b\mathbb{E}(Y_2|\mathcal{C}), \quad \mathbb{P}\text{-a.s.}$$

(2) *If $Y \geq 0$ $\mathbb{P}$-a.s., then $\mathbb{E}(Y|\mathcal{C}) \geq 0$ $\mathbb{P}$-a.s.*
(3) *If $\mathcal{F}(Y) \subset \mathcal{C}$, then $\mathbb{E}(Y|\mathcal{C}) = Y$ $\mathbb{P}$-a.s.*
(4) *If $\mathcal{C} \subset \mathcal{E} \subset \mathcal{F}$, then*

$$\mathbb{E}(\mathbb{E}(Y|\mathcal{C})|\mathcal{E}) = \mathbb{E}(\mathbb{E}(Y|\mathcal{E})|\mathcal{C}) = \mathbb{E}(Y|\mathcal{C}) \quad \mathbb{P}\text{-a.s.}$$

(5) *If* $\mathcal{F}(Y)$ *and* $\mathcal{C}$ *are independent, then* $\mathbb{E}(Y|\mathcal{C}) = \mathbb{E}Y$ $\mathbb{P}$*-a.s.*
(6) $\mathbb{E}(\mathbb{E}(Y|\mathcal{C})) = \mathbb{E}Y$.

Proof. (1)-(3) and (6) are obvious from the definition. For (4), note that $-|Y| \leq Y \leq |Y|$. By (2), this implies $-\mathbb{E}(|Y||\mathcal{C}) \leq \mathbb{E}(Y|\mathcal{C}) \leq \mathbb{E}(|Y||\mathcal{C})$ or equivalently $|\mathbb{E}(Y|\mathcal{C})| \leq \mathbb{E}(|Y||\mathcal{C})$ $\mathbb{P}$-a.s.. As a result of (6), this implies $\mathbb{E}|\mathbb{E}(Y|\mathcal{C})| \leq \mathbb{E}|Y| < \infty$. Observe that, for $C \in \mathcal{C} \subset \mathcal{E}$,

$$\int_C \mathbb{E}(\mathbb{E}(Y|\mathcal{E})|\mathcal{C})d\mathbb{P} = \int_C \mathbb{E}(Y|\mathcal{E})d\mathbb{P} = \int_C Y d\mathbb{P}.$$

This proves the desired identity. For (5), assume that $\mathcal{F}(Y)$ and $\mathcal{C}$ are independent. For $C \in \mathcal{C}$,

$$\int_C \mathbb{E}(Y|\mathcal{C})d\mathbb{P} = \mathbb{E}[Y\mathbf{1}_C] = \mathbb{E}Y\mathbb{P}(C).$$

□

Theorem 1.5 (Monotone convergence theorem)**.** *Let* $Y_1, Y_2, ...$ *be a sequence of non-negative random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$*. Suppose* $Y_n \leq Y_{n+1}$*,* Y_n *converge to* Y *a.s. and* $\mathbb{E}Y < \infty$*. Then, for any sub-*σ*-field* $\mathcal{C} \subset \mathcal{F}$*,*

$$\mathbb{E}(Y_n|\mathcal{C}) \to \mathbb{E}(Y|\mathcal{C}) \quad \textit{almost surely.}$$

Proof. Set $Z_n = Y - Y_n$. Then, $Z_n \geq 0$ and $Z_n \geq Z_{n+1}$. By Proposition 1.4, this implies, for $n \geq 1$, $\mathbb{E}(Z_n|\mathcal{C}) \geq \mathbb{E}(Z_{n+1}|\mathcal{C})$ a.s. and, thus, $\mathbb{E}(Z_n|\mathcal{C})$ converges a.s. to a non-negative random variable, say Z. Furthermore, using the fact of $Z_n \leq Y$, one may conclude $\mathbb{E}(Z_n|\mathcal{C}) \leq \mathbb{E}(Y|\mathcal{C})$ a.s. and $\mathbb{E}|\mathbb{E}(Y|\mathcal{C})| = \mathbb{E}Y < \infty$. By the Lebesgue bounded convergence theorem, we obtain

$$\mathbb{E}Z = \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(Z_n|\mathcal{C})) = \lim_{n\to\infty} \mathbb{E}Z_n = 0,$$

which proves $Z = 0$ a.s.. □

Theorem 1.6 (Fatou's lemma)**.** *Let* $Y_1, Y_2, ...$ *be non-negative random variables defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *with finite mean and let* $\mathcal{C} \subset \mathcal{F}$ *be a sub-*σ*-field. If* $\liminf_n Y_n$ *is integrable, then,*

$$\mathbb{E}\left(\liminf_{n\to\infty} Y_n \middle| \mathcal{C}\right) \leq \liminf_{n\to\infty} \mathbb{E}(Y_n|\mathcal{C}) \quad \textit{almost surely.}$$

Proof. Applying Theorem 1.5 to the sequence $\inf_{k\geq n} Y_k$ yields

$$\mathbb{E}\left(\liminf_{n\to\infty} Y_n \middle| \mathcal{C}\right) = \lim_{n\to\infty} \mathbb{E}\left(\inf_{k\geq n} Y_k \middle| \mathcal{C}\right) \leq \liminf_{n\to\infty} \mathbb{E}(Y_n|\mathcal{C}) \quad \text{almost surely.}$$

□

Theorem 1.7 (Lebesgue's bounded convergence theorem)**.** *Let* $Z, Y_1, Y_2, ...$ *be random variables satisfying* $|Y_n| \leq Z$ *and* $\mathbb{E}Z < \infty$*. If* Y_n *converges to* Y *a.s., then, for any* σ*-field* $\mathcal{C} \subset \mathcal{F}$*,*

$$\mathbb{E}(Y_n|\mathcal{C}) \to \mathbb{E}(Y|\mathcal{C}) \quad \textit{a.s.}$$

Proof. Consider the two sequences $(Z + \inf_{k\geq n} Y_k)_{n=1}^\infty$ and $(Z - \sup_{k\geq n} Y_k)_{n=1}^\infty$. By Theorem 1.5, one has

$$\mathbb{E}(Z + Y|\mathcal{C}) \overset{a.s.}{=} \lim_{n\to\infty} \mathbb{E}\left(Z + \inf_{k\geq n} Y_k \middle| \mathcal{C}\right) \overset{a.s.}{\leq} \mathbb{E}(Z|\mathcal{C}) + \liminf_{n\to\infty} \mathbb{E}(Y_n|\mathcal{C})$$

and

$$\mathbb{E}(Z - Y|\mathcal{C}) \overset{a.s.}{=} \lim_{n\to\infty} \mathbb{E}\left(Z - \sup_{k\geq n} Y_k \middle| \mathcal{C}\right) \overset{a.s.}{\leq} \mathbb{E}(Z|\mathcal{C}) - \limsup_{n\to\infty} \mathbb{E}(Y_n|\mathcal{C}).$$

Combining both inequalities gives the desired identity. □

Exercise 1.4. Let $X : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$ be a random variable and $\mathcal{C}$ be a sub-σ-field such that $\mathcal{F}(X) \subset \mathcal{C}$. Show that, for any random variable Y satisfying $\mathbb{E}|Y| < \infty$ and $\mathbb{E}|XY| < \infty$,

$$\mathbb{E}(XY|\mathcal{C}) = X\mathbb{E}(Y|\mathcal{C}) \quad a.s.$$

By Exercise 1.4, one has

Proposition 1.8. *Let X be a random element from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B})$, φ be a random variable on $(R, \mathcal{B})$ and Y be a random variable on $(\Omega, \mathcal{F})$. Suppose $\mathbb{E}|Y| < \infty$ and $\mathbb{E}|\varphi(X)Y| < \infty$. Then,*

(1) $\mathbb{E}(\varphi(X)Y|X) = \varphi(X)\mathbb{E}(Y|X)$ *a.s.*
(2) $\mathbb{E}(\varphi(X)Y|X = x) = \varphi(x)\mathbb{E}(Y|X = x)$ $\mathbb{P}_X$*-a.s.*

Proof. The first identity is immediate from Exercise 1.4. For (2), set $f(x) = \mathbb{E}(Y|X = x)$. By Proposition 1.3, $\mathbb{E}(Y|X) = f(X)$ a.s. and $\mathbb{E}(\varphi(X)Y|X) = \varphi(X)f(X)$ a.s.. Since φf is $\mathcal{B}$-measurable, $\varphi(x)f(x) = \mathbb{E}(\varphi(X)Y|X = x)$ $\mathbb{P}_X$-a.s. □

Remark 1.4. Note that the notation of $\mathbb{E}(\varphi(x)Y|X = x)$ makes no sense.

Exercise 1.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability. For any two random variables X, Y, define the L^2-distance between them by

$$d(X, Y) = \sqrt{\mathbb{E}[(X - Y)^2]}.$$

Suppose $\mathbb{E}Y^2 < \infty$. Prove that, for any σ-field $\mathcal{C} \subset \mathcal{F}$,

$$\inf \{d(X, Y) : X \text{ is } \mathcal{C}\text{-measurable}\} = d(\mathbb{E}(Y|\mathcal{C}), Y).$$

Exercise 1.6. Let Y be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}|Y| < \infty$. Assume that X_1, X_2 are random variables such that $\mathcal{F}(Y, X_1)$ and $\mathcal{F}(X_2)$ are independent. Prove that $\mathbb{E}(Y|X_1, X_2) = \mathbb{E}(Y|X_1)$ a.s..

Exercise 1.7. Let $X_1, X_2, ...$ be i.i.d. random variables with $\mathbb{E}|X_1| < \infty$ and set $S_n = X_1 + \cdots + X_n$. Prove that

$$\mathbb{E}(X_1|S_n, S_{n+1}, ...) = \frac{S_n}{n} \quad \text{a.s.}$$

1.3. **Regular conditional probabilities.**

Definition 1.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability and $\mathcal{F}_1, \mathcal{C}$ be sub-σ-fields of $\mathcal{F}$. $\mathbb{P}(A|\mathcal{C})$ is called a regular conditional probability on $\mathcal{F}_1$ given $\mathcal{C}$ if

(1) For $A \in \mathcal{F}_1$, $\mathbb{P}(A|\mathcal{C})$ is a conditional probability of A given $\mathcal{C}$.
(2) For $\omega \in \Omega$, $\mathbb{P}(\cdot|\mathcal{C})$ is a probability on $(\Omega, \mathcal{F}_1)$.

Proposition 1.9. *Let $\mathbb{P}(A|\mathcal{C})$ be a regular conditional probability on $\mathcal{F}_1$ given $\mathcal{C}$ and Y be a random variable on $(\Omega, \mathcal{F}_1)$ satisfying $\mathbb{E}|Y| < \infty$. Then,*

$$\mathbb{E}(Y|\mathcal{C}) = \int_\Omega Y(\omega)\mathbb{P}(d\omega|\mathcal{C}) \quad a.s.$$

Proof. By the linearity of conditional expectation and Theorem 1.7, we only need to consider the case $Y = \mathbf{1}_A$ where $A \in \mathcal{F}_1$, but this is obvious since

$$\mathbb{E}(\mathbf{1}_A|\mathcal{C}) \overset{a.s.}{=} \mathbb{P}(A|\mathcal{C}) = \int_\Omega \mathbf{1}_A(\omega)\mathbb{P}(d\omega|\mathcal{C}).$$

□

There are examples for which a regular conditional probability does not exist and this leads to the following definition.

Definition 1.6. Let $Y : (\Omega, \mathcal{F}) \to (R, \mathcal{B})$ be a random element and $\mathcal{C}$ be a sub-σ-field of $\mathcal{F}$. $\mathbb{P}^*(B|\mathcal{C})$ with $B \in \mathcal{B}$ is called a regular conditional distribution for Y given $\mathcal{C}$ if

(1) For $B \in \mathcal{B}$, $\mathbb{P}^*(B|\mathcal{C})$ is a conditional probability of $\{Y \in B\}$ given $\mathcal{C}$, that is, $\mathbb{P}^*(B|\mathcal{C}) = \mathbb{P}(Y \in B|\mathcal{C})$ a.s.
(2) For $\omega \in \Omega$, $\mathbb{P}^*(\cdot|\mathcal{C})$ is a probability on $(R, \mathcal{B})$.

Proposition 1.10. *Let Y be a random element taking values on $(R, \mathcal{B})$ and $\mathbb{P}^*(\cdot|\mathcal{C})$ be a regular conditional distribution for Y given $\mathcal{C}$. For any random variable φ on $(R, \mathcal{B})$ satisfying $\mathbb{E}|\varphi(Y)| < \infty$, one has*

$$\mathbb{E}(\varphi(Y)|\mathcal{C}) = \int_R \varphi(y)\mathbb{P}^*(dy|\mathcal{C}) \quad \textit{almost surely.}$$

Proof. The proof is similar to that of Proposition 1.9. Due to the linearity of conditional expectation and Theorem 1.7, one only needs to consider the case $\varphi = \mathbf{1}_B$ with $B \in \mathcal{B}$. This is in fact the case $\mathbb{P}(Y \in B|\mathcal{C}) = \mathbb{P}^*(B|\mathcal{C})$ a.s., which is exactly the definition of regular conditional distribution. □

As in the case of regular conditional probability, the regular conditional distribution might not exist.

Theorem 1.11. *For any random variable Y, there is a regular conditional distribution for Y given $\mathcal{C}$.*

Proof. **Step 1:** There exists a conditional distribution function $F(x|\mathcal{C})$, that is,

(1) For $y \in \mathbb{R}$, $F(y|\mathcal{C})$ is a conditional probability of $\{Y \le y\}$ given $\mathcal{C}$.
(2) For $\omega \in \Omega$, $F(y|\mathcal{C})$ is a distribution function.

To see this, let $\mathbb{Q} = \{q_i : i = 1, 2, ...\}$ be the set of all rational numbers and fix a version of $\mathbb{P}(Y \le q_i|\mathcal{C})$ for all $i \ge 1$. Define

$$M = \bigcup_{q_i < q_j} M_{i,j}, \quad M_{i,j} = \{\omega : \mathbb{P}(Y \le q_i|\mathcal{C}) > \mathbb{P}(Y \le q_j|\mathcal{C})\}$$

and

$$N = \bigcup_{i=1}^{\infty} N_i, \quad N_i = \left\{\omega \in M^c : \lim_{q\in\mathbb{Q},\, q\downarrow q_i} \mathbb{P}(Y \le q|\mathcal{C}) \ne \mathbb{P}(Y \le q_i|\mathcal{C})\right\}$$

and

$$L = \left\{\omega \in M^c : \lim_{q\in\mathbb{Q},\, q\uparrow\infty} \mathbb{P}(Y \le q|\mathcal{C}) \ne 1, \lim_{q\in\mathbb{Q},\, q\downarrow-\infty} \mathbb{P}(Y \le q|\mathcal{C}) \ne 0\right\}.$$

Then, $\mathbb{P}(M \cup N \cup L) = 0$.

Let $G(y)$ be any distribution function and, for $\omega \in \Omega \setminus (M \cup N \cup L)$, define

$$F(y|\mathcal{C}) = \begin{cases} G(y) & \text{if } \omega \in M \cup N \cup L \\ \lim_{r_j\downarrow y} \mathbb{P}(Y \le r_j|\mathcal{C}) & \text{otherwise} \end{cases}$$

Then, $F(y|\mathcal{C})$ is the desired distribution function.

Step 2: Let $F(y|\mathcal{C})$ be the conditional distribution chosen in Step 1. For $\omega \in \Omega$, let $\mathbb{P}^*(\cdot|\mathcal{C})$ be the unique probability on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying

$$\mathbb{P}^*((-\infty, y]|\mathcal{C}) = F(y|\mathcal{C}) \quad \forall y \in \mathbb{R}.$$

Set

$$\mathcal{D} = \{B \in \mathcal{B}(\mathbb{R}) : \mathbb{P}^*(B|\mathcal{C}) \overset{a.s.}{=} \mathbb{P}(Y \in B|\mathcal{C})\}.$$

Clearly, $\mathcal{D}$ is a λ-system containing $(-\infty, y]$ with $y \in \mathbb{R}$. By the $\pi - \lambda$ lemma, $\mathcal{D} = \mathcal{B}(\mathbb{R})$. □

Definition 1.7. A measurable space $(R, \mathcal{B})$ is called a *Borel space* if there exist $E \in \mathcal{B}(\mathbb{R})$ and a one-to-one correspondence $f : R \to E$ such that f and f^{-1} are respectively $\mathcal{B}$-measurable and $\mathcal{B}(\mathbb{R})$-measurable.

Theorem 1.12. *If $(R, \mathcal{B})$ is a Borel space and Y is a random element with values in R, then there exists a regular conditional distribution for Y given $\mathcal{C}$.*

Proof. Let E and $\varphi : R \to E$ be the Borel set and one-to-one correspondence in Definition 1.7 and $Z = \varphi(Y)$. Then, Z is a random variable and, hence, has a regular conditional distribution for Z given $\mathcal{C}$, say $\mathbb{P}^*(B|\mathcal{C})$ with $B \in \mathcal{B}(\mathbb{R})$. By defining

$$\widehat{\mathbb{P}}(D|\mathcal{C}) := \mathbb{P}^*(\varphi(D)|\mathcal{C}) \overset{a.s.}{=} \mathbb{P}(Z \in \varphi(D)|\mathcal{C}) \overset{a.s.}{=} \mathbb{P}(Y \in D|\mathcal{C}), \quad \forall D \in \mathcal{B}, \tag{1.3}$$

$\widehat{\mathbb{P}}(D|\mathcal{C})$ is a regular conditional distribution for Y given $\mathcal{C}$. □

Corollary 1.13. *If $(\Omega, \mathcal{F})$ is a Borel space, then, for any sub-σ-fields $\mathcal{F}_1, \mathcal{C}$ of $\mathcal{F}$, there exists a regular conditional probability on $\mathcal{F}_1$ given $\mathcal{C}$.*

Proof. The proof is obtained by applying Theorem 1.12 with $Y(\omega) = \omega$ for $\omega \in \Omega$. □

In a similar way, we may define a regular conditional probability (resp. distribution) given $X = x$.

Definition 1.8. Let X be a random element from $(\Omega, \mathcal{F})$ to $(S, \mathcal{E})$ and $\mathcal{F}_1 \subset \mathcal{F}$. $\mathbb{P}(A|X = x)$ is a regular conditional probability on $\mathcal{F}_1$ given $X = x$ if

(1) For $A \in \mathcal{F}_1$, $\mathbb{P}(A|X = x)$ is a conditional probability of A given $X = x$;
(2) For $x \in S$, $\mathbb{P}(\cdot|X = x)$ is a probability on $(\Omega, \mathcal{F}_1)$.

Let Y be a random element from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B})$. $\mathbb{P}^*(B|X = x)$ is a regular conditional distribution for Y given $X = x$ if

(1)' For $B \in \mathcal{B}$, $\mathbb{P}^*(B|X = x) = \mathbb{P}(Y \in B|X = x)$ $\mathbb{P}_X$-a.s.;
(2)' For $x \in S$, $\mathbb{P}^*(\cdot|X = x)$ is a probability on $(R, \mathcal{B})$.

Using a similar argument, Proposition 1.9, 1.10 and Theorem 1.12 hold in the above setting.

Proposition 1.14. *Consider a measurable space $(\Omega, \mathcal{F})$ and let X be a random element taking values on $(S, \mathcal{E})$.*

(1) *If $\mathbb{P}(A|X = x)$ is a regular conditional probability on $\mathcal{F}_1 \subset \mathcal{F}$ given $X = x$ and Y is a random variable on $(\Omega, \mathcal{F}_1)$ with $\mathbb{E}|Y| < \infty$, then*

$$\mathbb{E}(Y|X = x) = \int_\Omega Y(\omega)\mathbb{P}(d\omega|X = x) \quad \mathbb{P}_X\text{-}a.s.$$

(2) *If Y is a random element from $(\Omega, \mathcal{F})$ to $(R, \mathcal{B})$, φ is a random variable on $(R, \mathcal{B})$ with $\mathbb{E}|\varphi(Y)| < \infty$ and $\mathbb{P}^*(B|X = x)$ is a regular conditional distribution for Y given $X = x$, then*

$$\mathbb{E}(\varphi(Y)|X = x) = \int_R \varphi(y)\mathbb{P}^*(dy|X = x) \quad \mathbb{P}_X\text{-}a.s.$$

Theorem 1.15. *Let X, Y be random elements on $(\Omega, \mathcal{F})$ taking values respectively on $(S, \mathcal{E})$ and $(R, \mathcal{R})$. If $(R, \mathcal{B})$ is a Borel space, then there exists a regular conditional distribution for Y given $X = x$. In particular, if $(\Omega, \mathcal{F})$ is a Borel space, then, for any sub-σ-field $\mathcal{F}_1$ of $\mathcal{F}$, there exists a regular conditional probability on $\mathcal{F}_1$ given $X = x$.*

Exercise 1.8. Let X, Y be independent random elements taking values on $(R, \mathcal{B})$ and $(S, \mathcal{E})$ and φ be a random variable defined on $(R \times S, \mathcal{B} \otimes \mathcal{E})$. Prove that

(1) $\mathbb{P}(\varphi(x, Y) \in B)$ is a regular conditional distribution for $\varphi(X, Y)$ given $X = x$.

(2) If $\mathbb{E}|\varphi(X, Y)| < \infty$, then $\mathbb{E}(\varphi(X, Y)|X = x) = \mathbb{E}\varphi(x, Y)$ a.s..

Exercise 1.9. Let I be an interval, $\varphi : I \to \mathbb{R}$ be convex and Y is a random variable taking values on I. Suppose $\mathbb{E}|Y| < \infty$ and $\mathbb{E}|\varphi(Y)| < \infty$. Show that

(1) $\varphi(\mathbb{E}Y) \leq \mathbb{E}\varphi(Y)$.

(2) $\varphi(\mathbb{E}(Y|\mathcal{C})) \leq \mathbb{E}(\varphi(Y)|\mathcal{C})$ a.s..