Conditional probability and conditional expectation

Cunyi Nan

Abstract

This report introduces conditional probability and conditional expectation in probability theory which are especially necessary to develop martingale theory. Moreover, we can treat it as an operator between $L^1(\Omega, \mathcal{F}, P)$ and $L^1(\Omega, \mathcal{G}, P)$ where $\mathcal{G}$ is σ-subalgebra of $\mathcal{F}$.

1 Retrospect

We will apply Radon-Nikodým theorem known in measure theory to define conditional probability and conditional expectation. **Signed measure** is defined as measure which allowed to take finite values, that is, it cannot take value $\{+\infty, -\infty\}$.

Definition 1.1. (1) Let $(S, \mathcal{S})$ be a measure space, a set function $\nu : \mathcal{S} \to \mathbb{R}$ is called **finite signed measure** if it is σ-additive signed measure: $\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \nu(A_n)$ for disjoint $A_n \in \mathcal{S}$.

(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 all $A \in \mathcal{S}$ such that $\mu(A) = 0$.

The next Radon-Nikodým theorem shows that a finite signed measure which is absolutely continuous with respect to a finite measure can be represented as an integral, its proof can be found in any measure theory textbook.

Theorem 1.2 (Radon-Nikodým)**.** *Let μ be a finite measure on $(S, \mathcal{S})$ and 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

$$\nu(A) = \int_A f(x)\mu(\mathrm{d}x), \ A \in \mathcal{S}$$

(2) (Uniqueness) If another $\tilde{f}$ exists and represents ν as above, then $f(x) = \tilde{f}(x)$, μ-a.s..

2 Conditional probability and conditional expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space. $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$. X is a real-valued integrable r.v., that is, $E[|X|] < \infty$. For $A \in \mathcal{F}$, denote the integral of X on A by

$$E[X, A] := \int_A X \mathrm{d}P = \int_A X(\omega) P(\mathrm{d}\omega)$$

Definition 2.1. If Y satisfies the following two conditions:

(1) Y is $\mathcal{G}$-measurable real-valued r.v.

(2) For $A \in \mathcal{G}$, $E[X, A] = E[Y, A]$.

Then Y is the **conditional expectation** of X under $\mathcal{G}$, denoted by $E[X|\mathcal{G}]$. For $A \in \mathcal{F}$, $P(A|\mathcal{G}) = E[\mathbf{1}_A|\mathcal{G}]$ is called the **conditional probability** of A under $\mathcal{G}$.

Let us 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 $B \in \mathcal{F}$ that $P(B) > 0$, the conditional probability of A is

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

once we know B happens, the probability that A happens is given by it. The corresponding conditional expectation, that is, expectation under measure $P(\cdot|B)$ is defined by

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

When $X = \mathbf{1}_A$, it coincides with $P(A|B)$. For general r.v. X, we can 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 when $B_i \in \mathcal{F}$ are disjoint, $0 < P(B_i) < 1$ and $\Omega = \bigcup_{i=1}^n B_i$. We define the conditional probability of A under the division $\{B_i\}_{i=1}^n$ by

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

and the corresponding conditional expectation is defined by

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

After showing the easy case, we summarize properties of $E[X|\mathcal{G}]$.

Proposition 2.2. *Let X, Y be two real-valued integrable random variables in probability space $(\Omega, \mathcal{F}, P)$, taking values in measure space $(\mathbb{R}, \sigma_B)$. $\mathcal{G}$ is σ-subalgebra of $\mathcal{F}$.*

(1) For $\forall\ a, b \in \mathbb{R}$, $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$.

(2) If $X \geq 0$ a.s., then $E[X|\mathcal{G}] \geq 0$ a.s.

(3) If Y is bounded and $\mathcal{G}$-measurable univariate random variable, then $E[YX|\mathcal{G}] = YE[X|\mathcal{G}]$ a.s.

(4) (Tower property) Let $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ and $\mathcal{H}$ is also σ-algebra, then $E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$ a.s.

(5) If X is independent of $\mathcal{G}$, then $E[X|\mathcal{G}] = E[X]$ a.s.

Proof. The proof of (1) can be immediately finished according to linearity of expectation with respect to $\mathcal{F}$ and $\mathcal{G}$ is σ-subalgebra of $\mathcal{F}$.

(2) Look back to Definition 2.1, for every $A \in \mathcal{G}$, $E[E[X|\mathcal{G}], A] = E[X, A] = \int_A X \mathrm{d}P \geq 0$ since $X \geq 0$ a.s.. It means $E[X|\mathcal{G}] \geq 0$ a.s..

(3) It is immediate that the random variable $YE[X|\mathcal{G}]$ is $\mathcal{G}$-measurable. We only needs to show that for all $A \in \mathcal{G}$, the equation

$$\int_A YE[X|\mathcal{G}]\mathrm{d}P = \int_A YX\mathrm{d}P = \int_A E[YX|\mathcal{G}]\mathrm{d}P$$

holds. Starting by assuming $Y = \mathbf{1}_B, B \in \mathcal{G}$, then $A \cap B \in \mathcal{G}$ and

$$\int_A YE[X|\mathcal{G}]\mathrm{d}P = \int_{A\cap B} E[X|\mathcal{G}]\mathrm{d}P = \int_{A\cap B} X\mathrm{d}P = \int_A YX\mathrm{d}P$$

Since the equation holds for $Y = \mathbf{1}_B, B \in \mathcal{G}$, hence it also holds for $Y = \sum_{k=1}^n b_k \mathbf{1}_{B_k}$ where $b_k \in \mathbb{R}, B_k \in \mathcal{G}$. Now suppose X, Y are non-negative, then there exists a simple function sequence $\{Y_n\}$ increasing to Y. Then we have

$$\int_A Y_n E[X|\mathcal{G}]\mathrm{d}P = \int_A Y_n X\mathrm{d}P \tag{2.0.1}$$

By monotone convergence theorem, we have $\lim_{n\to\infty} Y_n X = YX$, then

$$\lim_{n\to\infty} Y_n E[X|\mathcal{G}] = YE[X|\mathcal{G}]$$

Let $n \to \infty$ in equation 2.0.1, using monotone convergence theorem yields

$$\int_A YE[X|\mathcal{G}]\mathrm{d}P = \int_A YX\mathrm{d}P$$

for every $A \in \mathcal{G}$. When X, Y are general univariate random variables, letting $X = X^+ - X^-$ and $Y = Y^+ - Y^-$ would complete the proof.

(4) To show the tower property, since $E[X|\mathcal{H}]$ is $\mathcal{H}$-measurable, it is enough to show

$$E[E[X|\mathcal{G}], B] = E[E[X|\mathcal{H}], B]$$

for every $B \in \mathcal{H}$. It is easy to see both sides are equal to $E[X, B]$ since $\mathcal{G}, \mathcal{H}$ are both σ-subalgebras of $\mathcal{F}$.

(5) When X is independent of $\mathcal{G}$, we have

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

for every $B \in \mathcal{G}$, so $E[X|\mathcal{G}] = E[X]$ a.s.. □

Proposition 2.3 (Jensen's inequality)**.** *X is random variable on $(\Omega, \mathcal{F}, P)$ taking values in $(S, \mathcal{S})$ and $\mathcal{G}$ is σ-subalgebra of $\mathcal{F}$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex lower semi-continuous function, then*

$$E[\varphi(X)|\mathcal{G}] \geq \varphi(E[X|\mathcal{G}])$$

Proof. For $\forall\ a \in \mathbb{R}$, there exists $c = c(a) \in \mathbb{R}$ such that $\varphi(x) \geq \varphi(a) + c(a)(x-a)$ due to property of convex function. Take $x = X$ and $a = E[X|\mathcal{G}]$, then take conditional expectation $E[\cdot|\mathcal{G}]$ on both sides, we could find this inequality. □

Corollary 2.4. *In particular, if $X \in L^p(\Omega, \mathcal{F}, P)$ for $p \geq 1$, then*

$$E[|X|^p|\mathcal{G}] \geq |E[X|\mathcal{G}]|^p \ a.s.$$

Proof. Note the function $\varphi(x) = |x|^p$ when $p \geq 1$ is convex, we can directly apply Proposition 2.3. □

After defining conditional probability and conditional expectation, we will move to a deeper level of it and can restate properties of conditional expectation with treating it as operator.

3 Conditional expectation as operator

This section introduces the second definition of conditional expectation. On probability space $(\Omega, \mathcal{F}, P)$, suppose $\mathcal{G}$ is σ-subalgebra of $\mathcal{F}$, for $p \geq 1$, define conditional expectation operator

$$E[\cdot|\mathcal{G}] : L^p(\Omega, \mathcal{F}, P) \to L^p(\Omega, \mathcal{G}, P)$$

For $X_1, X_2 \in L^p(\Omega, \mathcal{F}, P)$ and $a, b \in \mathbb{R}$, for every $A \in \mathcal{G}$,

$$\begin{aligned}
\int_A E[aX_1 + bX_2|\mathcal{G}]\mathrm{d}P &= \int_A (aX_1 + bX_2)\mathrm{d}P \\
&= a\int_A X_1\mathrm{d}P + b\int_A X_2\mathrm{d}P \\
&= a\int_A E[X_1|\mathcal{G}]\mathrm{d}P + b\int_A E[X_2|\mathcal{G}]\mathrm{d}P \\
&= \int_A aE[X_1|\mathcal{G}]\mathrm{d}P + \int_A bE[X_2|\mathcal{G}]\mathrm{d}P
\end{aligned}$$

These equations are well-defined since $E[X_1|\mathcal{G}]$ and $E[X_2|\mathcal{G}]$ are $\mathcal{G}$-measurable. We have shown that this operator is linear. Next, define norm of this linear operator as

$$\|E[\cdot|\mathcal{G}]\| := \sup_{X \in L^p(\Omega,\mathcal{F},P)} \frac{\|E[X|\mathcal{G}]\|_p}{\|X\|_p}$$

By Corollary 2.4, $|E[X|\mathcal{G}]|^p \leq E[|X|^p|\mathcal{G}]$ a.s., taking expectation on both sides to get

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

By property of conditional expectation, the right one is indeed $E[|X|^p]$. Hence the inequality

$$E[|E[X|\mathcal{G}]|^p] \leq E[|X|^p]$$

holds. For every $X \in L^p(\Omega, \mathcal{F}, P)$,

$$\frac{\|E[X|\mathcal{G}]\|_p}{\|X\|_p} = \left(\frac{E[|E[X|\mathcal{G}]|]^p}{E[|X|^p]}\right)^{\frac{1}{p}} \leq 1$$

Since its supreme is not greater than 1, that means the norm of conditional expectation operator $E[\cdot|\mathcal{G}] : L^p(\Omega, \mathcal{F}, P) \to L^p(\Omega, \mathcal{G}, P)$ is not greater than 1, that is, it is bounded linear operator.

We can reintroduce Proposition 2.2 with conditional expectation operator from $L^1(\Omega, \mathcal{F}, P)$ to $L^1(\Omega, \mathcal{G}, P)$.

Proposition 3.1. *$(\Omega, \mathcal{F}, P)$ is probability space, $\mathcal{G}$ is σ-subalgebra of $\mathcal{F}$. Then there exists conditional expectation operator*

$$E[\cdot|\mathcal{G}] : L^1(\Omega, \mathcal{F}, P) \to L^1(\Omega, \mathcal{G}, P)$$

satisfying the following properties:

(1) For every random variable $X \in L^1(\Omega, \mathcal{F}, P)$, $E[X|\mathcal{G}]$ can be described by the following two properties in almost surely meaning.

(i) $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable.

(ii) For $\forall\ A \in \mathcal{G}$, $\int_A E[X|\mathcal{G}]\mathrm{d}P = \int_A X\mathrm{d}P$.

(2) $E[\cdot|\mathcal{G}]$ is linear operator with norm 1. And $E[\cdot|\mathcal{G}]$ is positive operator, that is, for every $X \geq 0$ a.s., then $E[X|\mathcal{G}] \geq 0$ a.s.

(3) For every $X \in L^1(\Omega, \mathcal{F}, P)$ and $Y \in L^\infty(\Omega, \mathcal{G}, P)$, $E[YX|\mathcal{G}] = YE[X|\mathcal{G}]$ a.s.

(4) (Tower property) If $\mathcal{H} \subset \mathcal{G}$ is also σ-subalgebra, then $E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$ a.s.

(5) For every $X \in L^1(\Omega, \mathcal{F}, P)$, $|E[X|\mathcal{G}]| \leq E[|X||\mathcal{G}]$ a.s.

Proof. (1) For $X \in L^1(\Omega, \mathcal{F}, P)$, suppose $X \geq 0$ and $\int_\Omega X\mathrm{d}P > 0$. Define

$$\mu_X : \mathcal{F} \to [0,1],\ B \mapsto \frac{\int_B X\mathrm{d}P}{\int_\Omega X\mathrm{d}P}$$

then μ_X is probability measure. Note that $\mu_X|_\mathcal{G} \ll P|_\mathcal{G}$ where $P|_\mathcal{G}$ is restriction of probability measure P on $\mathcal{G}$, then by Radon-Nikodým theorem, there exists a unique $Y \in L^1(\Omega, \mathcal{G}, P)$ such that for every $A \in \mathcal{G}$,

$$\int_A Y\mathrm{d}P = \mu_X(A) = \frac{\int_B X\mathrm{d}P}{\int_\Omega X\mathrm{d}P}$$

holds. Let $E[X|\mathcal{G}] = Y\int_\Omega X\mathrm{d}P$. When $X = 0$, then define $E[X|\mathcal{G}] = 0$. Generally, if $X = X^+ - X^-$, then let $E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}]$. The existence is proved. To prove the uniqueness, suppose there are two random variables Y_1, Y_2 satisfying (i) and (ii), let $A = \{\omega \in \Omega : Y_1(\omega) < Y_2(\omega)\}$, clearly $A \in \mathcal{G}$, from (ii),

$$\int_A Y_1\mathrm{d}P = \int_A X\mathrm{d}P = \int_A Y_2\mathrm{d}P$$

hence $P(A) = 0$. Similarly we can prove $P(\{\omega \in \Omega : Y_1(\omega) > Y_2(\omega)\}) = 0$, hence $Y_1 = Y_2$ a.s.

(2) From (i) and (ii) in (1), one can see $E[\cdot|\mathcal{G}]$ is bounded linear operator and its norm is 1. Let $X \in L^1(\Omega, \mathcal{F}, P)$ and $X \geq 0$, set $A = \{\omega \in \Omega : E[X|\mathcal{G}] < 0\}$, then $A \in \mathcal{G}$ and

$$0 \leq \int_A X\mathrm{d}P = \int_A E[X|\mathcal{G}]\mathrm{d}P$$

meaning $P(A) = 0$, that is $E[X|\mathcal{G}] \geq 0$ a.s.

(3) We can know $YE[X|\mathcal{G}] \in L^1(\Omega, \mathcal{G}, P)$. If there exists $B \in \mathcal{G}$ such that $Y = \mathbf{1}_B$, then for every $A \in \mathcal{G}$,

$$\int_A YE[X|\mathcal{G}]\mathrm{d}P = \int_{A\cap B} E[X|\mathcal{G}]\mathrm{d}P = \int_{A\cap B} X\mathrm{d}P = \int_A YX\mathrm{d}P$$

from (1), $E[YX|\mathcal{G}] = YE[X|\mathcal{G}]$. Since $E[\cdot|\mathcal{G}]$ is linear, it is correct when Y is simple function. For general case $Y \in L^\infty(\Omega, \mathcal{G}, P)$, then there exsits uniformly bounded simple functions $\{Y_n\}$ on $(\Omega, \mathcal{G})$ that converges to Y almost surely. We soon get the result from dominated convergence theorem.

(4) For every $A \in \mathcal{H} \subset \mathcal{G}$, we have

$$\int_A E[E[X|\mathcal{G}]|\mathcal{H}]\mathrm{d}P = \int_A E[X|\mathcal{G}]\mathrm{d}P = \int_A X\mathrm{d}P = \int_A E[X|\mathcal{H}]\mathrm{d}P$$

From (1), $E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$.

(5) For $X \in L^1(\Omega, \mathcal{F}, P)$, then there exists $Y \in L^\infty(\Omega, \mathcal{G}, P)$ such that $|Y(\omega)| = 1, \omega \in \Omega$ and $|E[X|\mathcal{G}]| = YE[X|\mathcal{G}]$. From (3), $|E[X|\mathcal{G}]| = E[YX|\mathcal{G}]$. For every $A \in \mathcal{G}$,

$$\begin{aligned}\int_A |E[X|\mathcal{G}]|\mathrm{d}P &= \int_A E[YX|\mathcal{G}]\mathrm{d}P = \int_A YX\mathrm{d}P \\ &\leq \int_A |YX|\mathrm{d}P = \int_A |X|\mathrm{d}P \\ &= \int_A E[|X||\mathcal{G}]\mathrm{d}P\end{aligned}$$

hence $|E[X|\mathcal{G}]| \leq E[|X||\mathcal{G}]$. □

4 Freezing lemma

The useful lemma can be used to easily compute conditional expectation.

Lemma 4.1. *Let $(\Omega, \mathcal{F}, P)$ be probability space and let $\mathcal{G}_1, \mathcal{G}_2$ be two independent σ-subalgebras of $\mathcal{F}$. For $i = 1, 2$, let X_i be $\mathcal{G}_1$-measurable random variable from $(\Omega, \mathcal{F}, P)$ to measure space $(S_i, \mathcal{S}_i)$ and belonging to $L^1(\Omega, \mathcal{G}_i, P)$. Let $\Phi : S_1 \times S_2 \to \mathbb{R}$ be a measurable function when $S_1 \times S_2$ is endowed with σ-algebra $\mathcal{S}_1 \times \mathcal{S}_2$. Then*

$$E[\Phi(X_1, X_2)|\,\mathcal{G}_1] = (E[\Phi(\cdot, X_2)])(X_1) = E[\Phi(X_1, X_2)|X_1]$$

wherever the mapping from $\omega \mapsto \Phi(X_1(\omega), X_2(\omega))$ is absolutely integrable.

Proof. To take it simple we can assume $\Phi(x_1, x_2) = \varphi(x_1)\phi(x_2)$. Then

$$E[\varphi(X_1)\phi(X_2)|\,\mathcal{G}_1] = \varphi(X_1)E[\phi(X_2)|\,\mathcal{G}_1] = \varphi(X_1)E[\phi(X_2)]$$

We can find this is the value of $E[\varphi(x_1)\phi(X_2)]$ at $x_1 = X_1$ which can be described as $(E[\Phi(\cdot, X_2)])(X_1)$. Fix a set $C \in \mathcal{G}_1$, pick $\varphi(x_1) = \mathbf{1}_A(x_1)$ and $\phi(x_2) = \mathbf{1}_B(x_2)$ where $A \in \mathcal{S}_1, B \in \mathcal{S}_2$. Then

$$\int_C \mathbf{1}_{A\times B}(X_1, X_2)\mathrm{d}P = \int_C E[\mathbf{1}_{A\times B}(x_1, X_2)]|_{x_1=X_1}\mathrm{d}P$$

The right hand defines measure on product σ-algebra $\mathcal{S}_1 \times \mathcal{S}_2$. By linearity it becomes

$$\int_C \Phi(X_1, X_2)\mathrm{d}P = \int_C E[\Phi(x_1, X_2)]|_{x_1=X_1}\mathrm{d}P = (E[\Phi(\cdot, X_2)])(X_1)$$

on set $C \in \mathcal{G}_1$ for simple functions. By monotone convergence theorem and bounded convergence theorem we can get the result. □