7. Independence. The conditional expectation

In Linear Theory, the orthogonal property and the conditional expectation in the wide sense play a key role. Out of the framework of Linear Theory, a significant role plays the independence concept and conditional expectation.

7.1. **Independence concept.** Axiomatically, two random sets A and B for which probabilities $P(A)$ and $P(B)$ are well defined, are called independent, if

$$P(A \cap B) = P(A)P(B). \tag{7.1}$$

Following this axiom, random variables X, Y is said to be independent, if their joint distribution function $F(x,y) = P(X \le x, Y \le y)$ is expressed by the marginal distribution functions: $F(x,\infty), F(\infty,y)$ [1]

$$F(x,y) = F(x,\infty)F(\infty,y)), \ \forall\ x,y. \tag{7.2}$$

Particularly, if $F(x,y)$ obeys densities

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y},$$

the marginal distribution functions obey density as well, say $\phi(x), \psi(y)$, and (7.2) is transformed to

$$f(x,y) = \phi(x)\psi(y), \ \forall\ x,y. \tag{7.3}$$

An equivalent form of (7.2), (7.3) is the follows. For any bounded continuous (or piece wise constant functions) $h(x)$, $g(y)$

$$\boldsymbol{E}h(X)g(Y) = \boldsymbol{E}h(X)\boldsymbol{E}g(Y). \tag{7.4}$$

Taking $h(x) = e^{\imath\lambda x}$, $g(y) = e^{\imath\mu y}$, where $\imath = \sqrt{-1}$, $\lambda, \mu \in \mathbb{R}$, notice that (7.4) is equivalent to

$$\boldsymbol{E}e^{\imath(\lambda X + \mu Y)} = \boldsymbol{E}e^{\imath\lambda X}\boldsymbol{E}e^{\imath\mu Y}, \ \forall\ \lambda, \mu \in \mathbb{R}. \tag{7.5}$$

Since $\boldsymbol{E}e^{\imath(\lambda X + \mu Y)}$ and $\boldsymbol{E}e^{\imath\lambda X}$, $\boldsymbol{E}e^{\imath\mu Y}$ are the characteristic functions for (X,Y) and X, Y, respectively, the latter definition of the independence in words is: *random variables are independent if their joint characteristic function is equal to the product of marginal characteristic functions.*

The equivalence of all these definitions is verified by direct calculations. For example,

$$\begin{aligned}
\boldsymbol{E}h(X)g(Y) &= \int_{R^2} h(x)g(y)dF(x,y) \\
&= \int_R h(x)dF(x,\infty)\int_R g(y)dF(\infty,y) \\
&= \boldsymbol{E}h(X)\boldsymbol{E}g(Y).
\end{aligned}$$

[1] $F(x,\infty) := P(X \le x, Y \in \mathbb{R})$, $F(x,\infty) := P(X \in \mathbb{R}, Y \le y)$.

7.2. **Conditional expectation.** If (7.1) is lost, that is $P(A \cap B) \neq P(A)P(B)$, we replace (7.1) by the next equality

$$P(A \cap B) = P(A|B)P(B), \tag{7.6}$$

where, under $P(B) > 0$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{7.7}$$

and is called *conditional probability of A given B* and (7.7) is called *Bayes formula*. Notice that (7.7) solves the linear algebraic equation (7.6), which is solvable whatever $P(B)$ is ($P(B) = 0$ provides $P(A \cap B) = 0$). So, under $P(B) = 0$, any $P(A|B) \in [0, 1]$ can be taken.

Assume random variables X and Y take values in finite alphabets set spaces $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ respectively, that is [2]

$$X = \sum_{i=1}^{n} x_i I(X = x_i) \quad \text{and} \quad Y = \sum_{j=1}^{m} y_j I(Y = y_j).$$

Set $A = \{X = a_i\}$ and $B = \{Y = Y_j\}$. Then, by (7.7),

$$P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}. \tag{7.8}$$

By analogy with the expectation formula for X: $\boldsymbol{E}X = \sum\limits_{i=1}^{n} x_i P(X = x_i)$, we define the conditional expectation for X given $\{Y = y_j\}$:

$$\boldsymbol{E}(X|Y = y_j) = \sum_{i=1}^{n} x_i P(X = x_i|Y = y_j). \tag{7.9}$$

Definition 7.1. *The random variable*

$$\boldsymbol{E}(X|Y) = \sum_{j=1}^{m} \boldsymbol{E}(X|Y = y_j) I(Y = y_j) \tag{7.10}$$

is called the conditional expectation of X given Y. Obviously

$$\begin{aligned}\boldsymbol{E}(X|Y) &= \sum_{j=1}^{m} \boldsymbol{E}(X|Y = y_j) I(Y = y_j) \\ &= \sum_{j=1}^{m}\sum_{i=1}^{n} x_i \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} I(Y = y_j) := G(Y), \end{aligned} \tag{7.11}$$

where $G(z) = \sum_{j=1}^{m}\sum_{i=1}^{n} x_i \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)} I(z = y_j)$.

[2]Recall that $I(x = a) = \begin{cases} 1 & \text{if}, x = a \\ 0 & \text{otherwise.} \end{cases}$

It is directly verified that the conditional expectation defined above possesses the property

$$\boldsymbol{E}\big(X-\boldsymbol{E}(X|Y)\big)I(Y=y_j)=0,\ j=1,2,\dots,m, \tag{7.12}$$

or, in words: the difference $X-\boldsymbol{E}(X|Y)$ is orthogonal to any $I(Y=y_j)$ and so is orthogonal to any linear combination $\sum_{j=1}^m c_j I(Y=y_j)$ and, thus equivalently, is orthogonal to any bounded function $g(Y)$.

For arbitrary random variables X,Y the conditional expectation $\boldsymbol{E}(X|Y)$ is well defined, if

$$\boldsymbol{E}|X|<\infty, \tag{7.13}$$

with a help of properties, established for the above-mentioned simple case, that is according to

(i) $\boldsymbol{E}(X|Y)=G(Y)$

(ii) $\boldsymbol{E}\big(X-G(Y)\big)g(Y)=0$ (for every bounded function g).

The existence of $\boldsymbol{E}(X|Y)$ is readily verified when joint distribution function $F(x,y)$ obeys density $f(x,y)$. Indeed, under $g(y)=I(y\le z)$, **(ii)** is transformed into

$$\int_{-\infty}^{z}\int_{\mathbb{R}} xf(x,y)dxdy=\int_{-\infty}^{z} G(y)\int_{\mathbb{R}} f(x,y)dxdy,\ z\in\mathbb{R},$$

that is (with agreement $0/0=0$) we get (well known Bayes formula)

$$G(z)=\frac{\int_{\mathbb{R}} xf(x,z)dx}{\int_{\mathbb{R}} f(x,z)dx}. \tag{7.14}$$

If $F(x,y)$ even does not obey density, nevertheless

$$\int_{-\infty}^{z}\int_{R} xdF(x,y)=\int_{-\infty}^{z} G(y)\int_{R} dF(x,y)$$

provides (another type of the Bayes formula)

$$G(z)=\frac{d\int_{-\infty}^{z}\int_{R} xdF(x,y)}{dF(\infty,z)}(z). \tag{7.15}$$

The uniqueness of $\boldsymbol{E}(X|Y)$ ia argued as follows: if G_1, G_2 solve **(ii)**, then G_1-G_2 solves **(ii)** as well, that is $\boldsymbol{E}\big(G_1(Y)-G_2(Y)\big)g(Y)=0$, so that for $g(Y)=\mathsf{sign}\ \big(G_1(Y)-G_2(Y)\big)$ we have $\boldsymbol{E}|G_1(Y)-G_2(Y)|=0$.

Remark 7.2. For random vectors X,Y the conditional expectation is defined similarly.

7.2.1. Main properties of the conditional expectation.

1. $X\equiv c \implies \boldsymbol{E}(X|Y)=c$.

Proof. Take $g(Y)=\text{sign }(c-G(Y))$. □

2. $c_1X_1+c_2X_2 \implies c_1\boldsymbol{E}(X_1|Y)+c_2\boldsymbol{E}(X_2|Y)$.

Proof. It follows from

$$\begin{aligned}\boldsymbol{E}g(Y)\boldsymbol{E}(c_1X_1+c_2X_2|Y) = \boldsymbol{E}g(Y)[c_1X_1+c_2X_2]\\ = c_1\boldsymbol{E}g(Y)X_1+c_2\boldsymbol{E}g(Y)X_2\\ = c_1\boldsymbol{E}g(Y)\boldsymbol{E}(X_1|Y)+c_2\boldsymbol{E}g(Y)\boldsymbol{E}(X_2|Y).\end{aligned}$$

□

3. $X \geq 0 \implies \boldsymbol{E}(X|Y) \geq 0$.

Proof. For $g(Y) = I(G(Y) < 0)$, we get a contradiction

$$0 \leq \boldsymbol{E}Xg(Y) = \boldsymbol{E}G(Y)I(G(Y) < 0) < 0,$$

that is $P(G(Y) < 0) = 0$. □

4. $\big|\boldsymbol{E}(X|Y)\big| \leq \boldsymbol{E}(|X||Y)$.

Proof. For $a^+ = \max[a, 0]$ and $a^- = -\min[a, 0]$, the use of *2.* provides

$$\begin{aligned}\big|\boldsymbol{E}(X|Y)\big| &= \big|\boldsymbol{E}(X^+ - X^-|Y)\big|\\ &= \big|\boldsymbol{E}(X^+|Y) - \boldsymbol{E}(X^-|Y)\big|\\ &\leq \boldsymbol{E}(X^+|Y) + \boldsymbol{E}(X^-|Y)\\ &= \boldsymbol{E}(X^+ + X^-|Y)\\ &= \boldsymbol{E}\Big(|X|\big|Y\Big).\end{aligned}$$

□

5. $Y \equiv c \implies \boldsymbol{E}(X|Y) = \boldsymbol{E}X$.

Proof. Since $Y \equiv c$, also $h(Y) = h(c)$ and so

$$h(c)\boldsymbol{E}X = \boldsymbol{E}h(Y)X = \boldsymbol{E}h(Y)G(Y) = \boldsymbol{E}h(c)G(Y) = h(c)\boldsymbol{E}G(Y).$$

□

6. $X = h(Y) \implies \boldsymbol{E}(X|Y) = h(Y)$.

Proof. For $g(Y) = \mathsf{sign}\big(h(Y) - G(Y)\big)$, we have

$$0 = \boldsymbol{E}g(Y)\big(h(Y) - G(Y)\big) = \boldsymbol{E}\big|h(Y) - G(Y)\big|.$$

□

7. $\boldsymbol{E}\boldsymbol{E}(X|Y) = \boldsymbol{E}X$.

Proof. The result follows with $g(Y) \equiv 1$. □

8. $\boldsymbol{E}\Big(\boldsymbol{E}(X|Y_1, ...Y_n, Y_{n+1}, ...)|Y_1, ..., Y_n\Big) = \boldsymbol{E}(X|Y_1, ..., Y_n),\ \forall n \geq 1.$

Proof. Denote

$$\begin{aligned} G(Y_1, \ldots, Y_n, Y_{n+1}, ...) &= \boldsymbol{E}(X|Y_1, ..., Y_n, Y_{n+1}, ...) \\ \overline{G}(Y_1, \ldots, Y_n) &= \boldsymbol{E}(X|Y_1, ..., Y_n) \end{aligned}$$

and show that

$$\boldsymbol{E}\Big(G(Y_1, \ldots, Y_n, Y_{n+1}, ...)\Big|Y_1, ..., Y_n\Big) = \overline{G}(Y_1, \ldots, Y_n). \tag{7.16}$$

Taking into the consideration that $g(Y_1, \ldots, Y_n)$ is also a version of the function of arguments $Y_1, ..., Y_n, Y_{n+1}, ...$, write

$$\begin{aligned} \boldsymbol{E}\overline{G}(Y_1, \ldots, Y_n)g(Y_1, \ldots, Y_n) &= \boldsymbol{E}Xg(Y_1, \ldots, Y_n) \\ &= \boldsymbol{E}G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots)g(Y_1, \ldots, Y_n) \\ &= \boldsymbol{E}\boldsymbol{E}\Big(G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots)\Big|Y_1, \ldots, Y_n\Big)g(Y_1, \ldots, Y_n). \end{aligned}$$

It suffices now to choose

$$\begin{aligned} & g(Y_1, \ldots, Y_n) \\ & \quad = \mathsf{sign}\Big[\boldsymbol{E}\Big(G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots)\Big|Y_1, \ldots, Y_n\Big) - \overline{G}(Y_1, \ldots, Y_n)\Big]. \end{aligned}$$

to get the result. □

9. independence of $(X, Y) \Longrightarrow \boldsymbol{E}(X|Y) = \boldsymbol{E}X$.

Proof. The independence of X and Y implies

$$\boldsymbol{E}\big(\boldsymbol{E}X - \boldsymbol{E}(X|Y)\big)g(Y) = 0,$$

and, with $g(Y) = \mathsf{sign}\big(\boldsymbol{E}X - G(Y)\big)$, the result. □

10. $\boldsymbol{E}|h(Y)X| < \infty \Longrightarrow \boldsymbol{E}(h(Y)X|Y) = h(Y)\boldsymbol{E}(X|Y)$.

Proof. Write

$$\begin{aligned} & \boldsymbol{E}\big(\boldsymbol{E}(h(Y)X|Y) - h(Y)\boldsymbol{E}(X|Y)g(Y) \\ & \qquad = \boldsymbol{E}h(Y)Xg(Y) - \boldsymbol{E}h(Y)Xg(Y) = 0. \end{aligned}$$

Let $G(z), G^h(z)$ be such that $G(Y) = \boldsymbol{E}(X|Y), G^h(Y) = \boldsymbol{E}(h(Y)X|Y)$. Then, with $g(Y) = \mathsf{sign}\big([G^h(z) - G(z)\big)$, we get

$$\boldsymbol{E}|\boldsymbol{E}(h(Y)X|Y) - h(Y)\boldsymbol{E}(X|Y)| = 0$$

and the result.

□

11. $\boldsymbol{E}X^2 < \infty \Longrightarrow \big(\boldsymbol{E}(X|Y)\big)^2 \le E(X^2|Y)$ (Cauchy-Schwartz inequality).

Proof. By virtue of *3.* $\boldsymbol{E}\Big(\big[X - \boldsymbol{E}(X|Y)\big]^2\big|Y\Big) \geq 0$. By *6.*,

$$\begin{aligned}\boldsymbol{E}\Big(\big[X - \boldsymbol{E}(X|Y)\big]^2\big|Y\Big) \\ &= \boldsymbol{E}(X^2|Y) - 2\boldsymbol{E}\big(X\boldsymbol{E}(X|Y)|Y\big) + \big(\boldsymbol{E}(X|Y)\big)^2 \\ &= \boldsymbol{E}(X^2|Y) - \big(\boldsymbol{E}(X|Y)\big)^2\end{aligned}$$

and the result holds true. □

12. $\boldsymbol{E}X^2 < \infty \implies \boldsymbol{E}\big(X - \boldsymbol{E}(X|Y)\big)^2 \leq \boldsymbol{E}\big(X - h(Y)\big)^2$ for any $h(Y)$ with $\boldsymbol{E}h^2(Y) < \infty$.

Proof. Set $\delta(Y) = h(Y) - \boldsymbol{E}(X|Y)$. Then

$$\begin{aligned}\boldsymbol{E}(X - h(Y))^2 &= \boldsymbol{E}(\alpha - \boldsymbol{E}(X|Y) - \delta(Y))^2 \\ &= \boldsymbol{E}(X - \boldsymbol{E}(X|Y))^2 + \boldsymbol{E}\delta^2(Y) - 2\boldsymbol{E}[X - \boldsymbol{E}(X|Y)]\delta(Y) \\ &\geq \boldsymbol{E}(X - \boldsymbol{E}(X|Y))^2.\end{aligned}$$

□

13. $h(x)$ is convex, $\boldsymbol{E}|h(X)| < \infty \implies \boldsymbol{E}(h(X)|Y) \geq h\big(\boldsymbol{E}(X|Y)\big)$ (Jensen inequality).

Proof. Since $h(z)$ is the convex function, for fixed x there exists a number u_x, depending on x, such that $h(z) - h(x) \geq u_x(z - x)$ (in words: the tangent line lies below). Set $z = X$ and $x = \boldsymbol{E}(X|Y)$. Then,

$$h(X) - h(\boldsymbol{E}(X|Y) \geq u_{\boldsymbol{E}_{(X|Y)}}(X - \boldsymbol{E}(X|Y)).$$

So, $\boldsymbol{E}(.|Y)$ taken from both sides of the above inequality provides the result by *3.*, *6.*, and *10.* □

14. Let $X_n \to X$ and $|X_n| \leq \gamma$, $\boldsymbol{E}\gamma < \infty$. Then $\lim_n \boldsymbol{E}(X_n|Y) = \boldsymbol{E}(X|Y)$, P – a.s.

Proof. Put $\gamma_n = \max_{m \geq n} |X_m - X|$ and note that $\gamma_n \leq 2\gamma$, $\gamma_{n+1} \leq \gamma_n$, and $\lim_n \gamma_n = 0$. Therefore $\lim_n \boldsymbol{E}(\gamma_n|Y) = 0$ and

$$\begin{aligned}\lim_n |\boldsymbol{E}(X|Y) - \boldsymbol{E}(X_n|Y)| &\leq \lim_n \boldsymbol{E}(\max_{n \geq m} |X - X_n||Y) \\ &= \lim_n \boldsymbol{E}(\gamma_n|Y) = 0.\end{aligned}$$

□

15. If X, Y are independent of Z, then

$$\boldsymbol{E}(X|Y, Z) = \boldsymbol{E}(X|Y).$$

Proof. Set $g(Y,Z) = g_1(Y)g_2(Z)$. Then

$$\begin{aligned}\boldsymbol{E}Xg(Y,Z) = \boldsymbol{E}Xg_1(Y)g_2(Z) &= \boldsymbol{E}Xg_1(Y)\boldsymbol{E}g_2(Z)\\ &= \boldsymbol{E}g_1(Y)\boldsymbol{E}(X|Y)\boldsymbol{E}g_2(Z)\\ &= \boldsymbol{E}g_1(Y)g_2(Z)\boldsymbol{E}(X|Y) = \boldsymbol{E}g(Y,Z)\boldsymbol{E}(X|Y).\end{aligned}$$

□

Examples:

1. For independent and identically distributed X, Y with $\boldsymbol{E}|X| < \infty$,

$$\boldsymbol{E}(X|X+Y) = \boldsymbol{E}(Y|X+Y) = \frac{X+Y}{2}.$$

By symmetry, $\boldsymbol{E}(X|X+Y) = \boldsymbol{E}(Y|X+Y)$ and obvious equality $\boldsymbol{E}(X+Y|X+Y) = X+Y$ the result holds true.

2. For independent and identically distributed $X_1, \ldots, X_n$, in the same way is proved that for any i

$$\boldsymbol{E}X_i|X_1 + \ldots + X_n) = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

3. Let $X_1, \ldots, X_n, \ldots$ be independent and identically distributed random variables with $\boldsymbol{E}|X_1| < \infty$ and let $_k = \sum_{j=1}^{k} X_j$. Then for any $i \le n$

$$\boldsymbol{E}(X_i|S_n, S_{n+1}, \ldots,) = \frac{S_n}{n}. \tag{7.17}$$

Proof. Notice that $\{S_n, S_{n+1}, \ldots,\} = \{S_n, X_{n+1}, \ldots,\}$. So, it suffices to show that $\boldsymbol{E}(X_i|S_n, X_{n+1}, \ldots,) = \frac{S_n}{n}$. That result follows from *15.*, since X_i, S_n are independent of $(X_{n+1}, X_{n+2}, \ldots,)$. □