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 \( F(x, y) = P(X \leq x, Y \leq y) \) is expressed by the marginal distribution functions: \( F(x, \infty), F(\infty, y) \)

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

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) \( h(x), g(y) \)

\[ Eh(X)g(Y) = Eh(X)Eg(Y). \tag{7.4} \]

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

\[ Ee^{i(\lambda X+\mu Y)} = Ee^{i\lambda X}Ee^{i\mu Y}, \ \forall \ \lambda, \mu \in \mathbb{R}. \tag{7.5} \]

Since \( Ee^{i(\lambda X+\mu Y)} \) and \( Ee^{i\lambda X}, Ee^{i\mu Y} \) 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,

\[ Eh(X)g(Y) = \int_{\mathbb{R}^2} h(x)g(y)dF(x, y) \]
\[ = \int_{\mathbb{R}} h(x)dF(x, \infty)\int_{\mathbb{R}} g(y)dF(\infty, y) \]
\[ = Eh(X)Eg(Y). \]

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

\[
P(A \cap B) = P(A|B)P(B),
\]

(7.6)

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

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}
\]

(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

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

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

\[
P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}.
\]

(7.8)

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

\[
E(X|Y = y_j) = \sum_{i=1}^{n} x_i P(X = x_i|Y = y_j).
\]

(7.9)

**Definition 7.1.** *The random variable*

\[
E(X|Y) = \sum_{j=1}^{m} E(X|Y = y_j) I(Y = y_j)
\]

(7.10)

*is called the conditional expectation of \( X \) given \( Y \). Obviously*

\[
E(X|Y) = \sum_{j=1}^{m} E(X|Y = y_j) I(Y = y_j)
\]

\[
= \sum_{j=1}^{m} \sum_{i=1}^{n} x_i \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} I(Y = y_j) := G(Y),
\]

(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
\[ \mathbb{E}(X - \mathbb{E}(X|Y))I(Y = y_j) = 0, \quad j = 1, 2, \ldots, m, \quad (7.12) \]
or, in words: the difference \( X - \mathbb{E}(X|Y) \) 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 \( \mathbb{E}(X|Y) \) is well defined, if
\[ \mathbb{E}|X| < \infty, \quad (7.13) \]
with a help of properties, established for the above-mentioned simple case, that is according to
(i) \( \mathbb{E}(X|Y) = G(Y) \)
(ii) \( \mathbb{E}(X - G(Y))g(Y) = 0 \) (for every bounded function \( g \)).

The existence of \( \mathbb{E}(X|Y) \) is readily verified when joint distribution function \( F(x,y) \) obeys density \( f(x,y) \). Indeed, under \( g(y) = I(y \leq z) \),
(i) is transformed into
\[
\int_{-\infty}^{z} \int_{\mathbb{R}} x f(x,y) \, dx \, dy = \int_{-\infty}^{z} G(y) \int_{\mathbb{R}} f(x,y) \, dx \, dy, \quad z \in \mathbb{R},
\]
that is (with agreement \( 0/0 = 0 \)) we get (well known Bayes formula)
\[ G(z) = \frac{\int_{\mathbb{R}} x f(x,z) \, dx}{\int_{\mathbb{R}} f(x,z) \, dx}. \quad (7.14) \]
If \( F(x,y) \) even does not obey density, nevertheless
\[
\int_{-\infty}^{z} \int_{\mathbb{R}} x dF(x,y) = \int_{-\infty}^{z} G(y) \int_{\mathbb{R}} dF(x,y)
\]
provides (another type of the Bayes formula)
\[ G(z) = \frac{d \int_{-\infty}^{z} \int_{\mathbb{R}} x dF(x,y)}{dF(\infty,z)}(z). \quad (7.15) \]

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

**Remark 7.2.** For random vectors \( X, Y \) the conditional expectation is defined similarly.

**7.2.1. Main properties of the conditional expectation.**
1. \( X \equiv c \implies \mathbb{E}(X|Y) = c \).

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

2. \( c_1 X_1 + c_2 X_2 \implies c_1 \mathbb{E}(X_1|Y) + c_2 \mathbb{E}(X_2|Y) \).
Proof. It follows from
\[ Eg(Y)E(c_1X_1 + c_2X_2|Y) = Eg(Y)[c_1X_1 + c_2X_2] \]
\[ = c_1Eg(Y)X_1 + c_2Eg(Y)X_2 \]
\[ = c_1Eg(Y)E(X_1|Y) + c_2Eg(Y)E(X_2|Y). \]

\[ \square \]

3. \( X \geq 0 \implies E(X|Y) \geq 0. \)

Proof. For \( g(Y) = I(G(Y) < 0) \), we get a contradiction
\[ 0 \leq EXg(Y) = EG(Y)I(G(Y) < 0) < 0, \]
that is \( P(G(Y) < 0) = 0. \)

\[ \square \]

4. \( |E(X|Y)| \leq E(|X|Y) \).

Proof. For \( a^+ = \max[a, 0] \) and \( a^- = -\min[a, 0] \), the use of 2. provides
\[ |E(X|Y)| = |E(X^+ - X^-|Y)| \]
\[ = |E(X^+|Y) - E(X^-|Y)| \]
\[ \leq E(X^+|Y) + E(X^-|Y) \]
\[ = E(X^+ + X^-|Y) \]
\[ = E(\|X||Y^). \]

\[ \square \]

5. \( Y \equiv c \implies E(X|Y) = EX. \)

Proof. Since \( Y \equiv c \), also \( h(Y) = h(c) \) and so
\[ h(c)EX = Eh(Y)X = Eh(Y)G(Y) = Eh(c)G(Y) = h(c)EG(Y). \]

\[ \square \]

6. \( X = h(Y) \implies E(X|Y) = h(Y) \).

Proof. For \( g(Y) = \text{sign}(h(Y) - G(Y)) \), we have
\[ 0 = Eg(Y)(h(Y) - G(Y)) = E|h(Y) - G(Y)|. \]

\[ \square \]

7. \( EE(X|Y) = EX. \)

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

\[ \square \]

8. \( E\left(E(X|Y_1,...Y_n,Y_{n+1},...)|Y_1,...,Y_n\right) = E(X|Y_1,...,Y_n), \forall n \geq 1. \)
Proof. Denote
\[ G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots) = \mathbb{E}(X|Y_1, \ldots, Y_n, Y_{n+1}, \ldots) \]
and show that
\[ \mathbb{E}(G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots) \mid Y_1, \ldots, Y_n) = \overline{G}(Y_1, \ldots, Y_n). \quad (7.16) \]
Taking into the consideration that \( g(Y_1, \ldots, Y_n) \) is also a version of the function of arguments \( Y_1, \ldots, Y_n, Y_{n+1}, \ldots \), write
\[ \mathbb{E}G(Y_1, \ldots, Y_n)g(Y_1, \ldots, Y_n) = \mathbb{E}Xg(Y_1, \ldots, Y_n) = \mathbb{E}G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots)g(Y_1, \ldots, Y_n) \]
\[ = \mathbb{E}\left( G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots) \mid Y_1, \ldots, Y_n \right) g(Y_1, \ldots, Y_n). \]
It suffices now to choose
\[ g(Y_1, \ldots, Y_n) = \text{sign}\left[ \mathbb{E}\left( G(Y_1, \ldots, Y_n, Y_{n+1}, \ldots) \mid Y_1, \ldots, Y_n \right) - \overline{G}(Y_1, \ldots, Y_n) \right]. \]
to get the result. \( \square \)

9. independence of \((X, Y) \implies \mathbb{E}(X|Y) = \mathbb{E}X.\)

Proof. The independence of \( X \) and \( Y \) implies
\[ \mathbb{E}(\mathbb{E}X - \mathbb{E}(X|Y))g(Y) = 0, \]
and, with \( g(Y) = \text{sign}(\mathbb{E}X - G(Y)) \), the result. \( \square \)

10. \( \mathbb{E}|h(Y)X| < \infty \implies \mathbb{E}(h(Y)X|Y) = h(Y)\mathbb{E}(X|Y).\)

Proof. Write
\[ \mathbb{E}(\mathbb{E}(h(Y)X|Y) - h(Y)\mathbb{E}(X|Y)g(Y) = \mathbb{E}h(Y)Xg(Y) - \mathbb{E}h(Y)Xg(Y) = 0. \]
Let \( G(z) \), \( G^h(z) \) be such that \( G(Y) = \mathbb{E}(X|Y) \), \( G^h(Y) = \mathbb{E}(h(Y)X|Y). \)
Then, with \( g(Y) = \text{sign}([G^h(z) - G(z)]) \), we get
\[ \mathbb{E}|\mathbb{E}(h(Y)X|Y) - h(Y)\mathbb{E}(X|Y)| = 0 \]
and the result. \( \square \)

11. \( \mathbb{E}X^2 < \infty \implies (\mathbb{E}(X|Y))^2 \leq \mathbb{E}(X^2|Y) \) (Cauchy-Schwartz inequality).
Proof. By virtue of 3. \( \mathbb{E} \left( [X - \mathbb{E}(X|Y)]^2 | Y \right) \geq 0 \). By 6.,
\[
\mathbb{E} \left( [X - \mathbb{E}(X|Y)]^2 | Y \right) = \mathbb{E}(X^2|Y) - 2\mathbb{E}(X \mathbb{E}(X|Y)|Y) + (\mathbb{E}(X|Y))^2 \\
= \mathbb{E}(X^2|Y) - (\mathbb{E}(X|Y))^2
\]
and the result holds true. \( \square \)

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

Proof. Set \( \delta(Y) = h(Y) - \mathbb{E}(X|Y) \). Then
\[
\mathbb{E}(X - h(Y))^2 = \mathbb{E}(\alpha - \mathbb{E}(X|Y) - \delta(Y))^2 \\
= \mathbb{E}(X - \mathbb{E}(X|Y))^2 + \mathbb{E}\delta^2(Y) - 2\mathbb{E}[X - \mathbb{E}(X|Y)]\delta(Y) \\
\geq \mathbb{E}(X - \mathbb{E}(X|Y))^2.
\]

13. \( h(x) \) is convex, \( \mathbb{E}|h(X)| < \infty \implies \mathbb{E}(h(X)|Y) \geq h(\mathbb{E}(X|Y)) \) (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 \( x = \mathbb{E}(X|Y) \). Then,
\[
h(X) - h(\mathbb{E}(X|Y)) \geq u_{\mathbb{E}(X|Y)}(X - \mathbb{E}(X|Y)).
\]
So, \( \mathbb{E}(\cdot|Y) \) taken from both sides of the above inequality provides the result by 3., 6., and 10. \( \square \)

14. Let \( X_n \to X \) and \( |X_n| \leq \gamma \), \( \mathbb{E}\gamma < \infty \). Then
\[
\lim_n \mathbb{E}(X_n|Y) = \mathbb{E}(X|Y), \ P - a.s.
\]

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 \( \lim \gamma_n = 0 \). Therefore \( \lim_n \mathbb{E}(\gamma_n|Y) = 0 \) and
\[
\lim_n |\mathbb{E}(X|Y) - \mathbb{E}(X_n|Y)| \leq \lim_n \mathbb{E}(\max_{m \geq n} |X - X_n||Y) \\
= \lim_n \mathbb{E}(\gamma_n|Y) = 0.
\]

15. If \( X, Y \) are independent of \( Z \), then
\[
\mathbb{E}(X|Y, Z) = \mathbb{E}(X|Y).
\]

6
Proof. Set \( g(Y, Z) = g_1(Y)g_2(Z) \). Then

\[
\begin{align*}
    EXg(Y, Z) &= EXg_1(Y)g_2(Z) = EXg_1(Y)Eg_2(Z) \\
    &= Eg_1(Y)EXg_2(Z) \\
    &= Eg_1(Y)g_2(Z)EXg_2(Z) = Eg(Y, Z)EXg_2(Z).
\end{align*}
\]

□

Examples:
1. For independent and identically distributed \( X, Y \) with \( EX|X| < \infty \),
   \[
   EX|X + Y| = EY|X + Y| = \frac{X + Y}{2}.
   \]
   By symmetry, \( EX|X + Y| = EY|X + Y| \) and obvious equality \( EX + Y|X + Y| = X + Y \) the result holds true.

2. For independent and identically distributed \( X_1, \ldots, X_n \), in the same way is proved that for any \( i \)
   \[
   EX_i|X_1 + \ldots + X_n| = \frac{1}{n} \sum_{j=1}^{n} X_j.
   \]

3. Let \( X_1, \ldots, X_n, \ldots \) be independent and identically distributed random variables with \( EX_1|X_1| < \infty \) and let \( k = \sum_{j=1}^{k} X_j \). Then for any \( i \leq n \)
   \[
   EX_i|S_n, S_{n+1}, \ldots, \) = \( \frac{S_n}{n}. \)
   (7.17)

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