So far we have focussed on independent random variables, which are well suitable to deal with statistical inference, allowing to exploit the very powerful central limit theorem. However, for the purpose of studying time dependent random phenomena, it is necessary to consider a much wider class of random variables.

The treatment of mutually dependent random variables is based on the fundamental notion of *conditional probability* and on the more sophisticated *conditional expectation* which we will present in this chapter.

The conditional probability is presented in every textbook about probability; we sketch this topic briefly in the following section and then we turn to the conditional expectation, which will provide a very important tool to deal with stochastic processes.

I. CONDITIONAL PROBABILITY

Definizione 1 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $B \in \mathcal{F}$ *an event with non-zero probability; the* conditional probability of A with respect to B *is:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

Intuitively, $P(A|B)$ is the probability that the event A occurs when we know that the event B has occurred. It is very simple to show that the map:

$$\mathcal{F} \ni A \rightsquigarrow P(A|B) \tag{2}$$

defines a new probability measure on $(\Omega, \mathcal{F})$. If the events A and B are independent, $P(A|B) = P(A)$: the occurring of the event B does not provide any information about the occurring of A.

From the definition (1) of conditional probability the following important theorems easily follow:

Teorema 2 (Bayes) *Let* $(\Omega, \mathcal{F}, P)$ *a probability space and* $A, B \in \mathcal{F}$ *events,* $P(A) \neq 0$, $P(B) \neq 0$; *then:*

$$P(A|B)\,P(B) = P(B|A)\,P(A) \tag{3}$$

Proof. *Both members are equal to* $P(A \cap B)$.

Teorema 3 (Law of Total Probability) *Let* $(\Omega, \mathcal{F}, P)$ *a probability space and* $A \in \mathcal{F}$ *an event and* $\{B_i\}_{i=1}^n$ *mutually disjoint events,* $P(B_i) \neq 0$, *such that* $\cup_i B_i = \Omega$; *then:*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \tag{4}$$

Proof.

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

II. CONDITIONAL EXPECTATION

The remainder of the chapter is devoted to the presentation of the *conditional expectation*, which can be thought as a generalization of the concept of conditional probability: the idea of conditional expectation rises from the observation that the knowledge of one or more events, represented by a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, allows to "predict" values taken by a random variable X through another random variable $E[X|\mathcal{G}]$, the so-called conditional expectation of X with respect to $\mathcal{G}$.

Let X be a real **integrable** random variable on a probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Consider the map:

$$\mathcal{G} \ni B \rightsquigarrow Q^{X,\mathcal{G}}(B) := \int_B X(\omega)P(d\omega) \tag{5}$$

If $X \geq 0$, (5) defines a positive measure on $(\Omega, \mathcal{G})$, absolutely continuous with respect to P; by virtue of the Radon-Nikodym theorem, there exists a real random variable Z, **$\mathcal{G}$-measurable**, a.s. unique and such that:

$$Q^{X,\mathcal{G}}(B) = \int_B Z(\omega)P(d\omega) \quad \forall B \in \mathcal{G} \tag{6}$$

Such random variable will be denoted:

$$Z = E\left[X|\mathcal{G}\right] \tag{7}$$

and called **conditional expectation** of X given $\mathcal{G}$. If X is not positive, it can be represented as difference of two positive random variables $X = X^+ - X^-$ and its conditional expectation given $\mathcal{G}$ can be defined as follows:

$$Z = E\left[X|\mathcal{G}\right] = E\left[X^+|\mathcal{G}\right] - E\left[X^-|\mathcal{G}\right] \tag{8}$$

The conditional expectation $E\left[X|\mathcal{G}\right]$ is defined by the two conditions:

1. $E\left[X|\mathcal{G}\right]$ is $\mathcal{G}$-measurable

2. $E\left[1_B\, E\left[X|\mathcal{G}\right]\right] = E\left[1_B\, X\right], \quad \forall B \in \mathcal{G}$

With measure theory arguments basta una citazione?, it can be proved that the second condition is equivalent to:

$$E\left[W\, E\left[X|\mathcal{G}\right]\right] = E\left[W\, X\right] \tag{9}$$

for all **bounded** and **$\mathcal{G}$-measurable** random variables W.

The key point is the **$\mathcal{G}$-measurability**: if $\mathcal{G}$ represents the amount of information available, in general we cannot access all information about X; on the other hand, we can build up $E\left[X|\mathcal{G}\right]$, whose distribution is known, and use it to replace X whenever events belonging to $\mathcal{G}$ are considered.

We will often use the notation:

$$P\left(A|\mathcal{G}\right) \overset{def}{=} E\left[1_A|\mathcal{G}\right] \tag{10}$$

and call (10) the **conditional probability** of A given $\mathcal{G}$; we stress that (10), in contrast to (1), is a random variable. Moreover, it is $\mathcal{G}$-measurable and such that:

$$\int_B P\left(A|\mathcal{G}\right)(\omega)\, P(d\omega) = \int_B 1_A(\omega)P(d\omega) = P(A \cap B), \quad \forall B \in \mathcal{G} \tag{11}$$

III. AN ELEMENTARY CONSTRUCTION

We will now provide some intuitive and practical insight into the formal definition (6) of conditional expectation, giving an elementary construction of (6) under some simplifying hypotheses. Let $Y : \Omega \to \mathbb{R}$ be a random variable and, as usual:

$$\sigma(Y) \stackrel{def}{=} \left\{ A \subset \Omega | A = Y^{-1}(B), B \in \mathcal{B}(\mathbb{R}) \right\} \tag{12}$$

the σ-algebra **generated by** Y, i.e. the smallest sub-σ-algebra of $\mathcal{F}$ with respect to which Y is measurable. Intuitively, if our amount of information is $\sigma(Y)$, this means that, after an experiment, we know only the value of Y: we do not have access to other information.

Let us assume that Y be **discrete**, i.e. that it can assume at most countably infinite values $\{y_1, \dots, y_n, \dots\}$. For all events $A \in \mathcal{F}$, let us define:

$$P\left(A | Y = y_i\right) \stackrel{def}{=} \begin{cases} \frac{P(A \cap \{Y = y_i\})}{P(Y = y_i)} & \text{if } P\left(Y = y_i\right) > 0 \\ \\ 0 & otherwise \end{cases} \tag{13}$$

Equation (13) is nothing but the familiar conditional probability of A given the event $\{Y = y_i\}$. Due to the law of total probability:

$$P(A) = \sum_i P\left(A | Y = y_i\right) P\left(Y = y_i\right) \tag{14}$$

Let now $X : \Omega \to \mathbb{R}$ be an integrable random variable, which we assume **discrete** for simplicity, and consider the map:

$$\mathcal{B}(\mathbb{R}) \ni H \rightsquigarrow P\left(X \in H | Y = y_i\right) \tag{15}$$

(15) is a probability measure on $\mathbb{R}$, with expectation:

$$E\left[X | Y = y_i\right] \stackrel{def}{=} \sum_j x_j P\left(X = x_j | Y = y_i\right) \tag{16}$$

Consider now the function $h : \mathbb{R} \to \mathbb{R}$:

$$\mathbb{R} \ni y \rightsquigarrow h(y) \stackrel{def}{=} \begin{cases} E\left[X | Y = y_i\right], & \text{if } y = y_i,\ P\left(Y = y_i\right) > 0 \\ 0 & otherwise \end{cases} \tag{17}$$

h is clearly measurable, so that it makes perfectly sense to construct the random variable $Z : \Omega \to \mathbb{R}$

$$\omega \in \Omega \mapsto Z(\omega) \stackrel{def}{=} h\left(Y(\omega)\right) \tag{18}$$

that is, recalling (16):

$$Z(\omega) = \begin{cases} E\left[X | Y = y_i\right], & \text{if } Y(\omega) = y_i,\ P\left(Y = y_i\right) > 0 \\ arbitrary\, value & otherwise \end{cases} \tag{19}$$

Equation (19) has a straightforward interpretation: given a realization of Y, the expectation of the possible outcomes of X can be computed; this expectation is random, like Y. At a first sight, the arbitrary constant in (19) could seem disturbing: nevertheless, the subset of Ω on which Z takes an arbitrary value has probability equal to 0.

Incidentally, we remark that, for all $A \in \mathcal{F}$:

$$E\left[1_A|Y = y_i\right] = \sum_{a=0,1} a\, P\left(1_A = a|Y = y_i\right) = P\left(A|Y = y_i\right) \tag{20}$$

It remains to show that:

$$Z(\omega) = E\left[X|\sigma(Y)\right](\omega) \quad a.s. \tag{21}$$

First, we show that $Z : \Omega \to \mathbb{R}$ is $\sigma(Y)$-measurable; to this purpose, consider $H \in \mathcal{B}(\mathbb{R})$:

$$Z^{-1}(H) = (h \circ Y)^{-1}(H) = Y^{-1}\left(h^{-1}(H)\right) \in \sigma(Y) \tag{22}$$

since $h^{-1}(H) \in \mathcal{B}(\mathbb{R})$. Therefore, Z is $\sigma(Y)$-measurable. Let now be W a bounded and $\sigma(Y)$-measurable random variable (this includes the case $W = 1_B$, with $B \in \sigma(Y)$). By virtue of a theorem by J. L. Doob, which we state without proof reminding the interesting reader to [**?**], there exists a measurable function $w : \mathbb{R} \to \mathbb{R}$ such that $W = w(Y)$. Recalling that $Z = h(Y)$, we therefore have:

$$\begin{aligned} E\left[WZ\right] &= \textstyle\sum_i w(y_i) E\left[X|Y = y_i\right] P(Y = y_i) = \\ &= \textstyle\sum_i w(y_i) \sum_j x_j P\left(X = x_j|Y = y_i\right) P(Y = y_i) = \\ &= \textstyle\sum_{i,j} w(y_i) x_j P\left(\{X = x_j\} \cap \{Y = y_i\}\right) = E\left[WX\right] \end{aligned} \tag{23}$$

which is exactly the second condition defining $E\left[X|\sigma(Y)\right]$.

IV. AN EXPLICIT CALCULATION

The notion of conditional expectation $E\left[X|\sigma(Y)\right]$ is well defined also for continuous random variables X, Y. Several authors write, for the sake of simplicity, $E\left[X|Y\right]$ instead of $E\left[X|\sigma(Y)\right]$.

Remarkably, since $E\left[X|\sigma(Y)\right]$ is $\sigma(Y)$-measurable, by virtue of Doob's theorem there exists a measurable function g such that:

$$E[X|Y] = g(Y) \tag{24}$$

Equation (24) has the intuitive interpretation that to predict X given Y it is sufficient to apply a measurable "deterministic" function to Y. In the remainder of this section, we will present a practical way to compute explicitely $g(Y)$ in some special situations. The following discussion will be based on the simplifying assumptions:

1. that the random variables X, Y take values in $\mathbb{R}$.

2. that X and Y have joint law absolutely continuous with respect to the Lebesgue measure, and therefore admit joint probability density $p(x,y)$.

3. that the joint probability density $p(x,y)$ be a.e. non-zero.

Under these hypotheses, the marginal probability densities and the conditional probability density of X given Y can be defined with the formulas:

$$p_X(x) = \int dy\, p(x,y) \qquad p_Y(y) = \int dx\, p(x,y) \qquad p(x|y) = \frac{p(x,y)}{p_Y(y)} \tag{25}$$

By virtue of these definitions, and of Fubini's theorem, we find that for all bounded measurable functions $h : \mathbb{R} \to \mathbb{R}$ one has:

$$\begin{aligned} E[Xh(Y)] &= \int dx \int dy\, p(x,y)\, x\, h(y) = \\ &= \int dy\, p_Y(y)\, h(y) \int dx\, x\, p(x|y) = E[g(Y)h(Y)] \end{aligned}$$

where the measurable function:

$$g(y) = \int dx\, x\, p(x|y)$$

has appeared. Since, on the other hand:

$$E[Xh(Y)] = E[E[X|Y]h(Y)]$$

we conclude that $E[X|Y] = g(Y)$. As an example of remarkable importance, consider a bivariate normal random variable $(X,Y) \sim N(\mu, \Sigma)$ with joint density:

$$p(x,y) = \frac{e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2(1-\rho^2)} + \frac{\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y(1-\rho^2)} - \frac{(y-\mu_y)^2}{2\sigma_y^2(1-\rho^2)}}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

A straightforward calculation shows that:

$$p_Y(y) = \frac{e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}}{\sqrt{2\pi\sigma_y^2}}$$

which allows to write $p(x|y)$ and to compute $g(y)$. The result is:

$$g(y) = \mu_x + \frac{\rho\,\sigma_x}{\sigma_y}\,(y - \mu_y)$$

The conditional expectation of X given Y is therefore, in this special situation, *a linear function of* Y, explicitly depending on the elements of the covariance matrix Σ.

V. PROPERTIES OF CONDITIONAL EXPECTATION

Teorema 4 *Let X be a real integrable random variable, defined on a probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Then:*

1. *the map $X \rightsquigarrow E[X|\mathcal{G}]$ is a.s. linear*

2. *if $X \geq 0$ a.s., then $E[X|\mathcal{G}] \geq 0$ a.s.*

3. *$E[E[X|\mathcal{G}]] = E[X]$.*

4. *if X is $\mathcal{G}$-measurable, then $E[X|\mathcal{G}] = X$ a.s.*

5. *if X is independent on $\mathcal{G}$, i.e. is if $\sigma(X)$ and $\mathcal{G}$ are independent, then $E[X|\mathcal{G}] = E[X]$ a.s.*

6. *if $\mathcal{H} \subset \mathcal{G}$ is a σ-algebra, then $E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$ a.s.*

7. *if Y is bounded and $\mathcal{G}$-measurable, then $E[YX|\mathcal{G}] = YE[X|\mathcal{G}]$ a.s.*

Proof. *The first two points are obvious. To prove the third one, it is sufficient to recall that:*

$$E[1_B\, E[X|\mathcal{G}]] = E[1_B\, X], \quad \forall B \in \mathcal{G} \tag{26}$$

and choose $B = \Omega$. To prove the fourth point, it is sufficient to observe that, in such case, X itself satisfies the two conditions defining the conditional expectation.

To prove the fifth point, observe that since the random variable $\omega \mapsto E[X]$ is constant and therefore $\mathcal{G}$-measurable, and since for all $B \in \mathcal{G}$ the random variable 1_B is clearly $\mathcal{G}$-measurable and independent on X:

$$E[1_B\, X] = E[1_B]\, E[X] = E[1_B E[X]] \tag{27}$$

so that $\omega \mapsto E[X]$ satisfies the two conditions defining the conditional expectation.

To prove the sixth point, observe that by definition $E[E[X|\mathcal{G}]|\mathcal{H}]$ is $\mathcal{H}$-measurable; moreover, since $B \in \mathcal{H}$, 1_B é $\mathcal{H}$-measurable and also $\mathcal{G}$-measurable (since $\mathcal{G}$ contains $\mathcal{H}$). Therefore:

$$E[1_B\, E[E[X|\mathcal{G}]|\mathcal{H}]] = E[1_B\, E[X|\mathcal{G}]] = E[1_B\, X] \tag{28}$$

where the definition of conditional expectation has been applied twice.

To prove the last point observe that, as a product of $\mathcal{G}$-measurable random variables, $YE[X|\mathcal{G}]$ is $\mathcal{G}$-measurable. For all bounded, $\mathcal{G}$-measurable random variables Z, therefore:

$$E[Z\, Y E[X|\mathcal{G}]] = E[Z\, Y X] = E[Z\, E[YX|\mathcal{G}]] \tag{29}$$

since also ZY is bounded and $\mathcal{G}$-measurable.

VI. CONDITIONAL EXPECTATION AS PREDICTION

We are going now to put on a firm ground the intuitive idea of the conditional expectation as prediction: $E[X|\mathcal{G}]$ "predicts" X when the amount of information is $\mathcal{G}$. To this purpouse, we need a geometrical interpretation: let $L^2(\Omega, \mathcal{F}, P)$ be the Hilbert space of (complex valued) square-integrable random variables, endowed with the inner product:

$$\langle u|v\rangle \stackrel{def}{=} E[\overline{u}v] \tag{30}$$

Moreover, as usual $\mathcal{G}$ is sub-σ-algebra of $\mathcal{F}$. The space $L^2(\Omega, \mathcal{G}, P)$ is a closed subspace $L^2(\Omega, \mathcal{F}, P)$. Let us define the mapping:

$$L^2(\Omega, \mathcal{F}, P) \ni v \rightsquigarrow \hat{Q}v \stackrel{def}{=} E[v|\mathcal{G}] \tag{31}$$

We leave to the reader the simple proof of the fact that $\hat{Q}$ is a **linear operator** from $L^2(\Omega, \mathcal{F}, P)$ to $L^2(\Omega, \mathcal{G}, P)$.

Let us prove that $\hat{Q}$ is idempotent, that is $\hat{Q}^2 = \hat{Q}$:

$$\begin{aligned}\hat{Q}^2 v = \hat{Q}E[v|\mathcal{G}] = E[E[v|\mathcal{G}]|\mathcal{G}] = \\ = E[v|\mathcal{G}] = \hat{Q}v\end{aligned} \tag{32}$$

Moreover $\hat{Q}$ is self-adjoint, since:

$$\begin{aligned}\langle u|\hat{Q}v\rangle = E[\overline{u}E[v|\mathcal{G}]] = E[E[\overline{u}E[v|\mathcal{G}]|\mathcal{G}]] = \\ = E[E[v|\mathcal{G}]E[\overline{u}|\mathcal{G}]] = E[E[vE[\overline{u}|\mathcal{G}]|\mathcal{G}]] = \\ = E[vE[\overline{u}|\mathcal{G}]] = E\left[\overline{E[\overline{u}|\mathcal{G}]}v\right] = \langle\hat{Q}u|v\rangle\end{aligned} \tag{33}$$

Therefore $\hat{Q}$ is an **orthogonal projector** onto the subspace $L^2(\Omega, \mathcal{G}, P)$.

Let now be $X \in L^2(\Omega, \mathcal{F}, P)$ a real random variable; let us look for the element $Y \in L^2(\Omega, \mathcal{G}, P)$ such that $||X - Y||^2$ is minimum. The minimum is reached for $Y = E[X|\mathcal{G}]$. The key point is that $Y = \hat{Q}Y$, in fact:

$$\begin{aligned}E[(X-Y)^2] = ||X-Y||^2 = ||\hat{Q}X + (1-\hat{Q})X - \hat{Q}Y||^2 = \\ = ||\hat{Q}(X-Y)||^2 + ||(1-\hat{Q})X||^2 = ||Y - \hat{Q}X||^2 + ||(1-\hat{Q})X||^2\end{aligned} \tag{34}$$

and the minimum is achieved precisely at $Y = \hat{Q}X$.

In the sense of L^2, thus, $Y = E[X|\mathcal{G}]$ is the best approximation of X among the class of $\mathcal{G}$-measurable functions. This is the justification of the interpretation of $Y = E[X|\mathcal{G}]$ as a prediction: within the set of square-integrable $\mathcal{G}$-measurable random variables, $Y = E[X|\mathcal{G}]$ is the closest one to X in the topology of L^2.

VII. LINEAR REGRESSION AND CONDITIONAL EXPECTATION

There is a very interesting connection between conditional expectation and linear regression which we are going now to explore. Let's consider two real random variables Y and Z, representing two properties one wishes to measure during an experiment. Quite often it happens that the quantity Z can be measured with an high accuracy, while Y, a "response", contains a signal and a noise difficult to disentangle. In such situations, from a mathematical point of view, the experimentalist would like to work with $\sigma(Z)$-measurable random variables: such quantities, in fact, have a well defined value once the outcome of Z is known. The key point is that $E[Y|Z]$ is the best prediction for Z within the set of $\sigma(Z)$-measurable random variables.

Since the conditional expectation is a linear projector, we can always write the unique decomposition:

$$Y = E[Y|Z] + \varepsilon \tag{35}$$

where ε is a real random variable. It is immediate to show that:

$$E[\varepsilon] = E[\varepsilon|Z] = 0 \tag{36}$$

Moreover, since $Y - E[Y|Z]$ is orthogonal to all the random variables $\sigma(Z)$-measurable:

$$E[(Y - E[Y|Z])h(Z)] = 0 \tag{37}$$

in particular the following orthogonality property holds:

$$E[\varepsilon Z] = 0 \tag{38}$$

In order to proceed further, let's assume that the **two-dimensional random variable** (Z, X) **is normal**. In such case we know that the conditional expectation depends linearly on Z:

$$E[Y|Z] = a + bZ \tag{39}$$

with:

$$\begin{aligned} a &= \frac{Var(Z)E[Y] - E[Z]Cov(Z,Y)}{Var(Z)} \\ b &= \frac{cov(Z,Y)}{var(Z)} \end{aligned} \tag{40}$$

The "error" $\varepsilon = Y - (a + bZ)$ is also normal being a linear function of (X, Z) of zero mean. Moreover, since $E[\varepsilon Z] = 0$, ε is **independent** of Z. The variance is:

$$\sigma^2 = Var(\epsilon) = E\left[(Y - (a + bZ))^2\right] \tag{41}$$

and, from the geometrical interpretation of the conditional expectatio, we know that the parameters in (40) minimize such quantity.

To summarize, we have found that, whenever two quantities Z and Y have a joint normal law, the best prediction we can do for Y once we know the value of Z is a linear function of such value. The experimentalist collects a set of data $\{(z_1, y_1), \ldots, (z_n, y_n)\}$, interprets such data as realization of two-dimensional random variables (Z_i, Y_i) independent and identically distributed as (Z, Y), and uses such data to infer the values of a,b and σ^2, the last one providing the "accuracy" of the linear approximation.

For the statistical analysis, we refer to the previous chapter.

VIII. A USEFUL FORMULA

The conclusion of the present chapter is devoted to the presentation of a useful result, which will be used later.

We already know that if X is integrable and $\mathcal{G}$-measurable, Z is integrable an independent on $\mathcal{G}$, and the product XZ is integrable, then:

$$E\left[XZ|\mathcal{G}\right] = XE\left[Z|\mathcal{G}\right] = XE\left[Z\right] \tag{42}$$

If $X : \Omega \to E$ is a $\mathcal{G}$-measurable random variable taking values in the measurable space $(E, \mathcal{E})$, $f : E \to \mathbb{R}$ is a measurable function such that $f \circ X$ is integrable, $\mathcal{H}$ a sub-σ-algebra of $\mathcal{F}$ independent on $\mathcal{G}$ and Z a $\mathcal{H}$-measurable and integrable random variable, with $f(X)Z$ integrable, (42) implies the following equality:

$$E\left[f(X)Z|\mathcal{G}\right] = f(X)E\left[Z\right] \tag{43}$$

The latter can be rewritten introducing the function $\psi : E \times \Omega \to \mathbb{R}$:

$$E \times \Omega \ni (x, \omega) \rightsquigarrow \psi(x, \omega) \stackrel{def}{=} f(x)Z(\omega) \tag{44}$$

which is $\mathcal{E} \otimes \mathcal{H}$-measurable. If $\omega \rightsquigarrow \psi(X(\omega), \omega)$ is integrable, we have:

$$E\left[\psi(X, \cdot)|\mathcal{G}\right] = \Phi(X), \quad \Phi(x) \stackrel{def}{=} E\left[\psi(x, \cdot)\right] \tag{45}$$

Moving to functions ψ which are given by linear combinations of factorized functions and using suitable measure-theoretical arguments one arrives at the following general result, of which we omit the proof:

Teorema 5 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G}$ and $\mathcal{H}$* **mutually independent** *sub-σ-algebras of $\mathcal{F}$. let $X : \Omega \to E$ be a $\mathcal{G}$*-**measurable** *random variable taking values in the measurable space $(E, \mathcal{E})$ and ψ a function $\psi : E \times \Omega \to \mathbb{R}$ $\mathcal{E} \otimes \mathcal{H}$*-**measurable**, *such that $\omega \mapsto \psi(X(\omega), \omega)$ is integrable. Then:*

$$E\left[\psi(X, \cdot)|\mathcal{G}\right] = \Phi(X), \quad \Phi(x) \stackrel{def}{=} E\left[\psi(x, \cdot)\right] \tag{46}$$