11. CONDITIONAL EXPECTATION

DEFINITION.

Given $(\Omega, \mathcal{F}, \mathcal{P})$ and a σ-algebra $\mathcal{G} \subset \mathcal{F}$, a random variable $X(\omega)$ on Ω, we define the CONDITIONAL EXPECTATION OF X GIVEN $\mathcal{G}$ the unique (a.e.) random variable

$Y(\omega) = E\{X \mid \mathcal{G}\}(\omega)$ such that

- it is $\mathcal{G}$-measurable, i.e.

  $Y^{-1}(B) \in \mathcal{G} \quad \forall B \in \mathcal{B}(\mathbb{R})$

- $$\int_A X(\omega)\, d\mathcal{P}(\omega) = \int_A E\{X \mid \mathcal{G}\}(\omega)\, d\mathcal{P}(\omega)$$

  $\forall A \in \mathcal{G}$.

EXAMPLE. Given a random variable $X(\omega)$ on $(\Omega, \mathcal{F}, P)$, calculate its conditional expectation given $\mathcal{G} = \{\emptyset, A, A^c, \Omega\}$ where $A \in \mathcal{F}$. The events $\{ \; , A, A^c \; \}$ are the atoms of $\mathcal{G}$. Since any random variable which is $\mathcal{G}$-measurable must be constant over the atoms and $E\{X|\mathcal{G}\}$ must be $\mathcal{G}$-measurable, $E\{X|\mathcal{G}\}$ must have the form

$$E\{X|\mathcal{G}\}(\omega) = \alpha_1 \chi_A(\omega) + \alpha_2 \chi_{A^c}(\omega)$$

By the definition of $E\{X|\mathcal{G}\}$

$$\int_A E\{X|\mathcal{G}\} dP = \alpha_1 P(A) = \int_A X dP$$

$$\int_{A^c} E\{X|\mathcal{G}\} dP = \alpha_2 P(A_c) = \int_{A^c} X dP$$

It follows

$$E\{X|\mathcal{G}\}(\omega) = \begin{cases} \frac{1}{P(A)} \int_A X(\omega)\, dP(\omega) & \text{if } \omega \in A \\ \frac{1}{P(A^c)} \int_{A^c} X(\omega)\, dP(\omega) & \text{if } \omega \in A^c \end{cases}$$

In general, if $\mathcal{G}$ is generated by a countable family of atoms $\{A_i\}$, it is possible to see

$$E\{X|\mathcal{A}\}(\omega) = \sum_{i=1}^{\infty} \left(\frac{1}{P(A_i)} \int_{A_i} X(\omega)\, dP(\omega) \right) \chi_{A_i}(\omega)$$

Notice also that

$$\int_{A_i} X(\omega)\, dP(\omega) = \int_{\Omega} \chi_{A_i}(\omega) X(\omega)\, dP(\omega)$$

$$= E\{\chi_{A_i} X(\omega)\}$$ so that

$$E\{X|\mathcal{A}\}(\omega) = \sum_{i=1}^{\infty} \frac{1}{P(A_i)} E\{\chi_{A_i} X(\omega)\} \chi_{A_i}(\omega)$$

EXAMPLE Consider $(\Omega, \mathcal{F}, \lambda)$ with $\Omega = [0,1]$, $\lambda : \lambda([a,b]) = b-a$ and $\mathcal{F}$ is the σ-algebra generated by the following atoms

$$\{\phi, [0,1/3), (1/3, 1/2), [1/2, 4/5), [4/5, 1]\}$$

Any random variable $X(\omega)$ on $(\Omega, \mathcal{F}, \lambda)$ is constant on the atoms. Consider for example

$$X(\omega) = \begin{cases} 5 & \omega \in [0,1/3) \\ 2 & \omega \in [1/3, 1/2) \\ 3 & \omega \in [1/2, 4/5) \\ 0.5 & \omega \in [4/5, 1] \end{cases}$$

and the σ-algebra $\mathcal{G} \subset \mathcal{F}$ generata dagli atomi

$$\{\phi, [0,1/2), [1/2,1]\}$$

Let calculate $E\{X|\mathcal{G}\}$ on the atom $[0,1/2)$:

$$E\{X|\mathcal{G}\}(\omega) = \frac{1}{P([0,1/2))} \int_{[0,1/2)} X(\omega)\,dP(\omega)$$

$$= \frac{1}{1/2} \cdot \left[5 \cdot P([0,1/3)) + 2 \cdot P([1/3, 1/2)) \right]$$

$$= 4 \qquad \text{if } \omega \in [0,1/2)$$

and

$$E\{X|\mathcal{G}\}(\omega) = \frac{1}{P([1/2,1])} \int_{[1/2,1]} X(\omega)\,dP(\omega)$$

$$= \frac{1}{1/2} \cdot \left[3 \cdot P([1/2, 4/5)) + 0.5\, P([4/5, 1]) \right]$$

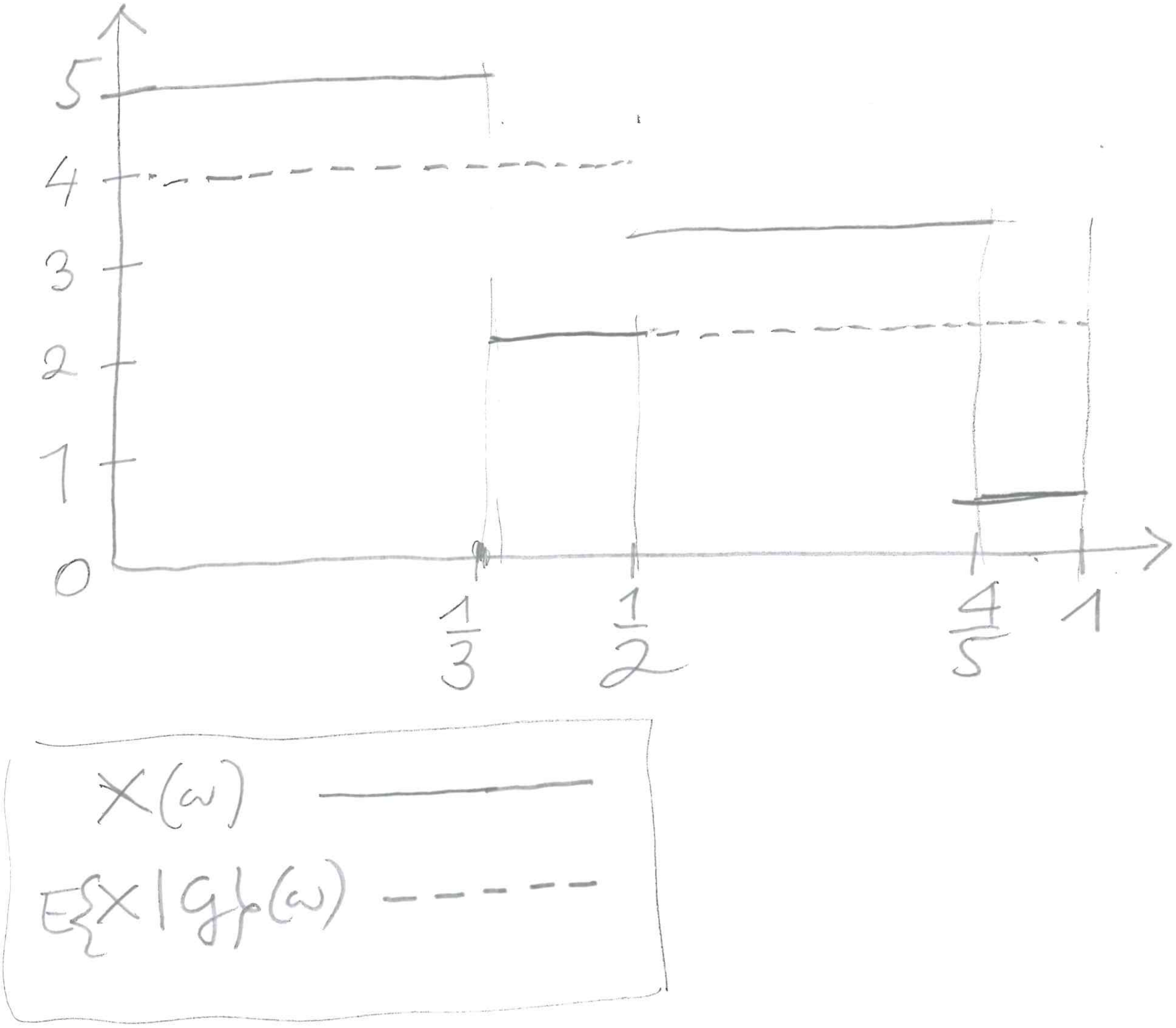
$$= 2 \qquad \text{if } \omega \in [1/2, 1]$$

Notice that

$$E\{X\} = \int_{\Omega} X(\omega)\,dP(\omega) = 5 \cdot P([0,1/3))$$

$$+ 2 \cdot P([1/3, 1/2)) + 3\, P([1/2, 4/5)) + \frac{1}{2} P([4/5, 1])$$

$$= 3$$



E{X|G} is the (weigted) mean of X over [0, 1/2) and [1/2, 1] respectively!

DEFINITION. For given random variables X and $\{Y_1, \dots, Y_n\}$ over $(\Omega, \mathcal{F}, \mathcal{P})$ we define the CONDITIONAL EXPECTATION OF X given $\{Y_1, \dots, Y_n\}$ the conditional expectation of X given the σ-algebra $\mathcal{F}^{Y_1, \dots, Y_n}$ generated by $\{Y_1, \dots, Y_n\}$. We write $E\{X \mid Y_1, \dots, Y_n\}$.

EXAMPLE. $\Omega = [0,1]$, $\mathcal{F}$ is the Borel σ-algebra on $\Omega = [0,1]$, $\mathcal{P}$: $\mathcal{P}([a,b]) = b-a$. Consider

$$X(\omega) = \omega$$

and

$$Y_1(\omega) = \begin{cases} 1 & \omega \in [0, 1/2) \\ 2 & \omega \in [1/2, 1] \end{cases} \qquad Y_2(\omega) = \begin{cases} 3 & \omega \in [0, 1/3] \\ 2 & \omega \in (1/3, 2/3) \\ 4 & \omega \in [2/3, 1] \end{cases}$$

Clearly, $\mathcal{F}^{Y_1, Y_2} \subset \mathcal{F}$ and $\mathcal{F}^{Y_1, Y_2}$ is the smallest σ-algebra containing

$$\{[0, 1/3], (1/3, 1/2), [1/2, 2/3), [2/3, 1]\}.$$

Since $E\{X \mid Y_1 Y_2\}$ is $\mathcal{F}^{Y_1, Y_2}$-misurable, it is constant on the atoms $\{[0,1/3], (1/3, 1/2), [1/2, 2/3), [2/3, 1]\}$.

We have

$$E\{X \mid Y_1 Y_2\} = \begin{cases} 1/6 & \omega \in [0, 1/3 \\ 5/12 & \omega \in (1/3, 1/2) \\ 7/12 & \omega \in [1/2, 2/3) \\ 5/6 & \omega \in [2/3, 1] \end{cases}$$

11.1 PROPERTIES OF CONDITIONAL EXPECTATION

i) if X is random variable which is also $\mathcal{G}$-misureable:

$$\boxed{E\{X \mid \mathcal{G}\} = X \quad a.e.}$$

ii) if $\mathcal{F}_m = \{\phi, \Omega\}$

$$\boxed{E\{X \mid \mathcal{G}\} = E\{X\}}$$

iii) $\forall a_1, a_2 \in \mathbb{R}$,

X_1, X_2 random variables :

$$\boxed{\begin{aligned} E\{a_1 X_1 + a_2 X_2 \mid \mathcal{G}\} &= a_1 E\{X \mid \mathcal{G}\} \\ &+ a_2 E\{X_2 \mid \mathcal{G}\} \end{aligned}}$$

iv) if X_1, X_2 are random variables

and $X_1 \leq X_2$ a.e. :

$$\boxed{E\{X_1 \mid \mathcal{G}\} \leq E\{X_2 \mid \mathcal{G}\} \quad \text{a.e.}}$$

v) if X is a random variable

$$E\{E\{X \mid \mathcal{G}\}\} = E\{X\}.$$

<u>REMARK</u> Hölder, Cauchy, Minkowski, Jensen inequalities straightforwardly extend to the case of conditional expectations

<u>FACT</u>. Given random variables X, Y on $(\Omega, \mathcal{F}, P)$ and Y is $\mathcal{G}$-measurable, $\mathcal{G} \subset \mathcal{F}$,

$$E\{XY \mid \mathcal{G}\} = Y E\{X \mid \mathcal{G}\} \text{ a.e.}$$

<u>FACT</u>. Given σ-algebras $\mathcal{G}_1, \mathcal{G}_2$ with $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ and X random variable on $(\Omega, \mathcal{F}, P)$:

$$E\{X \mid \mathcal{G}_1\} = E\{E\{X \mid \mathcal{G}_1\} \mid \mathcal{G}_2\}$$
$$= E\{E\{X \mid \mathcal{G}_2\} \mathcal{G}_1\} \quad \text{a.e.}$$

<u>FACT</u>. There exists a measurable function $f: \mathbb{R} \to \mathbb{R}$ such

$$E\{X \mid Y\}(\omega) = f(Y(\omega))$$

The above fact proves that from the values of $Y(\omega)$ it is possible to obtain the expectation of $X(\omega)$.

$$= 1 \cdot \mathbb{P}(A \cap \{\omega_1, \omega_5\}) / \mathbb{P}(A)$$
$$+ 3 \cdot \mathbb{P}(A \cap \{\omega_2, \omega_4\}) / \mathbb{P}(A)$$
$$+ 2 \cdot \mathbb{P}(A \cap \{\omega_3\}) / \mathbb{P}(A) = \frac{6}{5}$$

Finally, notice that X and Y are not independent:

$$\mathbb{P}\{X=1\} = \mathbb{P}\{\{\omega_1, \omega_5\}\} = \frac{9}{16}$$
$$\mathbb{P}\{Y=1\} = \mathbb{P}\{\{\omega_1, \omega_3\}\} = \frac{5}{8}$$
$$\mathbb{P}\{X=1, Y=1\} = \mathbb{P}\{\{\omega_1\}\} = \frac{1}{2}$$

But $\mathbb{P}\{X=1\}\mathbb{P}\{Y=1\} = \frac{45}{128}$

$$\neq \mathbb{P}\{X=1, Y=1\} = \frac{1}{2}$$

As already seen,

$$\boxed{\begin{array}{l} X, Y \text{ independent} \Rightarrow \\ E\{X|Y\} = E\{X\} \end{array}}$$

The converse is false but

$$\boxed{\begin{array}{l} E\{X|Y\} = E\{X\} \Rightarrow \\ X, Y \text{ uncorrelated} \end{array}}$$

Indeed,

$$\sigma_{XY} = E\{(Y - E\{Y\})(X - E\{X\})\} =$$

$$= E\{YX\} - E\{Y\}E\{X\} =$$

$$= E\{E\{YX | \mathcal{F}^Y\}\} - E\{Y\}E\{X\} =$$

$$= E\{YE\{X | \mathcal{F}^Y\}\} - E\{YE\{X\}\} =$$

$$= E\{Y(E\{X | \mathcal{F}^Y\} - E\{X\})\} = 0$$

It is also possible to write $E\{X|Y\}$ as some function of Y:

$$E\{X|Y\}(y) = \begin{cases} 2 & y=0 \\ 6/5 & y=1 \\ 3 & y=2 \end{cases}$$

If we take $P\{\cdot|Y\}$ as a probability measure on $(\Omega, \mathcal{F})$ we can calculate the conditional expectation $E\{X|Y\}$ exactly as we did for $E\{X\}$ with probability measure P:

$$E\{X|Y\} = \int_\Omega X(\omega)\, dP\{\omega|Y\}$$

For example, for $\omega \in A$:

$$E\{X|Y\}(\omega) = 1 \cdot P\{\{\omega_1, \omega_5\}|Y\}(\omega)$$
$$+ 3 \cdot P\{\{\omega_2, \omega_4\}|Y\}(\omega)$$
$$+ 2 \cdot P\{\{\omega_3\}|Y\}(\omega) =$$

Therefore, for calculating $E\{X|Y\}$ it is sufficient to calculate it on the atoms of $\mathcal{F}^Y$!

We obtain

$\omega \in A$: $\int_A X \, dP = \frac{3}{4} \Rightarrow E\{X|Y\} = \frac{6}{5}$

$\omega \in B$: $\int_B X \, dP = \frac{1}{4} \Rightarrow E\{X|Y\} = 3$

$\omega \in C$: $\int_C X \, dP = \frac{1}{4} \Rightarrow E\{X|Y\} = 2$

Let's see how to determine $P(D|Y)$ for any $D \in \mathcal{F}$. Also in this case, it is sufficient to calculate $P(D|Y)$ on the atoms of $\mathcal{F}^Y$:

$\omega \in A$: $\int_A X_D \, dP = P(A \cap D) \Rightarrow P(D|Y) = \dfrac{P(A \cap D)}{P(A)}$

$\omega \in B$: $\int_B X_D \, dP = P(B \cap D) \Rightarrow P(D|Y) = \dfrac{P(B \cap D)}{P(B)}$

$\omega \in C$: $\int_C X_D \, dP = P(C \cap D) \Rightarrow P(D|Y) = \dfrac{P(C \cap D)}{P(C)}$

EXAMPLE. $\Omega = \{\omega_1, \dots, \omega_5\}$ with $\mathcal{F} = \mathcal{F}_M$ and

$P\{\omega_1\} = \frac{1}{2}$, $P\{\omega_2\} = \frac{1}{4}$, $P\{\omega_3\} = \frac{1}{8}$,

$P\{\omega_4\} = \frac{1}{16}$, $P\{\omega_5\} = \frac{1}{16}$.

Let

$$X(\omega) = \begin{cases} 1 & \omega \in \{\omega_1, \omega_5\} \\ 3 & \omega \in \{\omega_2, \omega_4\} \\ 2 & \omega = \omega_3 \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \omega \in \{\omega_1, \omega_3\} \\ 2 & \omega = \omega_2 \\ 0 & \omega \in \{\omega_4, \omega_5\} \end{cases}$$

Notice that

$\mathcal{F}^Y = \{\emptyset, A, B, C, A^c, B^c, C^c, \Omega,$
$A \cup B, B \cup C, A \cup C\}$

with $A = \{\omega_1, \omega_3\}$, $B = \omega_2$, $C = \{\omega_4, \omega_5\}$.

The atoms of $\mathcal{F}^Y$ are A, B, C.

We calculate $E\{X|Y\}$.

For $\omega \in B = \{\omega_1, \omega_3\}$:

$$E\{X|Y\} = \frac{1}{P(B)} \int_B X \, dP = \frac{7}{3}$$

and $\omega \in B^c = \{\omega_2, \omega_4\}$:

$$E\{X|Y\} = \frac{1}{P(B^c)} \int_{B^c} X \, dP = \frac{7}{3}$$

Therefore, $E\{X|Y\} = \frac{7}{3} \; \forall \omega$. Moreover,

$$E\{X\} =$$

$$= P\{\omega_1, \omega_2\} \cdot 1 + P\{\omega_3, \omega_4\} \cdot 3$$

$$= \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 3 = \frac{7}{3} = E\{X|Y\}$$

It is straightforward to see that X and Y are independent since $\mathcal{F}^X$ and $\mathcal{F}^Y$ are

Therefore, in this sense the conditional expectation $E\{X|Y\}$ gives evidence of the correlation between X and Y. A fact which further support this interpretation is:

FACT. If X and Y are independent random variables:

$$E\{X|Y\}(\omega) = X(\omega) \quad a.e.$$

It is possible also to define a CONDITIONAL PROBABILITY measure CONDITIONAL PROBABILITY and a CONDITIONAL PROBABILITY DENSITY.

DEFINITION. Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a σ-algebra $\mathcal{G} \subset \mathcal{F}$, we define CONDITIONAL PROBABILITY OF $A \in \mathcal{F}$, GIVEN $\mathcal{G}$, as the random variable

$$\mathbb{P}(A|\mathcal{G}) \triangleq E\{X_A|\mathcal{G}\}.$$

Similarly, we define the CONDITIONAL PROBABILITY OF $A \in \mathcal{F}$, given the random variable Y, as

$$\boxed{\begin{aligned} \mathcal{P}(A \mid Y) &= \mathcal{P}(A \mid \mathcal{F}^Y) \\ &\triangleq E\{\chi_A \mid \mathcal{F}^Y\} \end{aligned}}$$

EXAMPLE. Consider $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $\mathcal{F} = \mathcal{F}_M = \{$ all subset of $\Omega\}$ with

$$\mathcal{P}\{\omega_1\} = \frac{1}{9}, \; \mathcal{P}\{\omega_2\} = \frac{2}{9}, \; \mathcal{P}\{\omega_3\} = \frac{2}{9}, \; \mathcal{P}\{\omega_4\} = \frac{4}{9}$$

Moreover,

$$X(\omega) = \begin{cases} 1 & \omega \in \{\omega_1, \omega_2\} \\ 3 & \omega \in \{\omega_3, \omega_4\} \end{cases}$$

$$Y(\omega) = \begin{cases} 1 & \omega \in \{\omega_1, \omega_3\} \\ 2 & \omega \in \{\omega_2, \omega_4\} \end{cases}$$

As usual,

$$\mathcal{F}^X = \{\phi, A, A^c, \Omega\}, \quad \mathcal{F}^Y = \{\phi, B, B^c, \Omega\}$$

where $A = \{\omega_1, \omega_2\}$ and $B = \{\omega_1, \omega_3\}$.

12. CONDITIONAL PROBABILITY DENSITY AND BAYES FORMULAS

We have introduced in the previous section the notion of conditional expectation of X given Y: this notion well-characterizes to what extent the a posteriori knowledge of the results of Y, improves our a priori knowledge of the results of X. As we did for expectation of X, we want to characterise a conditional probability density: to evaluate conditional expectations:

EXAMPLE $\Omega = \{\omega_1, .., \omega_n\}$,

$\mathcal{P}(\omega_i) = \frac{1}{n}$ $\forall \omega_i \in \Omega$, $\mathcal{F} = \mathcal{F}_M$.

Consider $A, B \in \mathcal{F}$ with cardinality α and β, respectively, such that $A \cap B$ contains γ points ω_i. Therefore,

$\mathcal{P}(A) = \frac{\alpha}{n}$, $\mathcal{P}(B) = \frac{\beta}{n}$, $\mathcal{P}(A \cap B) = \frac{\gamma}{n}$

If we know the result of our experiment is in B, what is the probability that the result be also in A? Clearly, we may naturally consider B (instead of Ω) as the new space of results with some probability

$$\mathcal{P}(\omega_i | B) = \frac{1}{\beta}$$

for each $\omega_i \in B$. We have on A :

$$P(A|B) = \frac{\gamma}{\beta}$$

$$= \frac{\gamma/n}{\beta/n} = \frac{P(A \cap B)}{P(B)}$$

DEFINITION. We define CONDITIONAL PROBABILITY of A given B, DENOTED BY $P(A|B)$:

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

with $P(B) \neq 0$.

If A and B are independent:

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

i.e. the a posteriori knowledge that the event B happened does not improve our a priori knowledge on the possible happening of A.

From our definition of $\mathcal{P}(A|B)$ it is possible to define a conditional expectation of a random variable X on $(\Omega, \mathcal{F}, \mathcal{P})$ given that $B \in \mathcal{F}$ happened:

$$\boxed{E\{X(\omega)|B\} \triangleq \frac{E\{X(\omega)\chi_B(\omega)\}}{\mathcal{P}(B)}, \quad \mathcal{P}(B) \neq 0.}$$

Notice if $X = \chi_A$

$$E\{\chi_A | B\} = \frac{E\{\chi_A \chi_B\}}{\mathcal{P}(B)} = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}$$

$$= \mathcal{P}(A|B)$$

Straightforwardly, we can extend our definition to the expectation of X given the values of another random variable Y in $B \in \mathcal{F}$:

$$E\{X(\omega) \mid Y \in B\} = \frac{E\{X\chi_{Y^{-1}(B)}(\omega)\}}{P(Y^{-1}(B))}$$

$$= \int_{\Omega} \frac{X(\omega)\chi_{Y^{-1}(B)}(\omega)}{P(Y^{-1}(B))}\, dP(\omega)$$

$$= \frac{\int_{Y^{-1}(B)} X(\omega)\, dP(\omega)}{P(Y^{-1}(B))} = \int_{Y^{-1}(B)} E\{X(\omega)|Y\}(\omega)\, dP(\omega|B)$$

where $P(\mathcal{D}|B) = \dfrac{P(\mathcal{D} \cap Y^{-1}(B))}{P(Y^{-1}(B))}$

if $P(Y^{-1}(B)) \neq 0$. This coincides with our previous definition of $E\{X(\omega)|Y\}$ (resp. $P\{A|Y\}$) (see previous section) in the sense that:

$$\boxed{\begin{array}{l} E\{X(\omega) \mid Y \in B\} = E\{X(\omega) \mid Y\}(\omega) \\ P\{A \mid B\} = P\{A \mid Y\}(\omega) \\ \forall \omega \in Y^{-1}(B), \quad A \in \mathcal{F}. \end{array}}$$

Notice that the above definitions depend on the fact that $\mathcal{P}(B) \neq 0$ or $\mathcal{P}(Y^{-1}(B)) \neq 0$. To circumvent this problem and obtain a general definition of conditional probability density we resort to a fundamental result of Bayes: given two random variables X and Y exists a unique (a.e.) function $h(x,y)$ such that

$$\boxed{E\{X(\omega) \mid Y\}_{Y(\omega)=y} = \int_{\mathbb{R}} x\, h(x,y)\, dx}$$

It is quite natural to take $h(x,y)$ as conditional probability density of X given Y.

From this it is possible to prove:

BAYES THEOREM. Given random variables X, Y and its joint probability density $p_{X,Y}(x,y)$, the conditional probability density $p_{X|Y}(x,y)$ of X given Y and, respectively, $p_{Y|X}(y,x)$ of Y given X satisfy:

$$p_{X|Y}(x,y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

$$p_{Y|X}(y,x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

Notice that

$$p_{X|Y}(x,y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_{Y|X}(y,x)\, p_X(x)}{\int_{\mathbb{R}} p_{Y|X}(y,x)\, p_X(x)\, dx}$$

using the formulas for marginal densities.

Conditional density can be used to evaluate conditional expectations given a posteriori values of the conditioning random variable:

$$E\{X \mid Y\}_{Y=y} = \int_{\mathbb{R}} x\, p_{X|Y}(x,y)\, dx$$

$$E\{Y \mid X\}_{X=x} = \int_{\mathbb{R}} y\, p_{Y|X}(y,x)\, dy$$

12. CONDITIONING AND ORTHOGONAL PROJECTIONS

Given $Y \in \mathcal{L}_2(\Omega, \mathcal{F}, P)$, consider the set of $\mathcal{F}^Y$-measurable random variables, i.e. the set of functions $Z = f(Y)$ for some $\mathcal{B}(\mathbb{R})$-measurable function $f: \mathbb{R} \to \mathbb{R}$. This is a linear space (which we denote by $\mathcal{M}^Y$).

We want to see that the orthogonal projection of a random variable X on $\mathcal{M}^Y$ (denoted $\Pi(X|\mathcal{M}^Y)$) is exactly $E\{X|\mathcal{F}^Y\}$. Since

$$X = X_{/\!/} + X_{\perp}$$

where $X_{/\!/}, X_{\perp}$ are the unique random variables such that $X_{/\!/} \in \mathcal{M}^Y$ and $X_{\perp} \in (\mathcal{M}^Y)^{\perp}$ (the orthogonal of $\mathcal{M}^Y$), it is sufficient to prove that

i) $X - E\{X|\mathcal{F}^Y\} \in (\mathcal{M}^Y)^\perp$

ii) $E\{X|\mathcal{F}^Y\} \in \mathcal{M}^Y$.

ii) is true because $E\{X|\mathcal{F}^Y\}$ is $\mathcal{F}^Y$-measurable.

We prove i). To prove this, we prove

$$<X - E\{X|\mathcal{F}^Y\}, Z> = 0 \quad \forall\, Z \in \mathcal{M}^Y.$$

Indeed if $Z \in \mathcal{M}^Y$:

$$<X - E\{X|\mathcal{F}^Y\}, Z> = E\{(X - E\{X|\mathcal{F}^Y\})Z\}$$

$$= E\{XZ - E\{XZ|\mathcal{F}^Y\}\} =$$

$$= E\{XZ\} - E\{E\{XZ|\mathcal{F}^Y\}\} =$$

$$= E\{XZ\} - E\{XZ\} = 0$$

The orthogonal projection has the following important property.

PROJECTION THEOREM. Let $\mathcal{H}$ a linear space with a scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, $\mathcal{M} \subset \mathcal{H}$ a closed subspace of $\mathcal{H}$. $\forall v \in \mathcal{H}$ $\exists!$ $m_0 \in \mathcal{M}$:

$$(*) \quad \| v - m_0 \|_{\mathcal{H}} \leq \| v - m \|_{\mathcal{H}}, \ \forall m \in \mathcal{M},$$

where $\| \cdot \|_{\mathcal{H}} = \langle \cdot, \cdot \rangle$. Necessary and sufficient condition for m_0 being the unique element of $\mathcal{M}$ satisfying $(*)$ is

$$\langle v - m_0, m \rangle_{\mathcal{H}} = 0 \quad \forall m \in \mathcal{M}.$$

In other words, $\forall v \in \mathcal{H}$

$$\arg\min_{m \in \mathcal{M}} \| v - m \| = \Pi(v | \mathcal{M})$$