# 1 Conditional Expectation

## 1.1 Conditional distribution

Consider two random variables $X$ and $Y$ with joint mass function or joint density function denoted by $f_{X,Y}$, i.e.,

$$f_{X,Y}(x,y) = \begin{cases} \Pr(X = x, Y = y), & X \text{ and } Y \text{ are discrete at points } x \text{ and } y \text{ respectively} \\ \frac{\partial^2}{\partial x \partial y} \Pr(X \leq x, Y \leq y), & X \text{ and } Y \text{ are continuous at points } x \text{ and } y \text{ respectively} \end{cases}$$

We define the following concepts.

- the **marginal mass or density function of** $X$

$$f_X(x) = \sum_y f_{X,Y}(x,y) \qquad \text{or} \qquad f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, \mathrm{d}y.$$

- the **marginal mass or density function of** $Y$

$$f_Y(y) = \sum_x f_{X,Y}(x,y) \qquad \text{or} \qquad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, \mathrm{d}x.$$

- the **conditional mass or density function of** $X$ **given** $Y = y$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad \text{provided } f_Y(y) > 0. \tag{1}$$

Using the conditional distribution of $X$ given $Y$, the marginal mass or density function of $X$ can be expressed as

$$f_X(x) = \int_{\mathbb{R}} f_{X|Y}(x|y)\, f_Y(y)\, \mathrm{d}y \quad \text{or} \quad f_X(x) = \sum_{y \in \mathbb{R}} f_{X|Y}(x|y)\, f_Y(y) \tag{2}$$

**Proposition 1.** *If the random variables $X$ and $Y$ are independent, we have*

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y).$$

*As an immediate consequence, we have*

$$f_{X|Y}(x|y) = \frac{f_X(x)\, f_Y(y)}{f_Y(y)} = f_X(x).$$

## 1.2 Conditional expectation w.r.t. random variables

Throughout this section, we assume that $X$ given $Y = y$ is a continuous random variable with density function $f_{X|Y}(\cdot|y)$ (if $X|Y$ is discrete, replace all the integral signs by summation signs). The conditional expectation of $X$ given $Y = y$ is given by the expected value with respect to the conditional density function

$$\mathbb{E}[X\,|Y = y] = \int_{\mathbb{R}} x f_{X|Y}(x|y)\, \mathrm{d}x.$$

This motivates the following definition:

**Definition 1.** The **conditional expectation of** $X$ **given** $Y$ is the random variable

$$\mathbb{E}[X|Y] = \int_{\mathbb{R}} x f_{X|Y}(x|Y)\, \mathrm{d}x.$$

**Remark 1.** The conditional expectation is a random variable since it takes elements in the range of $Y$ and assigns it to a number. In other words, if we define the function $g$ through

$$g(y) = \mathbb{E}\left[X \,|Y = y\right] = \int_{\mathbb{R}} x f_{X|Y}\left(x|y\right) \mathrm{d}x,$$

then

$$\mathbb{E}\left[X|Y\right] = g(Y).$$

We can interpret the conditional expected value as the "best" estimate for the value of $X$ given a realization of $Y$ (see Problem 1.5).

The conditional expectation obeys the following useful properties.

**Proposition 2.** *The conditional expectation has the following properties:*

1. *Law of total expectation:* $\mathbb{E}\left[\mathbb{E}\left[X|Y\right]\right] = \mathbb{E}\left[X\right]$
2. *Pulling out known factors: If $h$ is a function, then*

$$\mathbb{E}\left[h(Y)X|Y\right] = h(Y)\mathbb{E}\left[X|Y\right]$$

**Proof.** The properties follow directly from the definition

(a) We define $g(y) = \mathbb{E}\left[X|Y = y\right] = \int_{\mathbb{R}} x f_{X|Y}\left(x|y\right) \mathrm{d}x$. By the definition of the expected value,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[X|Y\right]\right] = \mathbb{E}\left[g(Y)\right] = \int_{\mathbb{R}} g(y) f_Y\left(y\right) \mathrm{d}y &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x f_{X|Y}\left(x|y\right) \mathrm{d}x\right) f_Y\left(y\right) \mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X|Y}\left(x|y\right) f_Y\left(y\right) \mathrm{d}x\, \mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}\left(x, y\right) \mathrm{d}x\, \mathrm{d}y \\
&= \int_{\mathbb{R}} x \left(\int_{\mathbb{R}} f_{X,Y}\left(x, y\right) \mathrm{d}y\right) \mathrm{d}x \\
&= \int_{\mathbb{R}} x f_X\left(x\right) \mathrm{d}x = \mathbb{E}\left[X\right].
\end{aligned}$$

(b) For any $y$ in the support of $Y$,

$$g(y) = \mathbb{E}\left[h(Y)X|Y = y\right] = \int_{\mathbb{R}} h(y) x f_{X|Y}\left(x|y\right) \mathrm{d}x = h(y) \int_{\mathbb{R}} x f_{X|Y}\left(x|y\right) \mathrm{d}x = h(y)\mathbb{E}\left[X|Y = y\right].$$

Therefore,

$$\mathbb{E}\left[h(Y)X|Y\right] = g(Y) = h(Y)\mathbb{E}\left[X|Y\right].$$

□

Likewise, one can define the conditional variance in the obvious way.

**Definition 2.** The **conditional variance of $X$ given $Y$** is defined as

$$\mathrm{Var}\left(X|Y\right) = \mathbb{E}\left[(X - \mathbb{E}[X|Y])^2|Y\right]$$

The conditional variance satisfies the following useful properties.

**Proposition 3.** *We have*

1. $\operatorname{Var}(X|Y) = \mathbb{E}\left[X^2 \,|\, Y\right] - (\mathbb{E}[X \,|\, Y])^2$
2. *Law of total variance:* $\operatorname{Var}(X) = \mathbb{E}[\operatorname{Var}(X|Y)] + \operatorname{Var}(\mathbb{E}[X|Y])$

**Proof.** (a) With $g(Y) = \mathbb{E}[X|Y]$ we have from Proposition 2 (b) that

$$\begin{aligned}
\operatorname{Var}(X|Y) &= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X|Y] + (\mathbb{E}[X|Y])^2 \,\middle|\, Y\right] \\
&= \mathbb{E}\left[X^2 \,|\, Y\right] - 2\mathbb{E}\left[X\mathbb{E}[X|Y] \,|\, Y\right] + \mathbb{E}\left[(\mathbb{E}[X|Y])^2 \middle| Y\right] \\
&= \mathbb{E}\left[X^2 \,|\, Y\right] - 2\mathbb{E}\left[Xg(Y) \,|\, Y\right] + \mathbb{E}\left[(g(Y))^2 \middle| Y\right] \\
&= \mathbb{E}\left[X^2 \,|\, Y\right] - 2g(Y) \cdot \mathbb{E}[X \,|\, Y] + (g(Y))^2\mathbb{E}[1|Y] \qquad \text{(by Proposition 2 (b))} \\
&= \mathbb{E}\left[X^2 \,|\, Y\right] - 2\mathbb{E}[X \,|\, Y] \cdot \mathbb{E}[X \,|\, Y] + (\mathbb{E}[X|Y])^2 \\
&= \mathbb{E}\left[X^2 \,|\, Y\right] - (\mathbb{E}[X \,|\, Y])^2
\end{aligned}$$

(b) It follows from (a) and Proposition 2 (a) that

$$\begin{aligned}
\mathbb{E}[\operatorname{Var}(X|Y)] &= \mathbb{E}\left[\mathbb{E}\left[X^2|Y\right]\right] - \mathbb{E}\left[(\mathbb{E}[X|Y])^2\right] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[(\mathbb{E}[X|Y])^2\right].
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\operatorname{Var}(\mathbb{E}[X|Y]) &= \mathbb{E}\left[(\mathbb{E}[X|Y])^2\right] - (\mathbb{E}[\mathbb{E}[X|Y]])^2 \\
&= \mathbb{E}\left[(\mathbb{E}[X|Y])^2\right] - (\mathbb{E}[X])^2.
\end{aligned}$$

Combining the preceding two relations implies

$$\mathbb{E}[\operatorname{Var}(X|Y)] + \operatorname{Var}(\mathbb{E}[X|Y]) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = \operatorname{Var}(X).$$

□

## 1.3 Example Problems

**Problem 1.1.** Suppose a fair coin is tossed 3 times. Define the random variables $X$ = "number of Heads", and

$$Y = \begin{cases} 1 & \text{Head occurs on the first toss,} \\ 0 & \text{Tail occurs on the first toss.} \end{cases}$$

1. Find the joint PMF for $(X, Y)$.
2. Are $X$ and $Y$ independent?
3. What is the conditional distribution of $X$ given $Y$?
4. What is the probability that $X + Y = 2$?

**Solution 1.1.**

**Part 1:** We can compute all the probabilities one by one and encode the joint PMF of $X$ and $Y$ in the table

| $f_{X,Y}(x,y)$ | $x$: 0 | 1 | 2 | 3 | $f_Y(y)$ |
|---|---|---|---|---|---|
| $y$ 0 | 1/8 | 2/8 | 1/8 | 0 | 1/2 |
| 1 | 0 | 1/8 | 2/8 | 1/8 | 1/2 |
| $f_X(x)$ | 1/8 | 3/8 | 3/8 | 1/8 | 1 |

**Part 2:** We can see

$$f_{X,Y}(0,1) = 0 \neq \frac{1}{8} \cdot \frac{1}{2} = f_X(0) f_Y(1)$$

which implies that $X$ and $Y$ are not independent (which makes perfect sense, as the number of heads we have should depend on whether we had heads in the first toss).

**Part 3:** Using the formula $f_{X \mid Y}(x \mid y) = f_{X,Y}(x,y)/f_Y(y)$ we find

| | $x$: 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f_{X \mid Y}(x \mid y=0)$ | 2/8 | 4/8 | 2/8 | 0 |
| $f_{X \mid Y}(x \mid y=1)$ | 0 | 2/8 | 4/8 | 2/8 |

**Part 4:** We have $X + Y = 2$ if and only if $X = 2, Y = 0$ or $X = 1, Y = 1$. We can sum these terms up in the joint PMF

$$\mathbb{P}(X+Y=2) = f(2,0) + f(1,1) + f(0,2) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

**Problem 1.2.** Suppose that $X$ and $\Theta$ are two random variables such that $X$ given $\Theta = \theta$ is Poisson distributed with mean $\theta$, i.e.,

$$f_{X|\Theta}(k|\theta) = e^{-\theta}\frac{\theta^k}{k!}, \quad k = 0, 1, 2, \ldots$$

and $\Theta$ is Gamma distributed with parameters $\alpha, \beta > 0$. That is, $\Theta$ has the density function

$$f_\Theta(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)}, \quad \theta > 0,$$

where $\Gamma$ denotes the Gamma function,

$$\Gamma(\alpha) = \int_0^\infty \theta^{\alpha-1} e^{-\theta}\, d\theta.$$

Compute the marginal mass function of $X$.

**Solution 1.2.** The marginal mass function of $X$ is given by

$$\begin{aligned}
\mathbb{P}(X=k) &= \int_0^\infty f_{X|\Theta}(k|\theta) f_\Theta(\theta)\,\mathrm{d}\theta \\
&= \int_0^\infty \frac{\theta^k e^{-\theta}}{k!} \cdot \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)}\,\mathrm{d}\theta \\
&= \frac{\beta^\alpha}{k!\Gamma(\alpha)} \int_0^\infty \theta^{k+\alpha-1} e^{-(\beta+1)\theta}\,\mathrm{d}\theta \\
&= \frac{\beta^\alpha}{k!\Gamma(\alpha)} \cdot \frac{1}{(\beta+1)^{k+\alpha}} \int_0^\infty x^{k+\alpha-1} e^{-x}\,\mathrm{d}x \\
&= \frac{1}{k!\Gamma(\alpha)} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^k \Gamma(k+\alpha) \\
&= \frac{(k+\alpha-1)(k+\alpha-2)\cdots(\alpha+1)\alpha}{k!} \left(1-\frac{1}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^k \\
&= \binom{k+\alpha-1}{k} \left(1-\frac{1}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^k.
\end{aligned}$$

Therefore, $X$ follows a negative binomial distribution with parameters $\alpha$ and $\frac{1}{\beta+1}$.

**Problem 1.3.** Suppose that $X$ given $\Theta = \theta$ is Poisson distributed with mean $\theta$ and $\Theta$ is Gamma distributed with density function

$$f_\Theta(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)}, \quad \theta > 0.$$

1. Compute $\mathbb{E}[X]$.
2. Compute $\mathrm{Var}[X]$.

**Solution 1.3.**

(a) Using the law of total expectation,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\Theta]] = \mathbb{E}[\Theta] = \frac{\alpha}{\beta}.$$

(b) By the law of total variance

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[\mathrm{Var}(X|\Theta)] + \mathrm{Var}(\mathbb{E}[X|\Theta]) \\
&= \mathbb{E}[\Theta] + \mathrm{Var}(\Theta) \\
&= \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} = \frac{\alpha(\beta+1)}{\beta^2}.
\end{aligned}$$

**Problem 1.4.** Suppose that

$$X = \begin{cases} \sum_{i=1}^N Y_i, & \text{if } N > 0, \\ 0, & \text{if } N = 0, \end{cases}$$

where $N$ is Poisson distributed with mean $\lambda$ and $Y_1, Y_2, \ldots$ is a sequence of iid random variables with mean $\mu$ and variance $\sigma^2$ that is independent of $N$. We say that $X$ is a **compound Poisson random variable**.

1. Compute $\mathbb{E}[X]$.
2. Compute $\mathrm{Var}[X]$.

**Solution 1.4.**

(a) By the law of total expectation

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{E}\left[X|N\right]\right] = \mathbb{E}\left[N\mu\right] = \lambda\mu,$$

(b) By the law of total variance

$$\begin{aligned}\mathrm{Var}\left(X\right) &= \mathbb{E}\left[\mathrm{Var}\left(X|N\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[X|N\right]\right)\\ &= \mathbb{E}\left[N\sigma^2\right] + \mathrm{Var}\left(N\mu\right)\\ &= \sigma^2\mathbb{E}\left[N\right] + \mu^2\mathrm{Var}\left(N\right)\\ &= \lambda\left(\sigma^2+\mu^2\right).\end{aligned}$$

**Problem 1.5.** For any measurable function $f$, show that

$$\mathbb{E}[(X - f(Y))^2] \geq \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2].$$

In particular, the conditional expectation minimizes the mean squared error.

**Solution 1.5.** This proof follows directly from the properties of conditional expected value. By adding and subtracting $\mathbb{E}[X \mid Y]$, we see that

$$\begin{aligned}\mathbb{E}[(X - f(Y))^2] &= \mathbb{E}[(X - \mathbb{E}[X \mid Y] + \mathbb{E}[X \mid Y] - f(Y))^2]\\ &= \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2] + \mathbb{E}[(\mathbb{E}[X \mid Y] - f(Y))^2] + 2\,\mathbb{E}[(X - \mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y] - f(Y))]\end{aligned}$$

Apply the law of total expectation and using the fact that $\mathbb{E}[X \mid Y]$ and $f(Y)$ are measurable functions of $Y$, we see that the cross terms vanish

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y] - f(Y))] &= \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X \mid Y])(\mathbb{E}[X \mid Y] - f(Y))] \mid Y]\\ &= \mathbb{E}[(\mathbb{E}[X \mid Y] - f(Y))\,\mathbb{E}[(X - \mathbb{E}[X \mid Y])] \mid Y]\\ &= \mathbb{E}[(\mathbb{E}[X \mid Y] - f(Y))(\mathbb{E}[X \mid Y] - \mathbb{E}[X \mid Y])]\\ &= 0.\end{aligned}$$

Since $\mathbb{E}[(\mathbb{E}[X \mid Y] - f(Y))^2] \geq 0$, we conclude that

$$\mathbb{E}[(X - f(Y))^2] = \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2] + \mathbb{E}[(\mathbb{E}[X \mid Y] - f(Y))^2] \geq \mathbb{E}[(X - \mathbb{E}[X \mid Y])^2]$$

as required.

# 2 Conditional expectations w.r.t. $\sigma$-fields

We now introduce general definition of conditional expectation that will allow us to condition on more general forms of (random) information. We will use $\sigma$-algebra $\mathscr{F}_0 \subset \mathscr{F}$ as a **model of information** and define the general notation of the conditional expectation of $X$ given information $\mathcal{F}_0$

$$\mathbb{E}[X|\mathscr{F}_0].$$

A $\sigma$-algebra is a natural model for the information because it contains both the negation and union of outcomes, which can easily deduced from existing information.

## 2.1 Constructing $\sigma$-algebras

We first take a closer look at possible constructions of $\sigma$-algebras.

**Definition 3.** Given a collection of sets $\mathcal{A}$ of $\Omega$, the $\sigma$**-algebra** generated by the collection of sets $\mathcal{A}$ is the smallest $\sigma$-algebra containing $\mathcal{A}$ and is often denoted by $\sigma(\mathcal{A})$.

**Example 1.** On $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ consider the following two partitions:

$$\mathscr{P}_1 = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}\}$$
$$\mathscr{P}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}.$$

In the first, we are able to distinguish between all elements of $\Omega$. In the second, we cannot distinguish between $\omega_1$ and $\omega_2$ and between $\omega_3$ and $\omega_4$. Thus, $\mathscr{P}_1$ is *finer* than $\mathscr{P}_2$. The $\sigma$-algebra $\sigma(\mathscr{P}_1)$ is equal to the power set of $\Omega$, i.e., it contains all subsets of $\Omega$. On the other hand,

$$\sigma(\mathscr{P}_2) = \big\{\emptyset, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\big\}.$$

### 2.1.1 $\sigma$-algebras generated by random variables

Suppose the $\sigma$-algebra $\mathscr{F}_0$ corresponds to the information from observing the values of a collection $Y_1, \dots, Y_n$ of $\mathscr{F}$-measurable random variables. Informally, $\mathscr{F}_0$ then consists of all events that can be described through the random variables $Y_1, \dots, Y_n$.

**Definition 4.** The $\sigma$-algebra $\mathscr{F}_0$ generated by $Y_1, \dots, Y_n$ is the $\sigma$-algebra generated by events of the form $\{Y_i \le x\}$ for all $x \in \mathbb{R}$ and $i = 1, \dots, n$. We write

$$\mathscr{F}_0 := \sigma(Y_1, \dots, Y_n).$$

**Remark 2.** Let $X$ be a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$. One can prove that the $\sigma$-algebra $\sigma(X)$ generated by $X$ is equivalent to

$$\sigma(X) = \big\{X^{-1}(B) \,|\, B \in \mathscr{B}(\mathbb{R})\big\},$$

where we recall that $\mathscr{B}(\mathbb{R})$ is the Borel $\sigma$-algebra on $\mathbb{R}$ and

$$X^{-1}(B) = \{\omega \in \Omega \,|\, X(\omega) \in B\} = \{X \in B\}$$

is the pre-image of $B$.

**Example 2.** Let $X$ be the number of heads obtained for a coin tossed twice. In this case, $\Omega = \{HH, HT, TH, TT\}$. Clearly, $X(HH) = 2$, $X(HT) = X(TH) = 1$ and $X(TT) = 0$. We have

$$\sigma(X) = \{\emptyset, \{HH\}, \{TT\}, \{TT, HH\}, \{HT, TH\}, \{HT, TH, HH\}, \{HT, TH, TT\}, \Omega\}.$$

Notice that this set is not equal to the power set of $\Omega$. In particular, the set $\{HT\}$ is not in $\sigma(X)$ since knowing the number of heads does not allow you to determine that $\{HT\}$ happened since it is indistinguishable from the event $\{TH\}$, while $\{HT, TH\}$ is in the set, since the events you flipped $HT$ or $TH$ corresponds to the event of flipping exactly 1 heads.

## 2.2 Independent $\sigma$-algebras

Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Recall that two **events** $A, B \in \mathscr{F}$ are called **independent** under $\mathbb{P}$ if

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

The notion of independence can be extended to $\sigma$-algebras in the obvious way.

**Definition 5.** Two $\sigma$-algebras $\mathscr{F}_1, \mathscr{F}_2 \subset \mathscr{F}$ are **independent** if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2), \quad \text{for any } A_1 \in \mathscr{F}_1 \text{ and } A_2 \in \mathscr{F}_2.$$

The notation of independence of random variables can also be stated with respect to $\sigma$-algebras.

**Definition 6.** Two **random variables** $X_1$ and $X_2$ on $(\Omega, \mathscr{F}, \mathbb{P})$ are **independent** if $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

**Remark 3.** This notion of independence is equivalent to the earlier notation defined in Week 1. That is the following statements are equivalent

1. $X_1$ and $X_2$ are independent,
2. The probabilities satisfy
$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2),$$
for any $B_1, B_2 \in \mathscr{B}(\mathbb{R})$.
3. The CDFs satisfy
$$\mathbb{P}(X_1 \le x_1, X_2 \le x_2) = \mathbb{P}(X_1 \le x_1)\mathbb{P}(X_2 \le x_2) = F_{X_1}(x_1)F_{X_2}(x_2) \qquad \forall x_1, x_2$$

The independence between a random variable and $\sigma$-algebra is also defined in the natural way.

**Definition 7.** A random variable $X$ is independent of a $\sigma$-algebra $\mathscr{F}_1 \subset \mathscr{F}$ if $\sigma(X)$ and $\mathscr{F}_1$ are independent.

## 2.3 Conditional expectations with respect to general $\sigma$-fields

**Definition 8.** Consider a random variable $X$ on $(\Omega, \mathscr{F}, \mathbb{P})$ and a $\sigma$-field $\mathscr{F}_0 \subset \mathscr{F}$. We define the **conditional expectation** of $X$ given $\mathscr{F}_0$ as a random variable $\mathbb{E}[X|\mathscr{F}_0]$ satisfying the following two conditions:

1. $\mathbb{E}[X|\mathscr{F}_0]$ is a $\mathscr{F}_0$-measurable random variable.
2. $\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}\big[\mathbb{1}_A \mathbb{E}[X|\mathscr{F}_0]\big]$ for any $A \in \mathscr{F}_0$.

The first condition is natural because we want to be able to define the conditional expectation with respect to the outcome of a random events: your best guess for a random variable should be able to adapt to a random event in $\mathcal{F}_0$. The second condition can be seen as a consistency condition: given that $A \subset \mathcal{F}_0$ occurred, then the average of $X$ given that $A$ happened must be equal to the average of $X$ restricted to the set $A$.

**Example 3.** One can show that the preceding definition gives the following special cases:

- Consider the case $\mathscr{F}_0 = \sigma(Y)$. In general, a random variable $Z$ is $\mathscr{F}_0$-measurable if and only if there is a function $h$ such that
$$Z = h(Y_1, \ldots, Y_n).$$
The conditional expectation is the function given by
$$\mathbb{E}[X|\mathscr{F}_0] = \mathbb{E}[X|Y]$$
where the right-hand side is the function of $Y$ defined in the same way as in Section 1.2.

- Consider the case $\mathscr{F}_0 = \sigma(Y_1, \ldots, Y_n)$. In general, a random variable $Z$ is $\mathcal{F}_0$ measurable if and only if there is a function $h$ such that
$$Z = h(Y_1, \ldots, Y_n).$$
The conditional expectation is the function given by
$$\mathbb{E}[X|\mathscr{F}_0] = \mathbb{E}[X|Y_1, \ldots, Y_n] = g(Y_1, \ldots, Y_n).$$
where the function $g$ can be defined in the same way as in Section 1.2. We denote by $f_{Y_1,\ldots,Y_n}$ the joint probability density (or probability mass function) of $Y_1, \ldots, Y_n$ and define
$$f_{X|Y_1,\ldots,Y_n}(x|y_1, \ldots, y_n) := \frac{f_{X,Y_1,\ldots,Y_n}(x, y_1, \ldots, y_n)}{f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)},$$
where $f_{X,Y_1,\ldots,Y_n}$ is the joint density of $X, Y_1, \ldots, Y_n$. Then we let
$$g(y_1, \ldots, y_n) = \int_{\mathbb{R}} x f_{X|Y_1,\ldots,Y_n}(x|y_1, \ldots, y_n) \mathrm{d}x.$$

- Let $\mathscr{P} = \{A_1, A_2, \ldots\}$ be a partition of $\Omega$ and let $\mathscr{F}_0 = \sigma(\mathscr{P})$. In general, a random variable $Z$ is $\mathscr{F}_0$-measurable if and only if $Z$ is of the form
$$Z = \sum_{i=1}^{\infty} z_i \mathbb{1}_{A_i}$$
for some real numbers $z_1, z_2, \ldots$. The conditional expectation is the function given by
$$\mathbb{E}[X|\mathscr{F}_0] = \sum_{i=1}^{\infty} \mathbb{E}[X|A_i] \mathbb{1}_{A_i}$$
where the coefficients are given by the **(elementary) conditional expectation**
$$\mathbb{E}[X|A_i] = \frac{\mathbb{E}[X \mathbb{1}_{A_i}]}{\mathbb{P}(A_i)}$$
whenever $\mathbb{P}(A_i) > 0$ and 0 if $\mathbb{P}[A_i] = 0$

The following proposition lists many useful propositions of the conditional expectation.

**Proposition 4.** *For a random variable $X$ on $(\Omega, \mathscr{F}, \mathbb{P})$ and a $\sigma$-field $\mathscr{F}_0 \subset \mathscr{F}$:*

1. *If $X$ is $\mathscr{F}_0$-measurable, then $\mathbb{E}[X|\mathscr{F}_0] = X$*
2. *If $\mathscr{G}$ is the trivial $\sigma$-field, i.e., $\mathscr{G} = \{\emptyset, \Omega\}$, then*
$$\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$$

3. ***Law of total expectation:*** $\mathbb{E}\left[\mathbb{E}[X|\mathscr{F}_0]\right] = \mathbb{E}[X]$
4. ***Linearity:*** $\mathbb{E}[aX + bY|\mathscr{F}_0] = a\mathbb{E}[X|\mathscr{F}_0] + b\mathbb{E}[Y|\mathscr{F}_0]$
5. ***Taking out known factors:*** *If $Y$ is $\mathscr{F}_0$-measurable, then*
$$\mathbb{E}[XY|\mathscr{F}_0] = Y\mathbb{E}[X|\mathscr{F}_0]$$
6. ***Tower property:*** *If $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}$ are $\sigma$-fields, then*
$$\mathbb{E}[X|\mathscr{F}_0] = \mathbb{E}\left[\mathbb{E}[X|\mathscr{F}_1]\right| \mathscr{F}_0]$$
7. ***Jensen's inequality:*** *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex, then*
$$\phi\left(\mathbb{E}[X|\mathscr{F}_0]\right) \leq \mathbb{E}[\phi\left(X\right)|\mathscr{F}_0]$$
8. ***Independence:*** *If $X$ is independent of $\mathscr{F}_0$, then*
$$\mathbb{E}[X|\mathscr{F}_0] = \mathbb{E}[X].$$

## 2.4 Example Problems

**Problem 2.1.** Prove all the statements in Proposition 4