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

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

We can interpret the conditional expected value as the "best" estimate for the value of X given a realization of Y. This motivates the following definition:

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

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

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

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

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

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

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

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

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

On the other hand,

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

Combining the preceding two relations implies

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

□

1.3 Example Problems

Problem 1.1. Suppose that X and Θ are two random variables such that X given $\Theta = \theta$ is Poisson distributed with mean θ, i.e.,

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

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

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

where Γ denotes the Gamma function,

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

Compute the marginal mass function of X.

Solution 1.1. 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 α and $\frac{1}{\beta+1}$.

Problem 1.2. Suppose that X given $\Theta = \theta$ is Poisson distributed with mean θ and Θ 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 1.2.

(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 1.3. 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 λ and $Y_1, Y_2, \ldots$ is a sequence of iid random variables with mean μ and variance σ^2 that is independent of N. We say that X is a **compound Poisson random variable**.

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

Solution 1.3.

(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

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