Probability and Stochastic Processes
Lecture 16: Conditional Expectation

Prof. Washim Uddin Mondal
Department of Electrical Engineering
Indian Institute of Technology Kanpur

Introduction

Definition

If the conditional PMF of X given $Y = y$ is $\mathbf{P}(X = \cdot | Y = y)$, then the conditional expectation of X given $Y = y$ is defined as:

$$\mathbf{E}[X|Y = y] = \sum_{x_i} x_i \mathbf{P}(X = x_i | Y = y)$$

where $\sum_{x_i}$ denotes the sum over the range of X. Alternatively, if the conditional PDF of X given $Y = y$ is $f_{X|Y}(x|y)$, then

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

Application

Example

Obtain $\mathbf{E}[X^3|Y=y], y>0$ if the joint PDF of X, Y is given as follows.

$$f_{X,Y}(x,y) = \frac{e^{-y}}{y}\mathbf{1}(0<x<y)$$

Solution

$$f_Y(y) = \int_0^y \frac{e^{-y}}{y}\mathrm{d}x = e^{-y} \implies f_{X|Y}(x|y) = \frac{1}{y}\mathbf{1}(0<x<y)$$

$$\implies \mathbf{E}[X^3|Y=y] = \int_0^y \frac{x^3}{y}\mathrm{d}x = \frac{y^3}{4}$$

Conditional Expectation as a Random Variable

Intuition

- Note that $\mathbf{E}[X|Y=y]$ is a function of y. Let $\mathbf{E}[X|Y=y] = g(y)$.
- We denote $\mathbf{E}[X|Y]$ as the random variable $g(Y)$, i.e., $\mathbf{E}[X|Y](\omega) = g(Y(\omega))$, $\forall \omega \in \Omega$.
- In particular, $\mathbf{E}[X|Y] : \Omega \to \mathbb{R}$ is such that $\mathbf{E}[X|Y](\omega) = \mathbf{E}[X|Y = Y(\omega)]$, $\forall \omega \in \Omega$

Example

Let $(\Omega, 2^\Omega, \mathbf{P})$ be a probability space where $\Omega = \{a, b, c\}$, and $\mathbf{P} : 2^\Omega \to [0, 1]$ is such that $\mathbf{P}(\{a\}) = \mathbf{P}(\{b\}) = 0.4$, and $\mathbf{P}(\{c\}) = 0.2$. If $X, Y : \Omega \to \mathbb{R}$ are such that $X(a) = 0$, $X(b) = X(c) = 1$, $Y(a) = 1$, $Y(b) = 2$, and $Y(c) = 3$, then obtain $\mathbf{E}[X|Y]$.

Conditional Expectation as a Random Variable

Solution

- Note the following expectations.

$$\mathbf{E}[X|Y](a) = \mathbf{E}[X|Y=1] = \frac{\mathbf{P}(X=1, Y=1)}{\mathbf{P}(Y=1)} = \frac{\mathbf{P}(\phi)}{\mathbf{P}(\{a\})} = 0$$

$$\mathbf{E}[X|Y](b) = \mathbf{E}[X|Y=2] = \frac{\mathbf{P}(X=1, Y=2)}{\mathbf{P}(Y=2)} = \frac{\mathbf{P}(\{b\})}{\mathbf{P}(\{b\})} = 1$$

$$\mathbf{E}[X|Y](c) = \mathbf{E}[X|Y=3] = \frac{\mathbf{P}(X=1, Y=3)}{\mathbf{P}(Y=3)} = \frac{\mathbf{P}(\{c\})}{\mathbf{P}(\{c\})} = 1$$

Law of Total Expectation

Proposition 16.1

Let X, Y be random variables such that $\mathbf{E}[X|Y]$ is a well-defined random variable. We have the following: $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$.

Proof of Proposition 16.1 (Continuous Case)

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

Law of Total Expectation

Example

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

Solution

- Let $Y \in \{1, 2, 3\}$ be the door chosen.
- Let X be the time (in hours) to reach safety.

Law of Total Expectation

Solution (Continued)

- We have the following.

$$\mathbf{E}[X|Y=1]=3,\ \mathbf{E}[X|Y=2]=5+\mathbf{E}[X],\ \mathbf{E}[X|Y=3]=7+\mathbf{E}[X]$$

- Applying the law of total expectation, we get:

$$\begin{aligned}\mathbf{E}[X] &= \sum_{y=1}^{3}\mathbf{E}[X|Y=y]\mathbf{P}(Y=y)\\ &= \frac{1}{3}(3)+\frac{1}{3}(\mathbf{E}[X]+5)+\frac{1}{3}(\mathbf{E}[X]+7)=5+\frac{2}{3}\mathbf{E}[X]\end{aligned}$$

- The desired expected length is $\mathbf{E}[X]=15$.

Law of Total Expectation

Example

Let (X_1, X_2) be a bivariate normal random vector whose joint PDF is given as follows.

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

where $(x_1, x_2) \in \mathbb{R}^2$ and $\rho \in (-1, 1)$. Find the correlation coefficient $\rho(X_1, X_2)$.

Solution

- Recall that $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Law of Total Expectation

Solution (Continued)

- Hence, $\mathbf{E}[X_1] = \mu_1$, $\mathrm{Var}(X_1) = \sigma_1^2$, $\mathbf{E}[X_2] = \mu_2$, and $\mathrm{Var}(X_2) = \sigma_2^2$.
- The conditional PDF of X_1 given $X_2 = x_2$ is $\mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$.
- Therefore, $\mathbf{E}[X_1|X_2] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)$. We have the following.

$$\begin{aligned}
\mathbf{E}[X_1X_2] = \mathbf{E}[\mathbf{E}[X_1X_2|X_2]] &= \mathbf{E}[\mu_1X_2 + \rho\frac{\sigma_1}{\sigma_2}(X_2^2 - \mu_2X_2)] \\
&= \mu_1\mu_2 + \rho\frac{\sigma_1}{\sigma_2}\left(\mathbf{E}[X_2^2] - \mu_2^2\right) = \mu_1\mu_2 + \rho\sigma_1\sigma_2
\end{aligned}$$

- The correlation coefficient is $\rho(X_1, X_2) = (\mathbf{E}[X_1X_2] - \mu_1\mu_2)/\sigma_1\sigma_2 = \rho$.

Law of Total Expectation

Example

Let $\{U_n\}_{n\in\mathbb{N}}$ be independent uniform $(0,1)$ random variables. Find $\mathbf{E}[N]$, where

$$N = \min\left\{n \in \mathbb{N} \middle| \sum_{i=1}^{n} U_i > 1\right\}$$

Solution

- Define $m(x) = \mathbf{E}[N(x)]$, where $x \in [0,1]$, and

$$N(x) = \min\left\{n \in \mathbb{N} \,\middle|\, \sum_{i=1}^{n} U_i > x\right\}$$

Law of Total Expectation

Solution (Continued)

- Note that $m(x) = \mathbf{E}[N(x)] = \int_0^1 \mathbf{E}[N(x)|U_1 = y]\mathrm{d}y$. The following holds $\forall y \in [0, 1]$.

$$\begin{aligned}
\mathbf{E}[N(x)|U_1 = y] &= \mathbf{E}\left[\min\left\{n \in \mathbb{N} \,\middle|\, y + \sum_{i=2}^{n} U_i > x\right\} \middle| U_1 = y\right] \\
&\overset{(a)}{=} \mathbf{E}\left[\min\left\{n \in \mathbb{N} \,\middle|\, \sum_{i=2}^{n} U_i > x - y\right\}\right] \\
&= \begin{cases} 1 + m(x-y) & \text{if } x \geq y \\ 1 & \text{if } x < y \end{cases}
\end{aligned}$$

Law of Total Expectation

Solution (Continued)

- Step (a) follows since U_i's are independent.
- Hence, $m(x) = 1 + \int_0^x m(x-y)\mathrm{d}y = 1 + \int_0^x m(z)\mathrm{d}z, \forall x \in [0,1]$.
- We get $m'(x) = m(x)$, which implies $m(x) = ke^x, \forall x \in [0,1]$.
- Since $m(0) = 1$, we get $k = 1$. We have $m(x) = e^x, \forall x \in [0,1]$.
- Our desired solution is $m(1) = \mathbf{E}[N] = e$.

Conditional Variance

Definition

- The conditional variance of X given $Y = y$ is defined as:

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

- We define $\mathrm{Var}(X|Y) = \mathbf{E}[(X - \mathbf{E}[X|Y])^2|Y]$. Note that

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

Conditional Variance

Proposition 16.2

$$\mathbf{E}\left[\mathrm{Var}(X|Y)\right] + \mathrm{Var}\left(\mathbf{E}[X|Y]\right) = \mathrm{Var}(X)$$

Proof of Proposition 16.2

- Note that $\mathbf{E}\left[\mathrm{Var}(X|Y)\right] = \mathbf{E}[X^2] - \mathbf{E}\left[\left(\mathrm{E}[X|Y]\right)^2\right]$.
- Moreover, $\mathrm{Var}(\mathbf{E}[X|Y]) = \mathbf{E}\left[\left(\mathbf{E}[X|Y]\right)^2\right] - \left(\mathbf{E}\left[\mathbf{E}[Y|X]\right]\right)^2 = \mathbf{E}\left[\left(\mathbf{E}[X|Y]\right)^2\right] - (\mathbf{E}[X])^2$, where the last step utilizes the law of total expectation.

Conditional Variance

Example

Suppose that by any time t, the number of people that have arrived at a train depot is a Poisson random variable with mean λt. If the initial train arrives at the depot at a time (independent of when the passengers arrive) that is uniformly distributed over $(0, T)$, what are the mean and variance of the number of passengers who enter the train?

Solution

- Let Y denote the time of arrival of the train.
- Let X be the number of passengers entering the train.
- The conditional PMF of X given $Y = t$ is $\mathrm{Poisson}(\lambda t)$.

Conditional Variance

Solution (Continued)

- We have $\mathbf{E}[X|Y=t] = \lambda t$ and $\mathrm{Var}(X|Y=t) = \lambda t$.
- Additionally, $\mathbf{E}[Y] = T/2$ and $\mathrm{Var}(Y) = T^2/12$.
- Clearly, $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[\lambda Y] = \lambda \mathbf{E}[Y] = \lambda T/2$.
- Also, $\mathrm{Var}(\mathrm{E}[X|Y]) = \mathrm{Var}(\lambda Y) = \lambda^2 \mathrm{Var}(Y) = \lambda^2 T^2/12$.
- Moreover, $\mathbf{E}\left[\mathrm{Var}(X|Y)\right] = \mathbf{E}[\lambda Y] = \lambda T/2$. Therefore,

$$\mathrm{Var}(X) = \frac{\lambda^2 T^2}{12} + \frac{\lambda T}{2}$$

Conditional Variance

Example

Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables with a common mean mean $\mathbf{E}[X]$ and a common variance $\mathrm{Var}(X)$, and $N \in \{1, 2, \cdots\}$ be a discrete random variable that is independent of the sequence $\{X_i\}_{i=1}^{\infty}$. Find (a) $\mathbf{E}[\sum_{i=1}^{N} X_i]$ and (b) $\mathrm{Var}(\sum_{i=1}^{N} X_i)$ in terms of $\mathbf{E}[X]$, $\mathrm{Var}(X)$, $\mathbf{E}[N]$, and $\mathrm{Var}(N)$.

Solution

- Note that $\mathbf{E}[\sum_{i=1}^{N} X_i | N = n] = \mathbf{E}\left[\sum_{i=1}^{n} X_i | N = n\right] \overset{(a)}{=} \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbf{E}[X]$. Relation (a) holds because N is independent of $\{X_i\}_{i=1}^{\infty}$.
- $\mathbf{E}[\sum_{i=1}^{N} X_i] = \mathbf{E}[\mathbf{E}[\sum_{i=1}^{N} X_i | N]] = \mathbf{E}[N\mathbf{E}[X]] = \mathbf{E}[N]\mathbf{E}[X]$.

Conditional Variance

Solution (Continued)

- Furthermore, $\mathrm{Var}\left(\mathbf{E}\left[\sum_{i=1}^{N} X_i | N\right]\right) = \mathrm{Var}\left(N\mathbf{E}[X]\right) = (E[X])^2 \mathrm{Var}(N)$. Also,

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i \middle| N = n\right) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i \middle| N = n\right) \stackrel{(a)}{=} \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \stackrel{(b)}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

- Step (a) holds because N is independent of $\{X_i\}_{i=1}^{\infty}$ and step (b) holds because X_i's are independent of each other.
- Finally, $\mathbf{E}[\mathrm{Var}(\sum_{i=1}^{N} X_i | N)] = \mathbf{E}\left[N \mathrm{Var}(X)\right] = \mathbf{E}[N]\mathrm{Var}(X)$. Hence,

$$\mathrm{Var}(X) = (\mathbf{E}[X])^2 \mathrm{Var}(N) + \mathbf{E}[N]\mathrm{Var}(X)$$

Conditional Variance

Alternative Solution

- The conditional MGF of $Y = \sum_{i=1}^{N} X_i$ given $N = n$ can be expressed as:

$$
\begin{aligned}
M_{Y|N}(t|n) &= \mathbf{E}\left[\exp\left(t\sum_{i=1}^{N} X_i\right) \middle| N = n\right] \\
&= \mathbf{E}\left[\exp\left(t\sum_{i=1}^{n} X_i\right) \middle| N = n\right] \overset{(a)}{=} \mathbf{E}\left[\exp\left(t\sum_{i=1}^{n} X_i\right)\right]
\end{aligned}
$$

where (a) holds because N is independent of $\{X_i\}_{i=1}^{\infty}$.

Conditional Variance

Alternative Solution (Continued)

- If $M_X(\cdot)$ is the common MGF of X_i's, then using the independence of X_i's, we get:

$$M_{Y|N}(t|n) = [M_X(t)]^n$$

- Hence, $M_Y(t) = \mathbf{E}\left[(M_X(t))^N\right]$. Differentiating, we obtain:

$$M_Y^{(1)}(t) = \mathbf{E}\left[N(M_X(t))^{N-1}M_X^{(1)}(t)\right]$$
$$M_Y^{(2)}(t) = \mathbf{E}\left[N(N-1)(M_X(t))^{N-2}(M_X^{(1)}(t))^2 + N(M_X(t))^{N-1}M_X^{(2)}(t)\right]$$

Conditional Variance

Alternative Solution (Continued)

- Recall that $M_X(0) = 1$, $M_X^{(1)}(0) = \mathbf{E}[X]$, and $M_X^{(2)}(0) = \mathbf{E}[X^2]$. Hence,

$$\begin{aligned}
\mathbf{E}[Y] &= M_Y^{(1)}(0) = \mathbf{E}[N]M_X^{(1)}(0) = \mathbf{E}[N]\mathbf{E}[X] \\
\mathbf{E}[Y^2] &= M_Y^{(2)}(0) = \mathbf{E}[N(N-1)](\mathbf{E}[X])^2 + \mathbf{E}[N]\mathbf{E}[X^2] \\
&= \mathbf{E}[N^2](\mathbf{E}[X])^2 + \mathbf{E}[N]\mathrm{Var}(X)
\end{aligned}$$

- Therefore, the variance can be computed as follows.

$$\mathrm{Var}(Y) = \mathbf{E}[Y^2] - (\mathbf{E}[Y])^2 = \mathrm{Var}(N)(\mathbf{E}[X])^2 + \mathbf{E}[N]\mathrm{Var}(X)$$