Lecture 13

1 Conditional Expectation and Conditional Variance

Definition 1.1. *Let X be a random variable in the probability space $(\Omega, \mathbb{P})$, and let A be an event such that $\mathbb{P}(A) > 0$. Define the* expectation of X conditioned on A *to be*

$$\mathbb{E}(X \mid A) = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega \mid A).$$

Claim 1.2. *It holds that*

$$\mathbb{E}(X \mid A) = \sum_{k} k \cdot \mathbb{P}(X = k \mid A).$$

Definition 1.3. *Let X be a random variable in the probability space $(\Omega, \mathbb{P})$, and let A be an event such that $\mathbb{P}(A) > 0$. Define the* variance of X conditioned on A *to be*

$$\mathrm{Var}(X \mid A) = \mathbb{E}\left((X - \mathbb{E}(X \mid A))^2 \mid A\right) = \mathbb{E}\left(X^2 \mid A\right) - (\mathbb{E}(X \mid A))^2.$$

Observe that all of the claims we prove for expectation and variance, hold when we condition on an event as well.

Definition 1.4. *Let X and Y be two random variables in the probability space $(\Omega, \mathbb{P})$. Define the* random variable $\mathbb{E}(X \mid Y)$ *to be*

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

Observe that

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(X \mid Y)) &= \sum_{\omega \in \Omega} \mathbb{E}(X \mid Y)(\omega) \cdot \mathbb{P}(\omega) = \sum_{\omega \in \Omega} \mathbb{E}(X \mid Y = Y(\omega))\, \mathbb{P}(\omega) \\
&= \sum_{y} \sum_{\substack{\omega \in \Omega \\ Y(\omega) = y}} \mathbb{E}(X \mid Y = Y(\omega))\, \mathbb{P}(\omega) = \sum_{y} \mathbb{E}(X \mid Y = y) \sum_{\substack{\omega \in \Omega \\ Y(\omega) = y}} \mathbb{P}(\omega) \\
&= \sum_{y} \mathbb{E}(X \mid Y = y) \cdot \mathbb{P}(\{\omega \in \Omega : Y(\omega) = y\}) = \sum_{y} \mathbb{E}(X \mid Y = y)\, \mathbb{P}(Y = y).
\end{aligned}$$

The following theorem provides a useful way of calculating the expectation of a random variable.

Theorem 1.5 (Law of Total Expectation)**.** *Let X and Y be two random variables in the probability space $(\Omega, \mathbb{P})$, where both have finite expectation. Then*

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y)).$$

Proof. It holds that

$$\begin{aligned}
\mathbb{E}\left(\mathbb{E}\left(X \mid Y\right)\right) &= \sum_y \mathbb{E}\left(X \mid Y = y\right) \cdot \mathbb{P}\left(Y = y\right) \\
&= \sum_y \left(\sum_x x \cdot \mathbb{P}\left(X = x \mid Y = y\right) \right) \cdot \mathbb{P}\left(Y = y\right) \\
&= \sum_y \left(\sum_x x \cdot \frac{\mathbb{P}\left(X = x, Y = y\right)}{\mathbb{P}\left(Y = y\right)} \right) \cdot \mathbb{P}\left(Y = y\right) \\
&= \sum_x x \cdot \left(\sum_y \mathbb{P}\left(X = x, Y = y\right) \right) \\
&= \sum_x x \cdot \mathbb{P}\left(X = x\right) \\
&= \mathbb{E}\left(X\right),
\end{aligned}$$

where the penultimate equality holds by the law of total probability. □

Example 1: A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 7 hours of travel. The second door leads to a tunnel that will take him back to the mine after 5 hours of travel and the third door leads to a tunnel that will take him back to the mine after 3 hours of travel. Every time he is in the mine, he chooses one of the three doors uniformly at random and goes through it. What is the expected time it will take the miner to reach safety?
Solution: Let X be the total amount of time it takes the miner to reach safety. Let Y be the random variable indicating his first choice, namely

$$Y \sim \begin{cases} 3 & 1/3 \\ 5 & 1/3 \\ 7 & 1/3 \end{cases}$$

where 3 corresponds to the 3 hours door, 5 corresponds to the 5 hours door, and 7 corresponds to the 7 hours door. Then, by the Law of Total Expectation we have

$$\begin{aligned}
\mathbb{E}\left(X\right) &= \mathbb{E}\left(\mathbb{E}\left(X \mid Y\right)\right) \\
&= \mathbb{E}\left(X \mid Y = 3\right) \cdot \mathbb{P}\left(Y = 3\right) + \mathbb{E}\left(X \mid Y = 5\right) \cdot \mathbb{P}\left(Y = 5\right) + \mathbb{E}\left(X \mid Y = 7\right) \cdot \mathbb{P}\left(Y = 7\right) \\
&= \frac{1}{3} \cdot \left(\left(3 + \mathbb{E}\left(X\right)\right) + \left(5 + \mathbb{E}\left(X\right)\right) + 7 \right) \\
&= 5 + \frac{2}{3} \cdot \mathbb{E}\left(X\right).
\end{aligned}$$

Therefore $\mathbb{E}\left(X\right) = 15$.

Example 2: Let $X \sim \text{Geom}\left(p\right)$. We will use the Law of Total Expectation to calculate $\mathbb{E}\left(X\right)$. Since $\mathbb{E}\left(X\right)$ does not depend on the definition of X (only on its distribution), we can assume that X is the total number of independent coin tosses until the first toss comes up heads, where the probability for heads in each toss is p. Let Y be the indicator for success in the first toss, i.e., $Y = 1$ if the first toss comes up heads and $Y = 0$ otherwise. Observe that $(X \mid Y = 1) = 1$ and $(X \mid Y = 0) \sim 1 + X$. Hence, by the Law of Total Expectation we have

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X \mid Y)) \\
&= \mathbb{E}(X \mid Y = 1) \cdot \mathbb{P}(Y = 1) + \mathbb{E}(X \mid Y = 0) \cdot \mathbb{P}(Y = 0) \\
&= p + \mathbb{E}(1 + X)(1 - p) \\
&= p + (1 + \mathbb{E}(X))(1 - p) \\
&= p + 1 + \mathbb{E}(X) - p - p\mathbb{E}(X)
\end{aligned}$$

implying that

$$p\mathbb{E}(X) = 1 \Longrightarrow \mathbb{E}(X) = 1/p.$$

Example 3: A fair coin is tossed until the first time it comes up heads, all coin tosses being mutually independent. Let N be the total number of tosses made. A new coin is made, with probability $1/N$ for heads, and is then tossed N times. Let R be the number of tosses of the new coin whose outcome is heads.

1. Are N and R independent?
2. What is $\mathbb{E}(R)$?

Solution:

1. The random variables N and R are dependent. Indeed, clearly $\mathbb{P}(R = 2, N = 1) = 0$ as one coin toss cannot result in heads twice. On the other hand, $\mathbb{P}(N = 1) = 1/2$ and

$$\begin{aligned}
\mathbb{P}(R = 2) \geq \mathbb{P}(R = 2, N = 2) &= \mathbb{P}(R = 2 \mid N = 2) \cdot \mathbb{P}(N = 2) \\
&= \binom{2}{2}(1/2)^2(1 - 1/2)^{2-2} \cdot 1/4 = 1/16.
\end{aligned}$$

   We conclude that $\mathbb{P}(R = 2, N = 1) \neq \mathbb{P}(R = 2) \cdot \mathbb{P}(N = 1)$ and thus N and R are dependent.

2. For every $n \in \mathbb{N}$ it holds that $(R \mid N = n) \sim \text{Bin}(n, 1/n)$ and thus $\mathbb{E}(R \mid N = n) = n \cdot 1/n = 1$. Therefore, it follows by the Law of Total Expectation that

$$\begin{aligned}
\mathbb{E}(R) &= \mathbb{E}(\mathbb{E}(R \mid N)) \\
&= \sum_{n=1}^{\infty} \mathbb{E}(R \mid N = n) \cdot \mathbb{P}(N = n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(N = n) \\
&= 1.
\end{aligned}$$

Definition 1.6. *Let* X *and* Y *be two random variables in the probability space* $(\Omega, \mathbb{P})$*. Define the* random variable $\mathrm{Var}\left(X \mid Y\right)$ *to be*

$$\mathrm{Var}\left(X \mid Y\right)(\omega) = \mathrm{Var}\left(X \mid Y = Y(\omega)\right).$$

Theorem 1.7 (Law of Total Variance)**.** *Let* X *and* Y *be two random variables in the probability space* $(\Omega, \mathbb{P})$*, where both have finite expectation. Then*

$$\mathrm{Var}\left(X\right) = \mathbb{E}\left(\mathrm{Var}\left(X \mid Y\right)\right) + \mathrm{Var}\left(\mathbb{E}\left(X \mid Y\right)\right).$$