EE 520: Random Processes **Fall 2021**

Lecture 4
Conditional Expectation

Instructor Name: John Lipor

Recommended Reading: Pishro-Nik: 5.1, 5.3.1, 6.2.4; Gubner: 2.4, 3.4, 3.5

1 Indicator Example

Before digging into new material, we'll do another example using an indicator random variable.

Example 1. Consider the set of integers $\{1, 2, \ldots, n\}$. A non-empty subset of $\{1, 2, \ldots, n\}$ is chosen at random. Let X denote the size of the subset. What is $\mathbb{E}[X]$?

First, we need to define an indicator RV such that summing all n of them will give us X. Let

$$X_i = \begin{cases} 1, & \text{if } i \text{ belongs to the subset} \\ 0, & \text{otherwise} \end{cases}$$

so that $X = \sum_{i=1}^{n} X_i$. There are $2^n - 1$ possible non-empty subsets, and 2^{n-1} of them contain the integer i (think HW1, problem 3a). Hence

$$\mathbb{E}[X_i] = P(X_i = 1) = \frac{2^{n-1}}{2^n - 1}.$$

We can then easily compute

$$\mathbb{E}[X] = n\frac{2^{n-1}}{2^n - 1}.$$

2 Expectation

We've already seen that two RVs are independent if their joint probabilities factor. Expectation can also be used to determine independence.

Fact 1. Two RVs X and Y are **independent** if and only if

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[k(Y)]$$

for all functions $h(\cdot)$ and $k(\cdot)$.

The "for all" condition in the above is an important distinction. When h and k are both the identity function (i.e., $h(x) = x$ and $k(y) = y$), we say the RVs are *uncorrelated*.

Definition 1. The **correlation** between to RVs X and Y is $\mathbb{E}[XY]$. If $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, we say X and Y are **uncorrelated**.

You should think for yourself about whether independence implies anything about correlation and likewise whether correlation implies anything about independence. In general, it is quite difficult to talk about expectations of products, so we often use tricks to bound them. One such trick is the Cauchy-Schwarz inequality.

Definition 2. The **Cauchy-Schwarz** inequality (for random variables) states that

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

Correlation does not account for the mean of either X or Y. This is done in the *covariance*.

Definition 3. The **covariance** between two RVs X and Y is

$$\text{cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right].$$

When we remove the mean from a RV, we call it *zero-mean* or *centered*, and the correlation and covariance become identical. A RV with zero mean and unit variance is called *standard* or *normalized*.

Fact 2. For uncorrelated RVs $X_1, X_2, \ldots, X_n$

$$\text{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{var}(X_i).$$

Example 2. What value of $\text{cov}(X,Y)$ implies X and Y are uncorrelated?

$$\begin{aligned}
\text{cov}(X,Y) &= \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] \\
&= \mathbb{E}\left[XY - \mathbb{E}[X]Y - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]\right] \\
&= \mathbb{E}[XY] - \mathbb{E}\left[\mathbb{E}[X]Y\right] - \mathbb{E}\left[X\mathbb{E}[Y]\right] + \mathbb{E}\left[\mathbb{E}[X]\mathbb{E}[Y]\right] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],
\end{aligned}$$

where we have used the fact that $\mathbb{E}[X]$ is not random, and hence $\mathbb{E}[\mathbb{E}[X]Y] = \mathbb{E}[X]\mathbb{E}[Y]$. This shows that X and Y are uncorrelated if and only if $\text{cov}(X,Y) = 0$.

3 Conditional Probability

For discrete RVs, the conditional probability follows directly from the definition of random events (see Lecture 2 notes). The conditional PMF is summarized via the following equation

$$p_{XY}(x_i, y_j) = p_{X|Y}(x_i \mid y_j) p_Y(y_j) = p_{Y|X}(y_i \mid x_j) p_X(x_j),$$

where

$$p_{X|Y}(x_i \mid y_j) = P(X = x_i \mid Y = y_j).$$

Using the PMF, the *law of total probability* (LTP) becomes

$$p_Y(y_j) = \sum_i p_{Y|X}(y_j \mid x_i) p_X(x_i).$$

Example 3. (Binary symmetric channel). Suppose we transmit $X \in \{0,1\}$ and receive a flipped bit with probability ε (called the "crossover" probability). Let Y be the received signal. What is $p_Y(y)$ if $X \sim \text{Ber}(p)$?

Using the LTP, we can easily find

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

and $P(Y=1) = 1 - P(Y=0)$.

Conditioning is done to "remove" randomness, making some terms easier to deal with. One takeaway is that when we condition on something, it is no longer random—this leads to the **substitution law**

$$P\left(g(X,Y) = z \mid X = x_i\right) = P\left(g(x_i, Y) = z \mid X = x_i\right).$$

Example 4. Let X and Y be RVs and $Z = X + Y = g(X,Y)$. Then

$$\begin{aligned}
P(Z = j \mid X = i) &= P(X + Y = j \mid X = i) \\
&= P(i + Y = j \mid X = i) \\
&= P(Y = j - i \mid X = i).
\end{aligned}$$

If we additionally have that X and Y are independent, then we can *drop the conditioning* to see that

$$P(Z = j \mid X = i) = P(Y = j - i).$$

4 Conditional Expectation

Note that $Y \mid X$ is a RV itself, so we can compute its expectation

$$\mathbb{E}[Y \mid X] = \sum_j y_j p_{Y|X}(y_j \mid x_i).$$

An important tool using conditional expectation is the *law of total expectation* (LTE). We'll use the LTE in two forms.

Definition 4. (LTE, version 1). Let $\{B_i\}_{i=1}^N$ partition the sample space Ω. Then

$$\mathbb{E}[X] = \sum_{i=1}^N \mathbb{E}[X \mid B_i] P(B_i).$$

Example 5. A biased coin with $P(H) = p$ is tossed sequentially until two consecutive heads appear. Let X be the number of required tosses. Find $\mathbb{E}[X]$.

For this problem, we create the partition

$$\begin{cases} B_1 = T \text{ on first toss} & P(B_1) = 1 - p \\ B_2 = HT & P(B_2) = p(1-p) \\ B_3 = HH & P(B_3) = p^2. \end{cases}$$

Using this partition, we compute

$$\mathbb{E}[X] = (1-p)\left(\mathbb{E}[X] + 1\right) + p(1-p)\left(\mathbb{E}[X] + 2\right) + 2p^2.$$

Note that this example may be useful for HW1, problem 5.

In the second variant of the LTE, we partition over the values a RV can take. This is of course not a *different* LTE, but I find it helpful to write it out separately.

Definition 5. (LTE, version 2). Let X and Y be two (possibly correlated) RVs. Then

$$\mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}_{X|Y}[X \mid Y]\right],$$

where $\mathbb{E}_Z[f(Z)]$ indicates that the expectation is with respect to the randomness in the RV Z.

Example 6. Let $\Omega = \{1, 2, \ldots, 6\}$, $\mathcal{F} = 2^\Omega$, and assume all outcomes are equiprobable. Define

$$X = \begin{cases} 1 & \omega \in \{2,3,4\} \\ 0 & \omega \in \{1,5,6\} \end{cases} \qquad Y = \begin{cases} 1 & \omega \in \{1,2,3\} \\ 0 & \omega \in \{4,5,6\}. \end{cases}$$

Clearly $\mathbb{E}[X] = 1/2$, but let's use LTE to check.

$$\mathbb{E}[X \mid Y = 0] = \mathbb{E}[X \mid \omega \in \{4,5,6\}] = \frac{1}{3} \times 1 + \frac{2}{3} \times 0 = \frac{1}{3}$$

$$\mathbb{E}[X \mid Y = 1] = \mathbb{E}[X \mid \omega \in \{1,2,3\}] = \frac{2}{3} \times 1 + \frac{1}{3} \times 0 = \frac{2}{3}$$

$$\mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}_{X|Y}\left[X \mid Y\right]\right] = \frac{1}{2} \times \frac{1}{3} + \frac{1}{2} \times \frac{2}{3} = \frac{1}{2}.$$