Chapter 5. Multiple Random Variables

5.3: Conditional Distributions

Slides (Google Drive) Video (YouTube)

5.3.1 Conditional PMFs/PDFs

Now that we've finished talking about joint distributions (whew), we can move on to conditional distributions and conditional expectation. This is actually just applying the concepts from 2.2 about conditional probability, generalizing to random variables (instead of events)!

Definition 5.3.1: Conditional PMFs and PDFs

If X, Y are discrete random variables, then the **conditional PMF** of X given Y is:

$$p_{X|Y}(a \mid b) = \mathbb{P}(X = a \mid Y = b) = \frac{p_{X,Y}(a,b)}{p_Y(b)} = \frac{p_{Y|X}(b \mid a)p_X(a)}{p_Y(b)}$$

Note that this should remind you of Bayes' Theorem (because that's what it is)!
If X, Y are continuous random variables, then the **conditional PDF** of X given Y is:

$$f_{X|Y}(u \mid v) = \frac{f_{X,Y}(u,v)}{f_Y(v)} = \frac{f_{Y|X}(v \mid u)f_X(u)}{f_Y(v)}$$

Again, this is just a generalization from discrete to continuous as we've been doing!
It's important to note that, for each *fixed value* of b, the probabilities that $X = a$ must sum to 1:

$$\sum_{a \in \Omega_X} p_{X|Y}(a \mid b) = 1$$

If X and Y are mixed (one discrete, one continuous), then a similar extension can be made where any discrete random variable has a p (a probability mass function) any continuous random variable has an f (a probability density function).

Example(s)

Back to our example of the blue and red die rolls from 5.1. Suppose we roll a fair blue 4-sided die and a fair red 4-sided die independently. Recall that $U = \min\{X, Y\}$ (the smaller of the two die rolls) and $V = \max\{X, Y\}$ (the larger of the two die rolls). Then, their joint PMF was:

U\V	1	2	3	4
1	1/16	2/16	2/16	2/16
2	0	1/16	2/16	2/16
3	0	0	1/16	2/16
4	0	0	0	1/16

What is $p_{U|V}(u \mid 3) = \mathbb{P}(U = u \mid V = 3)$ for each value of $u \in \Omega_U$ (these should sum to 1)!

Solution Well, we know by the definition of conditional probability that

$$p_{U|V}(u \mid 3) = \frac{p_{U,V}(u,3)}{p_V(3)}$$

We need to compute the denominator which is the marginal PMF of V (the sum of the third column):

$$p_V(3) = \sum_{a \in \Omega_U} p_{U,V}(a,3) = 2/16 + 2/16 + 1/16 + 0 = 5/16$$

Hence, our conditional PMF is

$$p_{U|V}(u \mid 3) = \begin{cases} \frac{2/16}{5/16} = \frac{2}{5} & u = 1 \\ \frac{2/16}{5/16} = \frac{2}{5} & u = 2 \\ \frac{1/16}{5/16} = \frac{1}{5} & u = 3 \\ \frac{0}{5/16} = 0 & u = 4 \end{cases}$$

□

5.3.2 Conditional Expectation

Just like conditional probabilities helped us compute "normal" (unconditional) probabilities in Chapter 2 (using LTP), we will learn about conditional expectation which will help us compute "normal" expectations!

Let's try to find out how we might define this idea of conditional expectation of a random variable X, given that we know some other RV Y takes on a particular value y. Well since $\mathbb{E}[X]$ (for discrete RVs) is defined to be:

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x\mathbb{P}(X = x) = \sum_{x \in \Omega_X} x p_X(x)$$

it's only fair that the conditional expectation of X, given knowledge that some other RV Y is equal to y is the same exact thing, EXCEPT the probabilities should be conditioned on $Y = y$ now:

$$\mathbb{E}[X \mid Y = y] = \sum_{x \in \Omega_X} x\mathbb{P}(X = x \mid Y = y) = \sum_{x \in \Omega_X} x p_{X,Y}(x \mid y)$$

Most notably, we are still summing over x and NOT y, since this expression should depend on y right? Given that $Y = y$, what is the expectation of X?

Definition 5.3.2: Conditional Expectation

Let X, Y be jointly distributed random variables.
If X is discrete (and Y is either discrete or continuous), then we define the **conditional expectation** of $g(X)$ given (the event that) $Y = y$ as:

$$\mathbb{E}[g(X) \mid Y = y] = \sum_{x \in \Omega_X} g(x) p_{X|Y}(x \mid y)$$

If X is continuous (and Y is either discrete or continuous), then we define the conditional expectation of $g(X)$ given (the event that) $Y = y$ as:

$$\mathbb{E}[g(X) \mid Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y) dx$$

Notice that these sums and integrals are **over** x (not y), since $\mathbb{E}[g(X) \mid Y = y]$ is a function of y. These formulas are exactly the same as $\mathbb{E}[g(X)]$, except the PMF/PDF of X is replaced with the conditional PMF/PDF of $X \mid Y = y$.

Example(s)

Suppose $X \sim \text{Unif}(0, 1)$ (continuous). We repeatedly draw independent $Y_1, Y_2, Y_3, \dots \sim \text{Unif}(0, 1)$ (continuous) until the first random time T such that $Y_T < X$. What is $\mathbb{E}[T]$?

The question is basically asking the following: we get some uniformly random decimal number X from $[0, 1]$. We keep drawing uniform random numbers until we get a value less than our initial value. What is the expected number of draws until this happens?

Solution We'll do this problem in a "bad" way (the only way we know how to know), and then learn the Law of Total Expectation next to see how this solution could be much simpler!

To find $\mathbb{E}[T]$, since T is discrete with range $\Omega_T = \{1, 2, 3, \dots\}$, we can find its PMF $p_T(t) = \mathbb{P}(T = t)$ for any value t and use the usual formula for expectation. However, T depends on the value of the initial number X right? If $X = 0.1$ it would take longer to get a number less than this than if $X = 0.99$. Let's try to find the probability $T = t$ given that $X = x$ first:

$$\mathbb{P}(T = t \mid X = x) = (1 - x)^{t-1} x$$

because the probability we get a number smaller than x is just x (Uniform CDF), and so we need to get $t - 1$ failures first before our first success. Actually, $(T \mid X = x) \sim \text{Geo}(x)$ so that's another way we could've computed this conditional PMF. Then, let's use the LTP to find $\mathbb{P}(T = t)$ (we need to *integrate* over all values of x because X is continuous, not discrete):

$$\mathbb{P}(T = t) = \int_0^1 \mathbb{P}(T = t \mid X = x) f_X(x) dx = \int_0^1 (1 - x)^{t-1} x \cdot 1 dx = \cdots = \frac{1}{t(t+1)}$$

after skipping some purely computational steps. Finally, since we have the PMF of T, we can compute expectation in the normal way:

$$\mathbb{E}[T] = \sum_{t=1}^{\infty} t p_T(t) = \sum_{t=1}^{\infty} t \frac{1}{t(t+1)} = \sum_{t=1}^{\infty} \frac{1}{t+1} = \infty$$

The reason this is ∞ is because this is like the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$ which is known to diverge to ∞. This is surprising right? The expected time until you get a number smaller than your first is infinite!
□

5.3.3 Law of Total Expectation (LTE)

Now we'll see how the Law of Total Expectation can make our lives easier! We'll also see an extremely cool application, which is to *elegantly* prove the expected value of a Geo(p) RV is $1/p$ (we did this algebraically in 3.5, but this was messy).

Theorem 5.3.1: Law of Total Expectation

Let X, Y be jointly distributed random variables.
If Y is discrete (and X is either discrete or continuous), then:

$$\mathbb{E}[g(X)] = \sum_{y \in \Omega_Y} \mathbb{E}[g(X) \mid Y = y] \, p_Y(y)$$

If Y is continuous (and X is either discrete or continuous), then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} \mathbb{E}[g(X) \mid Y = y] \, f_Y(y) dy$$

This looks exactly like the law of total probability we are used to. Basically to solve for $\mathbb{E}[g(X)]$, we need to take a weighted average of $\mathbb{E}[g(X) \mid Y = y]$ over all possible values of y.

Proof of LTE. Now we will prove the law of total expectation.

Suppose that X, Y are discrete (note that the same proof holds for any combination of X, Y being discrete or continous, but swapping sums to integrals as necessary). Then:

$$\begin{aligned}
\sum_{y \in \Omega_Y} \mathbb{E}[g(X) \mid Y = y] \, p_Y(y) &= \sum_{y \in \Omega_Y} \left(\sum_{x \in \Omega_X} g(x) p_{X|Y}(x \mid y) \right) p_Y(y) && \text{[def of conditional expectation]} \\
&= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} g(x) p_{X|Y}(x \mid y) p_Y(y) && \text{[swap sums]} \\
&= \sum_{x \in \Omega_X} g(x) \sum_{y \in \Omega_Y} p_{X,Y}(x, y) && \text{[def of conditional pmf]} \\
&= \sum_{x \in \Omega_X} g(x) p_X(x) && \text{[def of marginal pmf]} \\
&= \mathbb{E}[g(X)] && \text{[def of expectation/LOTUS]}
\end{aligned}$$

□

Example(s)

(This is the same example as earlier): Suppose $X \sim \text{Unif}(0, 1)$ (continuous). We repeatedly draw independent $Y_1, Y_2, Y_3, \dots \sim \text{Unif}(0, 1)$ (continuous) until the first random time T such that $Y_T < X$. What is $\mathbb{E}[T]$?

Solution Using the LTE now, we can solve this in a much simpler fashion. We know that $(T \mid X = x) \sim \text{Geo}(x)$ as stated earlier. By citing the expectation of a Geometric RV, we know that $\mathbb{E}[T \mid X = x] = \frac{1}{x}$. By the LTE, conditioning on x:

$$\mathbb{E}[T] = \int_0^1 \mathbb{E}[T \mid X = x] f_X(x) dx = \int_0^1 \frac{1}{x} 1 dx = [\ln(x)]_0^1 = \infty$$

This was a much faster way to getting to the answer than before! □

Example(s)

Let's finally prove that if $X \sim \text{Geo}(p)$, then $\mu = \mathbb{E}[X] = \frac{1}{p}$. Recall that the Geometric random variable is the number of independent Bernoulli trials with parameter p up to and including the first success.

Solution First, let's condition on whether our first flip was heads or tails (these events partition the sample space):

$$\mathbb{E}[X] = \mathbb{E}[X \mid H] \mathbb{P}(H) + \mathbb{E}[X \mid T] \mathbb{P}(T) [\text{Law of Total Expectation}]$$

What are those four values on the right though? We know $\mathbb{P}(H) = p$ and $\mathbb{P}(T) = 1 - p$, so that's out of the way.

What is $\mathbb{E}[X \mid H]$? If we got heads on the first try, then $\mathbb{E}[X \mid H] = 1$ since we are immediately done (i.e., the number of trials it took to get our first heads, given we got heads on the first trial, is 1).

What is $\mathbb{E}[X \mid T]$? This is a bit trickier: because the trials are independent, and we got a tail on the first try, we basically have to restart (memorylessness), and so our conditional expectation is just $\mathbb{E}[1 + X]$, since we are back to square one except with one additional trial!

Plugging these four values in gives a recursive formula ($\mathbb{E}[X]$ appears on both sides):

$$\mathbb{E}[X] = p + (1 + \mathbb{E}[X]) \cdot (1 - p)$$

We can solve this, using $\mu = \mathbb{E}[X]$ (for notational convenience):

$$\begin{aligned} \mu &= p + (1 + \mu)(1 - p) \\ \mu &= p + 1 - p + \mu - \mu p \\ \mu &= 1 + \mu - \mu p \\ 0 &= 1 - \mu p \\ \mu p &= 1 \\ \mu &= \frac{1}{p} \end{aligned}$$

This is a really "cute" proof of the expectation of a Geometric RV! See the notes in 3.5 to see the "ugly" calculus proof. $\square$

5.3.4 Exercises

1. What happens to linearity of expectation when you sum a *random* number of random variables? We know it holds for fixed values of n, but let's see what happens if we sum a random number N of them. It turns out, you get something very nice!

   Let $X_1, X_2, X_3, \ldots$ be a sequence of independent and identically distributed (iid) RVs, with common mean $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \ldots$. Let N be a random variable which has range $\Omega_N \subseteq \{0, 1, 2, \ldots\}$ (nonnegative integers), independent of all the X_i's. Show that $\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}[X_1]\,\mathbb{E}[N]$. That is, the expected sum of a random number of random variables is the expected number of random variables times the expected value of each (which you might think is intuitively true, but we have to prove it!).

   Solution: We have the following:

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] &= \sum_{n\in\Omega_N} \mathbb{E}\left[\sum_{i=1}^{N} X_i \mid N = n\right] p_N(n) && \text{[Law of Total Expectation]} \\
&= \sum_{n\in\Omega_N} \mathbb{E}\left[\sum_{i=1}^{n} X_i \mid N = n\right] p_N(n) && \text{[given } N = n : \text{substitute in the upper limit]} \\
&= \sum_{n\in\Omega_N} \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] p_N(n) && [N \text{ independent of } X_i' s] \\
&= \sum_{n\in\Omega_N} n\mathbb{E}[X_1]\, p_N(n) && \text{[Linearity of Expectation]} \\
&= \mathbb{E}[X_1] \sum_{n\in\Omega_N} n p_N(n) \\
&= \mathbb{E}[X_1]\,\mathbb{E}[N] && [\text{def of } \mathbb{E}[N]]
\end{aligned}$$