Joint probability mass functions of multiple random variables

We are often interested in **multiple random variables** resulting from the same experiment. For example, if you run a manufacturing facility, X may represent the number of failures in a batch of components provided by your supplier and Y may represent the number of systems you build that are returned from the customer.

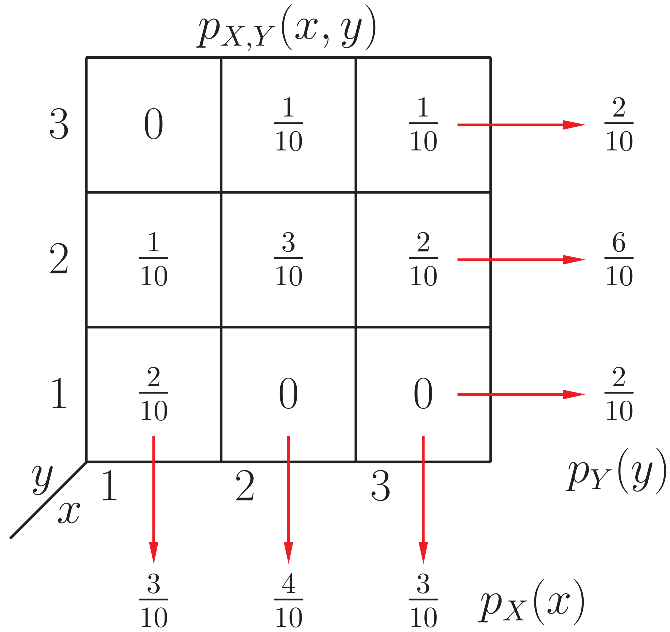
If we have two random variables, say X and Y, we of course have a pmf for each: $p_X(x)$ and $p_Y(y)$. But by themselves, these pmfs do not capture the **relationships** between the random variables. For this, we need the **joint pmf**, which has the compact notation:

$$p_{X,Y}(x,y) = \mathrm{P}\left(\{X = x\} \cap \{Y = y\}\right) = \mathrm{P}\left(X = x, Y = y\right).$$

Given the joint pmf, we can recover the **marginal distributions** $p_X(x)$ and $p_Y(y)$ by summing over all values of y and x respectively:

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$
$$p_Y(y) = \sum_x p_{X,Y}(x,y).$$

Example.



Similarly, we can define the **expectation** of any function of X and Y as:

$$\mathrm{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y).$$

Example. The **correlation** between X and Y is defined as

$$\mathrm{E}[XY] = \sum_x \sum_y xy \; p_{X,Y}(x,y).$$

Example. For any discrete random variables X and Y, we have that

$$\begin{aligned}
\mathrm{E}[X+Y] &= \sum_x \sum_y (x+y) p_{X,Y}(x,y) \\
&= \sum_x \sum_y x p_{X,Y}(x,y) + \sum_x \sum_y y p_{X,Y}(x,y) \\
&= \sum_x x \sum_y p_{X,Y}(x,y) + \sum_y y \sum_x p_{X,Y}(x,y) \\
&= \sum_x x p_X(x) + \sum_y y p_Y(y) \\
&= \mathrm{E}[X] + \mathrm{E}[Y].
\end{aligned}$$

Moreover, if $a, b, c \in \mathbb{R}$ are constants, then

$$\mathrm{E}[aX + bY + c] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y] + c.$$

More than two random variables

If we are interested in three random variables X, Y, Z, then we can define the joint pmf

$$p_{X,Y,Z}(x, y, z) = \mathrm{P}\left(X = x, Y = y, Z = z\right).$$

From this, we can calculate the marginals for any pair of random variables, e.g.,

$$p_{X,Y}(x, y) = \sum_{z} p_{X,Y,Z}(x, y, z),$$

or any single random variable, e.g.,

$$p_X(x) = \sum_{y}\sum_{z} p_{X,Y,Z}(x, y, z).$$

We can also compute expectations via

$$\mathrm{E}[g(X, Y, Z)] = \sum_{x}\sum_{y}\sum_{z} g(x, y, z) p_{X,Y,Z}(x, y, z)$$

and it is easy to show that

$$\mathrm{E}[aX + bY + cZ] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y] + c\,\mathrm{E}[Z].$$

Moreover, these notions are easily generalized to (possibly many) more than three variables. For random variables $X_1, \ldots, X_n$, we have the joint pmf

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \mathrm{P}\left(X_1 = x_1, \ldots, X_n = x_n\right),$$

we can calculate expectations via

$$\mathrm{E}[g(X_1, \ldots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(X_1, \ldots, X_n) p_{X_1,\ldots,X_n}(x_1, \ldots, x_n),$$

and if $a_1, \ldots, a_n \in \mathbb{R}$ are constants, we have

$$\mathrm{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \, \mathrm{E}[X_i].$$

Example. Suppose that I hand back n graded quizzes at random without actually looking at the names. Let X be the number of people who get the correct quiz returned to them. What is $E[X]$, that is, the expected number of people who get back their own quiz?

Begin by associating the random variable X_i with the i^{th} student. X_i will take the value of 1 if the i^{th} student gets back his own quiz, and 0 otherwise. Assuming that I am handing out the quizzes uniformly at random, we have that $\mathrm{P}(X_i = 1) = 1/n$ and $\mathrm{P}(X_i = 0) = 1 - 1/n$. Thus, $\mathrm{E}[X_i] = 1/n$. Using the fact that

$$X = X_1 + \cdots + X_n$$

we have that

$$\mathrm{E}[X] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n] = n \cdot \frac{1}{n} = 1.$$

Conditioning random variables on an event

It is straightforward to adjust the pmf of a random variable given knowledge of an event A (with $\mathrm{P}(A) > 0$):

$$p_{X|A}(x) = \mathrm{P}(X = x|A) = \frac{\mathrm{P}(\{X = x\} \cap A)}{\mathrm{P}(A)}$$

It is easy to see that

$$\sum_x \mathrm{P}(\{X = x\} \cap A) = \mathrm{P}(A),$$

since the events $\{X = x\}$ are disjoint. Thus, $p_{X|A}(x)$ is a valid pmf (it sums to 1 and is nonnegative for all x).

Exercise:
A student applied to college and knows decision letters will arrive sometime this week Monday-Friday with equal probability. Suppose that Monday and Tuesday have passed with no letter. Find the PMF describing the probability that the letter arrives on each of the remaining days.

Conditioning one random variable on another

The conditional pmf of ***X* given *Y*** is simply the special case of $A = \{Y = y\}$ above. But since this really captures what the random variable Y tells us about X we use the special notation

$$p_{X|Y}(x|y) = \mathrm{P}\left(X = x | Y = y\right)$$

For a fixed y, this is a valid pmf on X. Note that for a fixed x, this is ***not*** a valid PMF on Y. By the definitions above,

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{\text{joint pmf for } X \text{ and } Y}{\text{marginal pmf for } Y}.$$

Just as with the multiplication rule before, rewriting this definition gives us a way to use conditional distributions to construct unconditional distributions.

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y)$$
$$p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x)$$

Exercise. (B&T 2.15 on page 101) We are sending a message over a computer network from one point to another. Let X be the time it takes to transmit the message, which depends on the length Y of the message and the congestion present in the network. Suppose that Y has pmf

$$p_Y(y) = \begin{cases} 5/6 & y = 10^2 \\ 1/6 & y = 10^3. \end{cases}$$

Given Y, the time X is $Y/10^4$ seconds with probability $1/2$, $Y/10^3$ seconds with probability $1/3$, and $Y/10^2$ seconds with probability $1/6$. What is the pmf $p_X(x)$ for the transmission time?

Solution: We have

$$p_{X|Y}(x|Y = 10^2) = \begin{cases} & \\ & \end{cases}$$

and

$$p_{X|Y}(x|Y = 10^3) = \begin{cases} & \\ & \end{cases}$$

From here we can compute the unconditional pmf of X via the formula

$$p_X(x) = \sum_y p_{X|Y}(x|y) p_Y(y).$$

Specifically, we have

$$p_X(x) = \begin{cases} & \\ & \\ & \end{cases}$$

Conditional expectation

Since conditional pmfs behave just like ordinary pmfs (just defined and normalized over a new set) we can easily extend our notion of expectation to handle conditional pmfs. Specifically, we can calculate **conditional expectations** given an event A (with $\mathrm{P}(A) > 0$) via

$$\mathrm{E}[g(X)|A] = \sum_x g(x) p_{X|A}(x).$$

As before, a special case of this is where $A = \{Y = y\}$ for some random variable Y. In this case we use the special notation

$$\mathrm{E}[X|Y = y] = \sum_x x \, p_{X|Y}(x|y).$$

Recall that when we first introduced the notion of conditional probability, one of our motivations was to simplify some calculations using the total probability theorem. There is a similar result for expectation, which naturally goes by the name of the **total expectation theorem**. The simplest instantiation of this result is

$$\mathrm{E}[X] = \sum_y p_Y(y) \, \mathrm{E}[X|Y = y].$$

To see why this holds, note that since

$$\begin{aligned} p_X(x) &= \sum_y p_{X,Y}(x, y) \\ &= \sum_y p_Y(y) p_{X|Y}(x|y) \end{aligned}$$

we have

$$\begin{aligned}
\mathrm{E}[X] &= \sum_x x p_X(x) \\
&= \sum_x x \sum_y p_Y(y) p_{X|Y}(x|y) \\
&= \sum_y p_Y(y) \sum_x x p_{X|Y}(x|y) \\
&= \sum_y p_Y(y)\, \mathrm{E}[X|Y=y].
\end{aligned}$$

The most general variant of the total expectation theorem is as follows: Let $A_1, \ldots, A_n$ be disjoint events that form a partition of Ω with $\mathrm{P}(A_i) > 0$ for all i. Then

$$\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{P}(A_i)\, \mathrm{E}[X|A_i].$$

Furthermore, for any event B with $\mathrm{P}(A_i \cap B) > 0$ for all i,

$$\mathrm{E}[X|B] = \sum_{i=1}^{n} \mathrm{P}(A_i|B)\, \mathrm{E}[X|A_i \cap B].$$

Exercise:
Suppose that I cook dinner with probability 1/2, pick up takeout with probability 3/10, and go out to eat with probability 1/5. If I cook my expected time spent on dinner is 2 hours, if I pick up takeout the expected time is 45 minutes, and if I eat out the expected time is 1.5 hours. What is the expected amount of time I spend on dinner?

Exercise:
Your manufacturing plant gets parts from two different suppliers and therefore produces two versions of your product. Version 1 of the product has a 10% chance of failure in any given year, while version 2 has only a 5% chance of failure in any given year. 70% of the products are version 1.

Let X be the number of years until failure for an arbitrary device produced at your plant. What is the PMF of X?

What is the expected value of X?

Exercise Suppose that each day in class, you will get called on to answer a question with probability p. The probability that you are first asked a question on day x is

$$p_X(x) = (1-p)^{x-1}p.$$

Let A be the event that you are not called on the first day (i.e. the event $\{X > 1\}$).

1. What are $\mathrm{P}(A)$ and $\mathrm{P}(A^c)$?

2. What is the conditional pmf $p_{X|A^c}(x)$? (Hint: This is easy.)

3. What is the conditional mean $\mathrm{E}[X|A^c]$? (Hint: Given a correct answer to the previous question, this is even easier.)

4. What is the conditional pmf $p_{X|A}(x)$?

5. What is the conditional mean $\mathrm{E}[X|A]$ in terms of the unconditional mean $\mathrm{E}[X]$?

6. Calculate $\mathrm{E}[X]$ using the total expectation theorem.