Conditional Expectation

April 17, 2020

1 Expectation of a random variable, conditioned on an event

Let X and Y be random variables. We might be interested in computing the expectation of X, or more generally $g(X)$ for some function g, given information about Y. The most basic way to do this is to simply condition on the event that the random variable $Y = y$ for some fixed number y.

DEFINITION 1.1. *Let X and Y be continuous random variables and let $g : \mathbb{R} \to \mathbb{R}$ be a function. The* ***expectation of** $g(X)$ **conditional on the event*** $Y = y$, *is*

$$\mathbb{E}[g(X)|Y = y] = \int g(x) f_{X|Y}(x|y) dx.$$

Let X and Y be discrete random variables and let $g : \mathbb{R} \to \mathbb{R}$ be a function. The ***expectation of** $g(X)$ **conditional on the event*** $Y = y$, *is*

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

We're using our standard notation that, we're conditioning on what is after the | symbol. The above definition says that to compute a conditional expectation, you use the conditional distribution in place of the marginal distribution when you compute the normal expectation.

EXAMPLE 1.2. *Let Y be a geometric random variable with mean 10. Let X be uniformly distributed on the integers between 1 and Y. Compute the conditional expectation of X given $Y = 20$.*

Solution: The conditional pmf of X, given that $Y = 20$ is

$$p_{X|Y}(x|20) = 1/20, \text{ for } x = 1, \ldots, 20$$

and 0 otherwise. So

$$E[X|Y = 20] = \sum_x x p_{X|Y}(x|20) = \sum_{x=1}^{20} x \frac{1}{20} = \frac{21}{2}$$

EXAMPLE 1.3. *Let Y_1 and Y_2 be continuous random variables with jpdf:*

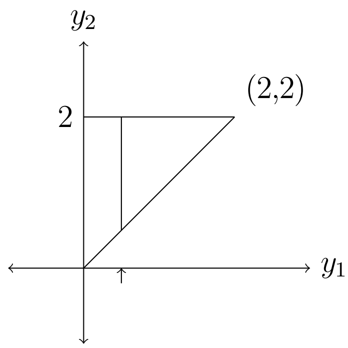
$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{3}{4}y_1^2, & \text{if } 0 \le y_1 \le y_2 \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Compute the conditional expectation of Y_2 given $Y_1 = y_1$.

(b) Given that $Y_1 = 1/2$, what is the expectation of Y_2?

Solution:

(a) The idea here is that we want to restrict the the density a segment where y_1 is constant. Like



in this picture:
Then using the density along this restriction we compute the expectation of Y_2

In a previous set of notes we computed:

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{1}{2-y_1}.$$

for $y_1 \le y_2 \le 2$, and zero otherwise.

So

$$\mathbb{E}[Y_2|Y_1 = y_1] = \int_{y_1}^{2} y_2 \frac{1}{2-y_1} dy_2 = \frac{1}{2-y_1} \frac{2^2 - y_1^2}{2} = \frac{2+y_1}{2}$$

(b) Here we simply evaluate the previous expression at $y_1 = 1/2$, and get

$$\mathbb{E}[Y_2|Y_1 = 1/2] = \frac{2+1/2}{2}$$

2 Conditional expectation with respect to a random variable

We can also compute conditional expectations where we condition on another random variable not the event that this other random variable takes on a value. This is a somewhat subtle difference. In the previous section, we considered the second case and our end results were numbers. If we condition on another random variable, the the result we get is again a random variable. We'll begin with the formal definition and then we'll repeat the last two examples in this new framework.

DEFINITION 2.1. *Let X and Y be continuous random variables and let $g : \mathbb{R} \to \mathbb{R}$ be a function. The* ***conditional expectation of*** $g(X)$ ***with respect to*** Y, *is*

$$\mathbb{E}[g(X)|Y] = \int g(x) f_{X|Y}(x|Y) dx.$$

Let X and Y be discrete random variables and let $g:\mathbb{R}\to\mathbb{R}$ be a function. The ***conditional expectation of*** $g(X)$ ***with respect to*** Y*, is*

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

EXAMPLE 2.2. *Let Y be a geometric random variable with mean 10. Let X be uniformly distributed on the integers between 1 and Y. Compute the conditional expectation of X given Y.*

Solution: The conditional pmf of X, given Y is

$$p_{X|Y}(x|Y)=1/Y, \text{ for } x=1,...,Y$$

and 0 otherwise. So

$$E[X|Y]=\sum_x xp_{X|Y}(x|Y)=\sum_{x=1}^{Y} x\frac{1}{Y}=\frac{Y+1}{2}$$

So the final result is a random variable, but only depends on the distribution of Y, not X, so it is some sense "less random".

EXAMPLE 2.3. *Let Y_1 and Y_2 be continuous random variables with jpdf:*

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases} \frac{3}{4}y_1^2, & \text{if } 0\le y_1\le y_2\le 2 \\ 0, & \text{otherwise.}\end{cases}$$

(a) Compute the conditional expectation of Y_2 given Y_1.

Solution:

Once again, we've seen:

$$f_{Y_2|Y_1}(y_2|Y_1)=\frac{1}{2-Y_1}.$$

for $y_1\le y_2\le 2$, and zero otherwise.

So

$$\mathbb{E}[Y_2|Y_1]=\int_{Y_1}^2 y_2\frac{1}{2-Y_1}dy_2=\frac{1}{2-Y_1}\frac{2^2-Y_1^2}{2}=\frac{2+Y_1}{2}$$

One of the main uses of this definition is it gives us shortcut for computing the expectations of random variables if we know their conditional distributions. This is known as **The law of iterated expectations**.

THEOREM 2.4 (**The law of iterated expectations**). *Let X and Y be random variables*

$$\mathbb{E}[X]=\mathbb{E}[\mathbb{E}[X|Y]]$$

This theorem says you can first compute the conditional expectation with respect to a variable, which gives a new random variable, but this new random variable has the same expectation as the original. Actually, quite a bit of information is being hidden in the notation above. The expectation on the left hand side is just with respect to the variable X and the outside expectation on the right side is just with respect to Y.

The proof of this theorem is in the book, but mostly just makes use of the fact the joint distribution of two random variables is equal to a conditional distribution times the marginal distribution of what you condition on.

Let's use this theorem to come the expectation of the random variables above.

EXAMPLE 2.5. *Let Y be a geometric random variable with mean 10. Let X be uniformly distributed on the integers between 1 and Y. Compute the (unconditional) expectation of X.*

Solution:

$$\mathbb{E}[X]=\mathbb{E}[\mathbb{E}[X|Y]]=\mathbb{E}[\frac{Y+1}{2}]=\mathbb{E}[Y/2]+1/2=11/2$$

So we never had to compute the (unconditional) pmf of X or the joint pmf of X and Y, knowing its conditional pmf with respect to Y and the marginal pmf of Y was enough to compute $\mathbb{E}[X]$.

In the other example, this trick is actually not so useful, because we start with the joint pmf, not a conditional pmf. We can still write in out to compare to the direct computation.

EXAMPLE 2.6. *Let Y_1 and Y_2 be continuous random variables with jpdf:*

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases} \frac{3}{4}y_1^2, & \text{if } 0\leq y_1\leq y_2\leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

Compute the (unconditional) expectation of Y_2.

Solution:

$$\mathbb{E}[Y_2]=\mathbb{E}[\mathbb{E}[Y_2|Y_1]]=\mathbb{E}[\frac{2+Y_1}{2}]=1+\frac{\mathbb{E}[Y_1]}{2}=1+6/10=8/5$$

Where we used from a previous notes that $\mathbb{E}[Y_1]=6/5$.

Note the direct computation would be:

$$\begin{aligned}\mathbb{E}[Y_2]&=\int_0^2\int_0^{y_2} y_2\left(\frac{3}{4}y_1^2\right)\,dy_1dy_2\\&=\int_0^2\int_0^{y_2} y_2\left(\frac{3}{4}y_1^2\right)\,dy_1dy_2\\&=\int_0^2 y_2\left(\frac{3}{4}\frac{y_1^3}{3}\right)_0^{y_2}\,dy_2\\&=\int_0^2\frac{1}{4}y_2^4\,dy_2\\&=\frac{1}{4}\frac{y_2^5}{5}\Big|_0^2\\&=\frac{8}{5}\end{aligned}$$

Which in this problem is the same amount of work as computing Y_1.

We can also define a conditional variance. We simply replace all the normal expectations in the definition of the variance with conditional variances.

Definition 2.7. *Let X and Y be random variables. The* ***conditional variance of*** X ***given*** Y *is*

$$\mathrm{Var}(X|Y)=\mathbb{E}\Big[(X-\mathbb{E}[X|Y])^2\Big|Y\Big]=\mathbb{E}[X^2|Y]-\mathbb{E}[X|Y]^2$$

We can also compute the variance of a random variable by a similar conditioning trick.

Theorem 2.8. *Let X and Y be random variables.*

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

The term $\mathrm{Var}(X|Y)$ gives how X varies for each observation of Y, so $\mathbb{E}[\mathrm{Var}(X|Y)]$ then gives the expectation of how X varies for each Y. The $\mathbb{E}[X|Y]$ encodes how X is related to Y, so the variance of this terms gives the contribution of how Y varies to the variance of X.

I don't see any intuition in the proof in the book, so here is a different one:

Proof. We begin by writing X as

$$X=\mathbb{E}[X|Y]+(X-\mathbb{E}[X|Y]).$$

We will see that the variance of each of the two terms is the two terms in the theorem, and that these two terms are uncorrelated.

Then we take the variance of both sides

$$\mathrm{Var}(X)=\mathrm{Var}(\mathbb{E}[X|Y])+\mathrm{Var}(X-\mathbb{E}[X|Y])+2\mathrm{Cov}(\mathbb{E}[X|Y],X-\mathbb{E}[X|Y]).$$

This is using the formula for the variance of the sum of random variables. The first term already appears in the theorem, so let's consider the second term.

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

We can insert a condition expectation inside to get $\mathbb{E}[(X-\mathbb{E}[X|Y])^2]=\mathbb{E}[\mathbb{E}[(X-\mathbb{E}[X|Y])^2|Y]]=\mathbb{E}[\mathrm{Var}(X|Y)]$ and then since $\mathbb{E}[\mathbb{E}[X|Y]]=\mathbb{E}[X]$ the second term is zero. We then have

$$\mathrm{Var}(X-\mathbb{E}[X|Y])=\mathbb{E}[\mathrm{Var}(X|Y)].$$

Now we just need to show the last term is zero:

$$\mathrm{Cov}(\mathbb{E}[X|Y],X-\mathbb{E}[X|Y])=\mathbb{E}[\mathbb{E}[X|Y](X-\mathbb{E}[X|Y])]-\mathbb{E}[\mathbb{E}[X|Y]]*\mathbb{E}[X-\mathbb{E}[X|Y]]$$

We've already seen $\mathbb{E}[X-\mathbb{E}[X|Y]]=0$ so we just need to expand the first term.

$$\mathbb{E}[\mathbb{E}[X|Y](X-\mathbb{E}[X|Y])]=\mathbb{E}[X\mathbb{E}[X|Y]]-E[\mathbb{E}[X|Y]^2]$$

$$=\mathbb{E}\Big[\mathbb{E}[X\mathbb{E}[X|Y]|Y]\Big]-\mathbb{E}\Big[\mathbb{E}[X|Y]^2\Big]=\mathbb{E}\Big[\mathbb{E}[X|Y]^2\Big]-E[\mathbb{E}[X|Y]^2]=0$$

where we inserted a conditional expectation into the first expression.

□

Example 2.9. *Let Y be a geometric random variable with mean 10. Let X be uniformly distributed on the integers between 1 and Y. Compute the (unconditional) variance of X.*

Note: as I finished this example, I realize it's a terrible one, but I'll leave it here because I worked it out. In general these are not the nicest computations to do, but this is still the nicest way to approach it.

Solution: Earlier we saw $\mathbb{E}[X|Y]=\frac{Y+1}{2}$, since Y is geometric with mean 10

$$\text{Var}(\mathbb{E}[X|Y])=\text{Var}(\frac{Y+1}{2})=\frac{1}{2^2}\text{Var}(Y)=\frac{1}{2^2}\frac{1-1/10}{1/10^2}=\frac{90}{4}$$

We now need to compute Var$(X|Y)$

$$\text{Var}(X|Y)=\mathbb{E}[X^2|Y]-\mathbb{E}[X|Y]^2$$

The second term we can get from above, the first term we compute from the conditional pmf:

$$\mathbb{E}[X^2|Y]=\sum_{x=1}^{Y}x^2 1/Y=\frac{(Y+1)(2Y+1)}{6}$$

where the last equality was copied from the internet.

$$\text{Var}(X|Y)=\mathbb{E}[X^2|Y]-\mathbb{E}[X|Y]^2=\frac{(Y+1)(2Y+1)}{6}-\left(\frac{Y+1}{2}\right)^2=\frac{Y^2-1}{12}$$

Then we take the expectation

$$\mathbb{E}[\text{Var}(X|Y)]=1/12(\mathbb{E}[Y^2]-1)=1/12(\frac{2-1/10}{1/10^2}-1)=1/12(189)$$

Where we used the that the second moment of a geometric random variable with mean $1/p$ is $(2-p)/p^2$.

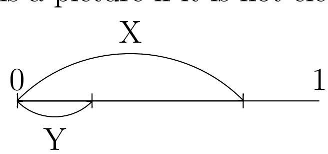
Combining these two expression we have

$$\text{Var}(X)=\text{Var}(\mathbb{E}[X|Y])+\mathbb{E}[\text{Var}(X|Y)]=\frac{90}{4}+\frac{189}{12}$$

Here is a nicer example:

EXAMPLE 2.10. *You begin with a stick of length 1 and break it at point, chosen uniformly at random. You then take the left piece and break it once again at a uniformly random chosen point. What is the expectation and variance of the length of left piece after the breaking.*

Solution: Let X be the length of the left piece after the first break and Y after the second. Here is a picture if it is not clear:



Where the location X was chosen uniformly between between 0 and 1 and the location Y is chosen uniformly between 0 and X.

Computing the pdf of Y would be hard, but it's conditional distribution given X is much simpler, so we will use condition expectations to compute the regular expectation and variance.

The random variable X is a uniform random variable on [0,1] so

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

The conditional distribution of Y is also uniform but on $[0,X]$ so

$$f_{Y|X}(y|x) = \begin{cases} 1/x & \text{if } 0 \le y \le x \\ 0, & \text{otherwise.} \end{cases}$$

So then we have

$$\mathbb{E}[Y|X] = \int y f_{Y|X}(y|X)dy = \int_0^X y 1/X dy = \frac{X}{2}$$

(This computation was not really necessary, since we already know the expectation of a uniform random variable.)

We can now compute $\mathbb{E}[Y]$

$$\mathbb{E}[Y] = \mathbb{E}\Big[\mathbb{E}[Y|X]\Big] = \mathbb{E}\left[\frac{X}{2}\right] = \frac{1}{4}$$

since the expectation of X is 1/2.

To compute the variance of Y we first compute the conditional variance

$$\text{Var}(Y|X) = \mathbb{E}[Y^2|X] - \mathbb{E}[Y|X]^2 = \int y^2 f_{Y|X}(y|X)dy - (X/2)^2$$

where the first term is just the definition of the conditional variance and the second was computed above. Let's expand the first term:

$$\int y^2 f_{Y|X}(y|X)dy = \int_0^X y^2 1/X dx = \frac{X^2}{3}$$

so

$$\text{Var}(Y|X) = \frac{X^2}{3} - \frac{X^2}{4} = \frac{X^2}{12}$$

(we actually have already computed this, it's just the variance of the of uniformly distributed random variable on the the interval $[0,Y]$).

Now we can consider the regular variance of Y.

$$\text{Var}(Y) = \text{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\text{Var}(Y|X)]$$

The first term is

$$\text{Var}(\mathbb{E}[Y|X]) = \text{Var}(\frac{X}{2}) = \frac{1}{4}\text{Var}(X) = 1/48$$

where we have used that $\text{Var}(X) = 1/12$, because it is uniformly distributed on [0,1].

The second term is

$$\mathbb{E}[\text{Var}(Y|X)] = \mathbb{E}[\frac{X^2}{12}] = 1/12\mathbb{E}[X^2] = 1/36$$

where we have used that $\mathbb{E}[X^2] = 1/3$, because it is uniformly distributed on [0,1].

So

$$\text{Var}(Y) = 1/48 + 1/36$$

Example 2.11. *Use the conditional expectation to compute the mean of geometric random variable with parameter p.*

Solution: We will condition on the first trial, because if we ignore the first trial we can get a new geometric r.v.

Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed (i.i.d) Bernoulli(p) random variables. Let Y be the location of the first success, so Y is a geometric random variable with parameter p.

Let $\tilde{Y}$ be the location of the first success if the first trial is ignored. So if the $X_2=1$ then $\tilde{Y}=1$, if $X_2=0, X_3=1$ then $\tilde{Y}=2$, and so on, regardless of the first trial. So $\tilde{Y}$ is also a geometric random variable with parameter p.

We now compute the conditional expectation of Y given X_1

$$\mathbb{E}[Y|X_1=1]=1$$

Because if the first trial is a success then $Y=1$.

$$\mathbb{E}[Y|X_1=0]=1+\mathbb{E}[\tilde{Y}]=1+\mathbb{E}[Y]$$

Because if the first trial is a failure, then the expectation of the Y is the same as $\tilde{Y}$ plus one, because Y counts the first trial but $\tilde{Y}$ doesn't.

So the random variable $\mathbb{E}[Y|X_1]$ equals 1 with probability p and $1+\mathbb{E}[Y]$ with probability $1-p$.

Now we compute $\mathbb{E}[Y]$ using the law of iterated expectations:

$$\mathbb{E}[Y]=\mathbb{E}[\mathbb{E}[Y|X_1]]=1*p+(1+\mathbb{E}[Y])(1-p)$$

Then solving for $\mathbb{E}[Y]$ we get

$$\mathbb{E}[Y]=1/p$$

as expected.