Lecture 13: Conditional Expectation and Variance

Anup Rao

February 2, 2018

We discuss the conditional expectation and variance.

Given an event E, the conditional expectation is exactly what you would expect after conditioning:

$$\mathbb{E}[X|E] = \sum_{x \in E} p(X = x|E) \cdot x.$$

Fact 1. *If E is an event and X is a random variable, then*

$$\mathbb{E}[X] = p(E) \cdot \mathbb{E}[X|E] + p(E^c) \cdot \mathbb{E}[X|E^c].$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_x p(X = x) \cdot x \\
&= \sum_x (p(E) \cdot p(X = x|E) + p(E^c) \cdot p(X = x|E^c)) \cdot x \\
&= p(E) \sum_x p(X = x|E) \cdot x + p(E^c) \sum_x p(X = x|E^c) \cdot x \\
&= p(E) \cdot \mathbb{E}[X|E] + p(E^c) \cdot \mathbb{E}[X|E^c].
\end{aligned}$$

□

Example

You toss a coin until you see heads. How many coin tosses do you expect to see?

Let X be the number of coin tosses. Then we see that

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} p(X = i) \cdot i = \sum_{i=1}^{\infty} (1/2)^i \cdot i.$$

We could calculate this using some identities from calculus.

We have

$$(1 - x)(1 + x + x^2 + \ldots + x^r) = (1 - x^{r+1}),$$

so we can express

$$1 + x + x^2 + \ldots + x^r = \frac{1 - x^{r+1}}{1 - x}.$$

If $|x| < 1$, then as r goes to infinity, we get

$$1 + x + x^2 + \ldots = \frac{1}{1 - x}.$$

Now, the expression we care about is

$$\mathbb{E}[X] = \sum_{i=1}^{\infty}(1/2)^i \cdot i = (1/2)\sum_{i=1}^{\infty}(1/2)^{i-1}\cdot i = x(1+2x+3x^2+4x^3+\dots),$$

where here $x = 1/2$. The infinite sum in this expression is exactly the derivative of

$$1 + x + x^2 + \dots .$$

So, taking the derivative of $(1/(1-x))$, we get that it should be $1/(1-x)^2$. Plugging this calculation in gives:

$$\mathbb{E}[X] = (1/2)\cdot\frac{1}{(1-(1/2))^2} = 2.$$

But there is an easier way to calculate all of this using conditional expectation. Let E denote the event that the first coin toss is heads. Then we have

$$\mathbb{E}[X] = p(E)\,\mathbb{E}[X|E] + p(E^c)\,\mathbb{E}[X|E^c].$$

Now $\mathbb{E}[X|E] = 1$, since under this event there are no more coin tosses. Conditioned on E^c, the expected number of heads is exactly 1 more than the expected number of heads for X. So, we get:

$$\mathbb{E}[X] = (1/2) + (1/2)(\mathbb{E}[X]+1),$$

which gives $\mathbb{E}[X] = 2$.

Variance

As we saw in the earlier examples, a random variable can be very far from its expectation. One way to measure how far a random variable is typically from its expectation is to measure its variance.

Suppose X is a random variable with $\mathbb{E}[X] = 0$. Then the variance of X is

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right].$$

When it is large then $|X|$ is typically large and the value of X is usually far from its expectation. For general X, we define its variance as

$$\mathrm{Var}[X] = \mathbb{E}\left[(X-\mu)^2\right],$$

where here $\mu = \mathbb{E}[X]$.

The variance of a random variable is always non-negative.

Example

For example, let X, Y, Z be random variables such that

$$X = \begin{cases} 1000 & \text{with probability } 1/2, \\ -1000 & \text{with probability } 1/2. \end{cases}$$

$$Y = \begin{cases} 1 & \text{with probability } \frac{n-1}{n}, \\ -(n-1) & \text{with probability } \frac{1}{n}. \end{cases}$$

$$Z = \begin{cases} 0 & \text{with probability } 1. \end{cases}$$

Then all have expectation 0. We have

$$\text{Var}[X] = (1/2)(1000)^2 + (1/2)(1000)^2 = 1000000,$$

$$\text{Var}[Y] = \frac{n-1}{n} \cdot 1^2 + \frac{1}{n}(n-1)^2 = \frac{n-1+(n-1)^2}{n} = \frac{n^2-n}{n} = n-1,$$

$$\text{Var}[Z] = 0.$$

Fact 2. *If X is a random variable, and $Y = aX + b$, for constants a, b, then* $\text{Var}[Y] = a^2 \cdot \text{Var}[X]$.

Proof.

$$\begin{aligned} \text{Var}[Y] &= \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right] \\ &= \mathbb{E}\left[(aX + b - a\,\mathbb{E}[X] - b)^2\right] \\ &= \mathbb{E}\left[a^2(X - \mathbb{E}[X])^2\right] = a^2 \cdot \text{Var}[X]. \end{aligned}$$

□

This means that we can always shift X to $X - \mathbb{E}[X]$ without changing its variance.

In general, if X, Y are two random variables, then $\text{Var}[X+Y]$ is not the same as $\text{Var}[X] + \text{Var}[Y]$. In fact, the variance of $X + Y$ might even be lower than the variance of either X or Y.

For example, if $X = -Y$, then $\text{Var}[X+Y] = 0$.

Fact 3. *When X and Y are independent,* $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$.

Proof.

$$\text{Var}[X+Y] = \text{Var}[X + Y - \mathbb{E}[X] - \mathbb{E}[Y]] = \text{Var}\left[X' + Y'\right],$$

where here $X' = X - \mathbb{E}[X]$ and $Y' = Y - \mathbb{E}[Y]$. Note that $\text{Var}[X] = \text{Var}[X']$ and $\text{Var}[Y] = \text{Var}[Y']$. Then we have

$$
\begin{aligned}
&= \mathbb{E}\left[(X' + Y')^2\right] && \text{by linearity of expectation} \\
&= \mathbb{E}\left[X'^2 + Y'^2 + 2X'Y'\right] \\
&= \mathbb{E}\left[X'^2\right] + \mathbb{E}\left[Y'^2\right] + \mathbb{E}\left[2X'Y'\right] && \text{by linearity of expectation} \\
&= \text{Var}[X] + \text{Var}[Y] + 2 \cdot \mathbb{E}\left[X'\right] \cdot \mathbb{E}\left[Y'\right] && \text{since } X, Y \text{ are independent} \\
&= \text{Var}[X] + \text{Var}[Y]. && \text{since } \mathbb{E}[X'] = 0
\end{aligned}
$$

□