Lecture 7: Expected value of a random variable

February 14, 2019

1 An example

Imagine a fair 4-faced die, with faces labeled 1,2,3,4. We roll two such dice. Let X be the value that appears on the first die, Y the value that appears on the second die, and $Z = \max(X, Y)$. You should be able to verify

$$P_Z(1) = \frac{1}{16}, P_Z(2) = \frac{3}{16}, P_Z(3) = \frac{5}{16}, P_Z(4) = \frac{7}{16}.$$

Suppose we perform this experiment many times (say N times) over. We would expect that 1 would be the maximum value in approximately $N/16$ trials, 2 the maximum value in $3N/16$ trials, *etc.* So the *average* maximum value over all these trials should be

$$\begin{aligned} \frac{1}{N} \cdot \left(1 \cdot \frac{N}{16} + 2 \cdot \frac{3N}{16} + 3 \cdot \frac{5N}{16} + 4 \cdot \frac{7N}{16}\right) &= \frac{1+6+15+28}{16} \\ &= \frac{50}{16} \\ &= 3.125. \end{aligned}$$

2 Definition.

We can perform the same calculation with any random variable. $E(X)$ denotes the *expected value.* or *expectation,* or *mean* of the random variable X. It really is just what we might expect as the average value of X if we perform the underlying experiment many times. The definition is just the weighted average of the values of X, where the weights are the probabilities:

$$E(X) = \sum_a a \cdot P_X(a).$$

What's the index set of the sum? We could say that we sum over all real numbers a—ordinarily this does not make sense, except in this case it does, because $P_X(a)$ is zero at all but a finite or discrete infinite set of values.

An alternative equivalent definition: If we view the X as a real-valued function defined on a sample space S, then $P_X(a)$ is the sum of $P(s)$ over all s for which $X(s) = a$. As a result,

$$E(X) = \sum_{s \in S} X(s) \cdot P(s).$$

There's a lot of fine print below in the derivation of expected values for standard distributions–read if you're interested, but feel free to skip.

3 Simple examples.

3.1 Bernoulli random variable with parameter p.

The definition gives

$$E(X) = 0 \cdot P_X(0) + 1 \cdot P_X(1) = P_X(1) = p.$$

Which only makes sense: if you flip the biased coin, with head probability p, a bunch of times, the proportion of tosses that yield heads should be p.

3.2 A single die (fair, 6-sided).

Here $P_X(i) = \frac{1}{6}$ for $i = 1, \ldots, 6$. So

$$E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6} \sum_{i=1}^{6} i = \frac{1}{6} \cdot 21 = 3.5.$$

All uniform distributions work this way: the expected value is just the ordinary mean of all the values of X.

3.3 Sum of two dice.

If you look at the graph of the PMF of the sum of two such dice, you can kind of just see from the symmetry that the answer is 7. More precisely, for any i between 0 and 5, $P(X = 7-i) = P(X = 7+i)$. In other words, 2 has the same probability as 12, 3 the same probability as 11, etc. So

$$\begin{aligned}
E(X) &= \sum_{2}^{12} i \cdot P(X = i) \\
&= 7 \cdot P(X = 7) + \sum_{i=1}^{5}((7 - i) + (7 - i)) \cdot P(X = 7 - i) \\
&= 7 \cdot \left[P(X = 7) + \sum_{i=1}^{5} 2 \cdot P(X = 7 - i)\right] \\
&= 7 \cdot \left[P(X = 7) + \sum_{i=1}^{5}(P(X = 7 - i) + P(X = 7 + i))\right] \\
&= 7 \cdot \sum_{i=2}^{12} P(X = i) \\
&= 7 \cdot 1 \\
&= 7.
\end{aligned}$$

(This was an effort to explain *why* the symmetry in the graph makes the value 7, without having to use any calculated values for the probability.) We'll see in a moment how to get this answer even more simply.

4 Linearity of Expectation

If X, Y are random variables defined on the same sample space S, then applying the second formula that we gave for the expected value gives

$$\begin{aligned} E(X+Y) &= \sum_{s\in S} P(s)\cdot(X+Y)(s) \\ &= \sum_{s\in S} P(s)\cdot(X(s)+Y(s)) \\ &= \sum_{s\in S}(P(s)\cdot X(s)+P(s)\cdot Y(s)) \\ &= \sum_{s\in S} P(s)\cdot X(s)+\sum_{s\in S} P(s)\cdot Y(s) \\ &= E(X)+E(Y). \end{aligned}$$

Observe that this does not require any additional assumptions on X and Y (in particular, it doesn't matter if they are independent or not). In a like manner, if $X_1,\ldots,X_n$ are all defined on S then

$$E(X_1+\cdots+X_n)=E(X_1)+\cdots+E(X_n),$$

and if $c\in\mathbf{R}$ is constant, then

$$E(cX)=c\cdot E(X).$$

This lets us redo our example with the pair of dice, using the earlier result for a single die: If X denotes the sum of the values on the two dice, then $X=X_1+X_2$, where X_i denotes the value of the i^{th} die. Thus

$$E(X)=E(X_1+X_2)=E(X_1)+E(X_2)=3.5+3.5=7.$$

5 Binomial Random Variable

As we saw, a binomial random variable X with parameters n,p is the sum of n Bernoulli random variables each with parameter p, so by our sum formula,

$$E(X)=p+\cdots+p=np.$$

If you tried to get this directly from the definition of expected value, you'd have to prove

$$\sum_{k=0}^{n} k\cdot\binom{n}{k}p^k(1-p)^{n-k}=np.$$

In fact, we just proved it! using our formula for the expected value of a sum, and the expected value of a Bernoulli random variable. It's not all that hard to prove this identity directly, but what we just did makes it *so* much easier.

Observe that this is the *obvious* answer: If you toss a biased coin n times, you'd expect, on average, to get np heads, where p is the probability of heads. (The theme of much of this lecture could be 'The expected value is what you expect it is.')

6 Geometric Random Variable

Again, there is a sort of obvious guess: If the probability of heads is, say, $\frac{1}{3}$, then the number of tosses until you get heads seems like it ought to be 3 on the average. In general, if X is a random variable with geometric distribution with parameter p, then the guess is $E(X) = \frac{1}{p}$. This is correct.

The textbook demonstrates this in two different ways, here is still another way. This is not the most rigorous argument, because it plays fast and loose extending the linearity for random variables to infinite sums, which would require some formal justification:

Let X be a geometrically distributed random variable with parameter p. For each $i > 0$, let X_i be the random variable that has the value 1 if the first $i-1$ tosses are tails. So, for example, $X_i = 0$ if the game terminates in $i-1$ or fewer tosses. Then

$$X = X_1 + X_2 + \cdots .$$

Each X_i is a Bernoulli random variable with parameter $(1-p)^{i-1}$, and thus $E(X_i) = (1-p)^{i-1}$. By linearity of expectation:

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} E(X_i) \\ &= \sum_{i=0}^{\infty} (1-p)^{i-1} \\ &= \frac{1}{1-(1-p)} \\ &= \frac{1}{p}. \end{aligned}$$

7 Hypergeometric Distribution

Recall that the underlying problem is choosing n balls from a bin that contains K red balls and N white balls. The random variable is the number of red balls selected. If N and K are very much larger than n, then this sampling-without-replacement problem starts to look like sampling-with-replacement, where the distribution is binomial with parameters n and $p = \frac{K}{N}$. This suggests that at least for large N we should have $E(X)$ approximately $n \cdot p = nK/N$, and in fact that is the exact answer, correct for all values of K, N, n.

Here is a somewhat informal proof of this fact by induction on N. If $N = 1$ then we have $E(X) = 0$ or 1, depending on whether the one ball is red or white. Now suppose that the formula is correct for $N-1$, where $N > 1$. We will use the notation $X_{N,K,n}$ to denote a hypergeometric random variable with parameters N, K, n. Now think of the underlying experiment as consisting of first, a draw of a single ball, followed by a draw of the remaining $n-1$ balls. K/N of the time, the first ball will be red. Since the second phase of the experiment consists of drawing $n-1$ balls from a bin with $N-1$ balls, $K-1$ of which are red, for these instances the average number of red balls will be

$$1 + E(X_{N-1,K-1,n-1}).$$

The alternative is that we draw a white ball first, and in the second phase we draw $n-1$ balls from a bin with $N-1$ balls and K red balls. We put these two results together and apply the inductive hypothesis:

$$\begin{aligned} E(X_{N,K,n}) &= \frac{K}{N} \cdot (1 + E(X_{N-1,K-1,n-1})) + (1 - \frac{K}{N}) \cdot E(X_{N-1,K,n-1}) \\ &= \frac{K}{N} \cdot \big(1 + (n-1) \cdot \frac{K-1}{N-1}\big) + (1 - \frac{K}{N}) \cdot (n-1) \cdot \frac{K}{N-1}. \end{aligned}$$

After a good deal of disagreeable algebra–but no clever tricks!–this all collapses to nK/N, as required.

8 Poisson Distribution

Remember that the Poisson distribution with $\lambda = np$ is a good approximation to the binomial distribution with small p and large n, at least for small values of the random variable X. So we might guess that the Poisson distribution has the same expected value as the binomial distribution, namely $np = \lambda$. Once again, the guess is correct.

The fine print: By definition,

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \cdot e^{-\lambda} \\ &= \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \cdot e^{-\lambda} \\ &= \lambda \cdot \sum_{k=0}^{\infty} \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda \cdot 1 \\ &= 1. \end{aligned}$$

9 Function of a random variable

It is *not* true in general that $E(X^2) = E(X)^2$, or $E(|X|) = |E(X)|$, or $E(\max(X,Y)) = \max(E(X), E(Y))$. If we have a random variable X and define a new random variable $Y = f(X)$, then there is no simple formula giving $E(f(X))$ in terms of $E(X)$. Instead, we have:

$$E(Y) = \sum_a f(a) \cdot P_X(a).$$

As an illustration, let's return to our example at the beginning of the lecture, with the two four-sided dice.

$$\begin{aligned} Z &= \max(X,Y) \\ &= \frac{X+Y+|X-Y|}{2}. \end{aligned}$$

$X - Y$ takes on the values ± 3 each with probability $\frac{1}{16}$, ± 2 each with probability $\frac{2}{16}$, ± 1 each with probability $\frac{3}{16}$ and 0 with probability $\frac{1}{4}$. It follows from this and the above that

$$\begin{aligned} E(|X-Y|) &= \sum_{j=-3}^{3} |j| \cdot P_{X-Y}(j) \\ &= 2 \cdot (3 \cdot \frac{1}{16} + 2 \cdot \frac{2}{16} + 1 \cdot \frac{3}{16}) \\ &= \frac{6+8+6}{16} \\ &= 1.25. \end{aligned}$$

From linearity of expectation we get

$$E(Z) = \frac{1}{2} \cdot (E(|X-Y|) + E(X) + E(Y)) = \frac{1}{2}(1.25 + 2.5 + 2.5) = 3.125,$$

which is what we found earlier.

10 St. Petersburg Paradox—not every random variable has an expected value

Imagine we play the coin-tossing game where we toss repeatedly until the coin shows heads, however in this version we get paid as a function of how long the game lasts. If the game ends on the first toss, we win nothing, but if ends on the second toss we win 1 dollar, on the third toss 2 dollars, on the fourth toss 4 dollars, and in general, if the game lasts exactly i tosses, where $i > 1$,we win 2^{i-2} dollars. Let Y denote the amount we win. Then $E(Y)$ is given by the infinite series

$$1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + \cdots = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \cdots .$$

This series has no limit, but grows to infinity as the number of summands increases Thus $E(Y)$ is not defined.

What does this mean if you actually play this game? It cannot mean, 'on average, I will win an infinite amount of money'. What will happen if you play this game repeatedly?