Discrete Random Variables: Expected Value
Class 4, 18.05, Spring 2014
Jeremy Orloff and Jonathan Bloom

1 Expected Value

In the R reading questions for this lecture, you simulated the average value of rolling a die many times. You should have gotten a value close to the exact answer of 3.5. To motivate the formal definition of the average, or *expected value*, we first consider some examples.

Example 1. Suppose we have a six-sided die marked with five 5 3's and one 6. (This was the red one from our non-transitive dice.) What would you *expect* the average of 6000 rolls to be?

answer: If we knew the value of each roll, we could compute the average by summing the 6000 values and dividing by 6000. Without knowing the values, we can compute the *expected* average as follows.

Since there are five 3's and one six we expect roughly 5/6 of the rolls will give 3 and 1/6 will give 6. Assuming this to be exactly true, we have the following table of values and counts:

value:	3	6
expected counts:	5000	1000

The average of these 6000 values is then

$$\frac{5000 \cdot 3 + 1000 \cdot 6}{6000} = \frac{5}{6} \cdot 3 + \frac{1}{6} \cdot 6 = 3.5$$

We consider this the expected average in the sense that we 'expect' each of the possible values to occur with the given frequencies.

Example 2. We roll two dice. You win \$1000 if the sum is 2 and lose \$100 otherwise. How much do you expect to win on average per trial?

answer: The probability of a 2 is 1/36. If you play N times, you can 'expect' $N/36$ of the trials to give a 2 and $35N/36$ of the trials to give something else. Thus your total expected winnings are

$$1000 \cdot \frac{N}{36} - 100 \cdot \frac{35N}{36}.$$

To get the expected average per trial we divide the total by N:

$$\text{expected average} \ = 1000 \cdot \frac{1}{36} - 100 \cdot \frac{35}{36} = -69.44.$$

Think: Would you be willing to play this game one time? Multiple times?

Notice that in both examples the sum for the expected average consists of terms which are a value of the random variable times its probabilitiy. This leads to the following definition.

Definition: Suppose X is a discrete random variable that takes values $x_1, x_2, \ldots, x_n$ with probabilities $p(x_1), p(x_2), \ldots, p(x_n)$. The *expected value* of X is denoted $E(X)$ and defined by

$$E(X) = \sum_{j=1}^{n} p(x_j)\, x_j = p(x_1)x_1 + p(x_2)x_2 + \ldots + p(x_n)x_n.$$

Notes:

1. The expected value is also called the **mean** or **average** of X and often denoted by μ ("mu").
2. As seen in the above examples, the expected value need not be a possible value of the random variable. Rather it is a weighted average of the possible values.
3. Expected value is a **summary statistic**, providing a measure of the *location* or *central tendency* of a random variable.
4. If all the values are equally probable then the expected value is just the usual average of the values.

Example 3. Find $E(X)$ for the random variable X with table:

values of X:	1	3	5
pmf:	1/6	1/6	2/3

<u>**answer:**</u> $E(X) = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 3 + \frac{2}{3}\cdot 5 = \frac{24}{6} = 4$

Example 4. Let X be a Bernoulli(p) random variable. Find $E(X)$.

<u>**answer:**</u> X takes values 1 and 0 with probabilities p and $1 - p$, so

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

Note: This is an important example. Be sure to remember that the expected value of a Bernoulli(p) random variable is p.

Think: What is the expected value of the sum of two dice?

1.1 Mean and center or mass

You may have wondered why we use the name 'probability mass function'. Here's the reason: if we place an object of mass $p(x_j)$ at position x_j for each j, then $E(X)$ is the position of the center of mass. Let's recall the latter notion via an example.

Example 5. Suppose we have two masses along the x-axis, mass $m_1 = 500$ at position $x_1 = 3$ and mass $m_2 = 100$ at position $x_2 = 6$. Where is the center of mass?

<u>**answer:**</u> Intuitively we know that the center of mass is closer to the larger mass.

From physics we know the center of mass is

$$\overline{x} = \frac{m_1x_1 + m_2x_2}{m_1 + m_2} = \frac{500 \cdot 3 + 100 \cdot 6}{600} = 3.5.$$

We call this formula a 'weighted' average of the x_1 and x_2. Here x_1 is weighted more heavily because it has more mass.

Now look at the definition of expected value $E(X)$. It is a weighted average of the values of X with the weights being probabilities $p(x_i)$ rather than masses! We might say that "The expected value is the point at which the distribution would balance". Note the similarity between the physics example and Example 1.

1.2 Properties of $E(X)$

When we add, scale or shift random variables the expected values do the same.

1. If X and Y are random variables on a sample space Ω then

$$E(X + Y) = E(X) + E(Y)$$

2. If a and b are constants then $E(aX + b) = aE(X) + b$.

We will think of $aX + b$ as scaling X by a and shifting it by b.

Before proving these properties, let's consider a few examples.

Example 6. Roll two dice and let X be the sum. Find $E(X)$.
answer: Let X_1 be the value on the first die and let X_2 be the value on the second die. Since $X = X_1 + X_2$ we have $E(X) = E(X_1) + E(X_2)$. Earlier we computed that $E(X_1) = E(X_2) = 3.5$, therefore $E(X) = 7$.

Example 7. Let $X \sim$ binomial(n, p). Find $E(X)$.

answer: Recall that X models the number of successes in n Bernoulli(p) random variables, which we'll call $X_1, \ldots X_n$. The key fact here is that

$$X = \sum_{j=1}^{n} X_j.$$

This is true because each X_j is either 0 or 1, so the sum just counts the number that are 1, i.e. the number of successes –exactly like X.

Now we can use property (1) to make the calculation simple.

$$X = \sum_{j=1}^{n} X_j \;\Rightarrow\; E(X) = \sum_j E(X_j) = \sum_j p = \boxed{np}.$$

We could have computed $E(X)$ directly as

$$E(X) = \sum_{k=0}^{n} kp(k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}.$$

It is possible to show that the sum of this series is indeed np. We think you'll agree that the method using property (1) much easier.

Example 8. (The mean does not always exist.)
Suppose X has an infinite number of values according to the following table

values x:	2	2^2	2^3	$\dots$	2^k	$\dots$
pmf $p(x)$:	$1/2$	$1/2^2$	$1/2^3$	$\dots$	$1/2^k$	$\dots$

Thus the mean is

$$E(X) = \sum_{k=1}^{\infty} 2^k \frac{1}{2^k} = \sum_{k=1}^{\infty} 1 = \infty.$$

1.3 Proofs of the properties of $E(X)$

The proof of property (1) is simple, but there is some subtlety in even understanding what it means to add two random variables. Recall that the value of random variable is a number determined by the outcome of an experiment. To add X and Y means to add the values of X and Y for the same outcome. In table form this looks like:

outcome ω:	ω_1	ω_2	ω_3	$\dots$	ω_n
value of X:	x_1	x_2	x_3	$\dots$	x_n
value of Y:	y_1	y_2	y_3	$\dots$	y_n
value of $X+Y$:	x_1+y_1	x_2+y_2	x_3+y_3	$\dots$	x_n+y_n
prob. $P(\omega)$:	$P(\omega_1)$	$P(\omega_2)$	$P(\omega_3)$	$\dots$	$P(\omega_n)$

The proof of (1) follows immediately:

$$E(X+Y) = \sum (x_i + y_i)P(\omega_i) = \sum x_i P(\omega_i) + \sum y_i P(\omega_i) = E(X) + E(Y).$$

The proof of property (2) only takes a line.

$$E(aX+b) = \sum p(x_i)(ax_i + b) = a\sum p(x_i)x_i + b\sum p(x_i) = aE(X) + b.$$

The b term in the last expression follows because $\sum p(x_i) = 1$.

Example 9. Mean of a geometric distribution
Let $X \sim \text{geo}(p)$. Recall this means X takes values $k = 0,\ 1,\ 2,\ \dots$ with probabilities $p(k) = (1-p)^k p$. (X models the number of tails before the first heads in a sequence of Bernoulli trials.) The mean is given by

$$E(X) = \frac{1-p}{p}.$$

To see this requires a clever trick. We hope you are able to follow the logic. We won't ask you to come up with something like this on your own. Here's the trick.

To compute $E(X)$ we have to sum the power series

$$E(X) = \sum_{k=0}^{\infty} k(1-p)^k p.$$

Here is the trick. We know the sum of the geometric series: $\displaystyle\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$.

Differentiate both sides: $\displaystyle\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$.

Multiply by x: $\displaystyle\sum_{k=0}^{\infty} kx^k = \frac{x}{(1-x)^2}$.

Replace x by $1-p$: $\displaystyle\sum_{k=0}^{\infty} k(1-p)^k = \frac{1-p}{p^2}$.

Multiply by p: $\displaystyle\sum_{k=0}^{\infty} k(1-p)^k p = \frac{1-p}{p}$.

This last expression is the mean.

$$E(X) = \frac{1-p}{p}.$$

Example 10. Flip a fair coin until you get heads for the first time. What is the expected number of times you flipped tails?
answer: The number of tails before the first head is modeled by $X \sim \text{geo}(1/2)$. From the previous example $E(X) = \dfrac{1/2}{1/2} = 1$. This is a surprisingly small number.

Example 11. Michael Jordan, the greatest basketball player ever, made 80% of his free throws. In a game what is the expected number he would make before his first miss.
answer: Here is an example where we want the number of successes before the first failure. Using the neutral language of heads and tails: success is tails (probability $1-p$) and failure is heads (probability $= p$). Therefore $p = .2$ and the number of tails (made free throws) before the first heads (missed free throw) is modeled by a $X \sim \text{geo}(.2)$. We saw in Example 9 that this is

$$E(X) = \frac{1-p}{p} = \frac{.8}{.2} = 4.$$

Functions of a Random Variable (*the change of variables formula*)
If X is a discrete random variable taking values $x_1, x_2, \ldots$ and h is a function the $h(X)$ is a new random variable. Its expected value is

$$E(h(X)) = \sum_j h(x_j)p(x_j).$$

We illustrate this with several examples.
Example 12. Let X be the value of a roll of one die and let $Y = X^2$. Find $E(Y)$.
answer: Since this is a discrete random variable we can make a table.

X	1	2	3	4	5	6
Y	1	4	9	16	25	36
prob	1/6	1/6	1/6	1/6	1/6	1/6

Notice the probability for each Y value is the same as that of the corresponding X value. So,

$$E(Y) = E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} = 15.167.$$

Example 13. Roll two dice and let X be the sum. Suppose the payoff function is given by $Y = X^2 - 6X + 1$. Is this a good bet?

answer: We have $E(Y) = \sum_{j=2}^{12} (j^2 - 6j + 1)p(j)$, where $p(j) = P(X = j)$.

We show the table, but really we'll use R to do the calculation.

X	2	3	4	5	6	7	8	9	10	11	12
Y	-7	-8	-7	-4	1	8	17	28	41	56	73
prob	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Here's the R code I used to compute $E(Y) = 13.833$.

```
x = 2:12
y = x^2 - 6*x + 1
p = c(1 2 3 4 5 6 5 4 3 2 1)/36
ave = sum(p*y)
```

It gave ave = 13.833.

To answer the question above: since the expected payoff is positive it looks like a bet worth taking.

(In this example, note that $P(Y = -7) = P(X = 2) + P(X = 4)$ because both $X = 2$ and $X = 4$ give $Y = -7$.).

Quiz: If $Y = h(X)$ does $E(Y) = h(E(X))$? **answer: NO!!!** This is not true in general!

Think: Is it true in the previous example?

Quiz: If $Y = 3X + 77$ does $E(Y) = 3E(X) + 77$?
answer: Yes. By property (2), scaling and shifting does behave like this.