CLARK UNIVERSITY 1887

Expectation
Math 217 Probability and Statistics
Prof. D. Joyce, Fall 2014

The definition of expected value in the discrete case. Let X be a discrete random variable. If you average the values that X takes on, weighted by the probabilities that X takes on those values, that is, weighted by the probability mass function, you get

$$E(X) = \sum_{x \in S} x\, P(X{=}x) = \sum_{x \in S} x\, f(x)$$

This is called the *expected value*, *mathematical expectation*, or *mean* of X, denoted both $E(X)$ and μ_X. When there's only one random variable under discussion, μ_X is usually abbreviated to just μ.

When X is a uniform discrete random varable, then $E(X)$ is just the arithmetic average of the values of X. For example, the expected value of a fair die is the average of the numbers 1, 2, 3, 4, 5, and 6, and that average is 3.5 since

$$3.5 = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6}.$$

Expectation of the binomial distribution. Let X_n be the number of successes in n Bernoulli trials, where the probability of success is p. This random variable X_n has a binomial distribution. We know its probability mass function is $f(x) = \binom{n}{x} p^x q^{n-x}$. So the expectation of X_n is

$$E(X_n) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}.$$

Let's write k for x just so that it looks a little more familiar to us.

$$E(X_n) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

We can evaluate that sum on the right by differentiating the equation of the binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Differentiate that equation with respect to x to get

$$n(x+y)^{n-1} = \sum_{k=0}^{n} k \binom{n}{k} x^{k-1} y^{n-k}.$$

Now replace x by p, y by q, and multiply both sides of the equation by p, and we find that

$$E(X_n) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = np.$$

Therefore, the expected number of successes in n Bernoulli trials is n times the probability of success on one trial.

Expectation of a geometric distribution. Consider the time T to the first success in a Bernoulli process. With p equal to the probability of success, we found that T has geometric distribution with the probability mass function $f(t) = pq^{t-1}$. Therefore, its expectation is

$$\begin{aligned} E(T) &= \sum_{t=1}^{\infty} tpq^{t-1} \\ &= p + 2pq + 3pq^2 + \cdots + npq^{n-1} + \cdots \end{aligned}$$

To evaluate it, we'll need to sum the infinite series somehow. We can do that using properties of power series.

We know that the geometric series

$$1 + x + x^2 + \cdots + x^n + \cdots$$

converges to $\dfrac{1}{1-x}$ when the ratio x lies between -1 and 1. We can differentiate power series, so differentiate

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots$$

to conclude

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \cdots + nx^{n-1} + \cdots$$

for x between -1 and 1. That's just the series we have for $E(T)$ except $E(T)$ has an extra factor of p, and q is replaced by x. Therefore,

$$E(T) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Thus, the expected time to the first success in a Bernoulli process is the reciprocal of the probability of success. So, on average, it takes 2 coin flips to get the first head. Also, on average, it takes 6 die tosses to roll the first 5.

The St. Petersburg paradox. This an example of a random variable with and infinite expectation. It's a game of flipping coins where your payoff doubles every time you get tails. If you get heads right off, you'll be paid \$1. If you get tails followed by heads, you'll get \$2. If two tails before the first heads, then \$4. Your payoff is $Y = 2^T$ where T has a geometric distribution with $p = 1/2$. Thus, $P(Y{=}2^n) = 1/2^n$, so the expectation of Y is

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \sum_{n=1}^{\infty} 1 = \infty.$$

The paradox comes when you try to figure out how much you should pay in order to play this game. Note that it has an infinite expected payoff.

Some properties of expectation. There are a number of these properties. Some are easier to prove than others.

Expectation of a function of a random variable. If a random variable Y is a function of another random variable X, $Y = \phi(X)$, then

$$E(Y) = \sum_{x \in S} \phi(x)\, P(X{=}x).$$

To show this, we'll expand the definition of the expectation of Y.

$$\begin{aligned} E(Y) &= \sum_y y\, P(Y{=}y) \\ &= \sum_y y \sum_{x \,|\, y=\phi(x)} P(X{=}x) \\ &= \sum_x \phi(x)\, P(X{=}x) \end{aligned}$$

Linearity of expectation. If X and Y are two random variables, not necessarily independent, then

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

Also, if c is a constant, then

$$E(cX) = cE(X).$$

Both of these properties are easy to prove. They only depend on linearity property of sums, that is,

$$\sum (x_k + y_k) = \sum x_k + \sum y_k$$

and

$$\sum c x_k = c \sum x_k.$$

Expectation preserves products when the variables are independent. Recall that random variables X and Y are *independent* when for all their outcomes x and y,

$$P(X{=}x \cap Y{=}y) = P(X{=}x)\, P(Y{=}y).$$

If X any Y are independent, then

$$E(XY) = E(X)\, E(Y),$$

but when they're not independent, that equation is usually false. Here's the proof when they're independent. Note the line that uses independence.

$$\begin{aligned}
E(XY) &= \sum_z z\,P(XY{=}z) \\
&= \sum_z \sum_{x,y\,|\,xy=z} xy\,P(X{=}x \text{ and } Y{=}y) \\
&= \sum_x \sum_y xyP(X{=}x \text{ and } Y{=}y) \\
&= \sum_x \sum_y xyP(X{=}x)\,P(Y{=}y) \\
&= \sum_x x \left(\sum_y yP(Y{=}y) \right) P(X{=}x) \\
&= \left(\sum_y yP(Y{=}y) \right) \left(\sum_x xP(X{=}x) \right) \\
&= E(Y)\,E(X)
\end{aligned}$$

Q.E.D.

Math 217 Home Page at `http://math.clarku.edu/~djoyce/ma217/`