Expectation Pt. 1

The Basics

Elliot Pickens

Nov. 10, 2021

1 Intro

Random variables, as powerful as they are, can be unwieldy tools. We've seen how we can use their distributions to find probabilities, quantiles, and other sorts of values, but it's hard to use those indicators alone to get a good understanding of a distribution. To remedy this we can use expectation calculations. Just as the name suggests, these calculations are used to find out what we can *expect* from a distribution without having to handle the extra baggage.

2 Expectation of Discrete Distributions

To calculate the *mean* of a random variable, which in this post we will refer to as the *expected value* of said random variable, we adapt our basic understanding of a mean to fit their properties. A simple way to reach our eventual definition is to begin with a random sample of integers for which we'd like to calculate the mean. Let's say that sample is $\{1, 1, 1, 2, 3, 4, 5, 7\}$. Then its mean is $(1 + 1 + 1 + 2 + 3 + 4 + 5 + 7)/8 = 24/8 = 3$. Now let's also say that this sample was taken from random variable X, and assert that each of these integers occurred at their true rate of occurrence. E.g., $P(X = 3) = 3/8$, $P(X = 2) = 1/8$, etc. Then we can calculate the mean of X as $\frac{3}{8}(1) + \frac{1}{8}(2 + 3 + 4 + 5 + 7) = \frac{3}{8} + \frac{21}{8} = \frac{24}{8} = 3$. Although this example may not be pretty, it shows how we translate the concept of a mean (or expected value) to a random variable. We multiply each possible value with it's rate of occurrence (probability) and then sum over all such products.

Thus, the definition of the *expectation* of X (where X is bounded and discrete) is

$$E(X) = \sum_{All\ x} xf(x) \tag{1}$$

where f is the pdf of X.

To generalize this definition to unbounded discrete random variables, we first split our sum over all x into sums over the negative and positive values. If either of these sums, which we can write as

$$\sum_{Positive\ x} xf(x), \qquad \sum_{Negative\ x} xf(x) \tag{2}$$

are finite then the expected value of X exists and is defined as

$$E(X) = \sum_{All\ x} xf(x) \tag{3}$$

which is equivalent to 1.

The reason why one of these sums must be finite is that if both are infinite then we either end up with a combined sum that fails to converge, or can be manipulated to converge to a number of values depending on how the two sums are reordered when they are merged into one. Under those difficult convergence conditions we cannot define 3. It should be noted, however, that when we say no expectation exists if both sums in 5 are infinite we are not saying that expectation is infinite. We can get infinite expected values where $E(X) = \pm\infty$ and either the positive or negative sum is unbounded, but when we say the expectation is undefined it simply cannot be calculated and does not exist.

I should also clarify that while we say $E(X)$ is the expectation of X, the expectation only really depends on the distribution of X. So long as any number of slightly different random variables have the same distribution they will all have the same expected value.

3 Expectation of Continuous Distributions

For continuous random variables we need to slightly alter our weighted sum of outcomes to properly calculate expected value. As you might expect, the change needed is the swapping of a sum for an integral. A bounded continuous random variable X with pdf f then has an expected value that is defined by the equation:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \tag{4}$$

For unbounded continuous random variables we once again have to split things up to ascertain the existence of expectation. We do this by breaking the integral in 4 into one from $-\infty$ to 0 and another from 0 to ∞.

$$\int_{-\infty}^{0} x f(x) dx, \quad \int_{0}^{\infty} x f(x) dx \tag{5}$$

If either of these integrals is finite then the expected value exists and is

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \tag{6}$$

Otherwise $E(X)$ cannot be determined and does not exist.

3.1 Univariate Functions of a Random Variable

If we can find the expectation of a random variable X then we should also be able to find the expectation of functions of X. To do this we need to substitute $R(X) = Y$ and $r(x) = y$ into 4 to get

$$E[r(X)] = E(Y) = \int_{-\infty}^{\infty} y g(y) dy \tag{7}$$

assuming the expectation of Y exists.

3.2 Law of the Unconscious Statistician

We can simplify 7 with the law of the unconscious statistician, which states that for a random variable X and a real valued function r then

$$E[r(X)] = \int_{-\infty}^{\infty} r(x) f(x) dx \tag{8}$$

if X is continuous and it has an expectation/mean that exists. When X is discrete (with a mean that exists) then we get

$$E[r(X)] = \sum_{All\ x} r(x) f(x) \tag{9}$$

Now let's build a little intuition as to how exactly this works by looking a few short proofs of special cases. First let's assume that we have some discrete distribution X . Then after applying a function to X our new distribution that we'll call Y will also be discrete. Then if we call the pf of Y g then we get:

$$\sum_y yg(y) = \sum_y yP(r(X) = y) \tag{10}$$
$$= \sum_y y \sum_{x:r(x)=y} f(x) \tag{11}$$
$$= \sum_y \sum_{x:r(x)=y} r(x)f(x) \tag{12}$$
$$= \sum_x r(x)f(x) \tag{13}$$

Which is exactly what we stated just a few short lines ago in 9.

Now for a "more special" special case that gives us some intuition about the continuous version of the law of the unconscious statistician. To set this up, let X be continuous and $r(x)$ be a strictly increasing or strictly decreasing differentiable function with inverse $s(y)$. Then

$$\int_{-\infty}^{\infty} r(x)f(x)dx = \int_{-\infty}^{\infty} yf[s(y)]\left|\frac{ds(y)}{dy}\right| dy \tag{14}$$
$$= \int_{-\infty}^{\infty} yg(y)dy \tag{15}$$

since if r is increasing we have

$$G(y) = P(Y \leq y) = P(r(X) \leq y) = P(X \leq s(y) = F(s(y))) \tag{16}$$

and

$$g(y) = \frac{dG(y)}{dy} = \frac{dF(s(y))}{dy} = f(s(y))\frac{ds(y)}{dy} \tag{17}$$

For a decreasing r things are almost the same. We can work them out as

$$G(y) = P(Y \leq y) = P(r(X) \leq y) = P(X \geq s(y) = 1 - F(s(y))) \tag{18}$$

which when differentiated becomes

$$g(y) = \frac{dG(y)}{dy} = -f(s(y))\frac{ds(y)}{dy} \tag{19}$$

Between these two examples (10, 14) we can see the contours of a full proof, or at the very least that it holds true in these few cases.

3.2.1 Multivariate Functions of Random Variables

Naturally the *Law of the Unconscious Statistician* also works when working with (potentially very complex) multivariate distributions. Say we have some set of random variables $X_1, ..., X_n$ that form a joint distribution with joint pdf $f(x_1, ..., x_n)$. Then if we have some function r that can transform n random variables we can create $Y = r(X_1, ..., X_n)$. The expectation of Y can then be found by using the following setup:

$$E(Y) = \int \underset{R^n}{\cdots} \int r(x_1, ..., x_n)f(x_1, ..., x_n)dx_1 \dots dx_n \tag{20}$$

or

$$E(Y) = \sum_{All\ x_1,...,x_n} r(x_1, ..., x_n)f(x_1, ..., x_n) \tag{21}$$

if our random variables form a discrete joint distribution.

4 Conclusion

In this post I covered the basic definition of the expectation of a random variable. Hopefully, this post didn't over complicate the basics too much. The real goal of this write up was to explain how expectation can be calculated under various circumstances.

In the next installation of this series I'll talk about something quite a bit more interesting (and important): properties of expectation.