Expectation Pt. 3

Properties of Expectation

Elliot Pickens

Jul 19, 2021

1 Intro

In this post I'll cover a few important properties of expectation. I my previous post (expectation pt. 1) I laid out the basics, but without understanding of a few additional properties it can be difficult to apply those definitions. Hopefully this post will help with that.

2 Properties

2.1 Linear Functions

To begin let's take a look at the expectation of a random variable Y defined as a linear function of another random variable X as $Y = aX + b$. Then the expectation of Y is

$$E(Y) = aE(X) + b \tag{1}$$

since[1]

$$
\begin{aligned}
E(Y) = E(aX+b) = \int_{-\infty}^{\infty}(ax+b)f(x)dx &= \int_{-\infty}^{\infty} axf(x) + bf(x)dx && (2)\\
&= a\int_{-\infty}^{\infty} xf(x)dx + b\int_{-\infty}^{\infty} f(x)dx && (3)\\
&= aE(X) + b && (4)
\end{aligned}
$$

2.2 Other Basic Theorems

Due to a jumbling of the originally intended order of my posts I realized that I referenced a few minor details about expectation in my previous post on variance. The theorems below should help fill in some of those details. I will, however, only be showing the continuous cases for these theorems for the sake of brevity given that the ideas are identical after replacing $\int_{-\infty}^{\infty}$ with $\sum_{All\ x}$.

First off we can say that if we have a random variable X for which $P(X \geq a) = 1$ then $E(X) \geq a$. Conversely if $P(X \leq b) = 1$ then $E(X) \leq b$. We can see this by applying two inequalities to the expectation calculation. For the greater than case

$$
\begin{aligned}
E(X) = \int_{-\infty}^{\infty} xf(x)dx &= \int_{a}^{\infty} xf(x)dx && (5)\\
&\geq \int_{a}^{\infty} af(x)dx = a && (6)
\end{aligned}
$$

and for the less than case

[1] The explanation in 2 is for the continuous case, but for the discrete case we can follow the same argument after replacing integrals with sums over all x.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{b} xf(x)dx \tag{7}$$
$$\leq \int_{-\infty}^{b} bf(x)dx = b \tag{8}$$

We can use this result to show a few other things like that $P(X = a) = 1$ if $E(X) = a$ and either $P(X \geq a) = 1$ or $P(X \leq a) = 1$. Plugging things in we get

$$E(X) = a = \int_{-\infty}^{\infty} xf(x)dx = \int_{a}^{\infty} xf(x)dx \tag{9}$$
$$\geq \int_{a}^{\infty} af(x)dx = a \tag{10}$$

for the $P(X \geq a) = 1$ case. Since the inequality above must actually be a strict equality then we know that $P(X > a) = 0$, because $\int_a^\infty xf(x)dx = \int_a^\infty af(x)dx$. Therefore $P(X = a) = 1$.

The opposite direction is much the same, as for $P(X \leq a) = 1$ we get

$$E(X) = a = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{a} xf(x)dx \tag{11}$$
$$\leq \int_{-\infty}^{a} af(x)dx = a \tag{12}$$

which implies $P(X < a) = 0$ (due the equality being strict) and $P(X \leq a) = 1$.

2.2.1 Linearity of Expectation

If we have a number of random variables $X_1, ..., X_n$ then the expectation of their sum $E(X_1 + ... + E_n) = E(X_1) + ... + E(X_n)$. To sketch the proof of this let's take a look at the sum of two random variables. The expectation of the sum of two random variables is

$$E(X_1 + X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) f(x_1, x_2) dx_1 dx_2 \tag{13}$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 f(x_1, x_2) dx_2 dx_1 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 f(x_1, x_2) dx_1 dx_2 \tag{14}$$
$$= \int_{-\infty}^{\infty} x_1 f(x_1) dx_1 + \int_{-\infty}^{\infty} x_2 f(x_2) dx_2 \tag{15}$$
$$= E(X_1) + E(X_2) \tag{16}$$

Using this result we can state that the base case of an inductive proof is $E(X_1) = E(X_1)$. We want to show $E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n)$, so let's put it in the form of the two RV case as

$$E(X_1 + ... + X_n) = E(X_n + \sum_{j=1}^{n-1} X_j) \tag{17}$$
$$= E(X_n) + E(\sum_{j=1}^{n-1} X_j) \tag{18}$$

where $E(\sum_{j=1}^{n-1} X_j)$ is the $n - 1^{th}$ case.

This is a bit of an interesting result, because it implies that the expectation of the sum of random variables is equal to the sum of their individual expectations even when the RVs are not independent. The fact that we can forego the joint distribution and use the marginal distributions is a neat detail to note.

Combining this with 1 we get that

$$E(a_1X_1 + ... + a_nX_n + b) = a_1E(X_1) + ... + a_nE(X_n) + b \tag{19}$$

2.3 Products of Independent Random Variables

For n independent random variables $X_1, ..., X_n$ where $E(X_i)$ is finite $\forall\, i$ the expectation of their product is

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i) \tag{20}$$

To show this let f be the joint distribution of $X_1, ..., X_n$ and f_i be the marginal distribution of each X_i. Then by independence we can say

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f_i(x_i) \tag{21}$$

$\forall$ points $(x_1, ..., x_n) \in \mathbb{R}^n$. Thus,

$$\begin{aligned}
E\left(\prod_{i=1}^{n} X_i\right) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} x_i\right) f(x_1, ..., x_n) dx_1 \dots dx_n && (22)\\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} x_i f_i(x_i)\right) dx_1 \dots dx_n && (23)\\
&= \prod_{i=1}^{n} \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i && (24)\\
&= \prod_{i=1}^{n} E(X_i) && (25)
\end{aligned}$$

2.4 Non-Negative Random Variables

2.4.1 Integer Valued Random Variables

For integer valued random variables we can say that

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) \tag{26}$$

To show this let's inspect its expectation

$$E(X) = \sum_{n=0}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} nP(X = n) \tag{27}$$

We can write out this sum as

$$\begin{matrix}
P(X=1) & P(X=2) & P(X=3) & \dots \\
 & P(X=2) & P(X=3) & \dots \\
 & & P(X=3) & \dots
\end{matrix} \tag{28}$$

By shifting the direction of our summation (from column-wise to row-wise) we can see that

$$\sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} P(X \geq n) \tag{29}$$

Therefore,

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) \tag{30}$$

2.4.2 General Non-negative Random Variables

As a minor note, for a non-negative random variable with a cdf F its expectation is

$$E(X) = \int_0^{\infty} [1 - F(x)]dx \tag{31}$$

3 Conclusion

I don't have much to says about the theorems presented in this post other than that they are quite useful and crop up often in just about any work where probability is heavily used. For that reason this post is something of a reference more than anything.

In my next installment of this series I'll get back on track and start on what I said I would: moments!

4 Acknowledgments

This post was based off *Probability and Statistics* by Degroot and Schervish.