Expectation Pt. 4

Moments

Elliot Pickens

Jan. 16, 2022

1 Intro

Today I'll be talking about moments. I discussed expectation and variance in a couple of recent posts and both are great ways to describe a distribution, but they aren't the only options out there. I wouldn't describe moments as an alternative, but they are a great addition to your statistics toolkit.

2 What is a Moment?

If we have a random variable X and a positive integer k then $E(X^k)$ is the k^{th} moment of X. The mean is the first moment, but certainly not only one. In fact, it is quite possible that a distribution will have a countable number of moments. This is true for all bounded random variables (where $P(a \leq X \leq b) = 1$ for some finite a, b). But it's also possible that the moments exist for unbounded random variables.

That's a little too ambiguous of a statement, so let's get into some details starting with existence. The moment of a distribution exists if $E(|X|^k) < \infty$. And if $E(|X|^k) < \infty$ for some k then the moment also exists for all $0 < j < k$. To show this let's assume we have some continuous RV X with a pdf f. For some j the j^{th} moment is then

$$\begin{aligned}
E(|X|^j) &= \int_{-\infty}^{\infty} |x|^j f(x)dx && (1)\\
&= \int_{|x|\leq 1} |x|^j f(x)dx + \int_{|x|>1} |x|^j f(x)dx && (2)\\
&= \int_{|x|\leq 1} 1 * f(x)dx + \int_{|x|>1} |x|^k f(x)dx && (3)\\
&= P(|X| \leq 1) + E(|X|^k) && (4)
\end{aligned}$$

since $E(|X|^k) < \infty$ then $P(|X| \leq 1) + E(|X|^k) < \infty$ and the j^{th} moment exists.

2.1 Central Moments

If we have an RV for which $E(X) = \mu$ then $E((X-\mu)^k)$ is the k^{th} central moment (or moment about the mean) of X for all positive integers k. By definition the first moment must be zero and the second moment is variance.[1]

Aside from variance, another central moment of interest is *skewness*. Skewness is defined as $E((X-\mu)^3)/\sigma^3$ and is used to describe the symmetry of a distribution.[2] For a symmetric distribution the skewness will be zero, but for all non symmetric distributions skewness with be non zero.

[1] If the distribution is symmetric then all odd moments about the mean will be zero (so long as they exist)

[2] The σ^3 is used to normalize the moment to isolate for symmetry and non the spread of a distribution

3 Moment Generating Functions

We can use moments to describe a distribution by using a moment generating function of mgf. We define such a function as

$$\psi(t) = E(e^{tX}) \tag{5}$$

where $\psi(t)$ is the mgf of X.

As was the case with both variance and expectation, the mgf of a random variable depends on its distribution alone. These functions also related to the moments of a distribution themselves. Although $E(X^k) \neq E(e^{tX})$, the bounded-ness of a random variable has a similar effect to what was described in 2.

But why would we want to use one of these? At first glance a moment generating function doesn't appear to have much in common with moments themselves, but it turns out that after a little work we can use them to do generate the moments of a random variable. To be more specific, on an open interval about 0 for which all values of t within the interval result in $\psi(t)$ being finite, the *nth* derivative of $\psi(t)$ equals the *nth* moment of X and is finite at $t = 0$. That is,

$$E(X^n) = \psi^{(n)}(0) \tag{6}$$

for $n \geq 1$, $n \in \mathbb{Z}$.

3.1 Properties of MGFs

3.1.1 Linear Functions

Let's say we have two random variables X, Y where $Y = aX + b$ with corresponding mgfs ψ_1 and ψ_2. Then for each t where $\psi_1(at)$ is finite, we can expand out ψ_2 using the definition of an mgf as

$$\begin{aligned}
\psi_2(t) &= E(e^{tY}) && (7)\\
&= E(e^{t(aX+b)}) && (8)\\
&= e^{bt}E(e^{atX}) && (9)\\
&= e^{bt}\psi_1(at) && (10)
\end{aligned}$$

thus, $\psi_2(t) = e^{bt}\psi_1(at)$.

3.1.2 Independent RVs

Assume we have a number of independent random variables $X_1, ..., X_n$ with mgfs $\psi_1, ..., \psi_2$. Then if we construct a new random variable $Y = X_1 + ... + X_n$ whose mgf is ψ we know

$$\begin{aligned}
\psi(t) &= E(e^{tY}) && (11)\\
&= E(e^{t(X_1+...+X_n)}) && (12)\\
&= E(\prod_{i=1}^{n} e^{tX_i}) && (13)\\
&= \prod_{i=1}^{n} E(e^{tX_i}) && (14)\\
&= \prod_{i=1}^{n} \psi_i(t) && (15)
\end{aligned}$$

So we have that $\psi(t) = \prod_{i=1}^{n} \psi_i(t)$.

3.1.3 Uniqueness of MGFs

If a pair of mgfs share a several properties then we can infer whether the distributions that define said mgfs are the same. More precisely, we can say that if a pair of mgfs are both finite and identical on an interval about $t = 0$ for every possible value of t then the distributions of the random variables behind the mgfs are identical.

4 Conclusion

In this post I only barely touched upon moments. There are plenty more things that can be said about them, and many more ways they can be used. I've chosen to keep this post short, so that I can move on to other things but it is likely that moments will come up again at some point in the future. They show up all over the place from descriptive statistics to MOM (method of moments) for model fitting.

Before I get around to that, however, I'm going to try and close out my set of posts on expectation. Next up: covariance.

5 Acknowledgments

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