Expectation Pt. 2

Variance

Elliot Pickens

Jul 19, 2021

1 Intro

We can only learn so much about a distribution from its mean. In fact, no two distributions with the same mean are necessarily equivalent (even when they are of the same general class). A uniform distribution and a normal distribution could, for example, both have means of 0. Furthermore, two normal or uniform distributions could each have a mean of zero while being drastically different. To better understand a dataset or distribution of interest we need additional information (in addition to visualization). Variance (and its sibling standard deviation) are one way we can glean extra info about our distribution of interest. In this post we'll explore what exactly variance is, and how it describes the "spread" of a distribution.

2 Variance and Standard Deviation

For a distribution X with mean $E[X] = \mu$ we define its variance as follows,

$$Var(X) = E[(X - \mu)^2] \tag{1}$$

Standard deviation is the square root of variance and often denoted as σ. Thus,

$$\sigma^2 = Var(X) \tag{2}$$

If the mean of a distributions does not exist then neither does the variance (or standard deviation) of a distribution. Similarly, if the expectation of a random variable is infinite then the variance and standard deviation are also infinite. It should also be noted that just as is the case for the mean, variance only depends upon the distribution meaning that any random variable following the same distribution will have the same variance even if there are minor differences between them (so long as those differences aren't serious e.g. a countable number of discontinuities).

By expanding the definition of variance shown in 1 we can express variance using the equivalent definition:

$$Var(X) = E(X^2) - (E(X))^2 \tag{3}$$

which we can show through the expansion:

$$\begin{aligned}
Var(X) &= E[(X - \mu)^2] && (4)\\
&= E(X^2 - 2\mu X + \mu^2) && (5)\\
&= E(X^2) - 2\mu E(X) + \mu^2 && (6)\\
&= E(X^2) - 2\mu^2 + \mu^2 && (7)\\
&= E(X^2) - \mu^2 && (8)\\
&= E(X^2) - (E(X))^2 && (9)
\end{aligned}$$

3 Properties of Variance

Now that we've covered the definition we can take a look at a few useful properties.

3.1 For Bounded Random Variables

First, we can say that for all bounded random random variables variance is ≥ 0 and finite. We can show this by noticing that $(X-\mu)^2$ defines a random variable based upon X, which we have asserted must be bounded. Thus, $(X-\mu)^2$ must also be bounded. And since $(X-\mu)^2 = X^2 - 2X\mu + \mu^2$ we know that $P((X-\mu)^2 \geq \mu^2) = 1$, which implies $E[(X-\mu)^2] \geq \mu^2$ and $Var(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2 \geq 0$.

3.2 Zero Variance RVs

We can also say that a distribution only has zero variance when all of its probability density is confined to a single point i.e. when $P(X = c) = 1$ for some c. This means that $E[X] = c$ and $P((X-c)^2 = 0) = 1$. Therefore,

$$E[(X-c)^2] = Var(X) = 0 \tag{10}$$

Coming from the other direction, if $Var(X) = 0$ then $P((X-\mu)^2 \geq 0) = 1$, but since we also have $E[(X-\mu)^2] = 0$ we get

$$P((X-\mu)^2 = 0) = 1 \tag{11}$$

and

$$P(X = \mu) = 1 \tag{12}$$

3.3 Effect of Constants

Let's say we define a random variable as $Y = aX + b$ where a and b are constants and X is a random variable. Then if $E(X) = \mu$ and $E(Y) = a\mu + b$, which means the variance of Y is

$$\begin{aligned}
Var(Y) &= E[(Y - E(Y))^2] && (13)\\
&= E[(aX + b - a\mu - b)^2] && (14)\\
&= E[(aX - a\mu)^2] && (15)\\
&= a^2 E[(X-\mu)^2] && (16)\\
&= a^2 Var(X) && (17)
\end{aligned}$$

This also means that $\sigma_Y = \sqrt{Var(X)} = |a|\sigma_X$.

It should be noted that $Var(X + b) = Var(X)$. Shifting a RV's outcomes in one direction or another does not change its "spread" so it does not effect variance. Similarly $Var(-X) = Var(X)$, because flipping values from positive to negative does not change the dispersion of a distribution.

3.4 Linearity of Variance

If we have a collection of independent random variables $X_1, ..., X_n$ then the variance of their sum is

$$Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n) \tag{18}$$

We can quickly show this by examining the case where we are only taking the sum of two independent random variables. For this $n = 2$ case we know that $E(X_1) = \mu_1$ and $E(X_2) = \mu_2$, so $E(X_1 + X_2) = \mu_1 + \mu_2$ by linearity of expectation. Thus,

$$\begin{aligned}
Var(X_1 + X_2) &= E[(X_1 + X_2 - \mu_1 - \mu_2)^2]) && (19)\\
&= E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)] && (20)\\
&= Var(X_1) + Var(X_2) + 2(E[X_1 - \mu_1]E[X_2 - \mu_2]) && (21)\\
&= Var(X_1) + Var(X_2) + 2((\mu_1 - \mu_1)(\mu_2 - \mu_2)) && (22)\\
&= Var(X_1) + Var(X_2) && (23)
\end{aligned}$$

since we can perform the transformation in step (21) of 19 due to independence. Now we can sketch an inductive proof. First we establish the base case of $Var(X_1) = Var(X_1)$. Then for some n we want $Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n)$. Since

$$Var(X_1 + ... + X_n) = Var(\sum_{i=0}^{n} X_i) \tag{24}$$

$$= Var(X_n + \sum_{i=0}^{n-1} X_i) \tag{25}$$

$$= Var(X_n + Y) \tag{26}$$

$$= Var(X_n) + Var(Y) \tag{27}$$

by 19 and $Var(Y)$ represents the $n-1$th step. From this we can see that $Var(X_1 + ... + X_2) = \sum_{i=1}^{n} Var(X_i)$ for all n.

3.4.1 Adding in Constants

By combining 3.3 and 3.4 we can say that for a group of independent random variables $X_1, ..., X_n$ with finite means and corresponding groups of constants $a_1, ..., a_n$ and $b_1, ..., b_n$ we have that

$$Var(a_1X_1 + b_1 + ... + a_nX_n + b_n) = Var(a_1X_1 + ... + a_nX_n + b) \tag{28}$$

$$= a_1^2 Var(X_1) + ... + a_n^2 Var(X_2) \tag{29}$$

4 Conclusion

The variance of a random variable is an important summary statistic that can help us understand the spread of a random variable X. We compute this value by finding the mean of the random variable $(X - E(X))^2$, which can also be written as $E(X^2) - E(X)^2$.

In the next post in this series we will examine "moments," which are special set of values associated with a distribution that can us understand it. From there we'll cover a few other miscellaneous topics related to expectation before moving onto other topics in probability.