CLARK UNIVERSITY 1887

Variance and standard deviation
Math 217 Probability and Statistics
Prof. D. Joyce, Fall 2014

Variance for discrete random variables. The variance of a random variable X is intended to give a measure of the spread of the random variable. If X takes values near its mean $\mu = E(X)$, then the variance should be small, but if it takes values from from μ, then the variance should be large.

The measure we'll use for distance from the mean will be the square of the distance from the mean, $(x-\mu)^2$, rather than the distance from the mean, $|x-\mu|$. There are three reasons for using the square of the distance rather than the absolute value of the difference. First, the square is easier to work with mathematically. For instance, x^2 has a derivative, but $|x|$ doesn't. Second, large distances from the mean become more significant, and it has been argued that this is a desirable property. Most important, though, is that the square is the right measure to use in order to derive the important theorems in the theory of probability, in particular, the Central Limit Theorem.

Anyway, we define the variance $\mathrm{Var}(X)$ of a random variable X as the expectation of the square of the distance from the mean, that is,

$$\mathrm{Var}(X) = E\big((X-\mu)^2\big).$$

As the square is used in the definition of variance, we'll use the square root of the variance to normalize this measure of the spread of the random variable. The square root of the variance is called the *standard deviation*, denoted in our text as $\mathrm{SD}(X)$

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{E\big((X-\mu)^2\big)}.$$

Just as we have a symbol μ for the mean, or expectation, of X, we denote the standard deviation of X as σ, and so the variance is σ^2.

Pair of dice. Let's take as an example the roll X of one fair die. We know $\mu = E(X) = 3.5$. What's the variance and standard deviation of X?

$$\begin{aligned}
& \mathrm{Var}(X) \\
= \;& E\big((X-\mu)^2\big) \\
= \;& \sum_{x=1}^{6} (x-3.5)^2\, P(X=x) \\
= \;& \tfrac{1}{6}\sum_{x=1}^{6} (x-3.5)^2 \\
= \;& \tfrac{1}{6}\big((-\tfrac{5}{2})^2 + (-\tfrac{3}{2})^2 + (-\tfrac{1}{2})^2 + (\tfrac{1}{2})^2 + (\tfrac{3}{2})^2 + (\tfrac{5}{2})^2\big) \\
= \;& \tfrac{35}{12}
\end{aligned}$$

Since the variance is $\sigma^2 = \frac{35}{12}$, therefore the standard deviation is $\sigma = \sqrt{35/12} \approx 1.707$.

Properties of variance. Although the definition works okay for computing variance, there is an alternative way to compute it that usually works better, namely,

$$\boxed{\sigma^2 = \mathrm{Var}(X) = E(X^2) - \mu^2}$$

Here's why that works.

$$\begin{aligned}
\sigma^2 &= \mathrm{Var}(x) \\
&= E\big((X-\mu)^2\big) \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu\mu + \mu^2 \\
&= E(X^2) - \mu^2
\end{aligned}$$

Here are a couple more properties of variance. First, if you multiply a random variable X by a constant c to get cX, the variance changes by a factor of the square of c, that is

$$\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X).$$

That's the main reason why we take the square root of variance to normalize it—the standard deviation of cX is c times the standard deviation of X:

$$\mathrm{SD}(cX) = |c|\;\mathrm{SD}(x).$$

(Absolute value is needed in case c is negative.) It's easy to show that $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$:

$$\begin{aligned}\mathrm{Var}(cX) &= E\big((cX-\mu)^2\big)\\ &= E\big(c^2(X-\mu)^2\big)\\ &= c^2E\big((X-\mu)^2\big)\\ &= c^2\,\mathrm{Var}(X)\end{aligned}$$

The next important property of variance is that it's translation invariant, that is, if you add a constant to a random variable, the variance doesn't change:

$$\mathrm{Var}(X+c) = \mathrm{Var}(X).$$

In general, the variance of the sum of two random variables is *not* the sum of the variances of the two random variables. But it is when the two random variables are independent.

Theorem. If X and Y are independent random variables, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Proof: This proof relies on the fact that $E(XY) = E(X)\,E(Y)$ when X and Y are independent.

$$\begin{aligned}&\mathrm{Var}(X+Y)\\ =\ & E((X+Y)^2) - \mu_{X+Y}^2\\ =\ & E(X^2+2XY+Y^2) - (\mu_X+\mu_Y)^2\\ =\ & E(X^2)+2E(X)E(Y)+E(Y^2)\\ & -\mu_X^2 - 2\mu_X\mu_Y - \mu_Y^2\\ =\ & E(X^2)+2\mu_X\mu_Y+E(Y^2)\\ & -\mu_X^2 - 2\mu_X\mu_Y - \mu_Y^2\\ =\ & E(X^2)-\mu_X^2+E(Y^2)-\mu_Y^2\\ =\ & \mathrm{Var}(X)+\mathrm{Var}(Y)\end{aligned}$$

Q.E.D.

Variance of the binomial distribution. Let S be the number of successes in n Bernoulli trials, where the probability of success is p. This random variable S has a binomial distribution, and we could use its probability mass function to compute it, but there's an easier way. The random variable S is actually a sum of n independent Bernoulli trials $S = X_1 + X_2 + \cdots + X_n$ where each X_i equals 1 with probability p and 0 with probability $q = 1-p$.

By the preceding theorem,

$$\mathrm{Var}(S) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n).$$

We can determine that if we can determine the variance of one Bernoulli trial X.

Now, $\mathrm{Var}(X) = E(X^2) - \mu^2$, and for a Bernoulli trial $\mu = p$. Let's compute $E(X^2)$. $E(X^2) = P(X{=}0)0 + P(X{=}1)1 = p$. Therefore, the variance of one Bernoulli trial is $\mathrm{Var}(X) = p - p^2 = pq$.

From that observation, we conclude the variance of the binomial distribution is

$$\mathrm{Var}(S) = n\,\mathrm{Var}(X) = npq.$$

Taking the square root, we see that the standard deviation of that binomial distribution is $\sqrt{n}pq$. That gives us the important observation that the spread of a binomial distribution is proportional to the square root of n, the number of trials.

The argument generalizes to other distributions: The standard deviation of a random sample is proportional to the square root of the number of trials in the sample.

Variance of a geometric distribution. Consider the time T to the first success in a Bernoulli process. Its probability mass function is $f(t) = pq^{t-1}$. We saw that its mean was $\mu = E(T) = \frac{1}{p}$. We'll compute its variance using the formula $\mathrm{Var}(X) = E(X^2) - \mu^2$.

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

The last power series we got when we evaluated $E(T)$ was

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

Multiply it by x to get

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

Differentiate that to get

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

Set x to q, and multiply the equation by p, and we get

$$\frac{1+q}{p^2} = p + 2^2pq + 3^2pq^2 + \cdots + n^2pq^{n-1} + \cdots$$

Therefore $E(T^2) = \dfrac{1+q}{p^2}$. Finally,

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

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