1 Variance of discrete random variables

We have already encountered *standard deviation* when we talked about the normal density. It was a measure of how much the distribution deviates from the mean. It is now time to get a more formal grasp on that quantity.

The square of standard deviation is called *variance,* and just like standard deviation is usually written as σ, variance is often written as σ^2. It is often much more convenient to talk about variance than about standard deviation. The chief reason for that is, that variance of a sum of two **independent** random variables is the sum of their variances. This is clearly not true of standard deviation, since it must remain a square root of the variance.

Definition 1 Let X be a numerically valued random variable with expected value $\mu = E(X)$. Then the *variance* of X, here denoted by $V(X)$, is

$$V(X) = E((X-\mu)^2).$$

Example 1 Let X be the result of rolling a six-sided die. Then $E(X) = 3.5$ and:

$$V(X) = E((X-3.5)^2) = \sum_{x=1}^{6}(x-3.5)^2 \times \frac{1}{6} = \frac{1}{6}((1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2)$$

$$= \frac{1}{6}(\frac{25}{4}+\frac{9}{4}+\frac{1}{4}+\frac{1}{4}+\frac{9}{4}+\frac{25}{4}) = \frac{70}{24} = \frac{35}{12}$$

Then the standard deviation is $\sigma = \sqrt{35/12}$.

Theorem 1 *For any random variable X with $E(X) = \mu$,*

$$V(X) = E(X^2) - \mu^2.$$

Proof By definition, $V(X) = E((X-\mu)^2)$. Then:

$$V(X) = E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2,$$

since μ is a numerical quantity. Further, $E(X) = \mu$ and so:

$$V(X) = E(X^2) - 2\mu^2 + \mu^2 = E(X) - \mu^2.$$

Therefore, there is an easier way to compute the variance of a roll of a die:

$$E(X^2) = \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}$$

$$\mu^2 = \frac{49}{4}$$

And so:

$$V(X) = \frac{182-147}{12} = \frac{35}{12}.$$

2 Worksheet problems

1. Knowing that $V(X) = E(X^2) - \mu^2$, find simple expressions for $V(cX)$ and $V(X+c)$, for a constant c.
2. If X and Y are independent random variables, then $V(X+Y) = V(X) + V(Y)$. What is $V(X)$ if X is the number of successes out of n Bernoulli trials with probability p?

$$V(cX) = c^2 V(X)$$
$$V(X+c) = V(X)$$

$V(cX) = E((cX)^2) - (c\mu)^2 = c^2(E(X^2 - \mu^2)) = c^2 V(X)$
$V(X+c) = E((X+c)^2) - (\mu+c)^2 = E(X^2 + 2cX + c^2) - (\mu^2 + 2\mu c + c^2)$
$\quad = E(X^2) + 2cE(X) + c^2 - \mu^2 - 2\mu c - c^2 = E(X^2) - \mu^2 = V(X)$

If X and Y are independent random variables,

$$V(X+Y) = V(X) + V(Y).$$

Example 2 (Bernoulli trials) Bernoulli trials are independent events. The variance of a single trial is:

$$E(X^2) - \mu^2 = p - p^2 = p(1-p)$$

There are n of them, so the combined variance is:

$$np(1-p) = npq$$

3 Expectation and variance of continuous random variables

For continuous random variables, expectation and variance work in a way analogous to the discrete case.

Definition 2 Let X be a real-valued random variable with density function $f(x)$. The expected value $\mu = E(X)$ is defined by:

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

provided that the integral $\int_{-\infty}^{\infty} |x| f(x) dx$ is finite.

For any real-valued random variables $X,\ Y$ with $E(X),\ E(Y)$:

$$E(X+Y) = E(X) + E(Y)$$
$$E(cX) = cE(X)$$

Example 3 (Uniform Density) recall from last week's homework, that for uniform density on (a,b), $f(x) = 1/(b-a)$:

$$\mu = \frac{1}{b-a}\int_a^b x dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

Example 4 (Exponential Density) Similarly, for exponential density $f(x) = \lambda e^{-\lambda x}$ on $(0,\infty)$:

$$\mu = \int_0^{\infty} x f(x) dx = \int_0^{\infty} x\lambda e^{-\lambda x} dx = \left[\frac{-e^{-\lambda x}}{\lambda} - xe^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda}$$

If $X,\ Y$ are mutually independent:

$$E(X \times Y) = E(X) \times E(Y)$$

Definition 3 Let X be a real-valued random variable with density function $f(x)$ and expected value $\mu = E(X)$. The variance $\sigma^2 = V(X)$ is:

$$\sigma^2 = V(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Just as in the discrete case:

$$V(X) = E(X^2) - \mu^2$$
$$V(cX) = c^2 V(X)$$
$$V(X+c) = V(X)$$

And if $X,\ Y$ are independent:

$$V(X+Y) = V(X) + V(Y)$$

4 Expected number of trials until first success

Suppose that we are going to flip a coin until the first time it comes up "tails." Each time we have 1/2 chance of success. Let N be a random variable equal to the total number of flips. What is $E(N)$?

Here is a brute force calculation:

$$E(n) = \sum_{i=1}^{\infty} \frac{i}{2^i} = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \dots = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \dots$$

$$= \underbrace{\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots}_{\sum_{i=1}^{\infty} \frac{1}{2^i}} + \underbrace{\frac{1}{4} + \frac{1}{8} + \dots}_{\sum_{i=2}^{\infty} \frac{1}{2^i}} + \underbrace{\frac{1}{8} + \dots}_{\sum_{i=3}^{\infty} \frac{1}{2^i}} + \dots$$

$$= \sum_{i=1}^{\infty} \frac{1}{2^i} + \sum_{i=2}^{\infty} \frac{1}{2^i} + \sum_{i=3}^{\infty} \frac{1}{2^i} + \dots = \sum_{i=1}^{\infty} \frac{1}{2^i} + \frac{1}{2}\sum_{i=1}^{\infty} \frac{1}{2^i} + \frac{1}{4}\sum_{i=1}^{\infty} \frac{1}{2^i} + \dots$$

$$= \underbrace{(1 + \frac{1}{2} + \frac{1}{4} + \dots)}_{2} \underbrace{\sum_{i=1}^{\infty} \frac{1}{2^i}}_{1} = 2$$

However, we can argue the following. When we flip the coin for the first time, with probability 1/2 it comes up tails. So in the sum that determines expected value, 1 has weight 1/2. Now, with probability 1/2 the flip is heads, so having made one flip we are exactly where we started, and have gained nothing - the expected number of flips FROM THIS POINT ON is exactly the same as the expected number of flips at the beginning. So:

$$E(N) = 1/2 + 1/2(1 + E(N))$$

we can multiply both sides by 2 and get:

$$2E(N) = 1 + 1 + E(N)$$
$$E(N) = 2$$

In general, if the chance of success on each trial is p, the expected number of trials until the first success is:

$$E(N) = p \times 1 + (1-p)(E(N) + 1)$$

- $E(N)$: expected # of trials
- p: probability of success on first trial
- $(1-p)$: probability of failure on first trial
- $(E(N)+1)$: expected # of trials if first trial fails

Now, here's a another way to get the same result. We have:

$$E(N) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}$$

We know that the sum of a geometric sequence works as follows:

$$\sum_{k=0}^{\infty} q^k = 1 + \sum_{k=1}^{\infty} q^k = \frac{1}{1-q}$$

To get the term in the middle it's enough to pull out the first term of the sum. Now, we can differentiate both sides of this equation:

$$\left(1 + \sum_{k=1}^{\infty} q^k\right)' = \left(\frac{1}{1-q}\right)'$$

$$\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$$

And going back to the expression for $E(N)$, we get:

$$E(N) = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{p}{p^2} = \frac{1}{p}$$

5 Coupon Collector's Problem

Alice would like to collect n different stickers from chips packets. In each packet, each of the n stickes can be found with probability $1/n$. How many packets, on average, does she need to buy to collect all stickers?

If $n = 2$, the first packet definitely contains a sticker she doesn't already have. Then, each next packet contains the other sticker with probability $1/2$. So the expected number of trials until first success is 2. We need 1 trial to get the first sticker, and another 2 to get the next. $E(X) = 1 + 2 = 3$

If $n = 3$, again, we get the first sticker on first trial. Then, with probability $2/3$ we get a different sticker. So it takes, on average 1.5 trials to get the second sticker. Then it takes another 3 to get the third. $E(X) = 1 + 1.5 + 3$

In general, it takes 1 trial to get the first sticker, $n/(n-1)$ to get the second, $n/(n-2)$ to get the third, etc, until once we have $n-1$ distinct stickers it takes another n trials to get the last.

$$E(X) = 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{2} + n$$

This is an incredibly important problem in computer science, since many algorithms collect things at random and this prblem describes their expected runtime.