

Geometric random variables

Remember that for a binomial random variable X , we're looking for the number of successes in a finite number of trials.

For a **geometric random variable**, most of the conditions we put on the binomial random variable still apply:

- each trial must be independent,
- each trial can be called a “success” or “failure,”
- the probability of success on each trial is constant.

The difference is that for a geometric random variable, we're looking at how many trials we have to use until we get a certain success. For a binomial random variable, we decided ahead of time on a certain number of trials. But for a geometric random variable, we'll run an infinite number of trials until we get a success.

For example, “flipping a coin until we get heads” could be described by a geometric random variable. It might take just one flip to get heads, but it could take us 5, 10, or (though very, very unlikely) 10,000 flips.

To find the probability that a success S occurs on the n th attempt, when a success has a probability of p , and therefore a failure has a probability of $1 - p$, we use this formula:

$$P(S = n) = p(1 - p)^{n-1}$$



If we look closely at this formula, we see that we're really just multiplying the probability of failure over and over again until the trial right before we have a success, and then multiplying by the probability of a success.

In other words, if we want to find the probability that we get our first success on the 7th trial, then the probability will be

$$P(\text{success on the 7th trial}) = (\text{probability of failure})^6(\text{probability of success})^1$$

Notice that the exponents add to 7, since we needed 7 trials to get the first success.

Example

I'm playing a game where the probability of winning a prize is 0.7. What is the probability that I don't win a prize until the 4th time I play the game, assuming each game is independent?

We're looking for the probability that I don't "succeed" until the 4th "trial," so we can represent this with a geometric random variable.

Since the probability of success is 0.7, it means the probability of failure is 0.3. Since I fail 3 times, and then succeed once on the 4th game, the probability of this happening is

$$P(S = 4) = (0.3)^3(0.7)^1$$

$$P(S = 4) = (0.027)(0.7)$$

$$P(S = 4) = 0.0189$$



$$P(S = 4) \approx 2\%$$

There's an approximately 2% chance that I don't win a prize until the fourth game.

More than and less than

Less than

Sometimes we can be asked to find the probability that it takes less than a specific number of trials in order to get our first success. For instance, continuing with the example we just worked through, we could be asked to find the probability that it takes us less than 4 games to win a prize.

This is the same as saying that we win a prize on game 1, 2, or 3. If we call a success S , that means we want either $S < 4$ or $S \leq 3$, which mean the same thing in the case of a geometric random variable.

$$P(S < 4) = P(S = 1) + P(S = 2) + P(S = 3)$$

The probability of success is 0.7 and the probability of failure is 0.3. When $S = 1$, that means we have 0 failures before we then have 1 success. When $S = 2$, that means we have 1 failure and then 1 success. When $S = 3$, that means we have 2 failures and then 1 success.

$$P(S < 4) = (0.3)^0(0.7)^1 + (0.3)^1(0.7)^1 + (0.3)^2(0.7)^1$$

$$P(S < 4) = (1)(0.7) + (0.3)(0.7) + (0.09)(0.7)$$



$$P(S < 4) = 0.7 + 0.21 + 0.063$$

$$P(S < 4) = 0.973$$

$$P(S < 4) = 97.3 \%$$

At most

This is slightly different than being asked the probability that it takes us less than 4 games to win a prize. If it takes less than 4 games to win, that means we get a prize in the third game, or earlier. But if it takes us at most 4 games to win, that means we could win a prize in the fourth game. We could write that as $S < 5$ or as $S \leq 4$. But either way, we fail no more than 3 times and then succeed in the fourth game, at the latest.

More than

Similarly, we'll be asked to find the probability that it takes more than a specific number of trials in order to get our first success. For instance, continuing with the same example, we could be asked to find the probability that it takes more than 2 games for us to win a prize.

Remember that all probability distributions add to 1. If we're looking for the probability that it takes more than 2 trials to win a prize, we can find the probability of winning on the first trial and the probability of winning on the second trial, and then subtract those probabilities from 1, which will give us all the total probability of all outcomes, other than winning on the first or second game.

So the probability that it takes more than 2 games to win is

$$P(S > 2) = 1 - P(S \leq 2)$$



$$P(S > 2) = 1 - [(0.3)^0(0.7)^1 + (0.3)^1(0.7)^1]$$

$$P(S > 2) = 1 - [(1)(0.7) + (0.3)(0.7)]$$

$$P(S > 2) = 1 - (0.7 + 0.21)$$

$$P(S > 2) = 1 - 0.91$$

$$P(S > 2) = 0.09$$

$$P(S > 2) = 9 \%$$

Keep in mind that we also could have written $S > 2$ as $S \geq 3$, or $S \leq 2$ as $S < 3$.

At least

This is slightly different than being asked the probability that it takes us more than 2 games to win a prize. If it takes more than 2 games to win, that means we don't get a prize until the third game. But if it takes us at least 2 games to win, that means we could win a prize in the second game. We could write that as $S > 1$ or as $S \geq 2$. But either way, we failed once and then succeeded sometimes in the second game or later.

$$P(S \geq 2) = 1 - P(S \leq 1)$$

$$P(S \geq 2) = 1 - P(S = 1)$$

$$P(S \geq 2) = 1 - (0.3)^0(0.7)^1$$

$$P(S \geq 2) = 1 - (1)(0.7)$$

$$P(S \geq 2) = 1 - 0.7$$



$$P(S \geq 2) = 0.3$$

$$P(S \geq 2) \approx 30\%$$

Mean, variance, and standard deviation

Mean

The mean μ_X of a geometric random variable, which can also be called the expected value $E(X)$ is given by

$$\mu_X = E(X) = \frac{1}{p}$$

where the probability of a success on a trial is p , and X is the number of independent trials required to get the first success.

So in our example from this section where we have a 70% chance of winning a prize, the mean is

$$\mu_X = \frac{1}{0.7} \approx 1.43$$

This means we should expect to win the game if we play about one or two times.

Variance and standard deviation

The variance σ_X^2 of a geometric random variable is given by

$$\sigma_X^2 = \frac{1-p}{p^2}$$



and standard deviation is the square root of the variance. Therefore, the variance of the geometric random variable we've been working with is

$$\sigma_X^2 = \frac{1 - 0.7}{0.7^2}$$

$$\sigma_X^2 = \frac{0.3}{0.49}$$

$$\sigma_X^2 \approx 0.61$$

and the standard deviation is

$$\sqrt{\sigma_X^2} \approx \sqrt{0.61}$$

$$\sigma_X \approx 0.78$$

