# 5.3 Geometric and Negative Binomial

## Unit 5: Averages

**Adapted from Blitzstein-Hwang Chapters 4, 5, and 10.**

We now introduce two more famous discrete distributions, the Geometric and Negative Binomial, and calculate their expected values.

### Story 5.3.1 (Geometric distribution).

Consider a sequence of independent Bernoulli trials, each with the same success probability $p \in (0,1)$, with trials performed until a success occurs. Let $X$ be the number of *failures* before the first successful trial. Then $X$ has the *Geometric distribution* with parameter $p$; we denote this by $X \sim \mathrm{Geom}(p)$.

For example, if we flip a fair coin until it lands Heads for the first time, then the number of Tails before the first occurrence of Heads is distributed as $\mathrm{Geom}(1/2)$. To get the Geometric PMF from the story, imagine the Bernoulli trials as a string of 0's (failures) ending in a single 1 (success). Each 0 has probability $q = 1 - p$ and the final 1 has probability $p$, so a string of $k$ failures followed by one success has probability $q^k p$.

> **Theorem 5.3.2 (Geometric PMF).**
>
> If $X \sim \mathrm{Geom}(p)$, then the PMF of $X$ is
>
> $$P(X = k) = q^k p$$
>
> for $k = 0, 1, 2, \ldots$, where $q = 1 - p$.

This is a valid PMF because

$$\sum_{k=0}^{\infty} q^k p = p \sum_{k=0}^{\infty} q^k = p \cdot \frac{1}{1-q} = 1.$$

Just as the binomial theorem shows that the Binomial PMF is valid, a geometric series shows that the Geometric PMF is valid!

☣ **WARNING 5.3.3 (CONVENTIONS FOR THE GEOMETRIC).**

There are differing conventions for the definition of the Geometric distribution; some sources define the Geometric as the total number of *trials*, including the success. In our convention, the Geometric distribution excludes the success, and the *First Success* distribution includes the success.

**DEFINITION 5.3.4 (FIRST SUCCESS DISTRIBUTION).**

In a sequence of independent Bernoulli trials with success probability $p$, let $Y$ be the number of *trials* until the first successful trial, including the success. Then $Y$ has the *First Success distribution* with parameter $p$; we denote this by $Y \sim \mathrm{FS}(p)$.

It is easy to convert back and forth between the two but important to be careful about which convention is being used. By definition, if $Y \sim \mathrm{FS}(p)$ then $Y - 1 \sim \mathrm{Geom}(p)$, and we can convert between the PMFs of $Y$ and $Y-1$ by writing $P(Y = k) = P(Y - 1 = k - 1)$. Conversely, if $X \sim \mathrm{Geom}(p)$, then $X + 1 \sim \mathrm{FS}(p)$.

**Example 5.3.5 (Geometric expectation).**

Let $X \sim \mathrm{Geom}(p)$. By definition,

$$E(X) = \sum_{k=0}^{\infty} kq^k p,$$

where $q = 1 - p$. This sum looks unpleasant; it's not a geometric series because of the extra $k$ multiplying each term. But we notice that each term looks similar to $kq^{k-1}$, the derivative of $q^k$ (with respect to $q$), so let's start there:

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

This geometric series converges since $0 < q < 1$. Differentiating both sides with respect to $q$, we get

$$\sum_{k=0}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}.$$

Finally, if we multiply both sides by $pq$, we recover the original sum we wanted to find:

$$E(X) = \sum_{k=0}^{\infty} kq^k p = pq \sum_{k=0}^{\infty} kq^{k-1} = pq \frac{1}{(1-q)^2} = \frac{q}{p}.$$

**Example 5.3.6 (First Success expectation).**

Since we can write $Y \sim \text{FS}(p)$ as $Y = X + 1$ where $X \sim \text{Geom}(p)$, we have

$$E(Y) = E(X+1) = \frac{q}{p} + 1 = \frac{1}{p}.$$

The Negative Binomial distribution generalizes the Geometric distribution: instead of waiting for just one success, we can wait for any predetermined number $r$ of successes.

**Story 5.3.7 (Negative Binomial distribution).**

In a sequence of independent Bernoulli trials with success probability $p$, if $X$ is the number of *failures* before the $r$th success, then $X$ is said to have the *Negative Binomial distribution* with parameters $r$ and $p$, denoted $X \sim \text{NBin}(r,p)$. Both the Binomial and the Negative Binomial distributions are based on independent Bernoulli trials; they differ in the *stopping rule* and in what they are counting: the Binomial counts the number of successes in a fixed number of *trials*, while the Negative Binomial counts the number of failures until a fixed number of *successes*.

**Theorem 5.3.8 (Negative Binomial PMF).**

If $X \sim \text{NBin}(r,p)$, then the PMF of $X$ is

$$P(X=n) = \binom{n+r-1}{r-1} p^r q^n$$

for $n = 0, 1, 2 \ldots,$ where $q = 1-p$.

**Proof**

Imagine a string of 0's and 1's, with 1's representing successes. The probability of any *specific* string of $n$ 0's and $r$ 1's is $p^r q^n$. How many such strings are there? Because we stop as soon as we hit the $r$th success, the string must terminate in a 1. Among the other $n+r-1$ positions, we choose $r-1$ places for the remaining 1's to go. So the overall probability of exactly $n$ failures before the $r$th success is

$$P(X=n) = \binom{n+r-1}{r-1} p^r q^n, \quad n = 0, 1, 2, \ldots.$$

Just as a Binomial r.v. can be represented as a sum of i.i.d. Bernoullis, a Negative Binomial r.v. can be represented as a sum of i.i.d. Geometrics.

**Theorem 5.3.9.**

Let $X \sim \text{NBin}(r,p)$, viewed as the number of failures before the $r$th success in a sequence of independent Bernoulli trials with success probability $p$. Then we can write $X = X_1 + \cdots + X_r$ where the $X_i$ are i.i.d. $\text{Geom}(p)$.

### Proof

Let $X_1$ be the number of failures until the first success, $X_2$ be the number of failures between the first success and the second success, and in general, $X_i$ be the number of failures between the $(i-1)$st success and the $i$th success. Then $X_1 \sim \mathrm{Geom}(p)$ by the story of the Geometric distribution. After the first success, the number of additional failures until the next success is still Geometric! So $X_2 \sim \mathrm{Geom}(p)$, and similarly for all the $X_i$. Furthermore, the $X_i$ are independent because the trials are all independent of each other. Adding the $X_i$, we get the total number of failures before the $r$th success, which is $X$. Using linearity, the expectation of the Negative Binomial now follows without any additional calculations.

### Example 5.3.10 (Negative Binomial expectation).

Let $X \sim \mathrm{NBin}(r,p)$. By the previous theorem, we can write $X = X_1 + \cdots + X_r$, where the $X_i$ are i.i.d. $\mathrm{Geom}(p)$. By linearity,

$$E(X) = E(X_1) + \cdots + E(X_r) = r \cdot \frac{q}{p}.$$

The next example is a famous problem in probability and an instructive application of the Geometric and First Success distributions. It is usually stated as a problem about collecting coupons, hence its name, but we'll use toys instead of coupons.

### Example 5.3.11 (Coupon collector).

Suppose there are $n$ types of toys, which you are collecting one by one, with the goal of getting a complete set. When collecting toys, the toy types are random (as is sometimes the case, for example, with toys included in cereal boxes or included with kids' meals from a fast food restaurant). Assume that each time you collect a toy, it is equally likely to be any of the $n$ types. What is the expected number of toys needed until you have a complete set?

### Solution

Let $N$ be the number of toys needed; we want to find $E(N)$. Our strategy will be to break up $N$ into a sum of simpler r.v.s so that we can apply linearity. So write

$$N = N_1 + N_2 + \cdots + N_n,$$

where $N_1$ is the number of toys until the first toy type you haven't seen before (which is always 1, as the first toy is always a new type), $N_2$ is the additional number of toys until the second toy type you haven't seen before, and so forth. Figure 5.3.12 illustrates these definitions with $n = 3$ toy types.

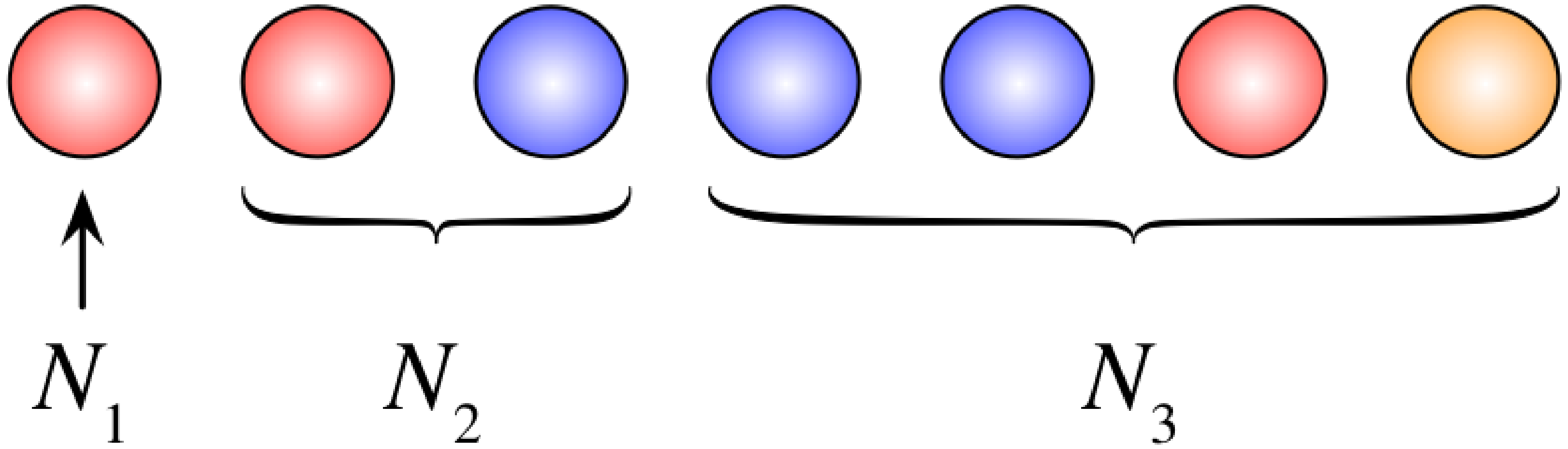


**Figure 5.3.12:** Coupon collector, $n = 3$. Here $N_1$ is the time (number of toys collected) until the first new toy type, $N_2$ is the additional time until the second new type, and $N_3$ is the additional time until the third new type. The total number of toys for a complete set is $N_1 + N_2 + N_3$.

View Larger Image

Image Description

By the story of the FS distribution, $N_2 \sim \textrm{FS}((n-1)/n)$: after collecting the first toy type, there's a $1/n$ chance of getting the same toy you already had (failure) and an $(n-1)/n$ chance you'll get something new (success). Similarly, $N_3$, the additional number of toys until the third new toy type, is distributed $\textrm{FS}((n-2)/n)$. In general,

$$N_j \sim \textrm{FS}((n-j+1)/n).$$

By linearity,

$$\begin{aligned} E(N) &= E(N_1) + E(N_2) + E(N_3) + \cdots + E(N_n) \\ &= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + n \\ &= n \sum_{j=1}^{n} \frac{1}{j}. \end{aligned}$$

For large $n$, this is very close to $n(\log n + 0.577)$.

> **Warning 5.3.13 (Expectation of a nonlinear function of an r.v.).**
>
> Expectation is linear, but in general we do *not* have $E(g(X)) = g(E(X))$ for arbitrary functions $g$. We must be careful not to move the $E$ around when $g$ is not linear. The next example shows a situation in which $E(g(X))$ is *very* different from $g(E(X))$.

**Example 5.3.14 (St. Petersburg paradox).**

Suppose a wealthy stranger offers to play the following game with you. You will flip a fair coin until it lands Heads for the first time, and you will receive $2 if the game lasts for 1 round, $4 if the game lasts for 2 rounds, $8 if the game lasts for 3 rounds, and in general, $\$2^n$ if the game lasts for $n$ rounds. What is the fair value of this game (the expected payoff)? How much would you be willing to pay to play this game once?

**Solution**

Let $X$ be your winnings from playing the game. By definition, $X = 2^N$ where $N$ is the number of rounds that the game lasts. Then $X$ is 2 with probability 1/2, 4 with probability 1/4, 8 with probability 1/8, and so on, so

$$E(X) = \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{8} \cdot 8 + \cdots = \infty.$$

The expected winnings are infinite! On the other hand, the number of rounds $N$ that the game lasts is the number of tosses until the first heads, so $N \sim \mathrm{FS}(1/2)$ and $E(N) = 2$. Thus $E(2^N) = \infty$ while $2^{E(N)} = 4$. Infinity certainly does not equal 4, illustrating the danger of confusing $E(g(X))$ with $g(E(X))$ when $g$ is not linear. This problem is often considered a paradox because although the game's expected payoff is infinite, most people would not be willing to pay very much to play the game (even if they could afford to lose the money). One explanation is to note that *the amount of money in the real world is finite*. The $\infty$ in the St. Petersburg paradox is driven by an infinite ''tail" of extremely rare events where you get extremely large payoffs. Cutting off this tail at some point, which makes sense in the real world, dramatically reduces the expected value of the game.

Learn About Verified Certificates