5.3 Geometric and Negative Binomial

We now introduce 2 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 success rate $p \in (0, 1)$, 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, denoted as $X \sim \text{Geom}(p)$
- For example, if we flip a fair coin until it lands Heads for the first time, then the number of Tails before the first occurence of Heads is distributed as $\text{Geom}(\frac{1}{2})$.
- To get the PMF, imagine the trials as a string of 0's end with 1, i.e:

$$\underbrace{0\ 0\ 0\ 0}_{\text{failures}}\ \underbrace{1}_{\text{success}}$$

- each 0 has probability $q = 1 - p$
- final 1 has probability p

So a string of k failures followed by 1 success has probability $q^k p$.

Theorem 5.3.2 (Geometric PMF)

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

$$P(X = k) = q^k p$$

for $k = 0, 1, 2, \dots$ where $q = 1 - p$

Warning 5.3.3 (Conventions for the Geometric)

There are different conventions for the definition of Geometric distribution:

- Some sources define Geometric as the total number of trials, including the success which is called First Success distribution.
- Our convention, the Geometric distribution excludes the success.

Definition 5.3.4 (First Success Distribution)

In a sequence of independent Bernoulli trials with success probability p, let Y be number of trials until the first successful trial, including the success.

Then Y has the First Success distribution with parameter p; denoted by $Y \sim FS(p)$

Convert between Geom(p) and FS(p):

- If $Y \sim FS(p)$, then $Y - 1 \sim Geom(p)$
- Convert between PMFs: $P(Y = k) = P(Y - 1 = k - 1)$
- Conversely, if $X \sim Geom(p)$, then $X + 1 \sim FS(p)$

Example 5.3.5 (Geometric expectation)

- Let $X \sim Geom(p)$, by definition: $E(X) = \sum_{k=0}^{\infty} k q^k p$, where $q = 1 - p$
- We could get a simpler version of this definition, notice that each term looks similar to kq^{k-1}, the derivative of q^k (with respect to q), so lets start there:

$$\sum_{k=0}^{\infty} q^k = q^0 + q^1 + \dots q^{\infty} = \frac{1}{1-q} \qquad \langle \text{ this converges since } 0 < q < 1 \rangle$$

$$\Leftrightarrow \sum_{k=0}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} \qquad \langle \text{ differentiating both sides respect to } q \rangle$$

$$\Leftrightarrow pq \sum_{k=0}^{\infty} kq^{k-1} = pq \frac{1}{(1-q)^2} \qquad \langle \text{ multiply both sides by } pq \rangle$$

$$\Leftrightarrow \sum_{k=0}^{\infty} k q^k q = \frac{q}{p}$$

So $\boxed{E(X) = \frac{q}{p} \text{ is the Geometric Expectation}}$

Example 5.3.6 (First Success Expectation)

Recall that we can convert $Y \sim FS(p)$ as $Y = X + 1$ with $X \sim Geom(p)$, we have:

$$\boxed{E(Y) = E(X+1) = \frac{q}{p} + 1 = \frac{1}{p} \text{ is the First Success Expectation}}$$

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 has Negative Binomial distribution with parameters r and p, denoted $X \sim NBin(r, p)$

- Comparison between Binomial and Negative Binomial:
  - Both distributions are based on independent Bernoulli trials.
  - They differ in stopping rule and what they are counting:
    - Binomial counts the number of success in fixed number of trials.
    - Negative Binomial counts the number of failures until a fixed number of successes.

Theorem 5.3.8 (Negative Binomial PMF)

If $X \sim NBin(r, p)$, 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 consists of n 0's (failures) and r 1's (successes). i.e:

$\underbrace{0\,1\,1\,0\,0}\,1$ < the string must terminate with a 1 >

- Pick $(r-1)$ places among the other $(n+r-1)$ positions for 1's to go.
- Multiply by probabilities of each Bernoulli trials.

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

Theorem 5.3.9 (Negative Binomial r.v can be represented as Sum of i.i.d Geometrics)

Let $X \sim NBin(r, p)$ 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 $X = X_1 + \ldots + X_r$ where the X_i are i.i.d $Geom(p)$

Proof: Imagine a string consists of n 0's (failures) and r 1's (successes). i.e:

$\underbrace{0\,0\,1}_{X_1 \sim Geom(p)}\,\underbrace{0\,0\,0\,1}_{X_2 \sim Geom(p)}$

$$\Rightarrow X = X_1 + X_2 \sim NBin(r, p)$$

Example 5.3.10 (Negative Binomial Expectation)

Let $X \sim NBin(r, p)$, we can also write $X = X_1 + X_2 \dots + X_r$ with X_i are i.i.d $Geom(p)$

By linearity: $$E(X) = E(X_1) + E(X_2) + \dots + E(X_r)$$
$$= \frac{q}{p} + \frac{q}{p} + \dots + \frac{q}{p}$$
$$= r \cdot \frac{q}{p}$$

So $\boxed{E(X) = r \cdot \frac{q}{p} \text{ is Negative Binomial Expectation}}$

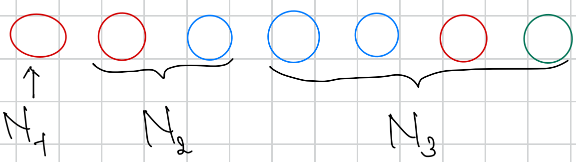
Example 5.3.11 (Coupon collector)

Suppose there are n types of toys, which you are collecting one by one. You aim to collect all types of toys, and the chance of you getting any of the n types are equally likely. What are the expected number of toys needed until you have a complete set?

Let N be number of toys needed, we can split the problem into smaller ones: $$N = N_1 + N_2 + N_3 + \dots + N_n$$

where
- N_1 is number of toys until the first new toy appears.
- N_2 is number of toys until the second new toy appears.
- N_3 is number of toys until the third new toy appears.

Visualization:



By the story of FS distribution:
- N_1 will alway equals 1
- $N_2 \sim FS\left(\frac{n-1}{n}\right)$ bc after getting the first toy. There are
  - $\frac{1}{n}$ chance you get the same toy (failure q)
  - $\frac{n-1}{n}$ chance you get smt new (success p)
- $N_3 \sim FS\left(\frac{n-2}{n}\right)$

Generalize this logic, we have:

$$N_j \sim FS\left(\frac{n-j+1}{n}\right)$$

By linearity: $E(N) = E(N_1) + E(N_2) + E(N_3) + \dots + E(N_n)$

$$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \dots + n$$

$$= n \sum_{j=1}^{n} \frac{1}{j}$$

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 for an arbitrary function g:

$$E(g(x)) \begin{cases} = g(E(X)) & \text{when } g \text{ is linear function} \\ \neq g(E(x)) & \text{otherwise} \end{cases}$$

Example 5.3.14 (St. Petersburg paradox)

Suppose you are playing a game of coin flipping. A fair coin will be flipped until it lands Heads for the first time, if the game lasts for n rounds, you will receive $\$2^n$. $\begin{cases} \text{What is the fair value of this game? (expected pay off)} \\ \text{How much would you be willing to pay to play this game?} \end{cases}$

- Let X be your winnings from playing the game. By definition, $X = 2^N$ with N being number of rounds the game lasts.
- Then X is 2 with probability $1/2$, 4 with probability $1/4$, 8 with probability $1/8$ etc... So: $E(X) = \frac{1}{2}\cdot 2 + \frac{1}{4}\cdot 4 + \frac{1}{8}\cdot 8 + \dots = \infty$
- Let N be number of rounds until the first Heads, so $\begin{cases} N \sim FS\left(\frac{1}{2}\right) \\ E(N) = 2 \end{cases}$

So $E(2^N) = \infty$ while $2^{E(N)} = 4$

Meaning $E(g(x)) \neq g(E(X))$ when g is not linear

- This is paradoxical because even though the reward is infinite, it is quite rare to get those result. For example, you only have $\frac{1}{4}$ chance of winning $\$4$.