

STAT 315: Geometric and Negative Binomial Random Variables

Luc Rey-Bellet

University of Massachusetts Amherst

luc@math.umass.edu

September 29, 2025

Geometric random variable

Geometric distribution

- An experiment consist of successive **identical and independent trials**
- Each trial results in **success S with probability p** or **failure F with probability $1 - p$** .
- The **geometric random variable N** is defined by

$N =$ number of trial until the first success

- The **pdf of a geometric random variable** is

$$p(k) = P(N = k) = (1 - p)^{k-1}p, \quad k = 1, 2, 3, \dots$$

We have $N = k$ if the first success occurs on the k^{th} trial i.e. we have the sequence of trials

$$\underbrace{F, F, \dots, F}_{k-1}, S \text{ which has probability } (1 - p)^{k-1}p$$

Math reminder: Geometric series

To compute the expected value and the variance of a geometric random variables we will need the following series (see you calculus class).

Geometric series

Provided $|x| < 1$ we have the following infinite series

$$\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + x^4 \dots = \frac{1}{1-x}$$

$$\sum_{k=1}^{\infty} kx^{k-1} = 1 + 2x + 3x^2 + 4x^3 + \dots = \frac{1}{(1-x)^2}$$

$$\sum_{k=2}^{\infty} k(k-1)x^{k-2} = 2 + 6x + 12x^2 + \dots = \frac{2}{(1-x)^3}$$

Differentiate term by term!

Survival probability and the memoryless property

Memoryless property

If N is a geometric random variable

$$\text{Survival probability} \quad P(N > n) = (1 - p)^n$$

$$\text{Memoryless property} \quad P(N > n + k | N > n) = P(N \geq k)$$

If p is the probability to "die" in any given time period $P(N > n)$ is the probability to survive at least n period of time.

$$\begin{aligned} P(N > n) &= P(N = n + 1) + P(N = n + 2) + \dots \\ &= (1 - p)^n p + (1 - p)^{n+1} p + (1 - p)^{n+2} p + \dots \\ &= p(1 - p)^n [1 + (1 - p) + (1 - p)^2 + \dots] \\ &= p(1 - p)^n \frac{1}{1 - (1 - p)} = (1 - p)^n \end{aligned}$$

$$P(N > n + k | N > n) = \frac{P(N > n + k)}{P(N > n)} = \frac{(1 - p)^{n+k}}{(1 - p)^n} = (1 - p)^k = P(N > k)$$

Expected value and variance of geometric random variables

Mean and Variance of a Geometric Random Variable

$$E[N] = \mu = \frac{1}{p} \qquad V[N] = \sigma^2 = \frac{1-p}{p^2}$$

Proof: The PDF is $p(k) = (1-p)^{k-1}p$ for $k = 1, 2, 3, \dots$

So

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

So the mean is $E[N] = \frac{1}{p}$. This is intuitive. If the probability of success is $1/10$ then on average it takes 10 trials to succeed!

For the variance instead of $E[N^2]$ compute again $E[N(N - 1)]$.

$$\begin{aligned} E[N(N - 1)] &= \sum_{k=1}^{\infty} k(k - 1)p(k) = p(1 - p) \sum_{k=2}^{\infty} k(k - 1)(1 - p)^{k-2} \\ &= p(1 - p) \frac{2}{(1 - (1 - p))^3} = p(1 - p) \frac{2}{p^3} = \frac{2(1 - p)}{p^2} \end{aligned}$$

So

$$\begin{aligned} V[N] &= E[N^2] - E[N]^2 \\ &= E[N(N - 1)] + E[N] - E[N]^2 \\ &= \frac{2(1 - p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{2(1 - p)}{p^2} + \frac{p - 1}{p^2} = \frac{(1 - p)}{p^2} \end{aligned}$$

Example

- A fair coin is tossed repeatedly until the first head appears. Let X be the trial number on which the first head occurs.
 - ▶ Write down the probability distribution function of X .
 - ▶ Compute $P(X = 1)$ and $P(X = 3)$.
 - ▶ Compute $P(X \leq 5)$
 - ▶ Find $E[X]$ and $V(X)$.
- An oil prospector is looking for oil in Oklahoma. It costs \$10,000 to drill a well and the probability to find oil in each well is $\frac{1}{8}$.
 - ▶ If the prospector has 50,000? What is the probability the prospector will go bankrupt?
 - ▶ What is the expectation and the variance of the amount of money you need to get one functioning well. What about 5 functioning wells?

Example: Bitcoin mining

- **Fun (or Scary) fact:** Bitcoin mining is using as much electricity as the entire country of Finland (circa 5 million people). Why?
- Bitcoin is a cryptocurrency where no central authority keeps track of transactions. To record transactions there is a **lottery** and whoever wins the lottery can record a block of transactions which are added to "the block chain"). He is **rewarded by earning 3.125 bitcoins (1 BTC = \$ 85,565.98 on 2/27/2025)**. The reward is halved every four years.

<https://bitinfocharts.com/bitcoin/>

- The winner of the lottery must produce a hash value for the hash function SHA256 which is less than some value, that is a 256bits starting with k 0's (this has probability $1/2^k$). This is done by farms of specialized computers ("the miners").
- The number of 0, k , needed is adjusted so that it takes about 10 minutes to get a winner. This depends on the number of players! If the entire sets of miners can do 2^k SHA256 computations every 10 minutes then it will take on average 10 minutes to produce a winner since $E[N] = \frac{1}{p}$.

Negative binomial random variables

We count the number of trials until $r = 2, 3, 4, 5, \dots$ successes.



Negative Binomial Random Variable

- An experiment consist of successive **identical and independent trials**
- Each trial results in **success S with probability p** or **failure F with probability $1 - p$** .
- A **negative binomial random variable N** with parameter r, p is
 $N =$ number of trial until r successes. $N = r, r + 1, r + 2, \dots$
- The **pdf of a negative binomial variable** is

$$p(k) = P(N = k) = \binom{k-1}{r-1} (1-p)^{k-r} p^r \quad k = r, r+1, r+2$$

The last, k^{th} , trial is a success (the r^{th} one) and the other $r - 1$ successes can be any of the other $k - 1$ trials, hence the binomial $\binom{r-1}{r-1}$

Example: NBA playoffs series

-  The Boston Celtics are playing the Denver Nuggets 
- The team which wins 4 games first wins the playoff series.
- Assume that each game is won independently by the Celtics with probability $p=.45$
- The probability the Boston Celtics wins the series is described by a **negative binomial RV Y with $r = 4$** since the Celtics must accumulate 4 wins before the 7 games.

$$\begin{aligned}P(\text{Celtics wins}) &= P(Y = 4) + P(Y = 5) + P(Y = 6) + P(Y = 7) \\&= p^4 + \binom{4}{0} p^4(1-p) + \binom{5}{2} p^4(1-p)^2 + \binom{6}{3} p^4(1-p)^3 \\&= 0.3917...\end{aligned}$$

Geometric and negative binomials PDF

The geometric RV is a negative binomial RV with $r = 1$.

$$E[N] = \frac{1}{p}, \quad V[N] = \frac{(1-p)}{p^2}$$

$$E[N] = \frac{r}{p}, \quad V[N] = \frac{r(1-p)}{p^2}$$

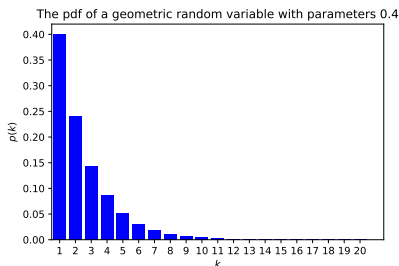


Figure: Geometric pdf
 $p(k) = (1-p)^{k-1}p$

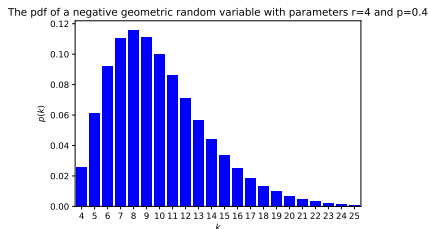


Figure: Negative binomial pdf
 $p(k) = \binom{k-1}{r-1} (1-p)^{k-1} p^r$

Remark: we will compute mean and variance of negative binomials later.