

11.3: The Geometric Distribution

Basic Theory

Definitions

Suppose again that our random experiment is to perform a sequence of Bernoulli trials $\mathbf{X} = (X_1, X_2, \dots)$ with success parameter $p \in (0, 1]$. In this section we will study the random variable N that gives the *trial number of the first success* and the random variable M that gives the number of failures before the first success.

Let $N = \min\{n \in \mathbb{N}_+ : X_n = 1\}$, the trial number of the first success, and let $M = N - 1$, the number of failures before the first success. The distribution of N is the *geometric distribution on \mathbb{N}_+* and the distribution of M is the *geometric distribution on \mathbb{N}* . In both cases, p is the *success parameter* of the distribution.

Since N and M differ by a constant, the properties of their distributions are very similar. Nonetheless, there are applications where it more natural to use one rather than the other, and in the literature, the term *geometric distribution* can refer to either. In this section, we will concentrate on the distribution of N , pausing occasionally to summarize the corresponding results for M .

The Probability Density Function

N has probability density function f given by $f(n) = p(1-p)^{n-1}$ for $n \in \mathbb{N}_+$.

Proof

Note first that $\{N = n\} = \{X_1 = 0, \dots, X_{n-1} = 0, X_n = 1\}$. By independence, the probability of this event is $(1-p)^{n-1}p$.

Check that f is a valid PDF

By standard results for geometric series

$$\sum_{n=1}^{\infty} \mathbb{P}(N = n) = \sum_{n=1}^{\infty} (1-p)^{n-1}p = \frac{p}{1-(1-p)} = 1 \quad (11.3.1)$$

A priori, we might have thought it possible to have $N = \infty$ with positive probability; that is, we might have thought that we could run Bernoulli trials forever without ever seeing a success. However, we now know this cannot happen when the success parameter p is positive.

The probability density function of M is given by $\mathbb{P}(M = n) = p(1-p)^n$ for $n \in \mathbb{N}$.

In the negative binomial experiment, set $k = 1$ to get the geometric distribution on \mathbb{N}_+ . Vary p with the scroll bar and note the shape and location of the probability density function. For selected values of p , run the simulation 1000 times and compare the relative frequency function to the probability density function.

Note that the probability density functions of N and M are decreasing, and hence have modes at 1 and 0, respectively. The geometric form of the probability density functions also explains the term *geometric distribution*.

Distribution Functions and the Memoryless Property

Suppose that T is a random variable taking values in \mathbb{N}_+ . Recall that the ordinary distribution function of T is the function $n \mapsto \mathbb{P}(T \leq n)$. In this section, the complementary function $n \mapsto \mathbb{P}(T > n)$ will play a fundamental role. We will refer to this function as the *right distribution function* of T . Of course both functions completely determine the distribution of T . Suppose again that N has the geometric distribution on \mathbb{N}_+ with success parameter $p \in (0, 1]$.

N has right distribution function G given by $G(n) = (1-p)^n$ for $n \in \mathbb{N}$.

Proof from Bernoulli trials

Note that $\{N > n\} = \{X_1 = 0, \dots, X_n = 0\}$. By independence, the probability of this event is $(1-p)^n$.

Direct proof

Using geometric series,

$$\mathbb{P}(N > n) = \sum_{k=n+1}^{\infty} \mathbb{P}(N = k) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p = \frac{p(1-p)^n}{1-(1-p)} = (1-p)^n \quad (11.3.2)$$

From the last result, it follows that the ordinary (left) distribution function of N is given by

$$F(n) = 1 - (1-p)^n, \quad n \in \mathbb{N} \quad (11.3.3)$$

We will now explore another characterization known as the *memoryless property*.

For $m \in \mathbb{N}$, the conditional distribution of $N - m$ given $N > m$ is the same as the distribution of N . That is,

$$\mathbb{P}(N > n + m \mid N > m) = \mathbb{P}(N > n); \quad m, n \in \mathbb{N} \quad (11.3.4)$$

Proof

From the [result above](#) and the definition of conditional probability,

$$\mathbb{P}(N > n + m \mid N > m) = \frac{\mathbb{P}(N > n + m)}{\mathbb{P}(N > m)} = \frac{(1-p)^{n+m}}{(1-p)^m} = (1-p)^n = \mathbb{P}(N > n) \quad (11.3.5)$$

Thus, if the first success has not occurred by trial number m , then the remaining number of trials needed to achieve the first success has the same distribution as the trial number of the first success in a fresh sequence of Bernoulli trials. In short, Bernoulli trials have no memory. This fact has implications for a gambler betting on Bernoulli trials (such as in the casino games roulette or craps). No betting strategy based on observations of past outcomes of the trials can possibly help the gambler.

Conversely, if T is a random variable taking values in \mathbb{N}_+ that satisfies the [memoryless property](#), then T has a geometric distribution.

Proof

Let $G(n) = \mathbb{P}(T > n)$ for $n \in \mathbb{N}$. The memoryless property and the definition of conditional probability imply that $G(m+n) = G(m)G(n)$ for $m, n \in \mathbb{N}$. Note that this is the *law of exponents* for G . It follows that $G(n) = G^n(1)$ for $n \in \mathbb{N}$. Hence T has the geometric distribution with parameter $p = 1 - G(1)$.

Moments

Suppose again that N is the trial number of the first success in a sequence of Bernoulli trials, so that N has the geometric distribution on \mathbb{N}_+ with parameter $p \in (0, 1]$. The mean and variance of N can be computed in several different ways.

$$\mathbb{E}(N) = \frac{1}{p}$$

Proof from the density function

Using the derivative of the geometric series,

$$\mathbb{E}(N) = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p \sum_{n=1}^{\infty} n(1-p)^{n-1} \quad (11.3.6)$$

$$= p \sum_{n=1}^{\infty} -\frac{d}{dp}(1-p)^n = -p \frac{d}{dp} \sum_{n=0}^{\infty} (1-p)^n \quad (11.3.7)$$

$$= -p \frac{d}{dp} \frac{1}{p} = -p \left(-\frac{1}{p^2} \right) = \frac{1}{p} \quad (11.3.8)$$

Proof from the right distribution function

Recall that since N takes positive integer values, its expected value can be computed as the sum of the right distribution function. Hence

$$\mathbb{E}(N) = \sum_{n=0}^{\infty} \mathbb{P}(N > n) = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{p} \quad (11.3.9)$$

Proof from Bernoulli trials

We condition on the first trial X_1 : If $X_1 = 1$ then $N = 1$ and hence $\mathbb{E}(N | X_1 = 1) = 1$. If $X_1 = 0$ (equivalently $N > 1$) then by the memoryless property, $N - 1$ has the same distribution as N . Hence $\mathbb{E}(N | X_1 = 0) = 1 + \mathbb{E}(N)$. In short

$$\mathbb{E}(N | X_1) = 1 + (1 - X_1)\mathbb{E}(N) \quad (11.3.10)$$

It follows that

$$\mathbb{E}(N) = \mathbb{E}[\mathbb{E}(N | X_1)] = 1 + (1 - p)\mathbb{E}(N) \quad (11.3.11)$$

Solving gives $\mathbb{E}(N) = \frac{1}{p}$.

This result makes intuitive sense. In a sequence of Bernoulli trials with success parameter p we would “expect” to wait $1/p$ trials for the first success.

$$\text{var}(N) = \frac{1-p}{p^2}$$

Direct proof

We first compute $\mathbb{E}[N(N-1)]$. This is an example of a factorial moment, and we will compute the general factorial moments below. Using derivatives of the geometric series again,

$$\mathbb{E}[N(N-1)] = \sum_{n=2}^{\infty} n(n-1)p(1-p)^{n-1} = p(1-p) \sum_{n=2}^{\infty} n(n-1)(1-p)^{n-2} \quad (11.3.12)$$

$$= p(1-p) \frac{d^2}{dp^2} \sum_{n=0}^{\infty} (1-p)^n = p(1-p) \frac{d^2}{dp^2} \frac{1}{p} = p(1-p) \frac{2}{p^3} = 2 \frac{1-p}{p^2} \quad (11.3.13)$$

Since $\mathbb{E}(N) = \frac{1}{p}$, it follows that $\mathbb{E}(N^2) = \frac{2-p}{p^2}$ and hence $\text{var}(N) = \frac{1-p}{p^2}$

Proof from Bernoulli trials

Recall that

$$\mathbb{E}(N | X_1) = 1 + (1 - X_1)\mathbb{E}(N) = 1 + \frac{1}{p}(1 - X_1) \quad (11.3.14)$$

and by the same reasoning, $\text{var}(N | X_1) = (1 - X_1)\text{var}(N)$. Hence

$$\text{var}(N) = \text{var}[\mathbb{E}(N | X_1)] + \mathbb{E}[\text{var}(N | X_1)] = \frac{1}{p^2}p(1-p) + (1-p)\text{var}(N) \quad (11.3.15)$$

Solving gives $\text{var}(N) = \frac{1-p}{p^2}$.

Note that $\text{var}(N) = 0$ if $p = 1$, hardly surprising since N is deterministic (taking just the value 1) in this case. At the other extreme, $\text{var}(N) \uparrow \infty$ as $p \downarrow 0$.

In the negative binomial experiment, set $k = 1$ to get the geometric distribution. Vary p with the scroll bar and note the location and size of the mean \pm standard deviation bar. For selected values of p , run the simulation 1000 times and compare the sample mean and standard deviation to the distribution mean and standard deviation.

the probability generating function P of N is given by

$$P(t) = \mathbb{E}(t^N) = \frac{pt}{1 - (1-p)t}, \quad |t| < \frac{1}{1-p} \quad (11.3.16)$$

Proof

This result follows from yet another application of geometric series:

$$\mathbb{E}(t^N) = \sum_{n=1}^{\infty} t^n p(1-p)^{n-1} = pt \sum_{n=1}^{\infty} [t(1-p)]^{n-1} = \frac{pt}{1-(1-p)t}, \quad |(1-p)t| < 1 \quad (11.3.17)$$

Recall again that for $x \in \mathbb{R}$ and $k \in \mathbb{N}$, the *falling power* of x of order k is $x^{(k)} = x(x-1)\cdots(x-k+1)$. If X is a random variable, then $\mathbb{E}[X^{(k)}]$ is the *factorial moment* of X of order k .

The factorial moments of N are given by

$$\mathbb{E}[N^{(k)}] = k! \frac{(1-p)^{k-1}}{p^k}, \quad k \in \mathbb{N}_+ \quad (11.3.18)$$

Proof from geometric series

Using derivatives of geometric series again,

$$\mathbb{E}[N^{(k)}] = \sum_{n=k}^{\infty} n^{(k)} p(1-p)^{n-1} = p(1-p)^{k-1} \sum_{n=k}^{\infty} n^{(k)} (1-p)^{n-k} \quad (11.3.19)$$

$$= p(1-p)^{k-1} (-1)^k \frac{d^k}{dp^k} \sum_{n=0}^{\infty} (1-p)^n = p(1-p)^{k-1} (-1)^k \frac{d^k}{dp^k} \frac{1}{p} = k! \frac{(1-p)^{k-1}}{p^k} \quad (11.3.20)$$

Proof from the generating function

Recall that $\mathbb{E}[N^{(k)}] = P^{(k)}(1)$ where P is the probability generating function of N . So the result follows from standard calculus.

Suppose that $p \in (0, 1)$. The skewness and kurtosis of N are

1. $\text{skew}(N) = \frac{2-p}{\sqrt{1-p}}$
2. $\text{kurt}(N) = \frac{p^2}{1-p}$

Proof

The factorial moments can be used to find the moments of N about 0. The results then follow from the standard computational formulas for skewness and kurtosis.

Note that the geometric distribution is always positively skewed. Moreover, $\text{skew}(N) \rightarrow \infty$ and $\text{kurt}(N) \rightarrow \infty$ as $p \uparrow 1$.

Suppose now that $M = N - 1$, so that M (the number of failures before the first success) has the geometric distribution on \mathbb{N} . Then

1. $\mathbb{E}(M) = \frac{1-p}{p}$
2. $\text{var}(M) = \frac{1-p}{p^2}$
3. $\text{skew}(M) = \frac{2-p}{\sqrt{1-p}}$
4. $\text{kurt}(M) = \frac{p^2}{1-p}$
5. $\mathbb{E}(t^M) = \frac{p}{1-(1-p)t}$ for $|t| < \frac{1}{1-p}$

Of course, the fact that the variance, skewness, and kurtosis are unchanged follows easily, since N and M differ by a constant.

The Quantile Function

Let F denote the distribution function of N , so that $F(n) = 1 - (1-p)^n$ for $n \in \mathbb{N}$. Recall that $F^{-1}(r) = \min\{n \in \mathbb{N}_+ : F(n) \geq r\}$ for $r \in (0, 1)$ is the quantile function of N .

The quantile function of N is

$$F^{-1}(r) = \left\lceil \frac{\ln(1-r)}{\ln(1-p)} \right\rceil, \quad r \in (0, 1) \quad (11.3.21)$$

Of course, the quantile function, like the probability density function and the distribution function, completely determines the distribution of N . Moreover, we can compute the median and quartiles to get measures of center and spread.

The first quartile, the median (or second quartile), and the third quartile are

$$\begin{aligned} 1. F^{-1}\left(\frac{1}{4}\right) &= \lceil \ln(3/4) / \ln(1-p) \rceil \approx \lceil -0.2877 / \ln(1-p) \rceil \\ 2. F^{-1}\left(\frac{1}{2}\right) &= \lceil \ln(1/2) / \ln(1-p) \rceil \approx \lceil -0.6931 / \ln(1-p) \rceil \\ 3. F^{-1}\left(\frac{3}{4}\right) &= \lceil \ln(1/4) / \ln(1-p) \rceil \approx \lceil -1.3863 / \ln(1-p) \rceil \end{aligned}$$

Open the special distribution calculator, and select the geometric distribution and CDF view. Vary p and note the shape and location of the CDF/quantile function. For various values of p , compute the median and the first and third quartiles.

The Constant Rate Property

Suppose that T is a random variable taking values in \mathbb{N}_+ , which we interpret as the first time that some event of interest occurs.

The function h given by

$$h(n) = \mathbb{P}(T = n \mid T \geq n) = \frac{\mathbb{P}(T = n)}{\mathbb{P}(T \geq n)}, \quad n \in \mathbb{N}_+ \quad (11.3.22)$$

is the *rate function* of T .

If T is interpreted as the (discrete) lifetime of a device, then h is a discrete version of the *failure rate function* studied in reliability theory. However, in our usual formulation of Bernoulli trials, the event of interest is success rather than failure (or death), so we will simply use the term *rate function* to avoid confusion. The constant rate property characterizes the geometric distribution. As usual, let N denote the trial number of the first success in a sequence of Bernoulli trials with success parameter $p \in (0, 1)$, so that N has the geometric distribution on \mathbb{N}_+ with parameter p .

N has constant rate p .

Proof

From the results above, $\mathbb{P}(N = n) = p(1-p)^{n-1}$ and $\mathbb{P}(N \geq n) = \mathbb{P}(N > n-1) = (1-p)^{n-1}$, so $\mathbb{P}(N = n) / \mathbb{P}(N \geq n) = p$ for $n \in \mathbb{N}_+$.

Conversely, if T has constant rate $p \in (0, 1)$ then T has the geometric distribution on \mathbb{N}_+ with success parameter p .

Proof

Let $H(n) = \mathbb{P}(T \geq n)$ for $n \in \mathbb{N}_+$. From the constant rate property, $\mathbb{P}(T = n) = p H(n)$ for $n \in \mathbb{N}_+$. Next note that $\mathbb{P}(T = n) = H(n) - H(n+1)$ for $n \in \mathbb{N}_+$. Thus, H satisfies the recurrence relation $H(n+1) = (1-p)H(n)$ for $n \in \mathbb{N}_+$. Also H satisfies the initial condition $H(1) = 1$. Solving the recurrence relation gives $H(n) = (1-p)^{n-1}$ for $n \in \mathbb{N}_+$.

Relation to the Uniform Distribution

Suppose again that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of Bernoulli trials with success parameter $p \in (0, 1)$. For $n \in \mathbb{N}_+$, recall that $Y_n = \sum_{i=1}^n X_i$, the number of successes in the first n trials, has the binomial distribution with parameters n and p . As before, N denotes the trial number of the first success.

Suppose that $n \in \mathbb{N}_+$. The conditional distribution of N given $Y_n = 1$ is uniform on $\{1, 2, \dots, n\}$.

Proof from sampling

We showed in the last section that given $Y_n = k$, the trial numbers of the successes form a random sample of size k chosen without replacement from $\{1, 2, \dots, n\}$. This result is a simple corollary with $k = 1$.

Direct proof

For $j \in \{1, 2, \dots, n\}$

$$\mathbb{P}(N = j \mid Y_n = 1) = \frac{\mathbb{P}(N = j, Y_n = 1)}{\mathbb{P}(Y_n = 1)} = \frac{\mathbb{P}(Y_{j-1} = 0, X_j = 1, Y_n - Y_j = 0)}{\mathbb{P}(Y_n = 1)} \quad (11.3.23)$$

In words, the events in the numerator of the last fraction are that there are no successes in the first $j-1$ trials, a success on trial j , and no successes in trials $j+1$ to n . These events are independent so

$$\mathbb{P}(N = j \mid Y_n = 1) = \frac{(1-p)^{j-1}p(1-p)^{n-j}}{np(1-p)^{n-1}} = \frac{1}{n} \quad (11.3.24)$$

Note that the conditional distribution does not depend on the success parameter p . If we know that there is exactly one success in the first n trials, then the trial number of that success is equally likely to be any of the n possibilities.

Another connection between the geometric distribution and the uniform distribution is given below in the [alternating coin tossing game](#): the conditional distribution of N given $N \leq n$ converges to the uniform distribution on $\{1, 2, \dots, n\}$ as $p \downarrow 0$.

Relation to the Exponential Distribution

The *Poisson process* on $[0, \infty)$, named for Simeon Poisson, is a model for random points in continuous time. There are many deep and interesting connections between the Bernoulli trials process (which can be thought of as a model for random points in *discrete* time) and the Poisson process. These connections are explored in detail in the chapter on the Poisson process. In this section we just give the most famous and important result—the convergence of the geometric distribution to the exponential distribution. The geometric distribution, as we know, governs the time of the first “random point” in the Bernoulli trials process, while the exponential distribution governs the time of the first random point in the Poisson process.

For reference, the exponential distribution with rate parameter $r \in (0, \infty)$ has distribution function $F(x) = 1 - e^{-rx}$ for $x \in [0, \infty)$. The mean of the exponential distribution is $1/r$ and the variance is $1/r^2$. In addition, the moment generating function is $s \mapsto \frac{1}{s-r}$ for $s > r$.

For $n \in \mathbb{N}_+$, suppose that U_n has the geometric distribution on \mathbb{N}_+ with success parameter $p_n \in (0, 1)$, where $np_n \rightarrow r > 0$ as $n \rightarrow \infty$. Then the distribution of U_n/n converges to the exponential distribution with parameter r as $n \rightarrow \infty$.

Proof

Let F_n denote the CDF of U_n/n . Then for $x \in [0, \infty)$

$$F_n(x) = \mathbb{P}\left(\frac{U_n}{n} \leq x\right) = \mathbb{P}(U_n \leq nx) = \mathbb{P}(U_n \leq \lfloor nx \rfloor) = 1 - (1 - p_n)^{\lfloor nx \rfloor} \quad (11.3.25)$$

But by a famous limit from calculus, $(1 - p_n)^n = \left(1 - \frac{np_n}{n}\right)^n \rightarrow e^{-r}$ as $n \rightarrow \infty$, and hence $(1 - p_n)^{nx} \rightarrow e^{-rx}$ as $n \rightarrow \infty$. But by definition, $\lfloor nx \rfloor \leq nx < \lfloor nx \rfloor + 1$ or equivalently, $nx - 1 < \lfloor nx \rfloor \leq nx$ so it follows that $(1 - p_n)^{\lfloor nx \rfloor} \rightarrow e^{-rx}$ as $n \rightarrow \infty$. Hence $F_n(x) \rightarrow 1 - e^{-rx}$ as $n \rightarrow \infty$, which is the CDF of the exponential distribution.

Note that the condition $np_n \rightarrow r$ as $n \rightarrow \infty$ is the same condition required for the convergence of the binomial distribution to the Poisson that we studied in the last section.

Special Families

The geometric distribution on \mathbb{N} is an infinitely divisible distribution and is a compound Poisson distribution. For the details, visit these individual sections and see the next section on the negative binomial distribution.

Examples and Applications

Simple Exercises

A standard, fair die is thrown until an ace occurs. Let N denote the number of throws. Find each of the following:

1. The probability density function of N .
2. The mean of N .
3. The variance of N .
4. The probability that the die will have to be thrown at least 5 times.

5. The quantile function of N .
6. The median and the first and third quartiles.

Answer

1. $\mathbb{P}(N = n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}$ for $n \in \mathbb{N}_+$
2. $\mathbb{E}(N) = 6$
3. $\text{var}(N) = 30$
4. $\mathbb{P}(N \geq 5) = 525/1296$
5. $F^{-1}(r) = \lceil \ln(1-r)/\ln(5/6) \rceil$ for $r \in (0, 1)$
6. Quartiles $q_1 = 2, q_2 = 4, q_3 = 8$

A type of missile has failure probability 0.02. Let N denote the number of launches before the first failure. Find each of the following:

1. The probability density function of N .
2. The mean of N .
3. The variance of N .
4. The probability of 20 consecutive successful launches.
5. The quantile function of N .
6. The median and the first and third quartiles.

Answer

1. $\mathbb{P}(N = n) = \left(\frac{49}{50}\right)^{n-1} \frac{1}{50}$ for $n \in \mathbb{N}_+$
2. $\mathbb{E}(N) = 50$
3. $\text{var}(N) = 2450$
4. $\mathbb{P}(N > 20) = 0.6676$
5. $F^{-1}(r) = \lceil \ln(1-r)/\ln(0.98) \rceil$ for $r \in (0, 1)$
6. Quartiles $q_1 = 15, q_2 = 35, q_3 = 69$

A student takes a multiple choice test with 10 questions, each with 5 choices (only one correct). The student blindly guesses and gets one question correct. Find the probability that the correct question was one of the first 4.

Answer

0.4

Recall that an American roulette wheel has 38 slots: 18 are red, 18 are black, and 2 are green. Suppose that you observe red or green on 10 consecutive spins. Give the conditional distribution of the number of additional spins needed for black to occur.

Answer

Geometric with $p = \frac{18}{38}$

The game of roulette is studied in more detail in the chapter on Games of Chance.

In the negative binomial experiment, set $k = 1$ to get the geometric distribution and set $p = 0.3$. Run the experiment 1000 times. Compute the appropriate relative frequencies and empirically investigate the memoryless property

$$\mathbb{P}(V > 5 \mid V > 2) = \mathbb{P}(V > 3) \quad (11.3.26)$$

The Petersburg Problem

We will now explore a gambling situation, known as the *Petersburg problem*, which leads to some famous and surprising results. Suppose that we are betting on a sequence of Bernoulli trials with success parameter $p \in (0, 1)$. We can bet any amount of money on a trial at *even stakes*: if the trial results in success, we receive that amount, and if the trial results in failure, we must pay that amount. We will use the following strategy, known as a *martingale strategy*:

1. We bet c units on the first trial.
2. Whenever we lose a trial, we double the bet for the next trial.
3. We stop as soon as we win a trial.

Let N denote the number of trials played, so that N has the geometric distribution with parameter p , and let W denote our net winnings when we stop.

$$W = c$$

Proof

The first win occurs on trial N , so the initial bet was doubled $N - 1$ times. The net winnings are

$$W = -c \sum_{i=0}^{N-2} 2^i + c2^{N-1} = c(1 - 2^{N-1} + 2^{N-1}) = c \quad (11.3.27)$$

Thus, W is not random and W is independent of p ! Since c is an arbitrary constant, it would appear that we have an ideal strategy. However, let us study the amount of money Z needed to play the strategy.

$$Z = c(2^N - 1)$$

The expected amount of money needed for the martingale strategy is

$$\mathbb{E}(Z) = \begin{cases} \frac{c}{2p-1}, & p > \frac{1}{2} \\ \infty, & p \leq \frac{1}{2} \end{cases} \quad (11.3.28)$$

Thus, the strategy is fatally flawed when the trials are unfavorable and even when they are fair, since we need infinite expected capital to make the strategy work in these cases.

Compute $\mathbb{E}(Z)$ explicitly if $c = 100$ and $p = 0.55$.

Answer

\$1000

In the negative binomial experiment, set $k = 1$. For each of the following values of p , run the experiment 100 times. For each run compute Z (with $c = 1$). Find the average value of Z over the 100 runs:

1. $p = 0.2$
2. $p = 0.5$
3. $p = 0.8$

For more information about gambling strategies, see the section on Red and Black. Martingales are studied in detail in a separate chapter.

The Alternating Coin-Tossing Game

A coin has probability of heads $p \in (0, 1]$. There are n players who take turns tossing the coin in round-robin style: player 1 first, then player 2, continuing until player n , then player 1 again, and so forth. The first player to toss heads wins the game.

Let N denote the number of the first toss that results in heads. Of course, N has the geometric distribution on \mathbb{N}_+ with parameter p . Additionally, let W denote the winner of the game; W takes values in the set $\{1, 2, \dots, n\}$. We are interested in the probability distribution of W .

For $i \in \{1, 2, \dots, n\}$, $W = i$ if and only if $N = i + kn$ for some $k \in \mathbb{N}$. That is, using modular arithmetic,

$$W = [(N - 1) \bmod n] + 1 \quad (11.3.29)$$

The winning player W has probability density function

$$\mathbb{P}(W = i) = \frac{p(1-p)^{i-1}}{1 - (1-p)^n}, \quad i \in \{1, 2, \dots, n\} \quad (11.3.30)$$

Proof

This follows from the previous exercise and the geometric distribution of N .

$$\mathbb{P}(W = i) = (1-p)^{i-1} \mathbb{P}(W = 1) \text{ for } i \in \{1, 2, \dots, n\}.$$

Proof

This result can be argued directly, using the memoryless property of the geometric distribution. In order for player i to win, the previous $i-1$ players must first all toss tails. Then, player i effectively becomes the first player in a new sequence of tosses. This result can be used to give another derivation of the probability density function in the previous exercise.

Note that $\mathbb{P}(W = i)$ is a decreasing function of $i \in \{1, 2, \dots, n\}$. Not surprisingly, the lower the toss order the better for the player.

Explicitly compute the probability density function of W when the coin is fair ($p = 1/2$).

Answer

$$\mathbb{P}(W = i) = 2^{n-1} / (2^n - 1), \quad i \in \{1, 2, \dots, n\}$$

Note from the [result above](#) that W itself has a *truncated geometric distribution*.

The distribution of W is the same as the conditional distribution of N given $N \leq n$:

$$\mathbb{P}(W = i) = \mathbb{P}(N = i \mid N \leq n), \quad i \in \{1, 2, \dots, n\} \quad (11.3.31)$$

The following problems explore some limiting distributions related to the alternating coin-tossing game.

For fixed $p \in (0, 1]$, the distribution of W converges to the geometric distribution with parameter p as $n \uparrow \infty$.

For fixed n , the distribution of W converges to the uniform distribution on $\{1, 2, \dots, n\}$ as $p \downarrow 0$.

Players at the end of the tossing order should hope for a coin biased towards tails.

Odd Man Out

In the game of *odd man out*, we start with a specified number of players, each with a coin that has the same probability of heads. The players toss their coins at the same time. If there is an *odd man*, that is a player with an outcome different than all of the other players, then the odd player is eliminated; otherwise no player is eliminated. In any event, the remaining players continue the game in the same manner. A slight technical problem arises with just two players, since different outcomes would make both players “odd”. So in this case, we might (arbitrarily) make the player with tails the odd man.

Suppose there are $k \in \{2, 3, \dots\}$ players and $p \in [0, 1]$. In a single round, the probability of an odd man is

$$r_k(p) = \begin{cases} 2p(1-p), & k = 2 \\ kp(1-p)^{k-1} + kp^{k-1}(1-p), & k \in \{3, 4, \dots\} \end{cases} \quad (11.3.32)$$

Proof

Let Y denote the number of heads. If $k = 2$, the event that there is an odd man is $\{Y = 1\}$. If $k \geq 3$, the event that there is an odd man is $\{Y \in \{1, k-1\}\}$. The result now follows since Y has a binomial distribution with parameters k and p .

The graph of r_k is more interesting than you might think.

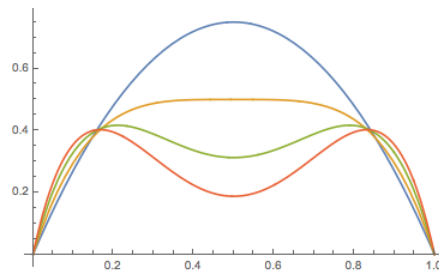


Figure 11.3.1: The graphs of r_k for $k \in \{3, 4, 5, 6\}$

For $k \in \{2, 3, \dots\}$, r_k has the following properties:

1. $r_k(0) = r_k(1) = 0$
2. r_k is symmetric about $p = \frac{1}{2}$
3. For fixed $p \in [0, 1]$, $r_k(p) \rightarrow 0$ as $k \rightarrow \infty$.

Proof

These properties are clear from the functional form of $r_k(p)$. Note that $r_k(p) = r_k(1-p)$.

For $k \in \{2, 3, 4\}$, r_k has the following properties:

1. r_k increases and then decreases, with maximum at $p = \frac{1}{2}$.
2. r_k is concave downward

Proof

This follows by computing the first derivatives: $r'_2(p) = 2(1-2p)$, $r'_3(p) = 3(1-2p)$, $r'_4(p) = 4(1-2p)^3$, and the second derivatives: $r''_2(p) = -4$, $r''_3(p) = -6$, $r''_4(p) = -24(1-2p)^2$.

For $k \in \{5, 6, \dots\}$, r_k has the following properties:

1. The maximum occurs at two points of the form p_k and $1-p_k$ where $p_k \in (0, \frac{1}{2})$ and $p_k \rightarrow 0$ as $k \rightarrow \infty$.
2. The maximum value $r_k(p_k) \rightarrow 1/e \approx 0.3679$ as $k \rightarrow \infty$.
3. The graph has a local minimum at $p = \frac{1}{2}$.

Proof sketch

Note that $r_k(p) = s_k(p) + s_k(1-p)$ where $s_k(t) = kt^{k-1}(1-t)$ for $t \in [0, 1]$. Also, $p \mapsto s_k(p)$ is the dominant term when $p > \frac{1}{2}$ while $p \mapsto s_k(1-p)$ is the dominant term when $p < \frac{1}{2}$. A simple analysis of the derivative shows that s_k increases and then decreases, reaching its maximum at $(k-1)/k$. Moreover, the maximum value is $s_k[(k-1)/k] = (1-1/k)^{k-1} \rightarrow e^{-1}$ as $k \rightarrow \infty$. Also, s_k is concave upward and then downward, with inflection point at $(k-2)/k$.

Suppose $p \in (0, 1)$, and let N_k denote the number of rounds until an odd man is eliminated, starting with k players. Then N_k has the geometric distribution on \mathbb{N}_+ with parameter $r_k(p)$. The mean and variance are

1. $\mu_k(p) = 1/r_k(p)$
2. $\sigma_k^2(p) = [1 - r_k(p)] / r_k^2(p)$

As we might expect, $\mu_k(p) \rightarrow \infty$ and $\sigma_k^2(p) \rightarrow \infty$ as $k \rightarrow \infty$ for fixed $p \in (0, 1)$. On the other hand, from the result [above](#), $\mu_k(p_k) \rightarrow e$ and $\sigma_k^2(p_k) \rightarrow e^2 - e$ as $k \rightarrow \infty$.

Suppose we start with $k \in \{2, 3, \dots\}$ players and $p \in (0, 1)$. The number of rounds until a single player remains is $M_k = \sum_{j=2}^k N_j$ where (N_2, N_3, \dots, N_k) are independent and N_j has the geometric distribution on \mathbb{N}_+ with parameter $r_j(p)$. The mean and variance are

1. $\mathbb{E}(M_k) = \sum_{j=2}^k 1/r_j(p)$
2. $\text{var}(M_k) = \sum_{j=2}^k [1 - r_j(p)] / r_j^2(p)$

Proof

The form of M_k follows from the previous result: N_k is the number of rounds until the first player is eliminated. Then the game continues independently with $k-1$ players, so N_{k-1} is the number of additional rounds until the second player is eliminated, and so forth. Parts (a) and (b) follow from the previous result and standard properties of expected value and variance.

Starting with k players and probability of heads $p \in (0, 1)$, the total number of coin tosses is $T_k = \sum_{j=2}^k jN_j$. The mean and variance are

1. $\mathbb{E}(T_k) = \sum_{j=2}^k j/r_j(p)$
2. $\text{var}(T_k) = \sum_{j=2}^k j^2 [1 - r_j(p)] / r_j^2(p)$

Proof

As before, the form of M_k follows from [result above](#): N_k is the number of rounds until the first player is eliminated, and each these rounds has k tosses. Then the game continues independently with $k-1$ players, so N_{k-1} is the number of additional rounds until the second player is eliminated with each round having $k-1$ tosses, and so forth. Parts (a) and (b) also follow from the [result above](#) and standard properties of expected value and variance.

Number of Trials Before a Pattern

Consider again a sequence of Bernoulli trials $\mathbf{X} = (X_1, X_2, \dots)$ with success parameter $p \in (0, 1)$. Recall that the number of trials M before the first success (outcome 1) occurs has the geometric distribution on \mathbb{N} with parameter p . A natural generalization is the random variable that gives the number of trials before a specific finite sequence of outcomes occurs for the first time. (Such a sequence is sometimes referred to as a *word* from the *alphabet* $\{0, 1\}$ or simply a *bit string*). In general, finding the distribution of this variable is a difficult problem, with the difficulty depending very much on the nature of the word. The problem of finding just the *expected* number of trials before a word occurs can be solved using powerful tools from the theory of renewal processes and from the theory of martingales.

To set up the notation, let \mathbf{x} denote a finite bit string and let $M_{\mathbf{x}}$ denote the number of trials before \mathbf{x} occurs for the first time. Finally, let $q = 1 - p$. Note that $M_{\mathbf{x}}$ takes values in \mathbb{N} . In the following exercises, we will consider $\mathbf{x} = 10$, a success followed by a failure. As always, try to derive the results yourself before looking at the proofs.

The probability density function f_{10} of M_{10} is given as follows:

1. If $p \neq \frac{1}{2}$ then

$$f_{10}(n) = pq \frac{p^{n+1} - q^{n+1}}{p - q}, \quad n \in \mathbb{N} \quad (11.3.33)$$

2. If $p = \frac{1}{2}$ then $f_{10}(n) = (n+1)\left(\frac{1}{2}\right)^{n+2}$ for $n \in \mathbb{N}$.

Proof

For $n \in \mathbb{N}$, the event $\{M_{10} = n\}$ can only occur if there is an initial string of 0s of length $k \in \{0, 1, \dots, n\}$ followed by a string of 1s of length $n-k$ and then 1 on trial $n+1$ and 0 on trial $n+2$. Hence

$$f_{10}(n) = \mathbb{P}(M_{10} = n) = \sum_{k=0}^n q^k p^{n-k} pq, \quad n \in \mathbb{N} \quad (11.3.34)$$

The stated result then follows from standard results on geometric series.

It's interesting to note that f is symmetric in p and q , that is, symmetric about $p = \frac{1}{2}$. It follows that the distribution function, probability generating function, expected value, and variance, which we consider below, are all also symmetric about $p = \frac{1}{2}$. It's also interesting to note that $f_{10}(0) = f_{10}(1) = pq$, and this is the largest value. So regardless of $p \in (0, 1)$ the distribution is bimodal with modes 0 and 1.

The distribution function F_{10} of M_{10} is given as follows:

1. If $p \neq \frac{1}{2}$ then

$$F_{10}(n) = 1 - \frac{p^{n+3} - q^{n+3}}{p - q}, \quad n \in \mathbb{N} \quad (11.3.35)$$

2. If $p = \frac{1}{2}$ then $F_{10} = 1 - (n+3)\left(\frac{1}{2}\right)^{n+2}$ for $n \in \mathbb{N}$.

Proof

By definition, $F_{10}(n) = \sum_{k=0}^n f_{10}(k)$ for $n \in \mathbb{N}$. The stated result then follows from the previous theorem, standard results on geometric series, and some algebra.

The probability generating function P_{10} of M_{10} is given as follows:

1. If $p \neq \frac{1}{2}$ then

$$P_{10}(t) = \frac{pq}{p-q} \left(\frac{p}{1-tp} - \frac{q}{1-tq} \right), \quad |t| < \min\{1/p, 1/q\} \quad (11.3.36)$$

2. If $p = \frac{1}{2}$ then $P_{10}(t) = 1/(t-2)^2$ for $|t| < 2$

Proof

By definition,

$$P_{10}(t) = \mathbb{E}(t^{M_{10}}) = \sum_{n=0}^{\infty} f_{10}(n)t^n \quad (11.3.37)$$

for all $t \in \mathbb{R}$ for which the series converges absolutely. The stated result then follows from the theorem above, and once again, standard results on geometric series.

The mean of M_{10} is given as follows:

1. If $p \neq \frac{1}{2}$ then

$$\mathbb{E}(M_{10}) = \frac{p^4 - q^4}{pq(p-q)} \quad (11.3.38)$$

2. If $p = \frac{1}{2}$ then $\mathbb{E}(M_{10}) = 2$.

Proof

Recall that $\mathbb{E}(M_{10}) = P'_{10}(1)$ so the stated result follows from calculus, using the previous theorem on the probability generating function. The mean can also be computed from the definition $\mathbb{E}(M_{10}) = \sum_{n=0}^{\infty} n f_{10}(n)$ using standard results from geometric series, but this method is more tedious.

The graph of $\mathbb{E}(M_{10})$ as a function of $p \in (0, 1)$ is given below. It's not surprising that $\mathbb{E}(M_{10}) \rightarrow \infty$ as $p \downarrow 0$ and as $p \uparrow 1$, and that the minimum value occurs when $p = \frac{1}{2}$.

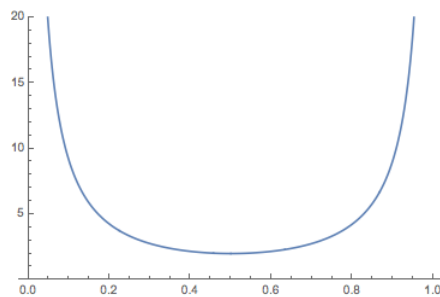


Figure 11.3.2: $\mathbb{E}(M_{10})$ as a function of p

The variance of M_{10} is given as follows:

1. If $p \neq \frac{1}{2}$ then

$$\text{var}(M_{10}) = \frac{2}{p^2 q^2} \left(\frac{p^6 - q^6}{p - q} \right) + \frac{1}{pq} \left(\frac{p^4 - q^4}{p - q} \right) - \frac{1}{p^2 q^2} \left(\frac{p^4 - q^4}{p - q} \right)^2 \quad (11.3.39)$$

2. If $p = \frac{1}{2}$ then $\text{var}(M_{10}) = 4$.

Proof

Recall that $P''_{10}(1) = \mathbb{E}[M_{10}(M_{10} - 1)]$, the second factorial moment, and so

$$\text{var}(M_{10}) = P''_{10}(1) + P'_{10}(1) - [P'_{10}(1)]^2 \quad (11.3.40)$$

The stated result then follows from calculus and the theorem above giving the probability generating function.

This page titled [11.3: The Geometric Distribution](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.