

4. The Negative Binomial Distribution

Basic Theory

Suppose again that our [random experiment](#) is to perform a sequence of [Bernoulli trials](#) $X = (X_1, X_2, \dots)$ with parameter $p \in (0, 1]$. Recall that the number of successes in the first n trials

$$Y_n = \sum_{i=1}^n X_i$$

has the [binomial distribution](#) with parameters n and p . In this section we will study the [random variable](#) that gives the **trial number of the k^{th} success**:

$$V_k = \min \{n \in \{1, 2, \dots\} : Y_n = k\}$$

Note that V_1 is the number of trials needed to get the first success, which we now know has the [geometric distribution](#) on \mathbb{N}_+ with parameter p .

The Probability Density Function

1. Show that $V_k = n$ if and only if $X_n = 1$ and $Y_{n-1} = k - 1$.

2. Use Exercise 1, independence, and the [binomial distribution](#) to show that

$$\mathbb{P}(V_k = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n \in \{k, k+1, k+2, \dots\}$$

The distribution defined by the density function in Exercise 2 is known as the **negative binomial distribution**; it has two parameters, the number of successes k and the success probability p .

3. In the **negative binomial experiment**, vary k and p with the scroll bars and note the shape of the density function. For selected values of k and p , run the experiment 1000 times with an update frequency of 10. Watch the apparent convergence of the relative frequency function to the density function.

4. Show that the binomial and negative binomial sequences are inverse to each other in the sense that

$$Y_n \geq k \text{ if and only if } V_k \leq n$$

This property is implicit in the definition of V_k given at the beginning of this section. In particular, argue that any event that can be expressed in terms of the negative binomial variables can also be expressed in terms of the binomial variables.

5. Show that $\mathbb{P}(V_k = n) > \mathbb{P}(V_k = n - 1)$ if and only if $n < 1 + \frac{k-1}{p}$. Thus, the density function at first

increases and then decreases, reaching its maximum value at $\lfloor 1 + \frac{k-1}{p} \rfloor$. This integer is the **mode** of the distribution and hence the negative binomial distribution is **unimodal**.

Times Between Successes

Next we will define the random variables that give the **number of trials between successive successes**. Let

$$U_1 = V_1 \text{ and } U_k = V_k - V_{k-1} \text{ for } k \in \{2, 3, \dots\}$$

6. Show that $U = (U_1, U_2, \dots)$ is a sequence of independent random variables, each having the geometric distribution on \mathbb{N}_+ with parameter p . Moreover,

$$V_k = \sum_{i=1}^k U_i$$

In statistical terms, U corresponds to sampling from the geometric distribution with parameter p , so that for each k , (U_1, U_2, \dots, U_k) is a random sample of size k from this distribution. The sequence of negative binomial variables V is the **partial sum process** corresponding to the sequence U . In statistical terms, $\frac{V_k}{k}$ is the sample mean corresponding to the random sample (U_1, U_2, \dots, U_k) ; this random variable gives the average number of trials between the first k successes. **Partial sum processes** are studied in more generality in the chapter on [Random Samples](#).

7. Use the partial sum representation to prove the following properties.

- If $j \leq k$ then $V_k - V_j$ has the same distribution as V_{k-j} , namely negative binomial with parameters $k - j$ and p . Thus, the process V has **stationary increments**.
- If $k_1 \leq k_2 \leq k_3 \leq \dots$ then $(V_{k_1}, V_{k_2} - V_{k_1}, V_{k_3} - V_{k_2}, \dots)$ is a sequence of independent random variables. Thus, the process V has **independent increments**.

Actually, any partial sum process corresponding to an independent, identically distributed sequence will have stationary, independent increments.

Moments

The **mean**, **variance** and **probability generating function** of V_k now follow easily from the representation as a sum of independent, identically distributed geometrically distributed variables..

8. Show that $\mathbb{E}(V_k) = k \frac{1}{p}$.

9. Show that $\text{var}(V_k) = k \frac{1-p}{p^2}$.

10. Show that $\mathbb{E}(t^{V_k}) = \left(\frac{p t}{1 - (1-p)t} \right)^k$ for $|t| < \frac{1}{1-p}$

11. In the **negative binomial experiment**, vary k and p with the scroll bars and note the location and size of the mean/standard deviation bar. For selected values of the parameters, run the experiment 1000 times with an update frequency of 10. Watch the apparent convergence of the sample mean and standard deviation to the distribution mean and standard deviation.
12. Verify the results of [Exercise 8](#), [Exercise 9](#), and [Exercise 10](#) directly using the probability density function. Note that this method requires significantly more work.
13. Suppose that V and W are independent random variables for an experiment, and that V has the negative binomial distribution with parameters j and p , and W has the negative binomial distribution with parameters k and p . Show that $V + W$ has the negative binomial distribution with parameters $k + j$ and p .
- Give a probabilistic proof, based on the partial sum representation.
 - Give an analytic proof, based on probability density functions
 - Give an analytic proof, based on probability generating functions.

Normal Approximation

14. In the **negative binomial experiment**, start with various values of p and $k = 1$. Successively increase k by 1, noting the shape of the probability density function each time.

Even though you are limited to $k = 5$, you can still see the characteristic bell shape. This is a consequence of the [central limit theorem](#) because the negative binomial variable can be written as a sum of k independent, identically distributed (geometric) random variables.

15. Show that the distribution of the standardized variable given below [converges](#) to the standard [normal distribution](#) as $k \rightarrow \infty$.

$$Z_k = \frac{p V_k - k}{\sqrt{k(1-p)}}$$

From a practical point of view, this result means that if k is “large”, the distribution of V_k is approximately normal with mean and variance given in [Exercise 8](#) and [Exercise 9](#), respectively. Just how large k needs to be for the approximation to work well depends on p . Also, when using the normal approximation, we should remember to use the [continuity correction](#), since the negative binomial is a discrete distribution.

Relation to Order Statistics

16. Suppose that $k \leq n$. Show that

$$\mathbb{P}(V_1 = j_1, V_2 = j_2, \dots, V_k = j_k | Y_n = k) = \frac{1}{\binom{n}{k}} \text{ for } (j_1, j_2, \dots, j_k) \in L$$

where $L = \{(j_1, j_2, \dots, j_k) \in \{1, 2, \dots, n\}^k : j_1 < j_2 < \dots < j_k\}$. Thus, given k successes in the first n

trials, the vector of success trial numbers is uniformly distributed on L . Equivalently, the vector of success trial numbers is distributed as the vector of **order statistics** corresponding to a sample of size k chosen at random and without replacement from $\{1, 2, \dots, n\}$.

17. Show that

$$\mathbb{P}(V_m = j | Y_n = k) = \frac{\binom{j-1}{m-1} \binom{n-j}{k-m}}{\binom{n}{k}}, \quad j \in \{m, m+1, \dots, n+(k-m)\}$$

Again, this is the distribution of the m^{th} **order statistic** when a sample of size k is selected at random and without replacement from the population $\{1, 2, \dots, n\}$.

Examples and Applications

18. A standard, fair die is thrown until 3 aces occur. Let V denote the number of throws.

- Find the probability density function of V .
- Find the mean of V
- Find the variance of V
- Find the probability that at least 20 throws will be needed.



19. A coin is tossed repeatedly. The 10th head occurs on the 25th toss.

- Find the probability density function of the trial number of the 5th head.
- Find the mean of the distribution in (a)
- Find the variance of the distribution in (a)



20. A certain type of missile has failure probability 0.02. Let N denote the launch number of the fourth failure.

- Find the probability density function of N .
- Find the mean of N
- Find the variance of N
- Find the probability that there will be at least 4 failures in the first 200 launches.



21. In the **negative binomial experiment**, set $p = 0.5$ and $k = 5$. Run the experiment 1000 times with an update frequency of 100. Compute and compare each of the following:

- $\mathbb{P}(8 \leq V_5 \leq 15)$

- b. The relative frequency of the event $\{8 \leq V_5 \leq 15\}$ in the simulation.
- c. The normal approximation to $\mathbb{P}(8 \leq V_5 \leq 15)$



22. A coin is tossed until the 50th head occurs.

- a. Assuming that the coin is fair, find the normal approximation of the probability that the coin is tossed at least 125 times.
- b. Suppose that you perform this experiment, and 125 tosses are required. Do you believe that the coin is fair?



The Banach Match Problem

Suppose that an absent-minded professor (is there any other kind?) has m matches in his right pocket and m matches in his left pocket. When he needs a match to light his pipe, he is equally likely to choose a match from either pocket. We want to compute the probability density function of the random variable W that gives the number of matches remaining when the professor first discovers that one of the pockets is empty. This is known as the **Banach match problem**, named for the mathematician **Stefan Banach**, who evidently behaved in the manner described.

We can recast the problem in terms of the negative binomial distribution. Clearly the match choices form a sequence of Bernoulli trials with parameter $p = \frac{1}{2}$. Specifically, we can consider a match from the right pocket as a win for player R , and a match from the left pocket as a win for player L . In a hypothetical infinite sequence of trials, let U denote the number of trials necessary for R to win $m + 1$ trials, and V the number of trials necessary for L to win $m + 1$ trials. Note that U and V each have the negative binomial distribution with parameters $m + 1$ and p .

23. For $k \in \{0, 1, \dots, m\}$, show that

- a. L has $m - k$ wins at the moment when R wins $m + 1$ games if and only if $U = 2m - k + 1$
- b. $\{U = 2m - k + 1\}$ is equivalent to the event that the professor first discovers that the right pocket is empty and that the left pocket has k matches
- c. $\mathbb{P}(U = 2m - k + 1) = \binom{2m-k}{m} \left(\frac{1}{2}\right)^{2m-k+1}$.

24. For $k \in \{0, 1, \dots, m\}$, show that

- a. R has $m - k$ wins at the moment when L wins $m + 1$ games if and only if $V = 2m - k + 1$
- b. $\{V = 2m - k + 1\}$ is equivalent to the event that the professor first discovers that the right pocket is empty and that the left pocket has k matches

$$c. \mathbb{P}(V = 2m - k + 1) = \binom{2m - k}{m} \left(\frac{1}{2}\right)^{2m - k + 1}.$$

25. Combine the results of the previous two exercises to conclude that

$$\mathbb{P}(W = k) = \binom{2m - k}{m} \left(\frac{1}{2}\right)^{2m - k}, \quad k \in \{0, 1, \dots, m\}$$

We can also solve the **non-symmetric Banach match problem**, using the same methods as above. Thus, suppose that the professor reaches for a match in his right pocket with probability p and in his left pocket with probability $1 - p$, where $0 < p < 1$. The essential change in the analysis is that U has the negative binomial distribution with parameters $m + 1$ and p , while V has the negative binomial distribution with parameters $m + 1$ and $1 - p$.

26. Show that

$$\mathbb{P}(W = k) = \binom{2m - k}{m} (p^{m+1} (1 - p)^{m-k} + (1 - p)^{m+1} p^{m-k}), \quad k \in \{0, 1, \dots, m\}$$

The Problem of Points

Suppose that two teams (or individuals) A and B play a sequence of Bernoulli trials, where p is the probability that player A wins a trial. For nonnegative integers n and m , let $A_{n,m}(p)$ denote the probability that A wins n points before B wins m points. Computing $A_{n,m}$ is an historically famous problem, known as **the problem of points**, that was solved by **Pierre de Fermat** and by **Blaise Pascal**.

27. Comment on the validity of the Bernoulli trial assumptions (independence of trials and constant probability of success) for games of sport that have a *skill* component as well as a *random* component.

There is an easy solution to the problem of points using the binomial distribution; this was essentially Pascal's solution. Let us pretend that $n + m - 1$ trials are played, regardless of the outcomes, and let Y_{n+m-1} denote the number of trials that A wins. By definition, Y_{n+m-1} has the binomial distribution with parameters $n + m - 1$ and p .

28. Show that A wins n trials before B wins m trials if and only if

$$Y_{n+m-1} \geq n$$

29. Use the result of the previous exercise to show that

$$A_{n,m}(p) = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k (1-p)^{n+m-1-k}$$

There is also an easy solution to the problem of points using the negative binomial distribution. In a sense, this has to be the case, given the equivalence between the binomial and negative binomial processes. First, let us pretend that the games go on forever, regardless of the outcomes, and let V_n denote the number of games needed for A to win n games. By definition, V_n has the negative binomial distribution with parameters n and p

30. Show that A wins n trials before B wins m trials if and only if

$$V_n \leq n + m - 1$$

31. Use the result of the previous exercise to show that

$$A_{n,m}(p) = \sum_{j=n}^{n+m-1} \binom{j-1}{n-1} p^n (1-p)^{j-n}$$

32. Show that for fixed n and m , $A_{n,m}(p)$ increases from 0 to 1 as p increases from 0 to 1.

33. Show that

- $A_{n,m}(p)$ decreases as n increases for fixed m and p .
- $A_{n,m}(p)$ increases as m increases for fixed n and p .

34. Show that $1 - A_{n,m}(p) = A_{m,n}(1-p)$ for any n , m , and p .

- Give an analytic proof.
- Give a probabilistic argument.

35. In the **problem of points experiments**, vary the parameters n , m , and p , and note how the probability changes. For selected values of the parameters, run the simulation 1000 times with an update frequency of 10. Note the apparent convergence of the relative frequency to the probability.

36. Condition on the outcome of the first trial to derive the following recurrence relation and boundary conditions (this was essentially Fermat's solution):

- $A_{n,m}(p) = p A_{n-1,m}(p) + (1-p) A_{n,m-1}(p)$ for $n \in \mathbb{N}_+$ and $m \in \mathbb{N}_+$
- $A_{n,0}(p) = 0$, $A_{0,m}(p) = 1$

Now let's study the number of trials needed for the problem of points experiment. Let $N_{n,m}$ denote the number of trials needed until either A wins n points or B wins m points, which occurs first. The negative binomial variables provide an easy derivation of the distribution $N_{n,m}$. Again, imagine that we continue the trials indefinitely. Let V_n denote the number of trials needed for A to win n points, and let W_m denote the number of trials needed for B to win m points.

37. Show that for $k \in \{\min\{m, n\}, \dots, n+m-1\}$

$$\mathbb{P}(N_{n,m} = k) = \mathbb{P}(V_n = k) + \mathbb{P}(W_m = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n} + \binom{k-1}{m-1} (1-p)^m p^{k-m}$$

Series of Games

The special case of the problem of points experiment with $m = n$ is important, because it corresponds to A and B playing a **best of $2n - 1$ game series**. That is, the first player to win n games wins the series. Such series, especially when $n \in \{2, 3, 4\}$ are frequently used in championship tournaments. We will let $A_n(p) = A_{n,n}(p)$ denote the probability that player A wins the series. From our general results with the problem of points, we have

$$A_n(p) = \sum_{k=n}^{2n-1} \binom{2n-1}{k} p^k (1-p)^{(2n-1)-k} = \sum_{j=n}^{2n-1} \binom{j-1}{n-1} p^n (1-p)^{j-n}$$

38. Suppose that $p = 0.6$. Explicitly find the probability that team A wins in each of the following cases:

- A best of 5 game series.
- A best of 7 game series.



39. In the **problem of points experiments**, vary the parameters n , m , and p (keeping $n = m$), and note how the probability changes. Now simulate a best of 5 series by selecting $n = m = 3$, $p = 0.6$. Run the experiment 1000 times, updating every 10 runs. Note the apparent convergence of the relative frequency to the true probability.

40. Show that $A_n(1-p) = 1 - A_n(p)$ for any n and p .

- Show that this condition means that the graph of A_n is symmetric with respect to $p = \frac{1}{2}$.
- Show that this condition implies that $A_n\left(\frac{1}{2}\right) = \frac{1}{2}$.

41. In the **problem of points experiments**, vary the parameters n , m , and p (keeping $n = m$), and note how the probability changes. Now simulate a best 7 series by selecting $n = m = 4$, $p = 0.45$. Run the experiment 1000 times, updating every 10 runs. Note the apparent convergence of the relative frequency to the true probability.

42. Suppose that $n > m$. Show that $A_n(p) > A_m(p)$ if and only if $p > \frac{1}{2}$. Interpret the result.

43. Let N_n denote the number of trials in the series. Use [Exercise 37](#) to show that

$$\mathbb{P}(N_n = k) = \binom{k-1}{n-1} (p^n (1-p)^{k-n} + (1-p)^n p^{k-n}), \quad k \in \{n, n+1, \dots, 2n-1\}$$

44. Explicitly compute the probability density function, expected value, and standard deviation for the number of games in a best of 7 series with the following values of p :

- 0.5
- 0.7
- 0.9



Division of Stakes

The problem of points originated from a question posed by **Chevalier de Mere**, who was interested in the fair division of stakes when a game is interrupted. Specifically, suppose that players A and B each put up C monetary units, and then play Bernoulli trials until one of them wins a specified number of trials. The winner then takes the entire $2C$ fortune.

45. If the game is interrupted when A needs to win n more trials and B needs to win m more trials, argue that the fortune should be divided between A and B , respectively, as follows:

- $2C A_{n,m}(p)$ for A ,
- $2C (1 - A_{n,m}(p))$ for B .

46. Suppose that players A and B bet \$50 each. The players toss a fair coin until one of them has 10 wins; the winner takes the entire fortune. Suppose that the game is interrupted by the gambling police when A has 5 wins and B has 3 wins. How should the stakes be divided?



[Virtual Laboratories](#) > [11. Bernoulli Trials](#) > [1](#) [2](#) [3](#) **[4](#)** [5](#) [6](#)

[Contents](#) | [Applets](#) | [Data Sets](#) | [Biographies](#) | [External Resources](#) | [Key words](#) | [Feedback](#) | ©