

11.4: The Negative Binomial Distribution

Basic Theory

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

$$Y_n = \sum_{i=1}^n X_i \quad (11.4.1)$$

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*:

$$V_k = \min \{n \in \mathbb{N}_+ : Y_n = k\} \quad (11.4.2)$$

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

The probability distribution of V_k is given by

$$\mathbb{P}(V_k = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n \in \{k, k+1, k+2, \dots\} \quad (11.4.3)$$

Proof

Note that $V_k = n$ if and only if $X_n = 1$ and $Y_{n-1} = k-1$. Hence, from independence and the binomial distribution,

$$\mathbb{P}(V_k = n) = \mathbb{P}(Y_{n-1} = k-1) \mathbb{P}(X_n = 1) = \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} p = \binom{n-1}{k-1} p^k (1-p)^{n-k} \quad (11.4.4)$$

The distribution defined by the density function in (1) is known as the *negative binomial distribution*; it has two parameters, the *stopping parameter* k and the *success probability* p .

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 and compare the relative frequency function to the probability density function.

The binomial and negative binomial sequences are inverse to each other in a certain sense.

For $n \in \mathbb{N}_+$ and $k \in \{0, 1, \dots, n\}$,

1. $Y_n \geq k \iff V_k \leq n$ and hence $\mathbb{P}(Y_n \geq k) = \mathbb{P}(V_k \leq n)$
2. $k \mathbb{P}(Y_n = k) = n \mathbb{P}(V_k = n)$

Proof

1. The events $\{Y_n \geq k\}$ and $\{V_k \leq n\}$ both mean that there are at least k successes in the first n Bernoulli trials.
2. From the formulas for the binomial and negative binomial PDFs, $k \mathbb{P}(Y_n = k)$ and $n \mathbb{P}(V_k = n)$ both simplify to $\frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$.

In particular, it follows from part (a) that any event that can be expressed in terms of the negative binomial variables can also be expressed in terms of the binomial variables.

The negative binomial distribution is unimodal. Let $t = 1 + \frac{k-1}{p}$. Then

1. $\mathbb{P}(V_k = n) > \mathbb{P}(V_k = n-1)$ if and only if $n < t$.
2. The probability density function at first increases and then decreases, reaching its maximum value at $\lfloor t \rfloor$.
3. There is a single mode at $\lfloor t \rfloor$ if t is not an integer, and two consecutive modes at $t-1$ and t if t is an integer.

Times Between Successes

Next we will define the random variables that give the *number of trials between successive successes*. Let $U_1 = V_1$ and $U_k = V_k - V_{k-1}$ for $k \in \{2, 3, \dots\}$

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

$$V_k = \sum_{i=1}^k U_i \quad (11.4.5)$$

In statistical terms, \mathbf{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 sample mean corresponding to this sample is V_k/k ; this random variable gives the average number of trials between the first k successes. In probability terms, the sequence of negative binomial variables \mathbf{V} is the *partial sum process* corresponding to the sequence \mathbf{U} . Partial sum processes are studied in more generality in the chapter on Random Samples.

The random process $\mathbf{V} = (V_1, V_2, \dots)$ has *stationary, independent increments*:

1. If $j < k$ then $V_k - V_j$ has the same distribution as V_{k-j} , namely negative binomial with parameters $k-j$ and p .
2. If $k_1 < k_2 < k_3 < \dots$ then $(V_{k_1}, V_{k_2} - V_{k_1}, V_{k_3} - V_{k_2}, \dots)$ is a sequence of independent random variables.

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

Basic Properties

The mean, variance and probability generating function of V_k can be computed in several ways. The method using the representation as a sum of independent, identically distributed geometrically distributed variables is the easiest.

V_k has probability generating function P given by

$$P(t) = \left(\frac{pt}{1 - (1-p)t} \right)^k, \quad |t| < \frac{1}{1-p} \quad (11.4.6)$$

Proof

Recall that the probability generating function of a sum of independent variables is the product of the probability generating functions of the variables. Recall also, the probability generating function of the geometric distribution with parameter p is $t \mapsto pt / [1 - (1-p)t]$. Thus, the result follows immediately from the [sum representation](#) above. A derivation can also be given directly from the probability density function.

The mean and variance of V_k are

1. $\mathbb{E}(V_k) = k \frac{1}{p}$.
2. $\text{var}(V_k) = k \frac{1-p}{p^2}$

Proof

The geometric distribution with parameter p has mean $1/p$ and variance $(1-p)/p^2$, so the results follow immediately from the [sum representation](#) above. Recall that the mean of a sum is the sum of the means, and the variance of the sum of *independent* variables is the sum of the variances. These results can also be proven directly from the probability density function or from the probability generating function.

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 and compare the sample mean and standard deviation to the distribution mean and standard deviation.

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 . Then $V + W$ has the negative binomial distribution with parameters $k+j$ and p .

Proof

Once again, the simplest proof is based on the representation as a sum of independent geometric variables. In the context of the [sum representation](#) above, we can take $V = V_j$ and $W = V_{k+j} - V_j$, so that $V + W = V_{k+j}$. Another simple proof uses probability generating functions. Recall again that the PGF of the sum of independent variables is the product of the PGFs. Finally, a difficult proof can be constructed using probability density functions. Recall that the PDF of a sum of independent variables is the convolution of the PDFs.

Normal Approximation

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 $k = 5$ in the app, 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.

The standard score of V_k is

$$Z_k = \frac{p V_k - k}{\sqrt{k(1-p)}} \quad (11.4.7)$$

The distribution of Z_k converges to the standard normal distribution as $k \rightarrow \infty$.

From a practical point of view, this result means that if k is “large”, the distribution of V_k is approximately normal with mean $k\frac{1}{p}$ and variance $k\frac{1-p}{p^2}$. 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

Suppose that $n \in \mathbb{N}_+$ and $k \in \{1, 2, \dots, n\}$, and let $L = \{(n_1, n_2, \dots, n_k) \in \{1, 2, \dots, n\}^k : n_1 < n_2 < \dots < n_k\}$. Then

$$\mathbb{P}(V_1 = n_1, V_2 = n_2, \dots, V_k = n_k \mid Y_n = k) = \frac{1}{\binom{n}{k}}, \quad (n_1, n_2, \dots, n_k) \in L \quad (11.4.8)$$

Proof

$$\mathbb{P}(V_1 = n_1, V_2 = n_2, \dots, V_k = n_k \mid Y_n = k) = \frac{\mathbb{P}(V_1 = n_1, V_2 = n_2, \dots, V_k = n_k, Y_n = k)}{\mathbb{P}(Y_n = k)} = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}} \quad (11.4.9)$$

Note that the event in the numerator of the first fraction means that in the first n trials, successes occurred at trials n_1, n_2, \dots, n_k and failures occurred at all other trials.

Thus, given exactly k successes in the first n trials, the vector of success trial numbers is uniformly distributed on the set of possibilities L , regardless of the value of the success parameter p . 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 $\{1, 2, \dots, n\}$.

Suppose that $n \in \mathbb{N}_+$, $k \in \{1, 2, \dots, n\}$, and $j \in \{1, 2, \dots, k\}$. Then

$$\mathbb{P}(V_j = m \mid Y_n = k) = \frac{\binom{m-1}{j-1} \binom{n-m}{k-j}}{\binom{n}{k}}, \quad m \in \{j, j+1, \dots, n+k-j\} \quad (11.4.10)$$

Proof

This follows immediately from the previous result and a theorem in the section on order statistics. However, a direct proof is also easy. Note that the event $\{V_j = m, Y_n = k\}$ means that there were $j-1$ successes in the first $m-1$ trials, a success on trial m and $k-j$ success in trials $m+1$ to n . Hence using the binomial distribution and independence,

$$\mathbb{P}(V_j = m \mid Y_n = k) = \frac{\mathbb{P}(V_j = m, Y_n = k)}{\mathbb{P}(Y_n = k)} = \frac{\binom{m-1}{j-1} p^{j-1} (1-p)^{(m-1)-(j-1)} p \binom{n-m}{k-j} p^{k-j} (1-p)^{(n-m)-(k-j)}}{\binom{n}{k} p^k (1-p)^{n-k}} \quad (11.4.11)$$

$$= \frac{\binom{m-1}{j-1} \binom{n-m}{k-j} p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{\binom{m-1}{j-1} \binom{n-m}{k-j}}{\binom{n}{k}}, \quad (11.4.12)$$

Thus, given exactly k successes in the first n trials, the trial number of the j th success has the same distribution as the j th order statistic when a sample of size k is selected at random and without replacement from the population $\{1, 2, \dots, n\}$. Again, this result does not depend on the value of the success parameter p . The following theorem gives the mean and variance of the conditional distribution.

Suppose again that $n \in \mathbb{N}_+$, $k \in \{1, 2, \dots, n\}$, and $j \in \{1, 2, \dots, k\}$. Then

1. $\mathbb{E}(V_j \mid Y_n = k) = j\frac{n+1}{k+1}$
2. $\text{var}(V_j \mid Y_n = k) = j(k-j+1)\frac{(n+1)(n-k)}{(k+1)^2(k+2)}$

Proof

These moment results follow immediately from the previous theorem and a theorem in the section on order statistics. However, there is also a nice heuristic argument for (a) using indicator variables. Given $Y_n = k$, the k successes divide the set of indices where the failures occur into $k+1$ disjoint sets (some may be empty, of course, if there are adjacent successes).

The red dots are successes and the green dots failures. The 8 successes in the 50 trials divide the set of failures into 9 disjoint sets.



Let I_i take the value 1 if the i th failure occurs before the j th success, and 0 otherwise, for $i \in \{1, 2, \dots, n-k\}$. Then given $Y_n = k$,

$$V_j = j + \sum_{i=1}^{n-k} I_i \quad (11.4.13)$$

Given $Y_n = k$, we know that the k successes and $n-k$ failures are randomly placed in $\{1, 2, \dots, n\}$ with each possible configuration having the same probability. Thus,

$$\mathbb{E}(I_i | Y_n = k) = \mathbb{P}(I_i = 1 | Y_n = k) = \frac{j}{k+1}, \quad i \in \{1, 2, \dots, n-k\} \quad (11.4.14)$$

Hence

$$\mathbb{E}(V_j | Y_n = k) = j + (n-k) \frac{j}{k+1} = j \frac{n+1}{k+1} \quad (11.4.15)$$

Examples and Applications

Coins, Dice and Other Gadgets

A standard, fair die is thrown until 3 aces occur. Let V denote the number of throws. Find each of the following:

1. The probability density function of V .
2. The mean of V .
3. The variance of V .
4. The probability that at least 20 throws will be needed.

Answer

1. $\mathbb{P}(V = n) = \binom{n-1}{2} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^{n-3}, \quad n \in \{3, 4, \dots\}$
2. $\mathbb{E}(V) = 18$
3. $\text{var}(V) = 90$
4. $\mathbb{P}(V \geq 20) = 0.3643$

A coin is tossed repeatedly. The 10th head occurs on the 25th toss. Find each of the following:

1. The probability density function of the trial number of the 5th head.
2. The mean of the distribution in (a).
3. The variance of the distribution in (a).

Answer

1. $\mathbb{P}(V_5 = m | V_{10} = 25) = \frac{\binom{m-1}{4} \binom{24-m}{4}}{\binom{24}{9}}, \quad m \in \{5, 6, \dots, 25\}$
2. $\mathbb{E}(V_5 | V_{10} = 25) = \frac{25}{2}$
3. $\text{var}(V_5 | V_{10} = 25) = \frac{375}{44}$

A certain type of missile has failure probability 0.02. Let N denote the launch number of the fourth 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 that there will be at least 4 failures in the first 200 launches.

Answer

1. $\mathbb{P}(N = n) = \binom{n-1}{3} \left(\frac{1}{50}\right)^4 \left(\frac{49}{50}\right)^{n-4}, \quad n \in \{4, 5, \dots\}$
2. $\mathbb{E}(N) = 200$
3. $\text{var}(N) = 9800$
4. $\mathbb{P}(N \leq 200) = 0.5685$

In the negative binomial experiment, set $p = 0.5$ and $k = 5$. Run the experiment 1000 times. Compute and compare each of the following:

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

Answer

1. $\mathbb{P}(8 \leq V_5 \leq 15) = 0.7142$
3. $\mathbb{P}(8 \leq V_5 \leq 15) \approx 0.7445$

A coin is tossed until the 50th head occurs.

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

Answer

1. 0.0072
2. No.

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 .

For $k \in \{0, 1, \dots, m\}$,

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

For $k \in \{0, 1, \dots, m\}$,

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

W has probability density function

$$\mathbb{P}(W = k) = \binom{2m-k}{m} \left(\frac{1}{2}\right)^{2m-k}, \quad k \in \{0, 1, \dots, m\} \quad (11.4.16)$$

Proof

This result follows from the previous two exercises, since $\mathbb{P}(W = k) = \mathbb{P}(U = 2m - k + 1) + \mathbb{P}(V = 2m - k + 1)$.

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 $1 - p$.

For the Banach match problem with parameter p , W has probability density function

$$\mathbb{P}(W = k) = \binom{2m-k}{m} [p^{m+1}(1-p)^{m-k} + (1-p)^{m+1}p^{m-k}], \quad k \in \{0, 1, \dots, m\} \quad (11.4.17)$$

The Problem of Points

Suppose that two teams (or individuals) A and B play a sequence of Bernoulli trials (which we will also call *points*), where $p \in (0, 1)$ is the probability that player A wins a point. 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 $A_{n,m}(p)$ is an historically famous problem, known as the *problem of points*, that was solved by Pierre de Fermat and by Blaise Pascal.

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. 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 in (3). First, let us pretend that the trials go on forever, regardless of the outcomes. Let Y_{n+m-1} denote the number of wins by player A in the first $n + m - 1$ points, and let V_n denote the number of trials needed for A to win n points. By definition, Y_{n+m-1} has the binomial distribution with parameters $n + m - 1$ and p , and V_n has the negative binomial distribution with parameters n and p .

Player A wins n points before B wins m points if and only if $Y_{n+m-1} \geq n$ if and only if $V_n \leq m+n-1$. Hence

$$A_{n,m}(p) = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k (1-p)^{n+m-1-k} = \sum_{j=n}^{n+m-1} \binom{j-1}{n-1} p^n (1-p)^{j-n} \quad (11.4.18)$$

$A_{n,m}(p)$ satisfies the following properties:

1. $A_{n,m}(p)$ increases from 0 to 1 as p increases from 0 to 1 for fixed n and m .
2. $A_{n,m}(p)$ decreases as n increases for fixed m and p .
3. $A_{n,m}(p)$ increases as m increases for fixed n and p .

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

Proof

A simple probabilistic proof is to note that both sides can be interpreted as the probability that a player with point probability $1-p$ wins m points before the opponent wins n points. An analytic proof can also be constructed using the [formulas above](#) for $A_{n,m}(p)$.

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 and note the apparent convergence of the relative frequency to the probability.

The win probability function for player A satisfies the following recurrence relation and boundary conditions (this was essentially Fermat's solution):

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

Proof

Condition on the outcome of the first trial.

Next let $N_{n,m}$ denote the number of trials needed until either A wins n points or B wins m points, whichever occurs first—the length of the problem of points experiment. The following result gives the distribution of $N_{n,m}$.

For $k \in \{\min\{m, n\}, \dots, n+m-1\}$

$$\mathbb{P}(N_{n,m} = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n} + \binom{k-1}{m-1} (1-p)^m p^{k-m} \quad (11.4.19)$$

Proof

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 denote the number of trials needed for B to win m points. Then $\mathbb{P}(N_{n,m} = k) = \mathbb{P}(V_n = k) + \mathbb{P}(W_m = k)$ for k in the indicated range.

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.

Let $A_n(p)$ denote the probability that player A wins the series. Then

$$A_n(p) = \sum_{k=n}^{2n-1} \binom{2n-1}{k} p^k (1-p)^{2n-1-k} = \sum_{j=n}^{2n-1} \binom{j-1}{n-1} p^n (1-p)^{j-n} \quad (11.4.20)$$

Proof

This follows directly from the [problem of points probability](#) above, since $A_n(p) = A_{n,n}(p)$.

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

1. A best of 5 game series.
2. A best of 7 game series.

Answer

1. 0.6825.
2. 0.7102

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 and compare the relative frequency to the true probability.

$A_n(1 - p) = 1 - A_n(p)$ for any $n \in \mathbb{N}_+$ and $p \in [0, 1]$. Therefore

1. The graph of A_n is symmetric with respect to $p = \frac{1}{2}$.
2. $A_n\left(\frac{1}{2}\right) = \frac{1}{2}$.

Proof

Again, there is a simple probabilistic argument for the equation: both sides represent the probability that a player with game probability $1 - p$ will win the series.

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 and compare the relative frequency to the true probability.

If $n > m$ then

1. $A_n(p) < A_m(p)$ if $0 < p < \frac{1}{2}$
2. $A_n(p) > A_m(p)$ if $\frac{1}{2} < p < 1$

Proof

The greater the number of games in the series, the more the series favors the stronger player (the one with the larger game probability).

Let N_n denote the number of trials in the series. Then N_n has probability density function

$$\mathbb{P}(N_n = k) = \binom{k-1}{n-1} [p^n(1-p)^{k-n} + (1-p)^n p^{k-n}], \quad k \in \{n, n+1, \dots, 2n-1\} \quad (11.4.21)$$

Proof

This result follows directly from the corresponding [problem of points result](#) above with $n = m$.

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 :

1. 0.5
2. 0.7
3. 0.9

Answer

1. $f(k) = \binom{k-1}{3} \left(\frac{1}{2}\right)^{k-1}$, $k \in \{4, 5, 6, 7\}$, $\mathbb{E}(N) = 5.8125$, $\text{sd}(N) = 1.0136$
2. $f(k) = \binom{k-1}{3} [(0.7)^4(0.3)^{k-4} + (0.3)^4(0.7)^{k-4}]$, $k \in \{4, 5, 6, 7\}$, $\mathbb{E}(N) = 5.3780$, $\text{sd}(N) = 1.0497$
3. $f(k) = \binom{k-1}{3} [(0.9)^4(0.1)^{k-4} + (0.1)^4(0.9)^{k-4}]$, $k \in \{4, 5, 6, 7\}$, $\mathbb{E}(N) = 4.4394$, $\text{sd}(N) = 0.6831$

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.

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

1. $2cA_{n,m}(p)$ for A
2. $2c[1 - A_{n,m}(p)] = 2cA_{m,n}(1-p)$ for B .

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?

Answer

A gets \$72.56, B gets \$27.44

Alternate and General Versions

Let's return to the formulation at the beginning of this section. Thus, suppose that we have a sequence of Bernoulli trials \mathbf{X} with success parameter $p \in (0, 1]$, and for $k \in \mathbb{N}_+$, we let V_k denote the trial number of the k th success. Thus, V_k has the negative binomial distribution with parameters k and p as we studied above. The random variable $W_k = V_k - k$ is the number of failures before the k th success. Let $N_1 = W_1$, the number of failures before the first success, and let $N_k = W_k - W_{k-1}$, the number of failures between the $(k-1)$ st success and the k th success, for $k \in \{2, 3, \dots\}$.

$\mathbf{N} = (N_1, N_2, \dots)$ is a sequence of independent random variables, each having the geometric distribution on \mathbb{N} with parameter p . Moreover,

$$W_k = \sum_{i=1}^k N_i \quad (11.4.22)$$

Thus, $\mathbf{W} = (W_1, W_2, \dots)$ is the partial sum process associated with \mathbf{N} . In particular, \mathbf{W} has stationary, independent increments.

Probability Density Functions

The probability density function of W_k is given by

$$\mathbb{P}(W_k = n) = \binom{n+k-1}{k-1} p^k (1-p)^n = \binom{n+k-1}{n} p^k (1-p)^n, \quad n \in \mathbb{N} \quad (11.4.23)$$

Proof

This result follows directly from the PDF of V_k , since $\mathbb{P}(W_k = n) = \mathbb{P}(V_k = k+n)$ for $n \in \mathbb{N}$.

The distribution of W_k is also referred to as the *negative binomial distribution* with parameters k and p . Thus, the term *negative binomial distribution* can refer either to the distribution of the trial number of the k th success or the distribution of the number of failures before the k th success, depending on the author and the context. The two random variables differ by a constant, so it's not a particularly important issue as long as we know which version is intended. In this text, we will refer to the alternate version as the negative binomial distribution on \mathbb{N} , to distinguish it from the original version, which has support set $\{k, k+1, \dots\}$.

More interestingly, however, the probability density function in the last result makes sense for any $k \in (0, \infty)$, not just integers. To see this, first recall the definition of the general binomial coefficient: if $a \in \mathbb{R}$ and $n \in \mathbb{N}$, we define

$$\binom{a}{n} = \frac{a^{(n)}}{n!} = \frac{a(a-1) \cdots (a-n+1)}{n!} \quad (11.4.24)$$

The function f given below defines a probability density function for every $p \in (0, 1)$ and $k \in (0, \infty)$:

$$f(n) = \binom{n+k-1}{n} p^k (1-p)^n, \quad n \in \mathbb{N} \quad (11.4.25)$$

Proof

Recall from the section on Combinatorial Structures that $\binom{n+k-1}{n} = (-1)^n \binom{-k}{n}$. From the general binomial theorem,

$$\sum_{n=0}^{\infty} f(n) = p^k \sum_{n=0}^{\infty} \binom{-k}{n} (-1)^n (1-p)^n = p^k [1 - (1-p)]^{-k} = 1 \quad (11.4.26)$$

Once again, the distribution defined by the probability density function in the last theorem is the *negative binomial distribution* on \mathbb{N} , with parameters k and p . The special case when k is a positive integer is sometimes referred to as the *Pascal distribution*, in honor of Blaise Pascal.

The distribution is unimodal. Let $t = |k-1| \frac{1-p}{p}$.

1. $f(n-1) < f(n)$ if and only if $n < t$.
2. The distribution has a single mode at $\lfloor t \rfloor$ if t is not an integer.
3. The distribution has two consecutive modes at $t-1$ and t if t is a positive integer.

Basic Properties

Suppose that W has the negative binomial distribution on \mathbb{N} with parameters $k \in (0, \infty)$ and $p \in (0, 1)$. To establish basic properties, we can no longer use the decomposition of W as a sum of independent geometric variables. Instead, the best approach is to derive the probability generating function and then use the generating function to obtain other basic properties.

W has probability generating function P given by

$$P(t) = \mathbb{E}(t^W) = \left(\frac{p}{1 - (1-p)t} \right)^k, \quad |t| < \frac{1}{1-p} \quad (11.4.27)$$

Proof

This follows from the general binomial theorem: for $|t| < 1/(1-p)$,

$$\mathbb{E}(t^W) = \sum_{n=0}^{\infty} f(n)t^n = p^k \sum_{n=0}^{\infty} \binom{-k}{n} (-1)^n (1-p)^n t^n = p^k [1 - (1-p)t]^{-k} \quad (11.4.28)$$

The moments of W can be obtained from the derivatives of the probability generating function.

W has the following moments:

1. $\mathbb{E}(W) = k \frac{1-p}{p}$
2. $\text{var}(W) = k \frac{1-p}{p^2}$
3. $\text{skew}(W) = \frac{2-p}{\sqrt{k(1-p)}}$
4. $\text{kurt}(W) = \frac{3(k+2)(1-p)+p^2}{k(1-p)}$

Proof

Recall that the factorial moments of W can be obtained from the derivatives of the probability generating function: $\mathbb{E}[W^{(k)}] = P^{(k)}(1)$. Then the various moments above can be obtained from standard formulas.

The negative binomial distribution on \mathbb{N} is preserved under sums of independent variables.

Suppose that V has the negative binomial distribution on \mathbb{N} with parameters $a \in (0, \infty)$ and $p \in (0, 1)$, and that W has the negative binomial distribution on \mathbb{N} with parameters $b \in (0, \infty)$ and $p \in (0, 1)$, and that V and W are independent. Then $V + W$ has the negative binomial on \mathbb{N} distribution with parameters $a + b$ and p .

Proof

This result follows from the probability generating functions. Recall that the PGF of $V + W$ is the product of the PGFs of V and W .

In the last result, note that the success parameter p must be the same for both variables.

Normal Approximation

Because of the decomposition of W when the parameter k is a positive integer, it's not surprising that a central limit theorem holds for the general negative binomial distribution.

Suppose that W has the negative binomial distribution with parameters $k \in (0, \infty)$ and $p \in (0, 1)$. The standard score of W is

$$Z = \frac{pW - k(1-p)}{\sqrt{k(1-p)}} \quad (11.4.29)$$

The distribution of Z converges to the standard normal distribution as $k \rightarrow \infty$.

Thus, if k is large (and not necessarily an integer), then the distribution of W is approximately normal with mean $k \frac{1-p}{p}$ and variance $k \frac{1-p}{p^2}$.

Special Families

The negative binomial distribution on \mathbb{N} belongs to several special families of distributions. First, It follows from the result above on [sums](#) that we can decompose a negative binomial variable on \mathbb{N} into the sum of an arbitrary number of independent, identically distributed variables. This special property is known as infinite divisibility, and is studied in more detail in the chapter on Special Distributions.

The negative binomial distribution on \mathbb{N} is infinitely divisible.

Proof

Suppose that V has the negative binomial distribution on \mathbb{N} with parameters $k \in (0, \infty)$ and $p \in (0, 1)$. It follows from the previous result that for any $n \in \mathbb{N}_+$, V can be represented as $V = \sum_{i=1}^n V_i$ where (V_1, V_2, \dots, V_n) are independent, and each has the negative binomial distribution on \mathbb{N} with parameters k/n and p .

A Poisson-distributed random sum of independent, identically distributed random variables is said to have a compound Poisson distributions; these distributions are studied in more detail in the chapter on the Poisson Process. A theorem of William Feller states that an infinite divisible distribution

on \mathbb{N} must be compound Poisson. Hence it follows from the previous result that the negative binomial distribution on \mathbb{N} belongs to this family. Here is the explicit result:

Let $p, k \in (0, \infty)$. Suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of independent variables, each having the logarithmic series distribution with shape parameter $1 - p$. Suppose also that N is independent of \mathbf{X} and has the Poisson distribution with parameter $-k \ln(p)$. Then $W = \sum_{i=1}^N X_i$ has the negative binomial distribution on \mathbb{N} with parameters k and p .

Proof

From the general theory of compound Poisson distributions, the probability generating function of W is $P(t) = \exp(\lambda[Q(t) - 1])$ where λ is the parameter of the Poisson variable N and $Q(t)$ is the common PGF of the terms in the sum. Using the PGF of the logarithmic series distribution, and the particular values of the parameters, we have

$$P(t) = \exp \left[-k \ln(p) \left(\frac{\ln[1 - (1-p)t]}{\ln(p)} - 1 \right) \right], \quad |t| < \frac{1}{1-p} \quad (11.4.30)$$

Using properties of logarithms and simple algebra, this reduces to

$$P(t) = \left(\frac{p}{1 - (1-p)t} \right)^k, \quad |t| < \frac{1}{1-p} \quad (11.4.31)$$

which is the PGF of the negative binomial distribution with parameters k and p .

As a special case ($k = 1$), it follows that the geometric distribution on \mathbb{N} is infinitely divisible and compound Poisson.

Next, the negative binomial distribution on \mathbb{N} belongs to the general exponential family. This family is important in inferential statistics and is studied in more detail in the chapter on Special Distributions.

Suppose that W has the negative binomial distribution on \mathbb{N} with parameters $k \in (0, \infty)$ and $p \in (0, 1)$. For fixed k , W has a one-parameter exponential distribution with natural statistic W and natural parameter $\ln(1 - p)$.

Proof

The PDF of W can be written as

$$f(n) = \binom{n+k-1}{n} p^k \exp[n \ln(1-p)], \quad n \in \mathbb{N} \quad (11.4.32)$$

so the result follows from the definition of the general exponential family.

Finally, the negative binomial distribution on \mathbb{N} is a power series distribution. Many special discrete distribution belong to this family, which is studied in more detail in the chapter on Special Distributions.

For fixed $k \in (0, \infty)$, the negative binomial distribution on \mathbb{N} with parameters k and $p \in (0, 1)$ is a power series distribution corresponding to the function $g(\theta) = 1/(1 - \theta)^k$ for $\theta \in (0, 1)$, where $\theta = 1 - p$.

Proof

In terms of the new parameter θ , the negative binomial pdf has the form $f(n) = \frac{1}{g(\theta)} \binom{n+k-1}{n} \theta^n$ for $n \in \mathbb{N}$, and $\sum_{n=0}^{\infty} \binom{n+k-1}{n} \theta^n = g(\theta)$.

Computational Exercises

Suppose that W has the negative binomial distribution with parameters $k = \frac{15}{2}$ and $p = \frac{3}{4}$. Compute each of the following:

1. $\mathbb{P}(W = 3)$
2. $\mathbb{E}(W)$
3. $\text{var}(W)$

Answer

1. $\mathbb{P}(W = 3) = 0.1823$
2. $\mathbb{E}(W) = \frac{5}{2}$
3. $\text{var}(W) = \frac{10}{3}$

Suppose that W has the negative binomial distribution with parameters $k = \frac{1}{3}$ and $p = \frac{1}{4}$. Compute each of the following:

1. $\mathbb{P}(W \leq 2)$
2. $\mathbb{E}(W)$
3. $\text{var}(W)$

Answer

1. $\mathbb{P}(W \leq 2) = \frac{11}{8\sqrt[3]{4}}$
2. $\mathbb{E}(W) = 1$
3. $\text{var}(W) = 4$

Suppose that W has the negative binomial distribution with parameters $k = 10\pi$ and $p = \frac{1}{3}$. Compute each of the following:

1. $\mathbb{E}(W)$
2. $\text{var}(W)$
3. The normal approximation to $\mathbb{P}(50 \leq W \leq 70)$

Answer

1. $\mathbb{E}(W) = 20\pi$
2. $\text{var}(W) = 60\pi$
3. $\mathbb{P}(50 \leq W \leq 70) \approx 0.5461$

This page titled [11.4: The Negative Binomial 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.