

# Negative binomial distribution

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **negative binomial distribution** is a discrete probability distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (denoted  $r$ ) occurs. For example, if we define a "1" as failure, all non-"1"s as successes, and we throw a die repeatedly until the third time "1" appears ( $r = \text{three failures}$ ), then the probability distribution of the number of non-"1"s that had appeared will be a negative binomial.

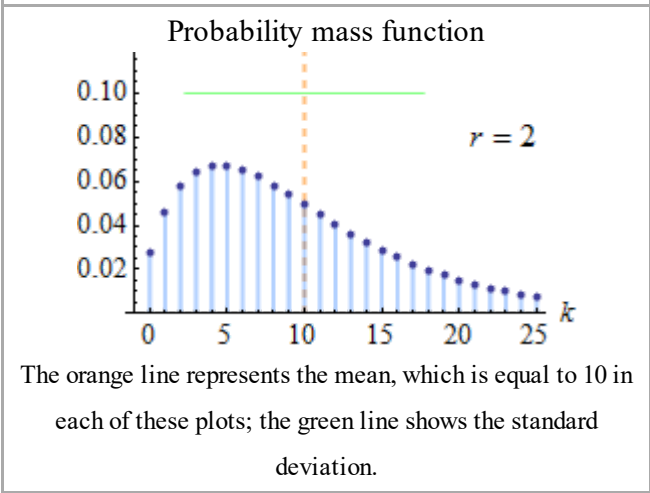
The **Pascal distribution** (after Blaise Pascal) and **Polya distribution** (for George Pólya) are special cases of the negative binomial. There is a convention among engineers, climatologists, and others to reserve "negative binomial" in a strict sense or "Pascal" for the case of an integer-valued stopping-time parameter  $r$ , and use "Polya" for the real-valued case.

For occurrences of "contagious" discrete events, like tornado outbreaks, the Polya distributions can be used to give more accurate models than the Poisson distribution by allowing the mean and variance to be different, unlike the Poisson. "Contagious" events have positively correlated occurrences causing a larger variance than if the occurrences were independent, due to a positive covariance term.

## Contents

- 1 Definition
  - 1.1 Probability mass function
  - 1.2 Expectation
  - 1.3 Variance
- 2 Alternative formulations
- 3 Occurrence
  - 3.1 Waiting time in a Bernoulli process
  - 3.2 Overdispersed Poisson
- 4 Related distributions
  - 4.1 Poisson distribution
  - 4.2 Gamma–Poisson mixture
  - 4.3 Sum of geometric distributions
  - 4.4 Representation as compound Poisson distribution
- 5 Properties
  - 5.1 Cumulative distribution function

Different texts adopt slightly different definitions for the negative binomial distribution. They can be distinguished by whether the support starts at  $k = 0$  or at  $k = r$ , whether  $p$  denotes the probability of a success or of a failure, and whether  $r$  represents success or failure,<sup>[1]</sup> so it is crucial to identify the specific parametrization used in any given text.



Notation	<b>NB(<math>r</math>, <math>p</math>)</b>
Parameters	$r > 0$ — number of failures until the experiment is stopped (integer, but the definition can also be extended to reals) $p \in (0,1)$ — success probability in each experiment (real)
Support	$k \in \{0, 1, 2, 3, \dots\}$ — number of successes

- 5.2 Sampling and point estimation of  $p$
- 5.3 Relation to the binomial theorem
- 6 Parameter estimation
  - 6.1 Maximum likelihood estimation
- 7 Examples
  - 7.1 Selling candy
  - 7.2 Length of hospital stay
- 8 See also
- 9 References

## Definition

Suppose there is a sequence of independent Bernoulli trials. Thus, each trial has two potential outcomes called “success” and “failure”. In each trial the probability of success is  $p$  and of failure is  $(1 - p)$ . We are observing this sequence until a predefined number  $r$  of failures has occurred. Then the random number of successes we have seen,  $X$ , will have the **negative binomial** (or **Pascal**) distribution:

$$X \sim \text{NB}(r; p)$$

When applied to real-world problems, outcomes of *success* and *failure* may or may not be outcomes we ordinarily view as good and bad, respectively. Suppose we used the negative binomial distribution to model the number of days a certain machine works before it breaks down. In this case “success” would be the result on a day when the machine worked properly, whereas a breakdown would be a “failure”. If we used the negative binomial distribution to model the number of goal attempts a sportsman makes before scoring  $r$  goals, though, then each unsuccessful attempt would be a “success”, and scoring a goal would be “failure”. If we are tossing a coin, then the negative binomial distribution can give the number of heads (“success”) we are likely to encounter before we encounter a certain number of tails (“failure”). In the probability mass function below,  $p$  is the probability of success, and  $(1-p)$  is the probability of failure.

## Probability mass function

The probability mass function of the negative binomial distribution is

$$f(k; r, p) \equiv \Pr(X = k) = \binom{k+r-1}{k} p^k (1-p)^r \quad \text{for } k = 0, 1, 2, \dots$$

<b>pmf</b>	$\binom{k+r-1}{k} \cdot (1-p)^r p^k$ , involving a binomial coefficient
<b>CDF</b>	$1 - I_p(k+1, r)$ , the regularized incomplete beta function
<b>Mean</b>	$\frac{pr}{1-p}$
<b>Mode</b>	$\begin{cases} \lfloor \frac{p(r-1)}{1-p} \rfloor & \text{if } r > 1 \\ 0 & \text{if } r \leq 1 \end{cases}$
<b>Variance</b>	$\frac{pr}{(1-p)^2}$
<b>Skewness</b>	$\frac{1+p}{\sqrt{pr}}$
<b>Ex. kurtosis</b>	$\frac{6}{r} + \frac{(1-p)^2}{pr}$
<b>MGF</b>	$\left( \frac{1-p}{1-pe^t} \right)^r$ for $t < -\log p$
<b>CF</b>	$\left( \frac{1-p}{1-pe^{it}} \right)^r$ with $t \in \mathbb{R}$
<b>PGF</b>	$\left( \frac{1-p}{1-pz} \right)^r$ for $ z  < \frac{1}{p}$
<b>Fisher information</b>	$\frac{r}{p^2(1-p)}$

where " $k$ " is the number of successes, " $r$ " is the number of failures, and " $p$ " is the probability of success. Here the quantity in parentheses is the binomial coefficient, and is equal to

$$\binom{k+r-1}{k} = \frac{(k+r-1)!}{k!(r-1)!} = \frac{(k+r-1)(k+r-2)\cdots(r)}{k!}.$$

This quantity can alternatively be written in the following manner, explaining the name “negative binomial”:

$$\frac{(k+r-1)\cdots(r)}{k!} = (-1)^k \frac{(-r)(-r-1)(-r-2)\cdots(-r-k+1)}{k!} = (-1)^k \binom{-r}{k}.$$

Note that by the last expression and the binomial series, for every  $0 \leq p < 1$ ,

$$(1-p)^{-r} = \sum_{k=0}^{\infty} \binom{-r}{k} (-p)^k = \sum_{k=0}^{\infty} \binom{k+r-1}{k} p^k,$$

hence the terms of the probability mass function indeed add up to one.

To understand the above definition of the probability mass function, note that the probability for every specific sequence of  $k$  successes and  $r$  failures is  $(1-p)^r p^k$ , because the outcomes of the  $k+r$  trials are supposed to happen independently. Since the  $r$ th failure comes last, it remains to choose the  $k$  trials with successes out of the remaining  $k+r-1$  trials. The above binomial coefficient, due to its combinatorial interpretation, gives precisely the number of all these sequences of length  $k+r-1$ .

The following recurrence relation holds:

$$\left\{ \begin{array}{l} (k+1) \Pr(k+1) - p \Pr(k)(k+r) = 0, \\ \Pr(0) = (1-p)^r \end{array} \right\}$$

## Expectation

When counting the number of " $k$ " successes given the number of " $r$ " failures, the expected number of successes is  $\frac{rp}{1-p}$ .

When counting the number of " $k+r$ " trials given " $r$ " failures, the expected total number of trials of a negative binomial distribution with parameters  $(r, p)$  is  $\frac{r}{1-p}$ .

To see this intuitively, imagine the above experiment is performed many times. That is, a set of trials is performed until  $r$  failures are obtained, then another set of trials, and then another etc. Write down the number of trials performed in each experiment:  $a, b, c, \dots$  and set  $a + b + c + \dots = N$ . Now we would expect about

$N(1-p)$  failures in total. Say the experiment was performed  $n$  times. Then there are  $nr$  failures in total. So we would expect  $nr = N(1-p)$ , so  $N/n = r/(1-p)$ . See that  $N/n$  is just the average number of trials per experiment. That is what we mean by "expectation". The average number of successes per experiment is  $N/n - r$ , which must have expected value equal to  $r/(1-p) - r = rp/(1-p)$ . This agrees with the mean given in the box on the right-hand side of this page.

## Variance

When counting the number of " $k$ " successes given the number of " $r$ " failures, the variance is  $\frac{rp}{(1-p)^2}$ .

## Alternative formulations

Some sources may define the negative binomial distribution slightly differently from the primary one here. The most common variations are where the random variable " $X$ " is counting different things. These variations can be seen in the table here:

	X is counting...	Probability Mass Function	Formula	Alternate Formula  (using equivalent binomial)	Alternate Formula  (simplified using: $n = k + r$ )	Support
1	" $k$ " successes, given " $r$ " failures	$f(k; r, p) \equiv \Pr(X = k) =$	$\binom{k+r-1}{k} p^k (1 - p)^r$	$\binom{k+r-1}{r-1} p^k (1 - p)^r$	$\binom{n-1}{k} p^k (1 - p)^r$	for $k = 0, 1, 2, \dots$
2	" $n$ " trials, given " $r$ " failures	$f(n; r, p) \equiv \Pr(X = n) =$	$\binom{n-1}{r-1} p^{n-r} (1 - p)^r$	$\binom{n-1}{n-r} p^{n-r} (1 - p)^r$		for $n = r, r + 1, r + 2, \dots$
3	" $r$ " failures, given " $k$ " successes	$f(r; k, p) \equiv \Pr(X = r) =$	$\binom{k+r-1}{r} p^k (1 - p)^r$ [2][3][4]	$\binom{k+r-1}{k-1} p^k (1 - p)^r$ [5][6][7][8]	$\binom{n-1}{r} p^k (1 - p)^r$	for $r = 0, 1, 2, \dots$
4	" $n$ " trials, given " $k$ " successes	$f(n; k, p) \equiv \Pr(X = n) =$	$\binom{n-1}{k-1} p^k (1 - p)^{n-k}$ [3][8][9][10][11]	$\binom{n-1}{n-k} p^k (1 - p)^{n-k}$		for $n = k, k + 1, k + 2, \dots$
	" $k$ " successes, given " $n$ " trials	$f(k; n, p) \equiv \Pr(X = k) =$	This is the binomial distribution: $\binom{n}{k} p^k (1 - p)^r$			for $k = 0, 1, 2, \dots$

Each of these definitions of the negative binomial distribution can be expressed in slightly different but equivalent ways. The first alternative formulation is simply an equivalent form of the binomial coefficient, that is:  $\binom{a}{b} = \binom{a}{a-b}$  for  $0 \leq b \leq a$ . The second alternate formulation somewhat simplifies the expression by recognizing that the total number of trials is simply the number of successes and failures, that is:  $n = k + r$ . These second formulations may be more intuitive to understand, however they are perhaps less practical as they have more terms.

1. This definition is where " $X$ " is the number of " $k$ " **successes** given a set of " $r$ " **failures**, and is the primary way the negative binomial distribution is defined in this article. The second alternative formula clearly shows the relationship of the negative binomial distribution to the binomial distribution. The only difference is that in the binomial coefficient of the negative binomial distribution, there are " $n-1$ " trials to choose from (instead of " $n$ ") when evaluating the number of ways that " $k$ " successes can occur. This is because when you are evaluating the number of ways you can obtain " $k$ " successes before you reach " $r$ " failures, the last trial must be a failure. As such, the other events have one fewer positions available when counting possible orderings.
  2. The second definition is where " $X$ " is the total number of " $n$ " **trials** needed to get " $r$ " **failures**. Since the total number of trials is equal to the number of successes plus the number of failures, the formulation is the same. The only difference in the distribution is the range is shifted by a factor of " $r$ ". As such, the Mean, the Median, and the mode are also shifted by a factor of " $r$ ".
  3. The definition where " $X$ " is the number of " $r$ " **failures** that occur for a given number of " $k$ " **successes**. This definition is very similar to the primary definition used in this article, only that " $k$ " successes and " $r$ " failures are switched when considering what is being counted and what is given. Note however, that " $p$ " still refers to the probability of "success".
  4. The definition where " $X$ " is the number of " $n$ " **trials** that occur for a given number of " $k$ " **successes**. This definition is very similar to definition #2, only that " $k$ " successes is given instead of " $r$ " failures. Note however, that " $p$ " still refers to the probability of "success".
- The definition of the negative binomial distribution can be extended to the case where the parameter " $r$ " can take on a positive real value. Although it is impossible to visualize a non-integer number of "failures", we can still formally define the distribution through its probability mass function. The problem of extending the definition to real-valued (positive)  $r$  boils down to extending the binomial coefficient to its real-valued counterpart, based on the gamma function:

$$\binom{k+r-1}{k} = \frac{(k+r-1)(k+r-2)\cdots(r)}{k!} = \frac{\Gamma(k+r)}{k! \Gamma(r)}$$

Now, after substituting this expression in the original definition, we say that  $X$  has a negative binomial (or **Pólya**) distribution if it has a probability mass function:

$$f(k; r, p) \equiv \Pr(X = k) = \frac{\Gamma(k+r)}{k! \Gamma(r)} p^k (1-p)^r \quad \text{for } k = 0, 1, 2, \dots$$

Here  $r$  is a real, positive number.

■

In negative binomial regression,<sup>[12]</sup> the distribution is specified in terms of its mean,  $m = \frac{pr}{1-p}$ , which is then related to explanatory variables as in linear regression or other Generalized linear models. From the expression for the mean  $m$ , one can derive  $p = \frac{m}{m+r}$  and  $1-p = \frac{r}{m+r}$ . Then, substituting these expressions in the one for the probability mass function when  $r$  is real-valued, yields this parametrization of the probability mass function in terms of  $m$ :

$$\Pr(X = k) = \frac{\Gamma(r+k)}{k! \Gamma(r)} \left( \frac{m}{r+m} \right)^k \left( \frac{r}{r+m} \right)^r \quad \text{for } k = 0, 1, 2, \dots$$

The variance can then be written as  $m + \frac{m^2}{r}$ . Some authors prefer to set  $\alpha = \frac{1}{r}$ , and express the variance as  $m + \alpha m^2$ . In this context, and depending on the author, either the parameter  $r$  or its reciprocal  $\alpha$  is referred to as the “dispersion parameter”, “shape parameter” or “clustering coefficient”,<sup>[13]</sup> or the “heterogeneity”,<sup>[12]</sup> or “aggregation” parameter.<sup>[7]</sup> The term “aggregation” is particularly used in ecology when describing counts of individual organisms. Decrease of the aggregation parameter  $r$  towards zero corresponds to increasing aggregation of the organisms; increase of  $r$  towards infinity corresponds to absence of aggregation, as can be described by Poisson regression.

- Sometimes the distribution is parameterized in terms of its mean  $\mu$  and variance  $\sigma^2$ . In that case,

$$p = \frac{\mu}{\sigma^2},$$

$$r = \frac{\mu^2}{\sigma^2 - \mu}, \text{ and}$$

$$\Pr(X = k) = \binom{k + \frac{\mu^2}{\sigma^2 - \mu} - 1}{k} \left( \frac{\mu}{\sigma^2} \right)^{\left( \frac{\mu^2}{\sigma^2 - \mu} \right)} \left( \frac{\sigma^2 - \mu}{\sigma^2} \right)^k.$$

## Occurrence

### Waiting time in a Bernoulli process

For the special case where  $r$  is an integer, the negative binomial distribution is known as the **Pascal distribution**. It is the probability distribution of a certain number of failures and successes in a series of independent and identically distributed Bernoulli trials. For  $k + r$  Bernoulli trials with success probability  $p$ , the negative binomial gives the probability of  $k$  successes and  $r$  failures, with a failure on the last trial. In other words, the negative binomial distribution is the probability distribution of the number of successes before the  $r$ th failure in a Bernoulli process, with probability  $p$  of successes on each trial. A Bernoulli process is a discrete time process, and so the number of trials, failures, and successes are integers.

Consider the following example. Suppose we repeatedly throw a dice, and consider a “1” to be a “failure”. The probability of success on each trial is 5/6. The number of successes before the third failure belongs to the infinite set  $\{0, 1, 2, 3, \dots\}$ . That number of successes is a negative-binomially distributed random variable.

When  $r = 1$  we get the probability distribution of number of successes before the first failure (i.e. the probability of the first failure occurring on the  $(k + 1)$ st trial), which is a geometric distribution:

$$f(k; r, p) = (1 - p) \cdot p^k$$

### Overdispersed Poisson

The negative binomial distribution, especially in its alternative parameterization described above, can be used as an alternative to the Poisson distribution. It is especially useful for discrete data over an unbounded positive range whose sample variance exceeds the sample mean. In such cases, the observations are overdispersed with respect to a Poisson distribution, for which the mean is equal to the variance. Hence a Poisson distribution is not an appropriate model. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean. See Cumulants of some discrete probability distributions.

An application of this is to annual counts of tropical cyclones in the North Atlantic or to monthly to 6-monthly counts of wintertime extratropical cyclones over Europe, for which the variance is greater than the mean.<sup>[14][15][16]</sup> In the case of modest overdispersion, this may produce substantially similar results to an overdispersed Poisson distribution.<sup>[17][18]</sup>

The negative binomial distribution is also commonly used to model gene expression in the form of discrete read count data from high-throughput RNA sequencing experiments.<sup>[19][20]</sup>

## Related distributions

- The geometric distribution (on  $\{0, 1, 2, 3, \dots\}$ ) is a special case of the negative binomial distribution, with

$$\text{Geom}(p) = \text{NB}(1, 1 - p).$$

- The negative binomial distribution is a special case of the discrete phase-type distribution.
- The negative binomial distribution is a special case of discrete Compound Poisson distribution.

## Poisson distribution

Consider a sequence of negative binomial random variables where the stopping parameter  $r$  goes to infinity, whereas the probability of success in each trial,  $p$ , goes to zero in such a way as to keep the mean of the distribution constant. Denoting this mean  $\lambda$ , the parameter  $p$  will be  $p = \lambda / (r + \lambda)$

$$\lambda = r \frac{p}{1 - p} \quad \Rightarrow \quad p = \frac{\lambda}{r + \lambda}.$$

Under this parametrization the probability mass function will be

$$f(k; r, p) = \frac{\Gamma(k + r)}{k! \cdot \Gamma(r)} p^k (1 - p)^r = \frac{\lambda^k}{k!} \cdot \frac{\Gamma(r + k)}{\Gamma(r) (r + \lambda)^k} \cdot \frac{1}{\left(1 + \frac{\lambda}{r}\right)^r}$$

Now if we consider the limit as  $r \rightarrow \infty$ , the second factor will converge to one, and the third to the exponent function:

$$\lim_{r \rightarrow \infty} f(k; r, p) = \frac{\lambda^k}{k!} \cdot 1 \cdot \frac{1}{e^\lambda},$$

which is the mass function of a Poisson-distributed random variable with expected value  $\lambda$ .

In other words, the alternatively parameterized negative binomial distribution converges to the Poisson distribution and  $r$  controls the deviation from the Poisson. This makes the negative binomial distribution suitable as a robust alternative to the Poisson, which approaches the Poisson for large  $r$ , but which has larger variance than the Poisson for small  $r$ .

$$\text{Poisson}(\lambda) = \lim_{r \rightarrow \infty} \text{NB}\left(r, \frac{\lambda}{\lambda + r}\right).$$

## Gamma–Poisson mixture

The negative binomial distribution also arises as a continuous mixture of Poisson distributions (i.e. a compound probability distribution) where the mixing distribution of the Poisson rate is a gamma distribution. That is, we can view the negative binomial as a  $\text{Poisson}(\lambda)$  distribution, where  $\lambda$  is itself a random variable, distributed as a gamma distribution with shape  $= r$  and scale  $\theta = p/(1 - p)$  or correspondingly rate  $\beta = (1 - p)/p$ .

Formally, this means that the mass function of the negative binomial distribution can be written as

$$\begin{aligned} f(k; r, p) &= \int_0^\infty f_{\text{Poisson}(\lambda)}(k) \cdot f_{\text{Gamma}\left(r, \frac{1-p}{p}\right)}(\lambda) \, d\lambda \\ &= \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} \cdot \lambda^{r-1} \frac{e^{-\lambda(1-p)/p}}{\left(\frac{p}{1-p}\right)^r \Gamma(r)} \, d\lambda \\ &= \frac{(1-p)^r p^{-r}}{k! \Gamma(r)} \int_0^\infty \lambda^{r+k-1} e^{-\lambda/p} \, d\lambda \\ &= \frac{(1-p)^r p^{-r}}{k! \Gamma(r)} p^{r+k} \Gamma(r+k) \\ &= \frac{\Gamma(r+k)}{k! \Gamma(r)} p^k (1-p)^r. \end{aligned}$$

Because of this, the negative binomial distribution is also known as the **gamma–Poisson (mixture) distribution**.



Note: The negative binomial distribution was originally derived as a limiting case of the gamma-Poisson distribution.<sup>[21]</sup>

## Sum of geometric distributions

If  $Y_r$  is a random variable following the negative binomial distribution with parameters  $r$  and  $p$ , and support  $\{0, 1, 2, \dots\}$ , then  $Y_r$  is a sum of  $r$  independent variables following the geometric distribution (on  $\{0, 1, 2, \dots\}$ ) with parameter  $1-p$ . As a result of the central limit theorem,  $Y_r$  (properly scaled and shifted) is therefore approximately normal for sufficiently large  $r$ .

Furthermore, if  $B_{s+r}$  is a random variable following the binomial distribution with parameters  $s + r$  and  $1 - p$ , then

$$\begin{aligned}\Pr(Y_r \leq s) &= 1 - I_p(s + 1, r) \\ &= 1 - I_p((s + r) - (r - 1), (r - 1) + 1) \\ &= 1 - \Pr(B_{s+r} \leq r - 1) \\ &= \Pr(B_{s+r} \geq r) \\ &= \Pr(\text{after } s + r \text{ trials, there are at least } r \text{ successes}).\end{aligned}$$

In this sense, the negative binomial distribution is the "inverse" of the binomial distribution.

The sum of independent negative-binomially distributed random variables  $r_1$  and  $r_2$  with the same value for parameter  $p$  is negative-binomially distributed with the same  $p$  but with "r-value"  $r_1 + r_2$ .

The negative binomial distribution is infinitely divisible, i.e., if  $Y$  has a negative binomial distribution, then for any positive integer  $n$ , there exist independent identically distributed random variables  $Y_1, \dots, Y_n$  whose sum has the same distribution that  $Y$  has.

## Representation as compound Poisson distribution

The negative binomial distribution  $NB(r, p)$  can be represented as a compound Poisson distribution: Let  $\{Y_n, n \in \mathbb{N}_0\}$  denote a sequence of independent and identically distributed random variables, each one having the logarithmic distribution  $\text{Log}(p)$ , with probability mass function

$$f(k; r, p) = \frac{-p^k}{k \ln(1 - p)}, \quad k \in \mathbb{N}.$$

Let  $N$  be a random variable, independent of the sequence, and suppose that  $N$  has a Poisson distribution with mean  $\lambda = -r \ln(1 - p)$ . Then the random sum

$$X = \sum_{n=1}^N Y_n$$

is  $\text{NB}(r,p)$ -distributed. To prove this, we calculate the probability generating function  $G_X$  of  $X$ , which is the composition of the probability generating functions  $G_N$  and  $G_{Y_1}$ . Using

$$G_N(z) = \exp(\lambda(z - 1)), \quad z \in \mathbb{R},$$

and

$$G_{Y_1}(z) = \frac{\ln(1 - pz)}{\ln(1 - p)}, \quad |z| < \frac{1}{p},$$

we obtain

$$\begin{aligned} G_X(z) &= G_N(G_{Y_1}(z)) \\ &= \exp\left(\lambda\left(\frac{\ln(1 - pz)}{\ln(1 - p)} - 1\right)\right) \\ &= \exp(-r(\ln(1 - pz) - \ln(1 - p))) \\ &= \left(\frac{1 - p}{1 - pz}\right)^r, \quad |z| < \frac{1}{p}, \end{aligned}$$

which is the probability generating function of the  $\text{NB}(r,p)$  distribution.

The following table describes four distributions related to the number of successes in a sequence of draws:

	With replacements	No replacements
Given number of draws	binomial distribution	hypergeometric distribution
Given number of failures	negative binomial distribution	negative hypergeometric distribution

## Properties

### Cumulative distribution function

The cumulative distribution function can be expressed in terms of the regularized incomplete beta function:

$$f(k; r, p) \equiv \Pr(X \leq k) = 1 - I_p(k + 1, r) = I_{1-p}(r, k + 1).$$

## Sampling and point estimation of $p$

Suppose  $p$  is unknown and an experiment is conducted where it is decided ahead of time that sampling will continue until  $r$  successes are found. A sufficient statistic for the experiment is  $k$ , the number of failures.

In estimating  $p$ , the minimum variance unbiased estimator is

$$\hat{p} = \frac{r - 1}{r + k - 1}.$$

The maximum likelihood estimate of  $p$  is

$$\tilde{p} = \frac{r}{r + k},$$

but this is a biased estimate. Its inverse  $(r + k)/r$ , is an unbiased estimate of  $1/p$ , however.<sup>[22]</sup>

## Relation to the binomial theorem

Suppose  $Y$  is a random variable with a binomial distribution with parameters  $n$  and  $p$ . Assume  $p + q = 1$ , with  $p, q \geq 0$ . Then the binomial theorem implies that

$$1 = 1^n = (p + q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}.$$

Using Newton's binomial theorem, this can equally be written as:

$$(p + q)^n = \sum_{k=0}^{\infty} \binom{n}{k} p^k q^{n-k},$$

in which the upper bound of summation is infinite. In this case, the binomial coefficient

$$\binom{n}{k} = \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!}.$$

is defined when  $n$  is a real number, instead of just a positive integer. But in our case of the binomial distribution it is zero when  $k > n$ . We can then say, for example

$$(p + q)^{8.3} = \sum_{k=0}^{\infty} \binom{8.3}{k} p^k q^{8.3-k}.$$

Now suppose  $r > 0$  and we use a negative exponent:

$$1 = p^r \cdot p^{-r} = p^r (1 - q)^{-r} = p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k.$$

Then all of the terms are positive, and the term

$$p^r \binom{-r}{k} (-q)^k$$

is just the probability that the number of failures before the  $r$ th success is equal to  $k$ , provided  $r$  is an integer. (If  $r$  is a negative non-integer, so that the exponent is a positive non-integer, then some of the terms in the sum above are negative, so we do not have a probability distribution on the set of all nonnegative integers.)

Now we also allow non-integer values of  $r$ . Then we have a proper negative binomial distribution, which is a generalization of the Pascal distribution, which coincides with the Pascal distribution when  $r$  happens to be a positive integer.

Recall from above that

The sum of independent negative-binomially distributed random variables  $r_1$  and  $r_2$  with the same value for parameter  $p$  is negative-binomially distributed with the same  $p$  but with "r-value"  $r_1 + r_2$ .

This property persists when the definition is thus generalized, and affords a quick way to see that the negative binomial distribution is infinitely divisible.

## Parameter estimation

### Maximum likelihood estimation

The maximum likelihood estimator only exists for samples for which the sample variance is larger than the sample mean.<sup>[23]</sup> The likelihood function for  $N$  iid observations  $(k_1, \dots, k_N)$  is

$$L(r, p) = \prod_{i=1}^N f(k_i; r, p)$$

from which we calculate the log-likelihood function

$$\ell(r, p) = \sum_{i=1}^N \ln(\Gamma(k_i + r)) - \sum_{i=1}^N \ln(k_i!) - N \ln(\Gamma(r)) + \sum_{i=1}^N k_i \ln(p) + Nr \ln(1 - p).$$

To find the maximum we take the partial derivatives with respect to  $r$  and  $p$  and set them equal to zero:

$$\frac{\partial \ell(r, p)}{\partial p} = \left[ \sum_{i=1}^N k_i \frac{1}{p} \right] - Nr \frac{1}{1 - p} = 0 \text{ and}$$

$$\frac{\partial \ell(r, p)}{\partial r} = \left[ \sum_{i=1}^N \psi(k_i + r) \right] - N\psi(r) + N \ln(1 - p) = 0$$

where

$$\psi(k) = \frac{\Gamma'(k)}{\Gamma(k)} \text{ is the digamma function.}$$

Solving the first equation for  $p$  gives:

$$p = \frac{\sum_{i=1}^N k_i}{Nr + \sum_{i=1}^N k_i}$$

Substituting this in the second equation gives:

$$\frac{\partial \ell(r, p)}{\partial r} = \left[ \sum_{i=1}^N \psi(k_i + r) \right] - N\psi(r) + N \ln \left( \frac{r}{r + \sum_{i=1}^N k_i / N} \right) = 0$$

This equation cannot be solved for  $r$  in closed form. If a numerical solution is desired, an iterative technique such as Newton's method can be used. Alternatively, the expectation–maximization algorithm can be used.<sup>[23]</sup>

# Examples

## Selling candy

Pat is required to sell candy bars to raise money for the 6th grade field trip. There are thirty houses in the neighborhood, and Pat is not supposed to return home until five candy bars have been sold. So the child goes door to door, selling candy bars. At each house, there is a 0.4 probability of selling one candy bar and a 0.6 probability of selling nothing.

*What's the probability of selling the last candy bar at the  $n$ th house?*

Recall that the  $\text{NegBin}(r, p)$  distribution describes the probability of  $k$  failures and  $r$  successes in  $k + r$  Bernoulli( $p$ ) trials with success on the last trial. Selling five candy bars means getting five successes. The number of trials (i.e. houses) this takes is therefore  $k + 5 = n$ . The random variable we are interested in is the number of houses, so we substitute  $k = n - 5$  into a  $\text{NegBin}(5, 0.4)$  mass function and obtain the following mass function of the distribution of houses (for  $n \geq 5$ ):

$$f(n) = \binom{(n-5) + 5 - 1}{n-5} 0.4^5 0.6^{n-5} = \binom{n-1}{n-5} 2^5 \frac{3^{n-5}}{5^n}.$$

*What's the probability that Pat finishes on the tenth house?*

$$f(10) = 0.1003290624.$$

*What's the probability that Pat finishes on or before reaching the eighth house?*

To finish on or before the eighth house, Pat must finish at the fifth, sixth, seventh, or eighth house. Sum those probabilities:

$$f(5) = 0.01024$$

$$f(6) = 0.03072$$

$$f(7) = 0.055296$$

$$f(8) = 0.0774144$$

$$\sum_{j=5}^8 f(j) = 0.17367.$$

*What's the probability that Pat exhausts all 30 houses in the neighborhood?*

This can be expressed as the probability that Pat does not finish on the fifth through the thirtieth house:

$$1 - \sum_{j=5}^{30} f(j) = 1 - I_{0.4}(5, 30 - 5 + 1) \approx 1 - 0.99849 = 0.00151.$$


## Length of hospital stay


Hospital length of stay is an example of real world data that can be modelled well with a negative binomial distribution.<sup>[24]</sup>

## See also

- Coupon collector's problem
- Beta negative binomial distribution
- Extended negative binomial distribution
- Negative multinomial distribution
- Binomial distribution
- Poisson distribution
- Exponential family
- Vector generalized linear model
- Compound Poisson distribution

## References

1. DeGroot, Morris H. (1986). *Probability and Statistics* (Second ed.). Addison-Wesley. pp. 258–259. ISBN 0-201-11366-X. LCCN 84006269. OCLC 10605205.
2. "Mathworks: Negative Binomial Distribution".
3. Cook, John D. "Notes on the Negative Binomial Distribution" (PDF).
4. Saha, Abhishek. "Introduction to Probability / Fundamentals of Probability: Lecture 14" (PDF).
5. <http://mathworld.wolfram.com/NegativeBinomialDistribution.html>
6. SAS Institute, "Negative Binomial Distribution (<http://support.sas.com/documentation/cdl/en/lefunctionsref/67960/HTML/default/viewer.htm#n0n7cce4a3gfknlvr0p1x0of99s.htm#n1olto4j49wrc3n11tnmuq2zibj6>)", *SAS(R) 9.4 Functions and CALL Routines: Reference, Fourth Edition*, SAS Institute, Cary, NC, 2016.
7. Crawley, Michael J. (2012). *The R Book*. Wiley. ISBN 978-1-118-44896-0.
8. "Set theory: Section 3.2.5 - Negative Binomial Distribution" (PDF).
9. "Random: The negative binomial distribution".
10. "Stat Trek: Negative Binomial Distribution".
11. Wroughton, Jacqueline. "Distinguishing Between Binomial, Hypergeometric and Negative Binomial Distributions" (PDF).
12. Hilbe, Joseph M. (2011). *Negative Binomial Regression* (Second ed.). Cambridge, UK: Cambridge University Press. ISBN 978-0-521-19815-8.
13. Lloyd-Smith, J. O. (2007). "Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly Overdispersed Data, with Applications to Infectious Diseases". *PLoS ONE*. **2** (2): e180. doi:10.1371/journal.pone.0000180. 
14. Villarini, G.; Vecchi, G.A.; Smith, J.A. (2010). "Modeling of the dependence of tropical storm counts in the North Atlantic Basin on climate indices". *Monthly Weather Review*. **138** (7): 2681–2705. doi:10.1175/2010MWR3315.1.
15. Mailier, P.J.; Stephenson, D.B.; Ferro, C.A.T.; Hodges, K.I. (2006). "Serial Clustering of Extratropical Cyclones". *Monthly Weather Review*. **134** (8): 2224–2240. doi:10.1175/MWR3160.1.

16. Vitolo, R.; Stephenson, D.B.; Cook, Ian M.; Mitchell-Wallace, K. (2009). "Serial clustering of intense European storms". *Meteorologische Zeitschrift*. **18** (4): 411–424. doi:10.1127/0941-2948/2009/0393.
17. McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models* (Second ed.). Boca Raton: Chapman and Hall/CRC. ISBN 0-412-31760-5.
18. Cameron, Adrian C.; Trivedi, Pravin K. (1998). *Regression analysis of count data*. Cambridge University Press. ISBN 0-521-63567-5.
19. Love, Michael; Anders, Simon (October 14, 2014). "Differential analysis of count data – the DESeq2 package" (PDF). Retrieved October 14, 2014.
20. Chen, Yunshun; Davis, McCarthy (September 25, 2014). "edgeR: differential expression analysis of digital gene expression data" (PDF). Retrieved October 14, 2014.
21. Greenwood, M.; Yule, G. U. (1920). "An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference of multiple attacks of disease or of repeated accidents". *J R Statist Soc.* **83** (2): 255–279. doi:10.2307/2341080.
22. Haldane, J. B. S. (1945). "On a Method of Estimating Frequencies". *Biometrika*. **33** (3): 222–225. JSTOR 2332299.
23. Aramidis, K. (1999). "An EM algorithm for estimating negative binomial parameters". *Australian & New Zealand Journal of Statistics*. **41** (2): 213–221. doi:10.1111/1467-842X.00075.
24. Carter, E.M., Potts, H.W.W. (4 April 2014). "Predicting length of stay from an electronic patient record system: a primary total knee replacement example". *BMC Medical Informatics & Decision Making*. **14**: 26. doi:10.1186/1472-6947-14-26. Retrieved 23 April 2014. 

Retrieved from "https://en.wikipedia.org/w/index.php?title=Negative\_binomial\_distribution&oldid=744269941"

Categories: Discrete distributions | Factorial and binomial topics | Exponential family distributions | Infinitely divisible probability distributions | Probability distributions

- 
- This page was last modified on 14 October 2016, at 05:26.
  - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.