

Negative binomial distribution

In probability theory and statistics, the **negative binomial distribution** is a discrete probability distribution that models the number of failures in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of successes (denoted *r*) occurs.^[2] For example, we can define rolling a 6 on a die as a success, and rolling any other number as a failure, and ask how many failure rolls will occur before we see the third success (*r* = 3). In such a case, the probability distribution of the number of failures that appear will be a negative binomial distribution.

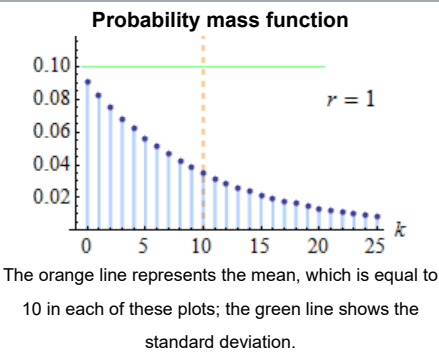
An alternative formulation is to model the number of total trials (instead of the number of failures). In fact, for a specified (non-random) number of successes (*r*), the number of failures (*n* - *r*) are random because the total trials (*n*) are random. For example, we could use the negative binomial distribution to model the number of days *n* (random) a certain machine works (specified by *r*) before it breaks down.

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

For occurrences of associated 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. The negative binomial distribution has a variance $\mu/(1 - p)$, with the distribution becoming identical to Poisson in the limit $p \rightarrow 0$ for a given mean μ . This can make the distribution a useful overdispersed alternative to the Poisson distribution, for example for a robust modification of Poisson regression. In epidemiology, it has been used to model disease transmission for infectious diseases where the likely number of onward infections may vary considerably from individual to individual and from setting to setting.^[3] More generally, it may be appropriate where events have positively correlated occurrences causing a larger variance than if the occurrences were independent, due to a positive covariance term.

The term "negative binomial" is likely due to the fact that a certain binomial coefficient that appears in the formula for the probability mass function of the distribution can be written more simply with negative numbers.^[4]

Different texts (and even different parts of this article) 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,^[1] so identifying the specific parametrization used is crucial in any given text.



Notation	NB(<i>r</i>, <i>p</i>)
Parameters	<i>r</i> > 0 — number of failures until the experiment is stopped (integer, but the definition can also be extended to reals) <i>p</i> ∈ [0,1] — success probability in each experiment (real)
Support	<i>k</i> ∈ {0, 1, 2, 3, ...} — number of successes
PMF	$k \mapsto \binom{k+r-1}{k} \cdot (1-p)^r p^k$, involving a binomial coefficient
CDF	$k \mapsto 1 - I_p(k+1, r)$, the regularized incomplete beta function
Mean	$\frac{pr}{1-p}$
Mode	$\begin{cases} \lfloor \frac{p(r-1)}{1-p} \rfloor & \text{if } r > 1 \\ 0 & \text{if } r \leq 1 \end{cases}$
Variance	$\frac{pr}{(1-p)^2}$
Skewness	$\frac{1+p}{\sqrt{pr}}$
Ex. kurtosis	$\frac{6}{r} + \frac{(1-p)^2}{pr}$
MGF	$\left(\frac{1-p}{1-pe^t}\right)^r$ for <i>t</i> < −log <i>p</i>
CF	$\left(\frac{1-p}{1-pe^{it}}\right)^r$ with <i>t</i> ∈ ℝ
PGF	$\left(\frac{1-p}{1-pz}\right)^r$ for <i>z</i> < $\frac{1}{p}$
Fisher information	$\frac{r}{(1-p)^2p}$
Method of Moments	$r = \frac{E[X]^2}{V[X] - E[X]}$

Contents

Definitions

- Probability mass function
- Cumulative distribution function
- Alternative formulations
- Alternative parameterizations
- Examples
 - Length of hospital stay
 - Selling candy

Properties

- Expectation
- Variance
- Relation to the binomial theorem
- Recurrence relation

Related distributions

- Poisson distribution
- Gamma–Poisson mixture
- Distribution of a sum of geometrically distributed random variables
- Representation as compound Poisson distribution
- (*a*,*b*,0) class of distributions

Statistical inference

- Parameter estimation

MVUE for p
Maximum likelihood estimation

Occurrence and applications

Waiting time in a Bernoulli process
Overdispersed Poisson

History

See also

References

$$p = 1 - \frac{E[X]}{V[X]}$$

Definitions

Imagine a sequence of independent Bernoulli trials: 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 observe this sequence until a predefined number r of successes occurs. Then the random number of observed failures, X , follows the **negative binomial** (or **Pascal**) distribution:

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

Probability mass function

The probability mass function of the negative binomial distribution is

$$f(k; r, p) \equiv \Pr(X = k) = \binom{k+r-1}{r-1} (1-p)^k p^r$$

where r is the number of successes, k is the number of failures, and p is the probability of success. Note that this formulation is an alternative formulation to the sidebar; in this formulation, the mean is $(1-p)r/p$ and the variance is $(1-p)r/p^2$.

Here, the quantity in parentheses is the binomial coefficient, and is equal to

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

There are k failures chosen from $k+r-1$ samples rather than $k+r$ because the last of the $k+r$ samples is by definition a success.

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

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

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

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

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

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

To understand the above definition of the probability mass function, note that the probability for every specific sequence of r successes and k failures is $p^r(1-p)^k$, because the outcomes of the $k+r$ trials are supposed to happen independently. Since the r th success always comes last, it remains to choose the k trials with failures 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 $k+r-1$.

Cumulative distribution function

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

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

It can also be expressed in terms of the cumulative distribution function of the binomial distribution:^[5]

$$F(k; r, p) = F_{\text{binomial}}(k; n = k+r, p).$$

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 failures, given r successes	$f(k; r, p) \equiv \Pr(X = k) =$	$\binom{k+r-1}{k} p^r (1-p)^k$ <small>[6][7][8]</small>	$\binom{k+r-1}{r-1} p^r (1-p)^k$ <small>[9][10][11][12]</small>	$\binom{n-1}{k} p^r (1-p)^k$	for $k = 0, 1, 2, \dots$
2	n trials, given r successes	$f(n; r, p) \equiv \Pr(X = n) =$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ <small>[7][12][13][14][15]</small>	$\binom{n-1}{n-r} p^r (1-p)^{n-r}$		for $n = r, r+1, r+2, \dots$
3	n trials, given r failures	$f(n; r, p) \equiv \Pr(X = n) =$	$\binom{n-1}{r-1} p^{n-r} (1-p)^r$	$\binom{n-1}{n-r} p^{n-r} (1-p)^r$	$\binom{n-1}{k} p^k (1-p)^r$	
4	r successes, given n trials	$f(r; n, p) \equiv \Pr(X = r) =$	This is the <u>binomial distribution</u> : $\binom{n}{r} p^r (1-p)^{n-r}$			for $r = 0, 1, 2, \dots, n$

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: $n = r + k$. These second formulations may be more intuitive to understand, however they are perhaps less practical as they have more terms.

1. The definition where X is the number of k **failures** that occur for a given number of r **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".
 2. The definition where X is the number of n **trials** that occur for a given number of r **successes**. This definition is very similar to definition #2, only that r successes is given instead of k 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:

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

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:

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

Here r is a real, positive number.

In negative binomial regression,^[16] 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 :

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

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 α is referred to as the "dispersion parameter", "shape parameter" or "clustering coefficient",^[17] or the "heterogeneity"^[16] or "aggregation" parameter.^[11] 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.

Alternative parameterizations

Sometimes the distribution is parameterized in terms of its mean μ and variance σ^2 :

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

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

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

Examples

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 via [Negative binomial regression](#).^{[18][19]}

Selling candy

Pat Collis is required to sell candy bars to raise money for the 6th grade field trip. Pat is (somewhat harshly) 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 0.6 probability of selling one candy bar and a 0.4 probability of selling nothing.

What's the probability of selling the last candy bar at the nth house?

Successfully selling candy enough times is what defines our stopping criterion (as opposed to failing to sell it), so k in this case represents the number of failures and r represents the number of successes. 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 $n \geq 5$):

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

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:

$$f(5) = 0.07776$$

$$f(6) = 0.15552$$

$$f(7) = 0.18662$$

$$f(8) = 0.17418$$

$$\sum_{j=5}^8 f(j) = 0.59408.$$

What's the probability that Pat exhausts all 30 houses that happen to stand in the neighborhood?

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

$$1 - \sum_{j=5}^{30} f(j) = 1 - I_{0.4}(5, 30 - 5 + 1) \approx 1 - 0.99999342 = 0.00000658.$$

Because of the rather high probability that Pat will sell to each house (60 percent), the probability of her NOT fulfilling her quest is vanishingly slim.

Properties

Expectation

The expected total number of successes in a negative binomial distribution with parameters (r, p) is $rp/(1-p)$. To see this, imagine an experiment simulating the negative binomial 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 Np successes 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 $N/n - r = r/(1 - p) - r = rp/(1 - p)$. This agrees with the mean given in the box on the right-hand side of this page.

Variance

When counting the number of successes given the number r of failures, the variance is $rp/(1 - p)^2$. When counting the number of failures before the r -th success, the variance is $r(1 - p)/p^2$.

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

$$1 = 1^n = (p + q)^n.$$

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.

Recurrence relation

The following recurrence relation holds:

$$\begin{cases} (k+1) \Pr(k+1) - p \Pr(k)(k+r) = 0, \\ \Pr(0) = (1-p)^r \end{cases}$$

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 as λ , the parameter p will be $p = \lambda/(r + \lambda)$

$$\text{Mean: } \lambda = \frac{pr}{1-p} \Rightarrow p = \frac{\lambda}{r + \lambda},$$

$$\text{Variance: } \lambda \left(1 + \frac{\lambda}{r}\right) > \lambda, \text{ thus always overdispersed.}$$

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 λ .

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 .

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

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 λ is itself a random variable, distributed as a gamma distribution with shape = r and scale $\theta = p/(1-p)$ or correspondingly rate $\beta = (1-p)/p$.

To display the intuition behind this statement, consider two independent Poisson processes, "Success" and "Failure", with intensities p and $1-p$. Together, the Success and Failure processes are equivalent to a single Poisson process of intensity 1, where an occurrence of the process is a success if a corresponding independent coin toss comes up heads with probability p ; otherwise, it is a failure. If r is a counting number, the coin tosses show that the count of successes before the r th failure follows a negative binomial distribution with parameters r and p . The count is also, however, the count of the Success Poisson process at the random time T of the r th occurrence in the Failure Poisson process. The Success count follows a Poisson distribution with mean pT , where T is the waiting time for r occurrences in a Poisson process of intensity $1-p$, i.e., T is gamma-distributed with shape parameter r and intensity $1-p$. Thus, the negative binomial distribution is equivalent to a Poisson distribution with mean pT , where the random variate T is gamma-distributed with shape parameter r and intensity $(1-p)$. The preceding paragraph follows, because $\lambda = pT$ is gamma-distributed with shape parameter r and intensity $(1-p)/p$.

The following formal derivation (which does not depend on r being a counting number) confirms the intuition.

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

Because of this, the negative binomial distribution is also known as the **gamma–Poisson (mixture) distribution**. The negative binomial distribution was originally derived as a limiting case of the gamma-Poisson distribution.^[20]

Distribution of a sum of geometrically distributed random variables

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 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 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 $\text{NB}(r, p)$ 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

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

which is the probability generating function of the $NB(r,p)$ 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

(a,b,0) class of distributions

The negative binomial, along with the Poisson and binomial distributions, is a member of the $(a,b,0)$ class of distributions. All three of these distributions are special cases of the [Panjer distribution](#). They are also members of the [natural exponential family](#).

Statistical inference

Parameter estimation

MVUE for 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

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

Maximum likelihood estimation

When r is known, 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.^[21]

When r is unknown, the maximum likelihood estimator for p and r together only exists for samples for which the sample variance is larger than the sample mean.^[22] 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

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

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

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

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

where

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

Solving the first equation for p gives:

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

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.^[22]

Occurrence and applications

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 die, 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.^{[23][24][25]} In the case of modest overdispersion, this may produce substantially similar results to an overdispersed Poisson distribution.^{[26][27]}

The negative binomial distribution is also commonly used to model data in the form of discrete sequence read counts from high-throughput RNA and DNA sequencing experiments.^{[28][29][30]}

History



This distribution was first studied in 1713, by Montmort, as the distribution of the number of trials required in an experiment to obtain a given number of successes.^[31] It had previously been mentioned by Pascal.^[32]

See also

- Coupon collector's problem
- Beta negative binomial distribution
- Extended negative binomial distribution
- Negative multinomial distribution
- Binomial distribution
- Poisson distribution

- [Compound Poisson distribution](#)
- [Exponential family](#)
- [Negative binomial regression](#)
- [Vector generalized linear model](#)

References

1. DeGroot, Morris H. (1986). *Probability and Statistics* (Second ed.). Addison-Wesley. pp. 258–259. ISBN 0-201-11366-X. LCCN 84006269 (<https://lccn.loc.gov/84006269>). OCLC 10605205 (<https://www.worldcat.org/oclc/10605205>).
2. Weisstein, Eric. "Negative Binomial Distribution" (<https://mathworld.wolfram.com/NegativeBinomialDistribution.html>). *Wolfram MathWorld*. Wolfram Research. Retrieved 11 October 2020.
3. e.g. J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, and W.M. Getz (2005), Superspreading and the effect of individual variation on disease emergence (<https://www.nature.com/articles/nature04153>), *Nature*, **438**, 355–359. doi:10.1038/nature04153 (<https://doi.org/10.1038/nature04153>)
The overdispersion parameter is usually denoted by the letter k in epidemiology, rather than r as here.
4. Casella, George; Berger, Roger L. (2002). *Statistical inference* (https://archive.org/details/statisticalinfer00case_045) (2nd ed.). Thomson Learning. p. 95 (https://archive.org/details/statisticalinfer00case_045/page/n59). ISBN 0-534-24312-6.
5. Morris K W (1963), A note on direct and inverse sampling, *Biometrika*, 50, 544–545.
6. "Mathworks: Negative Binomial Distribution" (<http://www.mathworks.com/help/stats/negative-binomial-distribution.html>).
7. Cook, John D. "Notes on the Negative Binomial Distribution" (http://www.johndcook.com/negative_binomial.pdf) (PDF).
8. Saha, Abhishek. "Introduction to Probability / Fundamentals of Probability: Lecture 14" (<http://www.stat.ufl.edu/~abhisheksaha/sta4321/lect14.pdf>) (PDF).
9. W., Weisstein, Eric. "Negative Binomial Distribution" (<http://mathworld.wolfram.com/NegativeBinomialDistribution.html>). *mathworld.wolfram.com*.
10. SAS Institute, "Negative Binomial Distribution" (<http://support.sas.com/documentation/cdl/en/lefuctionsref/67960/HTML/default/viewer.htm#n0n7cce4a3gfgkn1vr0p1x0of99s.htm#n1olto4j49wrc3n11tnmuq2zibj6>), *SAS(R) 9.4 Functions and CALL Routines: Reference, Fourth Edition*, SAS Institute, Cary, NC, 2016.
11. Crawley, Michael J. (2012). *The R Book* (<https://books.google.com/books?id=XYDI0mIH-moC>). Wiley. ISBN 978-1-118-44896-0.
12. "Set theory: Section 3.2.5 – Negative Binomial Distribution" (<http://www.math.ntu.edu.tw/~hchen/teaching/StatInference/notes/lecture16.pdf>) (PDF).
13. "Randomservices.org, Chapter 10: Bernoulli Trials, Section 4: The Negative Binomial Distribution" (<http://www.randomservices.org/random/bernoulli/NegativeBinomial.html>).
14. "Stat Trek: Negative Binomial Distribution" (<http://stattrek.com/probability-distributions/negative-binomial.aspx>).
15. Wroughton, Jacqueline. "Distinguishing Between Binomial, Hypergeometric and Negative Binomial Distributions" (<http://www.stat.purdue.edu/~zhanghao/STAT511/handout/Stt511%20Sec3.5.pdf>) (PDF).
16. Hilbe, Joseph M. (2011). *Negative Binomial Regression* (https://books.google.com/books?id=0Q_ijxOEBjMC) (Second ed.). Cambridge, UK: Cambridge University Press. ISBN 978-0-521-19815-8.
17. Lloyd-Smith, J. O. (2007). "Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly Overdispersed Data, with Applications to Infectious Diseases" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1791715>). *PLoS ONE*. **2** (2): e180. Bibcode:2007PLoSO...2..180L (<https://ui.adsabs.harvard.edu/abs/2007PLoSO...2..180L>). doi:10.1371/journal.pone.0000180 (<https://doi.org/10.1371/journal.pone.0000180>). PMC 1791715 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1791715>). PMID 17299582 (<https://pubmed.ncbi.nlm.nih.gov/17299582>). 
18. 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" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3992140>). *BMC Medical Informatics and Decision Making*. **14**: 26. doi:10.1186/1472-6947-14-26 (<https://doi.org/10.1186/1472-6947-14-26>). PMC 3992140 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3992140>). PMID 24708853 (<https://pubmed.ncbi.nlm.nih.gov/24708853>). 
19. Orooji, Arezoo; Nazar, Eisa; Sadeghi, Masoumeh; Moradi, Ali; Jafari, Zahra; Esmaily, Habibollah (2021-04-30). "Factors associated with length of stay in hospital among the elderly patients using count regression models" (<http://mjiri.iums.ac.ir/article-1-6183-en.html>). *Medical Journal of The Islamic Republic of Iran*. doi:10.47176/mjiri.35.5 (<https://doi.org/10.47176/mjiri.35.5>). PMC 8111647 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8111647>). PMID 33996656 (<https://pubmed.ncbi.nlm.nih.gov/33996656>).
20. 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" (<https://zenodo.org/record/1449492>). *J R Stat Soc*. **83** (2): 255–279. doi:10.2307/2341080 (<https://doi.org/10.2307/2341080>). JSTOR 2341080 (<https://www.jstor.org/stable/2341080>).
21. Haldane, J. B. S. (1945). "On a Method of Estimating Frequencies". *Biometrika*. **33** (3): 222–225. doi:10.1093/biomet/33.3.222 (<https://doi.org/10.1093/biomet/33.3.222>). hdl:10338.dmlcz/102575 (<https://hdl.handle.net/10338.dmlcz/102575>). JSTOR 2332299 (<https://www.jstor.org/stable/2332299>). PMID 21006837 (<https://pubmed.ncbi.nlm.nih.gov/21006837>).
22. 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 (<https://doi.org/10.1111/1467-842X.00075>).
23. 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" (<https://doi.org/10.1175/2F2010MWR3315.1>). *Monthly Weather Review*. **138** (7): 2681–2705. Bibcode:2010MWRv..138.2681V (<https://ui.adsabs.harvard.edu/abs/2010MWRv..138.2681V>). doi:10.1175/2010MWR3315.1 (<https://doi.org/10.1175/2010MWR3315.1>).
24. 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. Bibcode:2006MWRv..134.2224M (<https://ui.adsabs.harvard.edu/abs/2006MWRv..134.2224M>). doi:10.1175/MWR3160.1 (<https://doi.org/10.1175/2FMWR3160.1>).

25. Vitolo, R.; Stephenson, D.B.; Cook, Ian M.; Mitchell-Wallace, K. (2009). "Serial clustering of intense European storms" (<https://semanticscholar.org/paper/817e6732db9189e727a5138b957825f2b5715dea>). *Meteorologische Zeitschrift*. 18 (4): 411–424. Bibcode:2009MetZe..18..411V (<https://ui.adsabs.harvard.edu/abs/2009MetZe..18..411V>). doi:10.1127/0941-2948/2009/0393 (<https://doi.org/10.1127/0941-2948/2009/0393>). S2CID 67845213 (<https://api.semanticscholar.org/CorpusID:67845213>).
26. McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models* (Second ed.). Boca Raton: Chapman and Hall/CRC. ISBN 978-0-412-31760-6.
27. Cameron, Adrian C.; Trivedi, Pravin K. (1998). *Regression analysis of count data* (<https://archive.org/details/regressionanalys00c>). Cambridge University Press. ISBN 978-0-521-63567-7.
28. Robinson, M.D.; Smyth, G.K. (2007). "Moderated statistical tests for assessing differences in tag abundance" (<https://doi.org/10.1093/bioinformatics/btm453>). *Bioinformatics*. 23 (21): 2881–2887. doi:10.1093/bioinformatics/btm453 (<https://doi.org/10.1093/bioinformatics/btm453>). PMID 17881408 (<https://pubmed.ncbi.nlm.nih.gov/17881408>).
29. Love, Michael; Anders, Simon (October 14, 2014). "Differential analysis of count data – the DESeq2 package" (<http://www.bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.pdf>) (PDF). Retrieved October 14, 2014.
30. Chen, Yunshun; Davis, McCarthy (September 25, 2014). "edgeR: differential expression analysis of digital gene expression data" (<http://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf>) (PDF). Retrieved October 14, 2014.
31. Montmort PR de (1713) Essai d'analyse sur les jeux de hasard. 2nd ed. Quillau, Paris
32. Pascal B (1679) Varia Opera Mathematica. D. Petri de Fermat. Tolosae

Retrieved from "https://en.wikipedia.org/w/index.php?title=Negative_binomial_distribution&oldid=1108009070"

This page was last edited on 2 September 2022, at 02:41 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License 3.0; 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.