

Negative Binomial

Paul Johnson and Andrea Vieux

June 10, 2013

1 Introduction

The Negative Binomial is discrete probability distribution, one that can be used for counts or other integer-valued variables. It gives probabilities for integer values from 0 to ∞ . It is sometimes also known as the Pascal distribution or as Polya's distribution.

The Negative Binomial is best thought of as a variant of a Binomial distribution. The reader will recall that a *Bernoulli trial* is a "coin flip" experiment, one that returns "yes" or "no", "failure" or "success." The Binomial distribution describes the number of "successes" out of a given number of "Bernoulli trials". As such, its values are defined from 0 up to the number of trials.

The Negative Binomial describes the same sequence of Bernoulli trials that is described by the Binomial distribution. The difference is "what" is being described. Whereas Bernoulli tracks the number of successes, the Negative Binomial represents the number of failures that occurs at the beginning of the sequence as we wait for a given number of successes to be achieved.

The most frequent use of the Negative Binomial model has nothing to do with the property that it describes the chances of a string of failures. Rather, it is used to describe distributions that represent counts of events. It is a replacement for the commonly used Poisson distribution, which is often considered the default distribution for integer-valued count data. The Poisson has the property that its expected value equals its variance, a property that is not found in some empirically observed distributions. The Negative Binomial has a higher variance.

2 Mathematical Description

Recall that the Binomial distribution gives the probability of observing r successes out of N trials when the probability of observing a success is fixed at p .

$$B(r|N) = \binom{N}{r} p^r (1-p)^{N-r} = \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \quad (1)$$

The symbol $\binom{N}{r}$ is the "binomial coefficient" (from intermediate algebra) and it is spoken out loud as " N choose r ", meaning the number of ways one can choose subsets of r from a larger set of N .

$$\binom{N}{r} = \frac{N!}{N!(N-r)!} \quad (2)$$

Thinking of one particular sequence the N Bernoulli trials, one that has r successes, we can easily calculate the probabilities. The chance of observing one particular sequence, S, S, F, S, F, \dots, S , will be $p \cdot p \cdot (1-p) \cdot p \cdot (1-p) \dots p$. Regrouping indicates the probability of that particular sequence is $p^r (1-p)^{N-r}$. That is the probability of one particular sequence that has r successes. Put together the chances of all such sequences, of which there are $\binom{N}{r}$, and the formula in 1 should be clearly understood.

The Negative Binomial is defined with reference to the Binomial. The R help page for the `rbinom` function describes it as "the number of failures which occur in a sequence of Bernoulli trials before a target number of successes is reached" (R documentation).

We found the description on the Math World web site (<http://mathworld.wolfram.com/NegativeBinomialDistribution.html>) to have more intuitive appeal. The Negative Binomial Distribution gives “the probability of $r-1$ successes and x failures in $x+r-1$ trials, and success on the $(x+r)$ th trial.”

Take the first part of the definition, $(r-1)$ successes out of $(x+r-1)$ trials. By the Binomial distribution, that is

$$Prob(r-1 | x+r-1) = \frac{(x+r-1)!}{(r-1)!(x)!} p^{r-1} (1-p)^x \quad (3)$$

The probability of observing a success on the last trial, the $(x+r)$ th trial, is p . So, by the multiplication rule of probabilities, we multiply the previous equation 3 by p to obtain

$$NB(x | r, p) = \frac{(x+r-1)!}{(r-1)!(x)!} p^r (1-p)^x \quad (4)$$

Or, if you prefer to write it with the binomial coefficient,

$$NB(x | r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x \quad (5)$$

3 Central Tendency and Dispersion

If x has a Negative Binomial distribution, then x can take on integer values ranging from 0 to ∞ . The upper limit is infinity because we are letting the process run until there are exactly $(r-1)$ failures and then we conduct one additional experiment on which we obtain success.

The expected value and variance are:

$$E[x] = \frac{r(1-p)}{p} \quad (6)$$

$$Var(x) = \frac{r(1-p)}{p^2} \quad (7)$$

and the skewness and kurtosis are

$$Skewness(x) = \frac{2-p}{\sqrt{(1-p)r}}$$

$$Kurtosis(x) = \frac{p^2 - 6p + 6}{r(1-p)}$$

These are worth considering because of their dependence on r . As one would expect, the expected number of successes is increasing in the number of observed failures. However, the important effects of r are seen in the variance and skewness. As r increases, the variance increases. But the skewness shrinks. Hence, for small values of r , we should expect to see a very skewed distribution, whereas for higher values of r , the distribution should be more symmetrical.

4 Illustrations

The following code is for Figure 1, which displays example distributions as a function of a target number of successes and probabilities for successes on individual trials. Recall that the probabilities indicate the chances of observing a given number of failures after observing a given number of successes.

```
drawBinom <- function (S, p){
  x<- seq(0,200)
  xprob <- dnbinom(x, size=S, prob=p)
  mytitle <- paste("x",S,p)
```

```

    plot(x, xprob, type="s", main=mytitle, xlab=paste("number of failures before
", S, " successes"))
  }
par(mfcol=c(3,2))
sizes <- c(20,40,60)
pvals <- c(0.33,0.67)
for (i in 1:2){
  for (j in 1:3){
    drawBinom (sizes[j], pvals[i] )
  }
}

```

In the Figure 1 we can see how changes in the size and probability alter the Negative Binomial Distribution. The size moves the graph right to left, and the probability widens or shrinks the bell shape. For a graph which resembles the Normal Distribution, refer to Figure 2.

5 The Negative Binomial as a Mixture Distribution

From J. Scott Long's book (Citation ??):

The negative binomial distribution is a useful tool in the analysis of count data. It differs from the Poisson Distribution in that it allows for the conditional variance to exceed the conditional mean. The negative binomial distribution is the end result of a process which begins with a Poisson distribution with mean λ . The fixed (λ) is replaced with a random variable that has a mean of λ , but has nonzero variance. With this new framework, we are representing the possibility that there is unobserved heterogeneity.

Recall the definition of the Poisson gives the probability of observing x events as a function of a parameter λ .

$$Poisson(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (8)$$

Instead of thinking of λ as a constant, think of it as a random variable, representing the fact that the average count is likely to vary across experiments (counties, cities, etc). If we think of λ as a variable, then the probability of observing x failures has to be the result of a 2 part process. Part 1 chooses λ and part 2 chooses x . That means there are possibly many different routes through which we can observe a particular value x , since λ can vary from 0 to ∞ , and for each possible value of λ , there is "some chance" of observing any positive value.

$$\int P(x|\lambda)p(\lambda)d\lambda \quad (9)$$

If one chooses "just the right" distribution for λ , then the "mixture" distribution will be Negative Binomial.

Suppose we take the lucky guess that λ has a Gamma probability distribution

$$Prob(\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{(\alpha-1)} e^{-\lambda/\beta} \quad (10)$$

The "shape" parameter is α and the scale parameter is β , and the expected value is $\alpha\beta$ and the variance is $\alpha\beta^2$.

Inserting the given distributions into 9, we are able to calculate the marginal distribution of x .

Figure 1: Negative Binomial

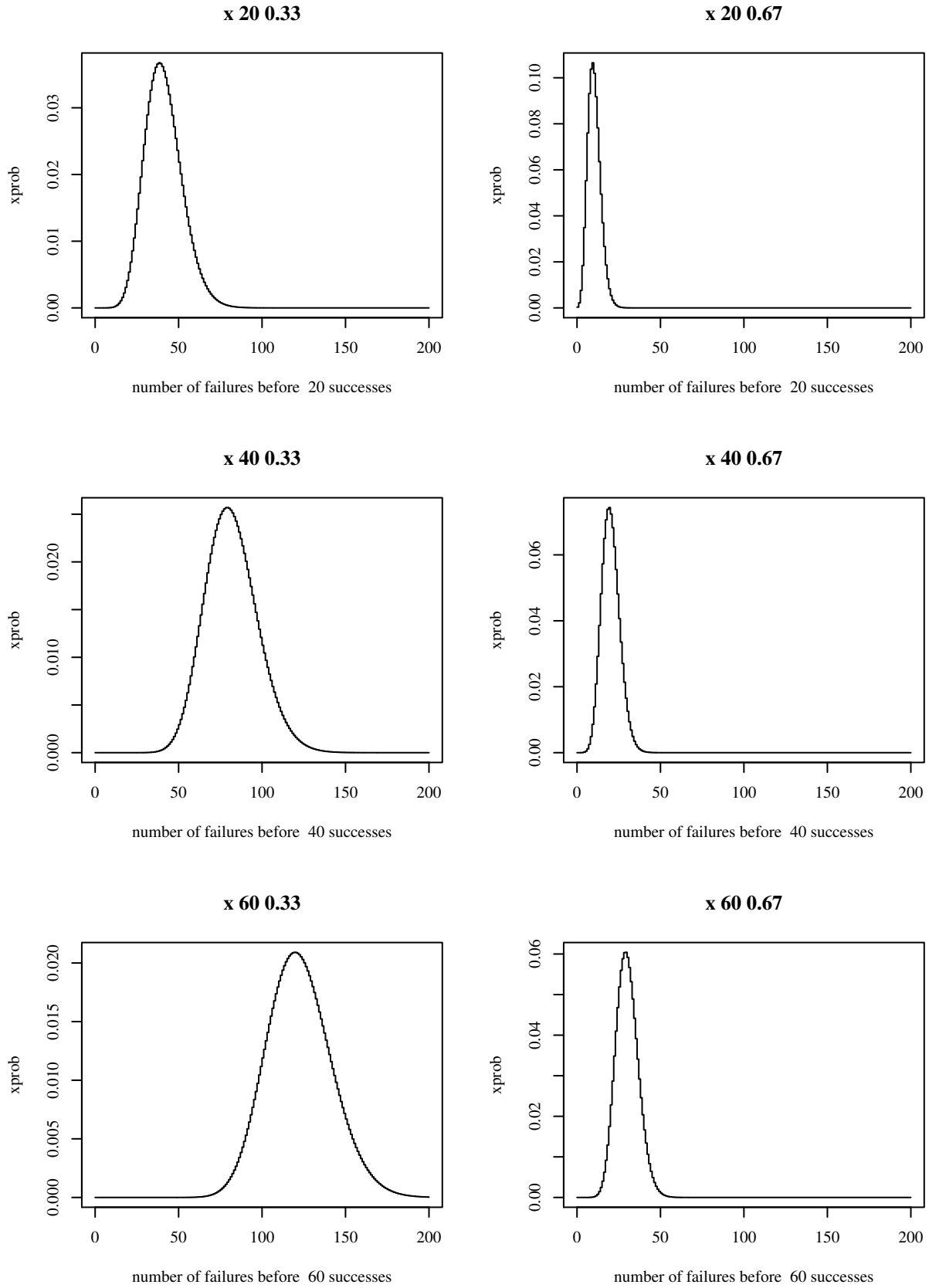
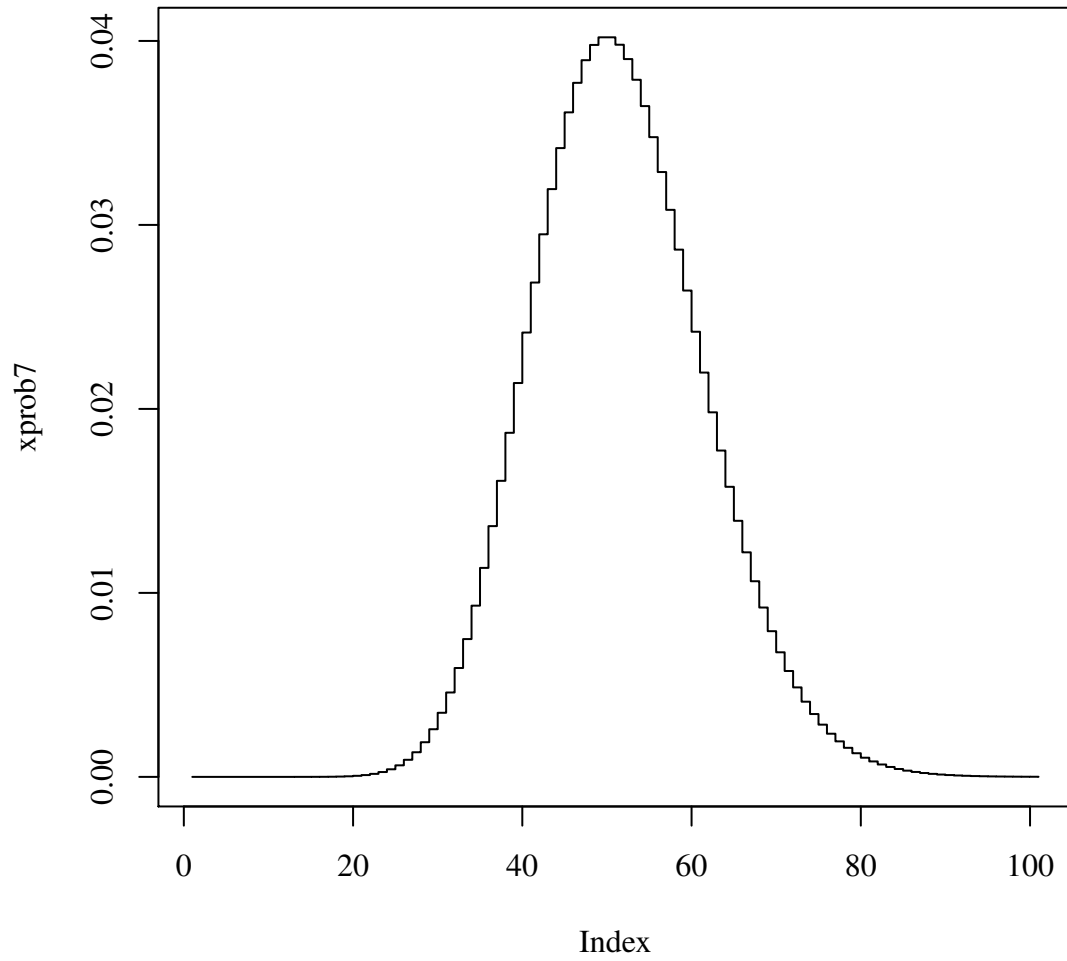


Figure 2: Negative Binomial 2

```
x7 <- seq(0, 100)
xprob7 <- dnbinom(x7, size = 50, prob = 0.5)
plot(xprob7, type="s")
```



$$Prob(x) = \int_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{(\alpha-1)} e^{-\lambda/\beta} d\lambda \quad (11)$$

$$= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \int_0^{\infty} e^{-\lambda} \lambda^x \lambda^{(\alpha-1)} e^{-\lambda/\beta} d\lambda \quad (12)$$

$$= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \int_0^{\infty} \lambda^{x+\alpha-1} e^{-\lambda} e^{-\lambda/\beta} d\lambda \quad (13)$$

$$= \frac{1}{x! \beta^\alpha \Gamma(\alpha)} \int_0^{\infty} \lambda^{x+\alpha-1} e^{-\lambda(1+1/\beta)} d\lambda \quad (14)$$

I have tried to simplify this through brute force, but have been unable to do so. I feel sure there is a specialized result in calculus that will help to use the Gamma function, which would be adapted to this case as:

$$\Gamma(x + \alpha) = \int_0^{\infty} \lambda^{x+\alpha-1} e^{-\lambda} d\lambda \quad (15)$$

And the end result, the Negative Binomial distribution, is stated either as

$$= \frac{\Gamma(\alpha + x)}{x! \Gamma(\alpha)} \left(\frac{1}{1 + \beta} \right)^\alpha \left(\frac{\beta}{1 + \beta} \right)^x \quad (16)$$

or

$$= \frac{\Gamma(\alpha + x)}{x! \Gamma(\alpha)} \left(\frac{1}{1 + \beta} \right)^\alpha \left(1 - \frac{1}{1 + \beta} \right)^x \quad (17)$$

This is the form of the Negative Binomial that was discussed earlier.

I have been unable to find the brute force method to make the jump from 14 to 17, but in Gelman, Carlin, Stern, and Rubin's Bayesian Data Analysis, 2ed (p. 53), an alternative derivation is offered. Recall Bayes's theorem

$$p(\lambda|x) = \frac{p(x|\lambda)p(\lambda)}{p(x)} \quad (18)$$

Rearrange that

$$p(x) = \frac{p(x|\lambda)p(\lambda)}{p(\lambda|x)} \quad (19)$$

The left hand side is our target—the marginal distribution of x . We already know formulae for the two components in the numerator, but the denominator requires a result from probability theory.

Most statistics books—and all Bayesian statistics books—will have a result about updating from a Poisson distribution. If x is a Poisson variable with parameter λ , and the conjugate prior for λ is the Gamma distribution, $Gamma(\lambda|\alpha, \beta)$, then Bayes's theorem (expression 18), leads to the conclusion that if x “events” are observed, then

$$p(\lambda|x) = Gamma(\lambda|\alpha + x, \beta + 1) \quad (20)$$

That is the last ingredient we need to solve expression 19.

Gelman et al put it this way:

$$p(x) = \frac{Poisson(x|\lambda)Gamma(\lambda|\alpha, \beta)}{Gamma(\lambda|\alpha + x, 1|\beta)} \quad (21)$$

Writing in the expressions obtained above,

$$p(x) = \frac{\frac{\lambda^x e^{-\lambda}}{x!} \cdot \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{(\alpha-1)} e^{-\lambda/\beta}}{\frac{1}{(1+\beta)^{(\alpha+x)} \Gamma(\alpha+x)} \lambda^{(\alpha+x-1)} e^{-\frac{\lambda}{1+\beta}}} \quad (22)$$

After cancellation and rearrangement, that simplifies to:

$$p(x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)x!} \cdot \frac{\beta^\alpha}{(1 + \beta)^{\alpha+x}} \quad (23)$$

Using the definition of $\Gamma(x)$ and rearranging again, we find:

$$= \frac{(\alpha + x - 1)!}{(\alpha - 1)!x!} \cdot \left(\frac{\beta}{1 + \beta}\right)^\alpha \left(\frac{1}{1 + \beta}\right)^x \quad (24)$$

Recognizing that the first term on the left is the Binomial coefficient, the derivation is complete.

We choose the shape and scale parameters so that the Gamma distributed value of λ will allow us to simplify 17. It appears obvious in this expression that the role of the probability of success on an individual trial is $p = \frac{1}{1+\beta}$ and that α is the “desired number of successes” in the Negative Binomial. That means that, if want to set the Gamma parameters equal to the proper levels, we choose $\beta = (1 - p)/p$ and $\alpha = n$ (the number of desired successes).

If n represents the number of successes and the number of failures is $x = N - n$

$$Prob(x) = \frac{\Gamma(x + n)}{\Gamma(n)x!} p^n (1 - p)^x$$

Another way to think of the mixture

Instead of replacing λ in the Poisson definition, suppose we think of multiplying it by a randomly chosen value δ . If δ has an expected value of 1, then $E(\lambda\delta) = \lambda$, so, “on average,” the original λ is reproduced. However, because δ might be higher or lower, then $\lambda\delta$ will have random variation, and so the number of observed events will fluctuate. Its average will still be λ , but it will have greater variance. In such a way, one can see the Poisson in this way (replacing λ by $\lambda\delta$).

$$P(x | \delta) = \frac{e^{-(\lambda\delta)}(\lambda\delta)^x}{x!}$$

Here is the question: how can one select a mathematically workable model for δ so that $E(\delta) = 1$?

I’ve seen this done in several ways. Recall that the Gamma distribution has expected value $\alpha\beta$ and variance $\alpha\beta^2$. So if we drew a variate y from any Gamma distribution and then divided by $\alpha\beta$, the result would be

$$x/\alpha\beta$$

and the expected value would be $E(x/\alpha\beta) = E(y)/\alpha\beta = 1$.

More commonly, attention is focused on a subset of possible Gamma distributions, the ones for which the $\beta = 1/\alpha$. When δ follows a Gamma distribution $Gamma(\delta|\alpha, 1/\alpha)$, then it has an expected value of 1 and its variance is $1/\alpha$. As α tends to 0, the variance tends toward infinity.

Think of λ as an “attribute” of an observation. If δ is a Gamma variate with mean 1 and variance $1/\alpha$, then $\lambda\delta$ is also a Gamma variate, with mean λ and variance λ^2/α . Maybe it is just a lucky guess, but it appears to me that $\lambda\delta$ would have the distribution $Gamma(\alpha, \lambda/\alpha)$.