Notes on the Negative Binomial Distribution

John D. Cook

October 28, 2009

Abstract

These notes give several properties of the negative binomial distribution.

1. Parameterizations
2. The connection between the negative binomial distribution and the binomial theorem
3. The mean and variance
4. The negative binomial as a Poisson with gamma mean
5. Relations to other distributions
6. Conjugate prior

1 Parameterizations

There are a couple variations of the negative binomial distribution.

The first version counts the number of the trial at which the rth success occurs. With this version,

$$P(X_1 = x \,|\, p, r) = \binom{x-1}{r-1} p^r (1-p)^{x-r}.$$

for integer $x \geq r$. Here $0 < p < 1$ and r is a positive integer.

The second version counts the number of failures before the rth success. With this version,

$$P(X_2 = x \,|\, p, r) = \binom{r+x-1}{x} p^r (1-p)^{x}.$$

for integer $x \geq 0$. If X_1 is a negative binomial random variable according to the first definition, then $X_2 = X_1 - r$ is a negative binomial according to the second definition.

We will standardize on this second version for the remainder of these notes. One advantage to this version is that the range of x is non-negative integers. Also, the definition can be more easily extended to all positive real values of r since there is no factor of r in the bottom of the binomial coefficient.

Rather than parameterizing the negative binomial in terms of r and p, considering μ and σ as derived quantities, we could consider μ and σ primary and r and p derived. In this case

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

and

$$p = \frac{r}{r + \mu}.$$

The factor $1/r$ is a sort of "clumping" parameter. As $r \to \infty$, the negative binomial converges in distribution to the Poisson as noted below. This fact is suggested by the variance approaching the mean as $r \to \infty$.

Viewing μ and σ as primary, we ignore the combinatorial motivation for defining the negative binomial and instead view it simply as a model for count data.

A little algebra shows that

$$\sigma^2 = \mu + \frac{1}{r}\mu^2.$$

Thus the variance is always larger than the mean for the negative binomial. Since the Poisson requires the mean and variance to be equal, it is unsuitable for data with greater variance than mean; the negative binomial may be appropriate in such settings.

2 What's negative about a negative binomial?

What's negative about the negative binomial distribution? What is binomial about it? This is not always clearly explained. The name comes from applying the general form of the binomial theorem with a negative exponent.

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

The xth term in the series above is

$$\binom{-r}{x} p^r (-q)^x = (-1)^x \binom{-r}{x} p^r q^x = \binom{r+x-1}{x} p^r (1-p)^x$$

which is the probability that $X = x$ where $X \sim$ negative binomial with parameters r and p.

3 Mean and variance

The negative binomial distribution with parameters r and p has mean

$$\mu = r(1-p)/p$$

and variance

$$\sigma^2 = r(1-p)/p^2 = \mu + \frac{1}{r}\mu^2.$$

4 Hierarchical Poisson-gamma distribution

In the first section of these notes we saw that the negative binomial distribution can be seen as an extension of the Poisson distribution that allows for greater variance. Here we examine another derivation of the negative binomial distribution that makes the connection with the Poisson more explicit.

Suppose $X \,|\, \Lambda$ is a Poisson random variable and Λ is a gamma(α, β) random variable. We create a new kind of random variable by starting with a Poisson but making it more variable by allowing the mean parameter to itself be random.

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

$$
\begin{aligned}
&= \frac{1}{x!\Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{\alpha+x+1} e^{-\lambda(1+1/\beta)}\, d\lambda \\
&= \frac{1}{\Gamma(x+1)\Gamma(\alpha)\beta^\alpha}\, \Gamma(\alpha+x) \left(\frac{\beta}{\beta+1}\right)^{\alpha+x} \\
&= \binom{\alpha+x-1}{x} \left(\frac{1}{\beta+1}\right)^{\alpha} \left(1-\frac{1}{\beta+1}\right)^{x}
\end{aligned}
$$

Therefore the marginal distribution of X is negative binomial with $r = \alpha$ and $p = 1/(\beta+1)$.

If we change the parameters in the distribution of Λ so that the mean stays the same but the distribution becomes more concentrated, we expect the distribution of X to become more like that of the Poisson. That is in fact what happens. The mean of Λ is $\lambda = \alpha\beta$ and the variance is $\alpha\beta^2$. If we let $\alpha \to \infty$ while keeping $\beta = \lambda/\alpha$, the variance of Λ goes to 0. The distribution on X converges to a Poisson distribution because as noted in Section 5.4 below, $r \to \infty$ and $p \to 1$ while keeping the mean constant.

5 Relation to other distributions

Throughout this section, assume X has a negative binomial distribution with parameters r and p.

5.1 Geometric

A negative binomial distribution with $r = 1$ is a geometric distribution. Also, the sum of r independent Geometric(p) random variables is a negative binomial(r, p) random variable.

5.2 Negative binomial

If each $X_i \sim$ is distributed as negative binomial(r_i, p) then $\sum X_i$ is distributed as negative binomial($\sum r_i$, p).

5.3 Binomial

Let $Y \sim \text{binomial}(n, p)$. Then $P(Y < r - 1) = P(X > n - r)$.

5.4 Poisson

If $r \to \infty$ and $p \to 1$ as μ stays constant, $P(X = x)$ converges to $e^{-\mu}\mu^x/x!$, the density for a Poisson(μ) distribution.

5.5 Beta

$F(x) = I_p(r, x + 1) = P(Y < p)$ where $Y \sim \text{beta}(r, x + 1)$.

5.6 Gamma

If $p \to 0$ as r stays constant, pX converges in distribution to a gamma distribution with shape r and scale 1.

6 Conjugate prior

If the likelihood function for an observation x is negative binomial(r, p) and p is distributed *a priori* as Beta(a, b) then the posterior distribution for p is Beta$(a + r, b + x)$. Note that this is the same as having observed r successes and x failures with a binomial$(r + x,\ p)$ likelihood. All that matters from a Bayesian perspective is that r successes were observed and x failures.

This document available at `http://www.johndcook.com/negative_binomial.pdf`.