Negative Binomial Distribution

The binomial distribution counts the number of successes in a fixed number of Bernoulli trials. Suppose that, instead, we count the number of Bernoulli trials required to get a fixed number of successes. This latter formulation leads to the negative binomial distribution.

In a sequence of independent Bernoulli(p) trials, let the random variable X denote the trial at which the rth success occurs, where r is a fixed integer. Then

$$P(X = x|r,p) = \binom{x-1}{r-1} p^r(1-p)^{x-r}, \quad x = r,\ r+1, \dots, \tag{3.2.9}$$

and we say that X has a *negative binomial(r,p) distribution.*

The derivation of (3.2.9) follows quickly from the binomial distribution. The event $\{X = x\}$ can occur only if there are exactly $r-1$ successes in the first $x-1$ trials, and a success on the xth trial. The probability of $r-1$ successes in $x-1$ trials is the binomial probability $\binom{x-1}{r-1} p^{r-1}(1-p)^{x-r}$, and with probability p there is a success on the xth trial. Multiplying these probabilities gives (3.2.9).

The negative binomial distribution is sometimes defined in terms of the random variable Y = number of failures before the rth success. This formulation is statistically equivalent to the one given above in terms of X = trial at which the rth success occurs, since $Y = X - r$. Using the relationship between Y and X, the alternative form of the negative binomial distribution is

$$P(Y = y) = \binom{r+y-1}{y} p^r(1-p)^y, \quad y = 0, 1, \dots. \tag{3.2.10}$$

Unless otherwise noted, when we refer to the negative binomial(r,p) distribution we will use this pmf.

The negative binomial distribution gets its name from the relationship

$$\binom{r+y-1}{y} = (-1)^y \binom{-r}{y} = (-1)^y \frac{(-r)(-r-1)(-r-2)\cdot \cdots \cdot(-r-y+1)}{(y)(y-1)(y-2)\cdot \cdots \cdot(2)(1)},$$

which is, in fact, the defining equation for binomial coefficients with negative integers (see Feller 1968 for a complete treatment). Substituting into (3.2.10) yields

$$P(Y = y) = (-1)^y \binom{-r}{y} p^r(1-p)^y,$$

which bears a striking resemblance to the binomial distribution.

The fact that $\sum_{y=0}^{\infty} P(Y = y) = 1$ is not easy to verify but follows from an extension of the Binomial Theorem, an extension that includes negative exponents. We will not pursue this further here. An excellent exposition on binomial coefficients can be found in Feller (1968).

The mean and variance of Y can be calculated using techniques similar to those used for the binomial distribution:

$$\begin{aligned}
\mathrm{E}Y &= \sum_{y=0}^{\infty} y \binom{r+y-1}{y} p^r (1-p)^y \\
&= \sum_{y=1}^{\infty} \frac{(r+y-1)!}{(y-1)!(r-1)!} p^r (1-p)^y \\
&= \sum_{y=1}^{\infty} r \binom{r+y-1}{y-1} p^r (1-p)^y.
\end{aligned}$$

Now write $z = y - 1$, and the sum becomes

$$\begin{aligned}
\mathrm{E}Y &= \sum_{z=0}^{\infty} r \binom{r+z}{z} p^r (1-p)^{z+1} \\
&= r\frac{(1-p)}{p} \sum_{z=0}^{\infty} \binom{(r+1)+z-1}{z} p^{r+1}(1-p)^z \qquad \left(\begin{array}{c}\text{summand is negative} \\ \text{binomial pmf}\end{array}\right) \\
&= r\frac{(1-p)}{p}.
\end{aligned}$$

Since the sum is over all values of a negative binomial$(r+1, p)$ distribution, it equals 1. A similar calculation will show

$$\operatorname{Var} Y = \frac{r(1-p)}{p^2}.$$

There is an interesting, and sometimes useful, reparameterization of the negative binomial distribution in terms of its mean. If we define the parameter $\mu = r(1-p)/p$, then $\mathrm{E}Y = \mu$ and a little algebra will show that

$$\operatorname{Var} Y = \mu + \frac{1}{r}\mu^2.$$

The variance is a quadratic function of the mean. This relationship can be useful in both data analysis and theoretical considerations (Morris 1982).

The negative binomial family of distributions includes the Poisson distribution as a limiting case. If $r \to \infty$ and $p \to 1$ such that $r(1-p) \to \lambda, 0 < \lambda < \infty$, then

$$\begin{aligned}
\mathrm{E}Y &= \frac{r(1-p)}{p} \to \lambda, \\
\operatorname{Var} Y &= \frac{r(1-p)}{p^2} \to \lambda,
\end{aligned}$$

which agree with the Poisson mean and variance. To demonstrate that the negative binomial$(r,p) \to$ Poisson(λ), we can show that all of the probabilities converge. The fact that the mgfs converge leads us to expect this (see Exercise 3.15).

Example 3.2.6 (Inverse binomial sampling) A technique known as inverse binomial sampling is useful in sampling biological populations. If the proportion of individuals possessing a certain characteristic is p and we sample until we see r such individuals, then the number of individuals sampled is a negative binomial random variable.

For example, suppose that in a population of fruit flies we are interested in the proportion having vestigial wings and decide to sample until we have found 100 such flies. The probability that we will have to examine at least N flies is (using (3.2.9))

$$\begin{aligned} P(X \geq N) &= \sum_{x=N}^{\infty} \binom{x-1}{99} p^{100}(1-p)^{x-100} \\ &= 1 - \sum_{x=100}^{N-1} \binom{x-1}{99} p^{100}(1-p)^{x-100}. \end{aligned}$$

For given p and N, we can evaluate this expression to determine how many fruit flies we are likely to look at. (Although the evaluation is cumbersome, the use of a recursion relation will speed things up.) ||

Example 3.2.6 shows that the negative binomial distribution can, like the Poisson, be used to model phenomena in which we are waiting for an occurrence. In the negative binomial case we are waiting for a specified number of successes.