Negative Binomial Distribution

The negative binomial distribution is a generalization of the Poisson distribution for count data that permits the variance to exceed the mean. The negative binomial distribution can be used to describe the distribution arising from an experiment consisting of a sequence of independent Bernoulli trials. The probability of success for each trial, *p*, is constant across the experiment. The experiment continues until a fixed number of successes are achieved.

Suppose we flip a coin repeatedly and count the number of heads (successes). If we continue flipping the coin until it has landed 2 times on heads, we are conducting a negative binomial experiment. The negative binomial random variable is the number of coin flips required to achieve 2 heads.

If the distribution of the response variable is Poisson, given the mean at a fixed setting of the predictors, and the mean itself follows a gamma distribution, then it follows that the marginal distribution for the response variable is a negative binomial. Unlike the Poisson distribution, the negative binomial distribution provides a way to model subject heterogeneity and account for overdispersion. It enables the model to have greater flexibility in modeling the relationship between the mean and the variance of the response variable than the Poisson distribution.

The negative binomial distribution is appropriate for aggregated events. For example, the distribution of an organism in space might be a negative binomial. Organisms tend to aggregate, either because aggregation enhances survival, or because individuals of the same species favor the same habitat or environmental conditions. For example, humans aggregate in towns and cities, bacteria aggregate in colonies, and cows aggregate in herds. The relative abundance of species in ecological communities results in overdispersion when you use a Poisson distribution. Instead, you can use the negative binomial distribution.

So, how does the negative binomial distribution account for overdispersion? The negative binomial model for counts expresses μ in terms of explanatory variables. The most common link function used is the Log link function, as in Poisson regression models. For a negative binomial distribution, the relationship between the variance and the mean has a dispersion parameter that must be estimated or set to a fixed value. The dispersion parameter, *k*, enables the variance to exceed the mean and enables the negative binomial distribution to account for overdispersion. The dispersion parameter *k* is not allowed to vary over observations. The limiting case when the parameter *k* is equal to 0 corresponds to a Poisson regression model.

It is always useful to examine the dispersion parameter to see how much greater than 0 it is. If the dispersion parameter is much greater than 0, overdispersion is evident and the standard errors increase. The increased standard errors reflect the overdispersion not captured with the Poisson model. In this case, the negative binomial model is more appropriate than the Poisson model, and the inferences from the negative binomial model are more accurate.

If the model for the mean is correctly specified, the parameter estimates from the negative binomial model are consistent. That is, as the sample size increases, the probability that the parameter estimate approach the true value increases, even if the true distribution is not negative binomial. For additional information on the negative binomial distribution, click the Information button.
