

FES 524: Natural Resources Data Analysis

Reading 8.1: Generalized linear models for count data

Contents

1 Generalized linear models for counts	1
1.1 Negative binomial distribution	1
1.2 Poisson distribution	2

1 Generalized linear models for counts

In this reading, we start by reviewing the count distributions we discussed in week 7 reading 1.

Counts are discrete, integer values that can include 0. For example, we could count the number of seedlings in a plot or the number of inflorescences on a plant. In practice, it is common to have counts where the effort to take the counts varies between observations. This effort could be the amount of time spent counting or the area counted in, for example. Any analysis should take into account varying effort in some way, often by making inference to a count per unit effort (e.g., count per square meter).

The most commonly used distributions for analyzing counts in natural resources are the negative binomial and the Poisson distributions.

1.1 Negative binomial distribution

The negative binomial distribution is a two parameter distribution, but unlike the normal distribution, the variance is related to the mean. In the mean-variance relationship shown in Table 1, you can see that we expect the variance to increase quadratically with the mean.

The negative binomial often works well for modeling counts in ecology. This is because the distribution tends to work well for “clustered” counts, where the presence of an organism is related to the presence of other organisms. Such data often have many zero or low counts as well as some very high counts. This results in high *dispersion* that distributions like the Poisson may not account for.

The example in reading 7.1 was a situation with clustered counts with very high counts of seedlings in areas with many parent trees but having no or few seedlings when there are few parent trees. In this example there are either a lot of seedlings together or there are very few seedlings present.

Since the the variance depends on the mean for this distribution, we can technically have problems where the variance in the data is larger or smaller than the variance based on the distribution. However, given this distribution has a scale parameter, ψ , overdispersion like in single parameter distributions isn’t possible. Any evidence of overdispersion (i.e., value of overdispersion > 1) will indicate a general lack of fit of the model to the data. We will discuss this a bit more later in this reading.

Table 1: Mean-variance relationships for distributions commonly used to model counts.

Distribution	Mean	Variance
Negative binomial	μ	$\mu + \mu^2/\psi$
Poisson	μ	μ

1.2 Poisson distribution