Gamma Distribution

April 11, 2016

Gamma distribution

The gamma random variable is defined on the interval $(0, \infty)$, so it is a useful distribution in ecology for modeling abundance or other positive valued random variables. It also has some properties that may allow more flexibility relative to lognormal. The probability density is defined as a function of two parameters : k the shape parameter and θ the scale parameter (to be consistent with Wikipedia).

In linear modeling, we are interested in modeling variation in the dependent variable Y. If we think about a single observation y_i, this can be modeled as a function of unique shape and scale parameters, thus $y_i \sim gamma(k_i, \theta_i)$ has the probability density function

$f(y_i) = \frac{1}{\theta_i^{\alpha_i}\Gamma(\alpha_i)} y_i^{(k_i-1)} e^{-(y_i/\theta_i)}$,

where $y_i, k_i, \theta_i > 0$

The first two moments of the distribution are:

$$E(y_i) = k_i * \theta_i$$

$$Var(y_i) = k_i * \theta_i^2$$

A couple things to note - 1) the ratio of the variance to the mean is constant for a given value of θ_i, so when the expected value is small the variance is also small, and vice versa. The Poisson distribution has this property too, but unlike the Poisson, the θ_i parameter describes how these two scale with each other; and 2) the coefficient of variation ($CV = \sqrt{Var(y_i)}/E(y_i)$ is equal to $1/\sqrt{k_i}$.

In regression modeling, we want to account for variation in outcomes, but we are really going to be focused on modeling θ_i with the gamma distribution and can assume that $k = k_i$ for all i since it is essentially a multiplier (technically, it is equal to the inverse of the 'dispersion' parameter which is in all exponential distributions and which we typically assume is the same across all observations, e.g. standard deviation of the normal distribution).

Using the relationships above, we can also see that

$$k = 1/CV^2$$

$$\theta_i = E(y_i) * CV^2 = \frac{k * \theta_i}{k}$$

which are the parameterizations that were suggested for homework #1.

The canonical link for the gamma distribution is the reciprocal $1/\mu_i$. So that implies that

$$\mu_i = \frac{1}{b_o + b_1 X_1 + ...}$$

But, this can lead to all sorts of weird behavior, the least of which is that it is not guaranteed to be positive. As a result, other link functions can be - and maybe should be - used. For example, we might use a log() link then to keep things positive. The interpretation of the parameters will vary with the choice of link function used, of course.