Notes on the Negative Binomial Distribution and the GLM Family

Genesis as a gamma mixture of Poissons

Define a random variable as a two-step process as

$$Y|G \sim \text{Po}(\mu G) \quad \text{and} \quad G \sim \gamma(\kappa^{-1}, \kappa)$$

Notice that the distribution of G has probability density of the form

$$f_G(g; \kappa) = \frac{\kappa^\kappa e^{-\kappa g} g^{\kappa-1}}{\Gamma(\kappa)}$$

with $\text{E}[G] = 1$ and $\text{var}[G] = 1/\kappa$. It is no restriction to assume that the mean of G is 1. If it were anything eles, say ω we could replace G by $G^\star = G/\omega$ and μ by $\mu^\star = \mu\omega$, and $G^\star$ does now have a gamma distribution with mean 1, and the distribution of Y is not changed.

The *marginal* distribution of Y then has probability function

$$
\begin{aligned}
p_Y(y; \mu, \kappa) &= \int_0^\infty \frac{\kappa^\kappa e^{-\kappa g} g^{\kappa-1}}{\Gamma(\kappa)} \times \frac{e^{-\mu g}(\mu g)^y}{y!} dg \\
&= \frac{\Gamma(\kappa+y)}{\Gamma(\kappa)y!} \frac{\kappa^\kappa \mu^y}{(\mu+\kappa)^{\kappa+y}} \times \int_0^\infty \frac{(\mu+\kappa)^{\kappa+y} e^{-(\mu+\kappa)g} g^{y+\kappa-1}}{\Gamma(\kappa+y)} dg \\
&= \frac{\Gamma(\kappa+y)}{\Gamma(\kappa)y!} \frac{\kappa^\kappa \mu^y}{(\mu+\kappa)^{\kappa+y}} \times 1
\end{aligned}
$$

since the integral term in the second last line above is the integral of a normalized gamma density over its complete range.

The probability function defines the Negative Binomial distribution. It is a discrete distribution frequently used for modelling processes with a response count for which the data are overdispersed relative to the Poisson distribution.

Membership of the GLM family

The Negative Binomial distribution belongs to the GLM family, but *only if the parameter κ is known*. To see this, write the probability function in the form:

$$p_Y(y; \mu, \kappa) = \exp\left[\left\{y \log\left(\frac{\mu}{\mu+\kappa}\right) - \kappa \log(\mu+\kappa)\right\} + \kappa \log\kappa + \log\Gamma(\kappa+y) - \log\Gamma(\kappa) - \log y!\right]$$

This belongs to the GLM family with

$$
\begin{aligned}
A = \varphi &= 1 \quad \text{both known,} \\
\theta(\mu) &= \log\left(\frac{\mu}{\mu+\kappa}\right) \\
b(\theta) &= \kappa \log(\mu+\kappa) = \kappa \log\left(\frac{\kappa}{1-e^\theta}\right) \quad \text{and note that} \\
\mu = b'(\theta) &= \frac{\kappa e^\theta}{1-e^\theta} = \mu \\
\sigma^2 = b''(\theta) &= \frac{\kappa e^\theta}{1-e^\theta} + \frac{\kappa e^{2\theta}}{(1-e^\theta)^2} = \mu + \mu^2/\kappa
\end{aligned}
$$

The extra term in the variance, μ^2/κ, accounts for the ability of the Negative Binomial to model situations which are overdispersed relative to the Poisson. Note that as $\kappa \to \infty$ this term goes to zero and in fact the whole distribution approaches the Poisson as a limit.

The natural link depends on κ, which in practice is rarely known, so the natural link is not very much used in practice. The common link to use is the log link, which the natural link in fact approaches as $\kappa \to \infty$, the natural link for the Poisson case.

It can be shown that if κ is unknown and we fit a model like a GLM, with $\mu = \ell^{-1}(\eta)$, then the MLEs $\hat{\kappa}$ and $\hat{\boldsymbol{\beta}}$ are asymptotically independent. In this respect κ is similar to a dispersion parameter for a true GLM.

Software, fitting algorithm and statistical tests

The function `glm.nb` in the `MASS` library estimates both κ (which is referred to in the software as "theta") and $\boldsymbol{\beta}$ by an alternating process. Starting with an initial estimate of κ, it estimates $\boldsymbol{\beta}$ assumint κ fixed and using the same process as the function `glm` itself. Then fixing $\boldsymbol{\beta}$, it refined the estimate of κ, and the process continues until convergence.

It is important to notice that tests of models fitted using `glm.nb` do not use deviances at all. They use exact likelihood ratio tests based on the optimised log-likelihoods. The deviance is reported, but has no statistical value since it has been adjusted by the fitting process for κ.

An alternative way of proceeding is to estimate κ from, say, the largest model to be fitted. Then in future model fits the estimate of κ may be held fixed and treated as if known, rather like a transformation parameter. In this case all models fitted with the fixed value of κ would use the fitting function `glm` with family `negative.binomial(theta = <value>)`. If the fact that the fixed value for κ has been estimated is ignored, and the value treated as a known quantity, then models would be compared using differences of deviance and Chi-squared tests. In practice the outcome is usually very similar to that obtained by estimating κ separately for every model and using likelihood ratio tests, but the computational load is much less.