Over/Under-dispersion, zero-inflation and Poisson-Tweedie models

John Hinde, National University of Ireland, Galway, Ireland

john.hinde@nuigalway.ie

The standard distribution for the analysis of count data is the Poisson. Frequently, in practice it is too restrictive in that the variability in the data is either significantly greater (overdispersed) or less (underdispersed) than that implied by the models variance function. For the analysis of count data, Nelder and McCullagh (1989) says that overdispersion is the norm and not the exception and this has been well studied, see Hinde and Demétrio (1999) and many subsequent articles presenting a wide range of distributions. An associated phenomena is zero-inflation where the data exhibit more zeros than expected under the Poisson model. Models allowing for zero-inflation include the zero-inflated mixture model and the two-stage hurdle model. Although less common, underdispersion can also arise, typically from dependent responses. Here we will consider how underdispersion can occur as a result of features of the underlying counting, or data collection, process. The range of distributions for modelling underdispersed count data is relatively limited, although models can be derived in specific situations.

A class of general models is presented based on Poisson-Tweedie factorial dispersion models with variance $\mu + \phi\mu^p$, where μ is the mean, ϕ and p are the dispersion and Tweedie power parameters, respectively. This class of models provides a flexible and comprehensive family including many standard discrete models. The family provides for modelling of overdispersed count data and can also accommodate zero-inflation and underdispersion. For a general approach we consider an extended version of the Poisson-Tweedie model and discuss estimation of regression, dispersion and Tweedie power (variance function) parameters, with illustrative applications.