

Probability distributions to model under-dispersed count data in a Bayesian regression context: evaluation of existing distributions and proposal of new ones to analyze biodiversity data.

Frédéric Gosselin (IRSTEA, France; frederic.gosselin@irstea.fr)

Statistical likelihood-based analysis of count data usually only allows over-dispersion relative to the Poisson distribution. Yet, as we show for a biodiversity data set, some situations might warrant the consideration of under-dispersed distributions as well. A literature review proposed two modifications of the Generalized Poisson distribution as the only appropriate existing probability distributions that allowed for under-dispersion and that were easily applicable in a Bayesian regression setting.

Here we propose relevant alternatives to these two distributions, which involve the Bernoulli, Pòlya or Poisson distribution. The application of these distributions to our biodiversity data set showed that all the ecological groups had an under-dispersed species richness at at least one scale. In these cases, models that allowed for under-dispersion gave more precise estimators of parameters or more accurate estimators of the variance of random effects than models that allowed only over-dispersion. No under-dispersed distribution outperformed the other ones in all situations.

These new families of distributions are very flexible since they allow for both under- and over-dispersion by including the Poisson and negative binomial distributions in the over-dispersed range.

will consider delivering lightning talk