

Contributed Discussion on Article by Müller and Mitra

Comment by Murray Aitkin¹ and Julia Polak²

Space restrictions limit our discussion to the first example.

The example analysis extrapolates from a zero-truncated observed count sample to predict the zero count. Prediction outside the data range is always hazardous. As the authors note, a general multinomial distribution on the observed data cannot predict the zero count: a parametric model is essential for this, with the consequent *strong model-dependence* of the prediction.

Nakatani and Sato (2008) have given a survey and discussion of the zero-truncated Poisson, negative binomial and other discrete distributions for this extrapolation, and a Bayesian analysis of the Poisson and negative binomial distributions can be found in Vergne, Calavas, Cazeau, Durand, Dufour and Grosbois (2012). These authors point out the limitation of the small sample size of their collected data that prevents them from fitting more complicated models. Moreover, they explain why using alternative nonparametric estimates is not suitable.

They used conventional parametric Bayesian methods, not the Dirichlet Process (DP). We follow their analysis for the first example. With only four counts, the data could be analysed by a truncated Poisson(μ) distribution. The MLE of μ is 0.86, and with a flat prior on μ the posterior distribution of μ is easily computed. The median is 0.876 and the 95% central credible interval is (0.60, 1.22). Transforming from μ to $55/(1 - e^{-\mu})$ gives a posterior (Figure 1) for the total number N of T-cell types very close to the authors' Figure 1(b), with median 94 and 95% central credible interval (78,122). What additional information does the DP analysis provide? The authors aim to find a modeling approach between a “misleadingly precise” parametric model like this one, and a fully general multinomial model which could not provide information about the unobserved zero class.

The Dirichlet process with a Poisson base mass function leads to a truncated mixed Poisson distribution for the observed data. With only four support points in the data, no more than two components can be identified from the mixture likelihood, with two extra parameters over the truncated Poisson. The authors give no details of the complexity of their DP model, so it is unclear how it is related to the truncated Poisson or two-component mixed Poisson models, or to the truncated negative binomial distribu-

¹Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia

²Centre for Molecular, Environmental, Genetic and Analytic Epidemiology (MEGA), University of Melbourne, Melbourne, Australia