Modeling Bimodal Count Data Using Com-poisson Mixtures

Smarajit Bose
Indian Statistical Institute, India

Abstract

Bimodal truncated count distributions are frequently observed in aggregate surveys and ratings when respondents are mixed in their opinion. They also arise in censored count data, where the highest category might create an additional mode. Modeling bi-modal behaviour in count data is useful for various purposes, from comparing shapes of different samples (or questions) to predicting future ratings by new raters. The Poisson distribution is the most common distribution for fitting count data and can be modified to achieve mixtures of truncated Poisson distributions. However, it is suitable only for modelling equi-dispersed distributions. Real life data often exhibit over- or underdispersion. In such cases, the Poisson distribution typically does not provide good approximations. Also, the Poisson distribution and even Poisson mixtures are limited in their ability to capture bimodality.

A more flexible alternative is the Conway-Maxwell-Poisson (COM-Poisson or CMP) distribution, a two-parameter generalization of the Poisson distribution, that allows for over- and under-dispersion. Despite its higher flexibility, CMP distribution cannot capture bi-modality. In this paper, we propose a mixture of CMPs for capturing a wide range of truncated count data, which can exhibit unimodal as well as bimodal behaviour (with individual components exhibiting equi-, under- or over- dispersion). We present methods for estimating the parameters of a mixture of two CMP distributions using an Expectation-Maximization(EM) approach. The methods are illustrated using simulated and real data.

(Joint work with G. Shmueli, P. Sur and P. Dubey)