Multilevel modelling of tabular counts via Deconstructed Maximum Likelihood

Jarod Lee*
University of Technology Sydney, Sydney, Australia – Yan.Lee@uts.edu.au

James Brown
University of Technology Sydney, Sydney, Australia – James.Brown@uts.edu.au

Louise Ryan
University of Technology Sydney, Sydney, Australia – Louise.M.Ryan@uts.edu.au

A multilevel Poisson regression model is proposed for modelling spatially clustered tabular count data. Counts in tabular data are usually perturbed to reduce disclosure risk, and this can have great impact on small cell counts. We propose a Poisson mixed model with multiplicative random effect that is Gamma distributed. Compared to the existing model using log-Normal random effect, our model utilizes summary statistics of individual with events, which has benefits in terms of privacy protection, data quality improvement and small area analysis, since perturbation of event counts is no longer required. The parameters in our model are estimated via Deconstructed Maximum Likelihood, a new framework where dataset of fundamental different nature, namely individual and area level variables in our context, are used for inference without having to combine them into a single analysis file. The modularity approach is representative of the structure of multilevel data and is scalable as it avoids the repetition of area level variables over individuals living in the common geographical region. Also, our model has closed form likelihood, rendering inference exact without the use of numerical integration. When the variance of the log-Normal random effect is small, the two models produce similar estimates. We apply our model to the Australian unemployment data and show the potential benefits of incorporating it into data extraction tools.

Keywords: Perturbation; Sufficient statistics; Privacy; Modularity.