Compound Mixed Models for Over-dispersed Count Data: An Examination with Insurance Claim Dataset

Mr. R. Tharshan

Department of Mathematics and Statistics,
University of Jaffna, Sri Lanka.

tharshan10684@gmail.com

Abstract

In practice, the proportion format data analyzing under the binomial assumption; homogeneous success probability at each trial or count data analyzing under the Poisson assumption; *equi-dispersion* constraint, the observed variance simply exceeds from the expected variance. This context is explained by the *over-dispersion* and mostly it is common in biomedical and criminology studies. In the potential solution part, two-stage models that lead to compound mixed probability models for the responses allowing *over-dispersion*, proposed to overcome this phenomenon. Our principle goal in this research is to examine the work efficiency of the mixture of the Poisson and gamma: negative binomial (NB) and the mixture of beta and NB: beta negative binomial (BNB) into the insurance claim dataset that vary in the value of the sample index dispersion (ϕ). The Poisson, NB, and BNB models fit by the maximum likelihood estimation (MLE), tested, and compared based on the p-value of the χ^2 goodness of fit test on fifteen different sets of insurance claim frequency data that obtained from R packages covering the ϕ ranges from 1.053 to 3.154. This study finds that NB and BNB fit better than Poisson handling *over-dispersion* in the insurance claim datasets. It is observed that work efficiency of NB and BNB do not consistent with ϕ values and comparatively for large value of ϕ, the BNB is a better fit than NB.

Keywords - Over-dispersion, Two-stage model, Equi-dispersion, Negative binomial, Sample index dispersion