BINOMIAL-MIXTURE MODELS FOR OVERDISPERSED BINOMIAL OUTCOME DATA AND THEIR APPLICATIONS

C. Manoj

Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka

The binomial outcome data are widely encountered in many real world applications. The Binomial distribution often fails to model the binomial outcome data since the variance of the observed binomial outcome data exceeds the assumed theoretical binomial variance, a phenomenon known as Overdispersion. This research investigates the problem of overdispersion in detail. Among the two reasons for overdispersion, "variation among probability of success parameter" and "correlations between binary trials which make up the binomial random variable", we mainly focus on the first reason. This variation is modeled by mixing a continuous distribution defined on the standard unit interval to the probability of success parameter. The resultant general family of univariate discrete distributions is known as the family of Binomial mixture distributions. The Beta-Binomial distribution is a prominent member of this family of distributions. Available literatures on the studies of theoretical and practical importance of Beta-Binomial distribution in modeling overdispersed binomial outcome data are numerous. On the other hand, currently little attention is paid to generalizations of this distribution and other recently developed related distributions.

The primary purpose of this research work is the introduction of a new, generalized and more flexible member to the family of Binomial mixture distributions. This is achieved by mixing the McDonald's Generalized Beta distribution of the first kind to the success probability of the Binomial distribution. Consequently, a new probability distribution called McDonald Generalized Beta-Binomial distribution (McGBB) is introduced to the statistical and probability literature. Several theoretical properties of McGBB are discussed. The parameters of the McGBB distribution are estimated via maximum likelihood estimation methods. It is shown that this distribution model gives better fit than the previous models when modeling real world overdispersed binomial outcome data. An extended simulation study is presented to compare the McGBB distribution with its related distributions in handling overdispersed binomial outcome data. It is shown that McGBB model is superior to its nested models at many parameter settings.

We also focus on investigating the properties of maximum likelihood based point and interval estimates of the inverse dispersion parameter θ of the Beta-Binomial distribution in modeling severely overdispersed binomial outcome data. We encounter a re-parameterized Beta-Binomial distribution which has been previously described in Morris (1997) and Bolker (2008). In addition to examining the stability of maximum likelihood point estimates of θ, the two interval estimates namely the Profile likelihood based confidence interval (CI_{PL}) and proposed Quadratic approximation to the log likelihood based confidence interval (CI_{QA}) are compared by means of coverage probabilities. It is evident that CI_{QA} outperforms CI_{PL} at many parameter combinations.