$C_{(\alpha)}$ Tests for Testing Homogeneity of Proportions in Presence of McDONALD Generalized Beta-Binomial Over-Dispersion

Bichanga Lawrence Areba[1], Ali Islam[2], Orawo Luke[3]

[1, 2, 3]Department of Mathematics, Egerton University, Njoro, Kenya

Abstract: ***An important problem in toxicology, teratology, consumers purchasing behavior, drinking behavior of alcohol, in studies of dental caries in children and other similar fields is to compare proportions of certain characteristic in several groups. A special case is to compare the proportions in a control group with that in a treatment group. However, these proportions often exhibit variation greater than predicted by a simple binomial model. Continuous distribution defined on the standard unit interval is used to test homogeneity of proportion as one way of handling over-dispersion of the binomial distribution. The Kumaraswamy-Binomial (KB) distribution, Beta-Binomial (BB) distribution and the new McDonald Generalized Beta-Binomial (McGBB) distributions are prominent members of Binomial mixture distribution. The new McDonald Generalized Beta-Binomial distribution model has shown to give better fit than the Kumaraswamy-Binomial distribution and Beta-Binomial distribution on both the simulation study and the real data set in handling binomial outcome data. In this paper we focus on testing homogeneity of proportions in presence the new McGBB distribution over-dispersion by deriving the $C(\alpha)$ tests using the Quasi-likelihood and the Extended Quasi-likelihood estimating functions. The performance of the derived $C(\alpha)$ tests are better when compared, through simulations, with the Likelihood ratio test.***

Keywords: Common Over-dispersion, Likelihood Ratio statistic, Simulation, Quasi-likelihood

1. Introduction

Data in form of proportions arise in Toxicology and other similar fields. An important problem is to compare proportion of a certain characteristic in several groups. However, these proportions often exhibit variation greater than predicted by a simple binomial model (Williams, 1975).The superiority of the McGBB distribution to BB distribution in handling over-dispersion has been shown (Chandbroseet *al*., 2013).A number of procedures are available for testing homogeneity of the proportions in presence of over-dispersion. Of these, the Likelihood ratio (LR) test has found prominence in literature.

The purpose of this paper is to derive $C(\alpha)$ (Neyman, 1959)statistics C_Q *and* C_{Q^+} based onQuasi-likelihood and Extended Quasi-likelihood estimates respectively,for testing homogeneity of the proportions in presence of McGBBover-dispersion. $C(\alpha)$ test is based on the residual of a regression of the score function for the parameter(s) of interest on the nuisance parameters. The nuisance parameters are then replaced by $\sqrt{n}$ consistent estimators.If the nuisance parameters are replaced by their maximum likelihood estimators (mle's), which are $\sqrt{n}$ consistent, the $C(\alpha)$ statistic reduces to the score statistic (Rao, 1948). The $C(\alpha)$ statistichas been widely used as a test statistics (Neyman and Scott, 1966; Moran, 1973; Paul, 1982; Tarone, 1985; Barnwal and Paul, 1988; Boos, 1992; Paul and Islam, 1992, 1994). This is because $C(\alpha)$ statistics require estimates under the null hypothesis, it often produces a statistic which is simple to calculate, it has been found useful for detecting over-dispersion in binomial and poisson data (Paul *et al.*, 1989; Dean and Lawless, 1990). It also often maintains at least approximately, a pre-assigned level of significance (Bartoo and Puri, 1967). It is locally asymptotically most powerful (Bühler and Puri, 1966; Moran, 1970).

The paper is organized as follows:In section 2 we present the derivation of the $C(\alpha)$ statistics, section 3 simulation and section 4 results and discussion for comparing the size and power of the $C(\alpha)$ statistics with the LR statistic.

2. The $C(\alpha)$ Test Statistics

Suppose that there are S treatment groups and that the i^{th} group has k_i litters. The proportion responding in the j-th litter of i-th group is $\frac{y_{ij}}{n_{ij}}$, $j = 1,\dots,k_i$; $i = 1,\dots,s$.

2.1 McDonald Generalized Beta-Binomial distribution

The probability mass function of the new McGBB (α, β, γ) distribution is given by,

$$P_{McGBB}(y;\alpha,\beta,\gamma) = \binom{n}{y}\frac{\gamma}{B(\alpha,\beta)}\sum_{j=0}^{n-y}(-1)^j\binom{n-y}{j}B\left(\frac{y}{\gamma}+\alpha+\frac{j}{\gamma},\beta\right) \quad (1)$$

where $y = 0,1,\dots,n$ and $\alpha,\beta,\gamma > 0$

Then the mean and variance of the new McGBB (n, α, β, γ) distribution are given by,

$E(Y) = n\pi$ and $\text{var}(Y) = n\pi(1-\pi)\{1+(n-1)\rho\}$, respectively where

$$\pi = \frac{B\left(\alpha+\beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)} \text{ and}$$

$$\rho = \frac{\left(\frac{B\left(\alpha+\beta, \frac{2}{\gamma}\right)}{B\left(\alpha, \frac{2}{\gamma}\right)}\right) - \left(\frac{B\left(\alpha+\beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}\right)^2}{\left(\frac{B\left(\alpha+\beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}\right) - \left(\frac{B\left(\alpha+\beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}\right)^2} \quad (2)$$

where ρ is the overdispersion parameter of the new McGBB distribution

2.2 The C(α) test statistic based on the quasi-likelihood(C_Q)

The Quasi-likelihood is based on the knowledge of the first two moments of the random variable $Z = \frac{y}{n}$ (Wedderburn, 1974).

$$E(Z) = \pi,\ \text{var}(Z) = \frac{\pi(1-\pi)}{n}\{1+(n-1)\phi\}, 0 \le \pi \le 1 \textit{ and } \left(\frac{-1}{n-1}\right) < \phi < 1$$

This specification of mean and variance coincides with those based on the new McGBB model. The Quasi-likelihood for an observation Z with the above mean and variance is given by $Q(Z,\pi,\phi) = \int_Z^{\pi} \frac{(Z-t)n}{t(1-t)\{1+(n-1)\phi\}} dt$ which for the data under consideration becomes

$$Q = \sum_{i=0}^{n} \frac{1}{\{1+(n-1)\phi\}}\left[y\log\left(\frac{\pi}{Z}\right) + (n-y)\log\left\{\frac{(1-\pi)}{(1-Z)}\right\}\right]$$

where $\pi = \frac{B\left(\alpha+\beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)}$ and $\hat{\phi} = \frac{\left(\frac{\text{var}(y)}{\bar{y}\left(1-\frac{\bar{y}}{n}\right)} - 1\right)}{(n-1)}$

(3)

where y is a discrete random variable, n is the number of trials, ϕ is the over-dispersion parameter.

Define $\lambda = (\lambda_1, \lambda_2, \lambda_3) = (\alpha, \beta, \gamma)$. Then let

$$\Psi_i = \frac{\partial Q}{\partial \alpha_i},\ i = 1, \ldots, S-1$$

and

$$\varphi_k = \frac{\partial Q}{\partial \lambda_k},\ k = 1,2,3$$

To make things simple we assume homogenous and under this assumption we wish to test the hypothesis $H_0 : \pi_1 = \ldots = \pi_S$ against H_1 : not all π_i's are the same. Now, let $\hat{\lambda}$ be some $\sqrt{m}$ consistent estimator of λ under the null hypothesis. Then the $C(\alpha)$ test is based on

$$S_i(\hat{\lambda}) = \Psi_i(\hat{\lambda}) - \beta_{1i}\varphi_1(\hat{\lambda}) - \beta_{2i}\varphi_2(\hat{\lambda}) - \beta_{3i}\varphi_3(\hat{\lambda}),\ i = 1, \ldots, S$$

where β_{1i}, β_{2i} and β_{3i} are the partial regression coefficients of Ψ_i and φ_1, Ψ_i and φ_2 and Ψ_i and φ_3 respectively. The variance-covariance matrix of $S(\lambda) = \{S_1(\lambda), \ldots, S_{s-1}(\lambda)\}'$ is $D - AB^{-1}A'$ and the regression coefficients $\beta = (\beta_1, \beta_2, \beta_3) = AB^{-1}$ where $\beta_1 = (\beta_{11}, \ldots, \beta_{1S-1})$, $\beta_2 = (\beta_{21}, \ldots, \beta_{2S-1})$, $\beta_3 = (\beta_{31}, \ldots, \beta_{3S-1})$.

$$D_{it} = E\left(\frac{-\partial Q}{\partial \lambda_i \partial \lambda_t}\right),\ i,t = 1, \ldots, S-1$$

$$A_{ik} = E\left(\frac{\partial Q}{\partial \lambda_i \partial \lambda_k}\right) \quad \begin{matrix} i = 1, \ldots, S-1 \\ k = 1,2,3 \end{matrix}$$

$$B_{ks} = E\left(\frac{-\partial Q}{\partial \lambda_k \partial \lambda_s}\right),\quad k,s = 1,2,3$$

Using $\hat{\lambda}$ in S, A, B and D, the $C(\alpha)$ test statistic is given by $S'(D - AB^{-1}A')^{-1} S$, which is approximately distributed as chi-square with $S-1$ degrees of freedom.

Using the Quasi log-likelihood (3)and taking partial derivatives, we obtain

$$\frac{\partial Q}{\partial \alpha} = \sum_{y=0}^{n} \frac{1}{\{1+(n-1)\phi\}}\left[\left(y - \frac{\pi(n-y)}{1-\pi}\right)\left(\psi(\alpha+\beta) + \psi\left(\alpha + \frac{1}{\gamma}\right) - \psi(\alpha) - \psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] (4)$$

$$\frac{\partial Q}{\partial \beta}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(y-\frac{\pi(n-y)}{1-\pi}\right)\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{5}$$

$$\frac{\partial Q}{\partial \gamma}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{y}{\gamma^{2}}-\frac{\pi(n-y)}{\gamma^{2}(1-\pi)}\right)\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\right] \tag{6}$$

Maximum quasi-likelihood estimates of β_j 's are obtained by equating (4), (5) and (6) to zero and solving simultaneously. Denote the estimates by $\hat{\lambda}_{QL}$. The second derivatives of Q are given below

$$\frac{\partial^2 Q}{\partial \alpha^2}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(y-\frac{\pi(n-y)}{1-\pi}\right)\left(\psi'(\alpha+\beta)+\psi'\left(\alpha+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right.$$
$$\left.-\left(\frac{(n-y)\pi}{(1-\pi)^{2}}\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^{2}\right] \tag{7}$$

$$\frac{\partial^2 Q}{\partial \beta^2}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(y-\frac{\pi(n-y)}{1-\pi}\right)\left(\psi'(\alpha+\beta)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)-\left(\frac{(n-y)\pi}{(1-\pi)^{2}}\right)\times\right.$$
$$\left.\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^{2}\right] \tag{8}$$

$$\frac{\partial^2 Q}{\partial \gamma^2}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{y}{\gamma^{4}}-\frac{\pi(n-y)}{\gamma^{4}(1-\pi)}\right)\left(\psi'\left(\alpha+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)-\left(\frac{(n-y)\pi}{\gamma^{4}(1-\pi)^{2}}\right)\times\right. \tag{9}$$
$$\left.\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^{2}-\left(\frac{2ny}{\gamma^{3}}-\frac{2\pi(n-y)}{\gamma^{3}(1-\pi)}\right)\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right]$$

$$\frac{\partial^2 Q}{\partial \alpha \partial \beta}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(y-\frac{\pi(n-y)}{1-\pi}\right)\left(\psi'(\alpha+\beta)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)-\left(\frac{\pi(n-y)}{(1-\pi)^{2}}\right)\times\right.$$
$$\left.\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{10}$$

$$\frac{\partial^2 Q}{\partial \alpha \partial \gamma}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{y}{\gamma^{2}}-\frac{\pi(n-y)}{\gamma^{2}(1-\pi)}\right)\left(\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)-\left(\frac{\pi(n-y)}{\gamma^{2}(1-\pi)^{2}}\right)\times\right.$$
$$\left.\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{11}$$

$$\frac{\partial^2 Q}{\partial \beta \partial \gamma}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{y}{\gamma^{2}}-\frac{\pi(n-y)}{\gamma^{2}(1-\pi)}\right)\left(\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\frac{1}{\gamma}\right)\right)-\left(\frac{\pi(n-y)}{\gamma^{2}(1-\pi)^{2}}\right)\times\right.$$
$$\left.\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{12}$$

where $\psi(.)$ *and* $\psi'(.)$ are digamma and trigamma functions respectively.

Expectations of the minus the second derivatives are given below,

$$D_{11} = B_{11} = A_{11} = \sum_{y=0}^{n} \frac{n}{\{1+(n-1)\phi\}} \left[\left(\frac{\pi}{1-\pi} \right) \left(\psi(\alpha+\beta) + \psi\left(\alpha+\frac{1}{\gamma}\right) - \psi(\alpha) - \psi\left(\alpha+\beta+\frac{1}{\gamma}\right) \right)^2 \right] \tag{13}$$

$$B_{22} = \sum_{y=0}^{n} \frac{n}{\{1+(n-1)\phi\}} \left[\left(\frac{\pi}{1-\pi} \right) \left(\psi(\alpha+\beta) - \psi\left(\alpha+\beta+\frac{1}{\gamma}\right) \right)^2 \right] \tag{14}$$

$$B_{33} = \sum_{y=0}^{n} \frac{n}{\{1+(n-1)\phi\}} \left[\left(\frac{\pi}{\gamma^4(1-\pi)} \right) \left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right) - \psi\left(\alpha+\frac{1}{\gamma}\right) \right)^2 \right] \tag{15}$$

$$\begin{aligned} A_{12} = B_{12} = \sum_{y=0}^{n} \frac{n}{\{1+(n-1)\phi\}} \Bigg[\left(\frac{\pi}{1-\pi} \right) \left(\psi(\alpha+\beta) - \psi\left(\alpha+\beta+\frac{1}{\gamma}\right) \right) \times \\ \left(\psi(\alpha+\beta) + \psi\left(\alpha+\frac{1}{\gamma}\right) - \psi(\alpha) - \psi\left(\alpha+\beta+\frac{1}{\gamma}\right) \right) \Bigg] \end{aligned} \tag{16}$$

$$\begin{aligned} A_{13} = B_{13} = \sum_{y=0}^{n} \frac{n}{\{1+(n-1)\phi\}} \Bigg[\left(\frac{\pi}{\gamma^2(1-\pi)} \right) \left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right) - \psi\left(\alpha+\frac{1}{\gamma}\right) \right) \times \\ \left(\psi(\alpha+\beta) + \psi\left(\alpha+\frac{1}{\gamma}\right) - \psi(\alpha) - \psi\left(\alpha+\beta+\frac{1}{\gamma}\right) \right) \Bigg] \end{aligned} \tag{17}$$

$$B_{23} = \sum_{y=0}^{n} \frac{n}{\{1+(n-1)\phi\}} \left[\left(\frac{\pi}{\gamma^2(1-\pi)} \right) \left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right) - \psi\left(\alpha+\frac{1}{\gamma}\right) \right) \left(\psi(\alpha+\beta) - \psi\left(\alpha+\beta+\frac{1}{\gamma}\right) \right) \right] \tag{18}$$

Denote the Quasi-likelihood estimates of $\lambda = (\alpha, \beta, \gamma)$ by $\hat{\lambda}_{QL}$. If $\hat{\lambda}_{QL}$ is used in S, A, B and D, which is $\sqrt{m}$ consistent estimates of λ under the null hypothesis, then $S(\hat{\lambda}_{QL}) = \Psi(\hat{\lambda}_{QL})$. Then the quasi-likelihood score or the $C(\alpha)$ statistic is

$$C_Q = \Psi'\left(D - AB^{-1}A'\right)^{-1} \Psi \tag{19}$$

2.3 The C(α) test statistic based on the Extended quasi-likelihood(C_{Q^+})

The extended quasi-likelihood(Nelder and Pregibon, 1987) can be used for the simultaneous estimation of the β_j and ϕ. The extended quasi-log-likelihood for an observation z with mean and variance specified is

$$Q^+(z,\pi,\phi) = -\frac{1}{2}\log(2k) - \frac{1}{2}\log\left[\frac{z(1-z)\{1-(n-1)\phi\}}{n} \right] + \int_z^{\pi} \frac{(z-t)n}{t(1-t)\{1+(n-1)\phi\}} dt \tag{20}$$

The Extended quasi-log-likelihood for the data under consideration, then, is

$$Q^+(z,\pi,\phi) = C - \frac{1}{2}\sum_{y=0}^{n} \left[\log\{1+(n-1)\phi\} + \frac{2}{\{1+(n-1)\phi\}} \left[y\log\left(\frac{\pi}{z}\right) + (n-y)\log\left(\frac{1-\pi}{1-z}\right) \right] \right] \tag{21}$$

where C is term not involving the parameters.Define $\lambda = (\lambda_1, \lambda_2, \lambda_3) = (\alpha, \beta, \gamma)$. Then let

$$\Psi_i = \frac{\partial Q^+}{\partial \alpha_i},\ i = 1,\dots,S-1 \text{ and } \varphi_k = \frac{\partial Q^+}{\partial \lambda_k},\ k = 1,2,3$$

We wish to test the hypothesis $H_0 : \pi_1 = \dots = \pi_S$ against H_1 : not all π_i's are the same.Now, let $\hat{\lambda}$ be some $\sqrt{m}$ consistent estimator of λ under the null hypothesis. Then the $C(\alpha)$ test is based on

$$S_i(\hat{\lambda}) = \Psi_i(\hat{\lambda}) - \beta_{1i}\varphi_1(\hat{\lambda}) - \beta_{2i}\varphi_2(\hat{\lambda}) - \beta_{3i}\varphi_3(\hat{\lambda}),\ i = 1,\dots,S-1$$

where β_{1i}, β_{2i} and β_{3i} are the partial regression coefficients of Ψ_i and φ_1, Ψ_i and φ_2 and Ψ_i and φ_3 respectively. The variance-covariance matrix of $S(\lambda)=\{S_1(\lambda),\ldots,S_{s-1}(\lambda)\}'$ is $D-AB^{-1}A'$ and the regression coefficients $\beta=(\beta_1,\beta_2,\beta_3)=AB^{-1}$ where $\beta_1=(\beta_{11},\ldots,\beta_{1S-1})$, $\beta_2=(\beta_{21},\ldots,\beta_{2S-1})$, $\beta_3=(\beta_{31},\ldots,\beta_{3S-1})$.

$$D_{it}=E\left(\frac{-\partial Q^+}{\partial\lambda_i\partial\lambda_t}\right),\ i,t=1,\ldots,S-1$$

$$A_{ik}=E\left(\frac{\partial Q^+}{\partial\lambda_i\partial\lambda_k}\right)\quad \begin{array}{l} i=1,\ldots,S-1\\ k=1,2,3\end{array}$$

$$B_{ks}=E\left(\frac{-\partial Q^+}{\partial\lambda_k\partial\lambda_s}\right),\quad k,s=1,2,3$$

Using $\hat{\lambda}$ in S, A, B and D, the $C(\alpha)$ test statistic is given by $S'\left(D-AB^{-1}A'\right)^{-1}S$, which is approximately distributed as chi-square with $S-1$ degrees of freedom.

The unbiased estimating equations for β_j obtained from Q^+ are

$$\frac{\partial Q^+}{\partial\alpha}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{1-\pi}-y\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right]=0 \tag{22}$$

$$\frac{\partial Q^+}{\partial\beta}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{1-\pi}-y\right)\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right]=0 \tag{23}$$

$$\frac{\partial Q^+}{\partial\gamma}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{\gamma^2(1-\pi)}-\frac{y}{\gamma^2}\right)\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\right]=0 \tag{24}$$

Maximum extended quasi-likelihood estimates of β_j's are obtained by solving (22), (23) and (24) simultaneously. Denote the estimates by $\hat{\lambda}_{EQL}$ The second derivatives of Q^+ are given below

$$\frac{\partial^2 Q^+}{\partial\alpha^2}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{1-\pi}-y\right)\left(\psi'(\alpha+\beta)+\psi'\left(\alpha+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right.$$
$$\left.+\left(\frac{(n-y)\pi}{(1-\pi)^2}\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^2\right] \tag{25}$$

$$\frac{\partial^2 Q^+}{\partial\beta^2}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{1-\pi}-y\right)\left(\psi'(\alpha+\beta)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)+\left(\frac{(n-y)\pi}{(1-\pi)^2}\right)\times\right.$$
$$\left.\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^2\right] \tag{26}$$

$$\frac{\partial^2 Q^+}{\partial\gamma^2}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{\gamma^4(1-\pi)}-\frac{y}{\gamma^4}\right)\left(\psi'\left(\alpha+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)-\left(\frac{(n-y)\pi}{\gamma^4(1-\pi)^2}\right)\times\right.$$
$$\left.\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^2+\left(\frac{2ny}{\gamma^3}-\frac{2\pi(n-y)}{\gamma^3(1-\pi)}\right)\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{27}$$

$$\frac{\partial^2 Q^+}{\partial\alpha\partial\beta}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{1-\pi}-y\right)\left(\psi'(\alpha+\beta)-\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)+\left(\frac{\pi(n-y)}{(1-\pi)^2}\right)\times\right.$$
$$\left.\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{28}$$

$$\frac{\partial^2 Q^+}{\partial\alpha\partial\gamma}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{\gamma^2(1-\pi)}-\frac{y}{\gamma^2}\right)\left(\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)+\left(\frac{\pi(n-y)}{\gamma^2(1-\pi)^2}\right)\times\right.$$
$$\left.\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{29}$$

$$\frac{\partial^2 Q^+}{\partial\beta\partial\gamma}=\sum_{y=0}^{n}\frac{1}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi(n-y)}{\gamma^2(1-\pi)}-\frac{y}{\gamma^2}\right)\left(\psi'\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi'\left(\alpha+\frac{1}{\gamma}\right)\right)+\left(\frac{\pi(n-y)}{\gamma^2(1-\pi)^2}\right)\times\right.$$
$$\left.\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{30}$$

Expectations of the minus the second derivatives are given below,

$$D_{11}=A_{11}=B_{11}=\sum_{y=0}^{n}\frac{-n}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi}{1-\pi}\right)\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^2\right] \tag{31}$$

$$B_{22}=\sum_{y=0}^{n}\frac{-n}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi}{1-\pi}\right)\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)^2\right] \tag{32}$$

$$B_{33}=\sum_{y=0}^{n}\frac{-n}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi}{\gamma^4(1-\pi)}\right)\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)^2\right] \tag{33}$$

$$A_{12}=B_{12}=\sum_{y=0}^{n}\frac{-n}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi}{1-\pi}\right)\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\times\right.$$
$$\left.\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{34}$$

$$A_{13}=B_{13}=\sum_{y=0}^{n}\frac{-n}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi}{\gamma^2(1-\pi)}\right)\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\times\right.$$
$$\left.\left(\psi(\alpha+\beta)+\psi\left(\alpha+\frac{1}{\gamma}\right)-\psi(\alpha)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{35}$$

$$B_{23}=\sum_{y=0}^{n}\frac{-n}{\{1+(n-1)\phi\}}\left[\left(\frac{\pi}{\gamma^2(1-\pi)}\right)\left(\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)-\psi\left(\alpha+\frac{1}{\gamma}\right)\right)\left(\psi(\alpha+\beta)-\psi\left(\alpha+\beta+\frac{1}{\gamma}\right)\right)\right] \tag{36}$$

Denote the Extended Quasi-likelihood estimates of $\lambda=(\alpha,\beta,\gamma)$ by $\hat{\lambda}_{EQL}$. If $\hat{\lambda}_{EQL}$ is used in S, A, B and D, which is $\sqrt{m}$ consistent estimates of λ under the null hypothesis, then $S(\hat{\lambda}_{EQL})=\Psi(\hat{\lambda}_{EQL})$. Then the quasi-likelihood score or the $C(\alpha)$ statistic is

$$C_{Q^+}=\Psi'\left(D-AB^{-1}A'\right)^{-1}\Psi \tag{37}$$

Note that under the null hypothesis the parameters α,β and γ are common across groups. In this study we consider $S=2$ groups. So the estimation of α,β and γ from the S groups can be considered to be estimation from a single group consisting of the combined data in the S groups. The null hypothesis is rejected if C_Q or $C_{Q^+}>\chi^2_{\alpha,p}$. where p is the degrees of freedom.

3. Simulation

In this section we report on a simulation study conducted to compare the performance, in terms of size and power, of likelihood ratio statistic (LR) and $C(\alpha)$ statistics.The simulated data was generated based on the new McGBB distribution. The 1000 over-dispersed data set were simulated using the algorithm developed (Ahn and Chen, 1995). The open source statistical software R (version 3.1.1) was used in the study to simulate data. In the simulation study, empirical levels were calculated based on 1000 replications for each combination of varying valuesof

$\alpha_1 = \alpha_2$ =0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95 and 1.00 and for values $\beta_1 = \beta_2 = \beta = 0.30$ and $\gamma_1 = \gamma_2 = \gamma = 1$ parameters were chosen. For power, we considered varying values of α_2 =0.22, 0.24, 0.26, 0.28, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.70, 0.71, 0.72, 0.73, 0.74 and 0.75. For each value of α_2 empirical powers were calculated for $\beta_1 = \beta_2 = \beta = 0.70$ and $\gamma_1 = \gamma_2 = \gamma = 1$.

4. Results and Discussion

4.1 Results

Table 1: Empirical levels; $\alpha = 0.05$; based on 1000 simulated data sets for $\beta_1 = \beta_2 = \beta = 0.30$, $\gamma_1 = \gamma_2 = \gamma = 1$ and $\alpha_1 = \alpha_2$ varied.

	Estimated Empirical levels		
$\alpha_1 = \alpha_2$ varied	LR Test	C_{Q^+} Test	C_Q Test
0.10	0.044	0.046	0.048
0.15	0.046	0.044	0.046
0.20	0.042	0.042	0.046
0.25	0.029	0.048	0.048
0.30	0.038	0.045	0.041
0.35	0.027	0.039	0.044
0.40	0.022	0.048	0.049
0.45	0.023	0.046	0.050
0.50	0.038	0.057	0.052
0.55	0.099	0.043	0.049
0.60	0.129	0.046	0.049
0.65	0.085	0.054	0.045
0.70	0.101	0.038	0.044
0.75	0.076	0.047	0.050
0.80	0.058	0.049	0.049
0.85	0.073	0.052	0.050
0.90	0.058	0.042	0.041
0.95	0.051	0.056	0.049
1.00	0.065	0.054	0.048

(a) Empirical Levels of C(QL) test, C(EQL) test and LR test based on 1000 simulated data sets

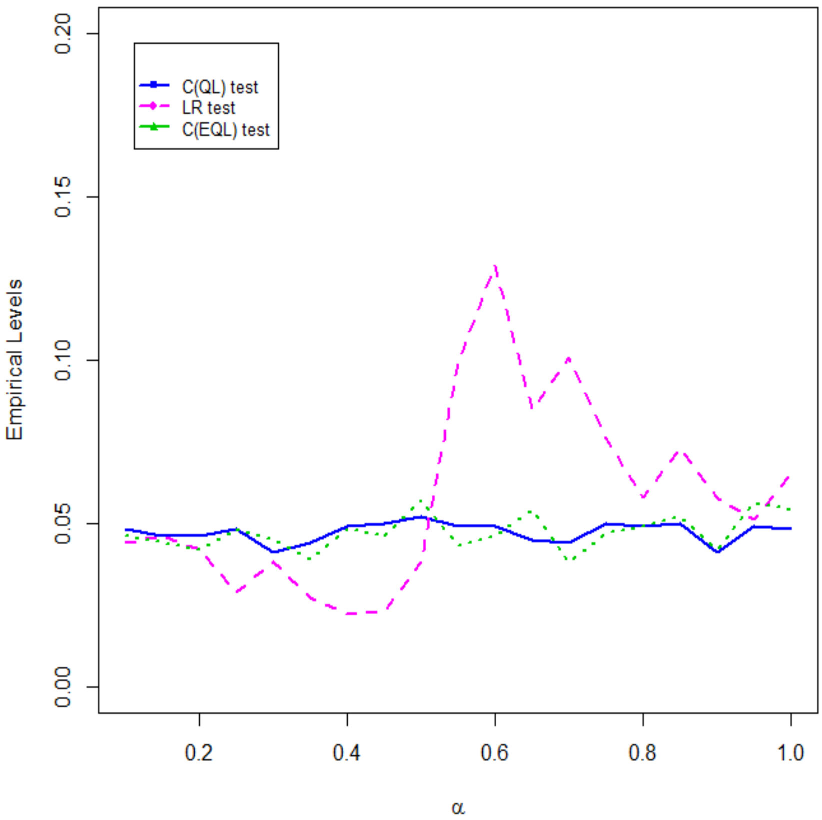

Figure 1: Plot of empirical level comparison for C_Q test, C_{Q^+} test and LR testunder the McGBB model for varied $\alpha_1 = \alpha_2$ and for values of $\beta_1 = \beta_2 = \beta = 0.30$ and $\gamma_1 = \gamma_2 = \gamma = 1$ for all procedures.

Table 2: Empirical Power; $\alpha = 0.05$; based on 1000 simulated data sets for $\alpha_1 = 0.20$ $\beta_1 = \beta_2 = \beta = 0.70$, $\gamma_1 = \gamma_2 = \gamma = 1$ and α_2 varied

Estimated Empirical Power

α_2 varied	LR Test	C_{Q^*} Test	C_Q Test
0.22	0.058	0.156	0.165
0.24	0.071	0.189	0.247
0.26	0.108	0.231	0.340
0.28	0.125	0.285	0.381
0.30	0.147	0.341	0.452
0.35	0.270	0.505	0.554
0.40	0.350	0.581	0.672
0.45	0.522	0.662	0.748
0.50	0.644	0.716	0.803
0.55	0.698	0.775	0.856
0.60	0.857	0.824	0.886
0.61	0.879	0.919	0.895
0.62	0.910	0.932	0.966
0.63	0.925	0.952	0.974
0.64	0.942	0.953	0.975
0.65	0.956	0.964	0.975
0.66	0.968	0.967	0.977
0.67	0.970	0.972	0.979
0.68	0.981	0.977	0.986
0.69	0.982	0.983	0.989
0.70	0.987	0.992	0.992
0.71	0.989	0.992	0.993
0.72	0.990	0.994	0.995
0.73	0.992	0.998	0.996
0.74	0.996	0.998	0.998
0.75	0.998	0.999	0.998

(b) Empirical powers of C(QL) test, C(EQL) test and LR test based on 1000 simulated data sets

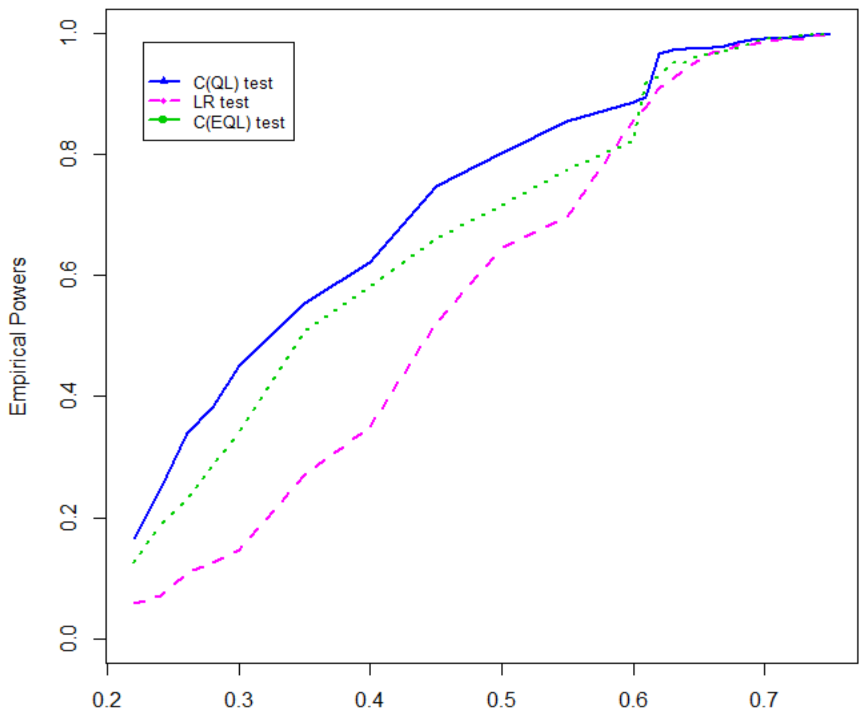

Figure 2: Plot of empirical power comparison for C_Q test, C_{Q^+} test and LR test under McGBB model for varied α_2 and for values of $\alpha_1 = 0.20$, $\beta_1 = \beta_2 = \beta = 0.70$ and $\gamma_1 = \gamma_2 = \gamma = 1$ for all procedures.

4.2 Discussion

The results given in table 1 and figure 1 shows that, for all varying values of $\alpha_1 = \alpha_2$ the $C(\alpha)$ test statistics C_Q and C_{Q^+} shows conservative behavior. For small $\alpha_1 = \alpha_2$ ($\alpha_1 = \alpha_2 =$0.10, 0.15 and 0.20)the LR and the $C(\alpha)$ tests shows some conservative behavior, otherwise all the statistics produce consistent empirical levels close to the nominal level.At $\alpha_1 = \alpha_2$ ($\alpha_1 = \alpha_2 =$0.55, 0.60, 0.65, 0.70 and 0.75), the LR test shows liberal behavior and produce empirical levels that are far away to the nominal level hence not consistent. The $C(\alpha)$ test C_Q and C_{Q^+} produce empirical levels very close to the nominal level while LR test shows liberal behavior hence the $C(\alpha)$ tests are preferred since it shows consistency on all varying values of $\alpha_1 = \alpha_2$. The performanceof the $C(\alpha)$ test is better in that it holds nominal level quite well and also has a simple form.

The results given in table 2 and figure 2.For α_2 ($\alpha_2 =$0.22, 0.24, 0.26, 0.28, 0.30, 0.35, 0.40, 0.45 and 0.50), the power of the LR test is to some extent smaller than those of the other two statistics C_Q and C_{Q^+}. $C(\alpha)$ tests shows a higher power consistency than LR test hence its a better test statistics preferable, as they require estimates of the parameters only under the null hypothesis. C_Q isconsistent for all varying values of α_2 with the highest empirical power.Maximum likelihood estimates (mle's) of the parameters under the null and alternative hypothesis were obtained by maximizing log-likelihood of McGBB distribution (1) using the R-language subroutine. The quasi-likelihood and Extended quasi-likelihood estimates of the parameters under the null hypothesis were obtained by maximizing the Quasi log-likelihood (3) and Extended Quasi log-likelihood (21)using the R-language subroutine

4.3 Comparison and Conclusion of the $C(\alpha)$ statistics and LR test statistics

Performance evaluation measures empirical level (size) and the empirical power were obtained for the simulated data. The comparison of the test statistics based on empirical level and empirical power is as given in table 1 and 2 respectively. Based on the results from the tables and figure 1 and 2, $C(\alpha)$ tests perform better than LR test since they are consistent and holds nominal level quite well and have higher power. The C_Q test is the best since it shows the highest conservative behavior and the highest empirical power.

References

[1] Ahn, H., and Chen, J. J. (1995).Generation of over-dispersed and under-dispersed binomial variates.*Journal of Computational and Graphical Statistics*, **4**(1), 55-64.

[2] Bartoo, J.B., and Puri, P.S. (1967).On optimal asympotic tests of composite statistical hypothesis.*Ann. Math, Statist.*,**38**,1845-1852.

[3] Boos, D.D (1992). On the generalized score tests.*The American Statistician*, **46**,327-333.

[4] Bŏhlur, W. J., and Puri, P.S (1966).On optimal asymptotic tests of composite hypothesis with several constraints. Z. *Wahrsch*, **5**,71-88.

[5] Chandrabose, M., Pushpa, W., and Roshan D. (2013). The McDonald Generalized Beta- Binomial distribution: A New Binomial Mixture Distribution and simulation based comparison with its nested distributions in handling overdispersion. *International journal of statistics and probability*; **vol. 2**,No. 2;2013.

[6] Dean, C., and Lawless, J.F. (1990). Tests for detecting over-dispersion in poisson regression models.*J. Am. Statist. Assoc.*, **84**,467-472.

[7] Moran, P.A.P. (1970). On asymptotically optimal tests of composite hypotheses tests. *Biometrika*,**57**,45- 75.

[8] Moran, P.A.P. (1973).Asymptotic properties of homogeneity tests.*Biometrika*,**60**,79-85.

[9] Nelder, J.A. and Pregion, D.(1987). An extended quasi-likelihood function.

[10] *Biometrika*, 74, 221-232.

[11] Neyman, J. (1959). Optimal asymptotic test for composite hypotheses.In probability and statistics, ed. U. Grenander, 213-234.The Harold Cramer Volume. Uppsala: Almgrist Wiksell.

[12] Neyman, J., and Scott, E.L. (1966). On the use of C(α) optimal tests of composite hypotheses. *Bull. Int. Statist. Inst.*, **41**, 477-497.

[13] Paul, S. R. (1982). Analysis of proportions of affected foetuses in teratological experiments. *Biometrics*, **38**(2),361-370.

[14] Paul, S. R., and Islam, A. S. (1992).C(α) tests for homogeneity of proportions in taxocology in presence of Beta-Binomial over-dispersion. Accepted in David Williams Volume: *Statistics in Toxicology*.Oxford University Press.

[15] Islam, A. S.(1994). Parametric and semi-parametric models for the analysis of proportions in presence of over/under dispersion.PhD thesis, University of Windsor

[16] Paul, S. R., Liang, K. Y., and Self, S. A (1989).On testing departure from the binomial and multinomial assumptions.*Biometrics*, **45**,231-236.

[17] Rao, C.R.(1948). Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proc. Camb. Phil. Soc.*, 44, 50-57

[18] Tarone, R.E. (1985). On heterogeneity tests based on efficient scores.*Biometrika*, **72**,91-95.

[19] Wedderburn, R. W. M. (1974).Quasi-likelihood function, generalized linear models and the gauss-newton method. *Biometrika*, **61**, 439-447.

[20] Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity.*Biometrics*, 31, 949-952