Chilean Journal of Statistics
Vol. 5, No. 1, April 2014, 73–85

Generalized Linear Models
Research Paper

Goodness-of-fit Tests for Modified Multinomial Logit Models

Marcelo Angelo Cirillo[1,*] and Patrícia de Siqueira Ramos[2]

[1]Exact Sciences Department, Federal University of Lavras, Brazil,
[2]Exact Sciences Institute, Federal University of Alfenas, Brazil.

(Accepted in final form: 20 December 2010)

Abstract

Since the performance of Pearson's χ^2 and deviance tests typically used to evaluate goodness of fit of multinomial models depends on sample size and number of categories, the resulting p-values may become distorted. Having that fact as a basis, this article explored a modification in the construction of the above cited tests by replacing the estimates of maximum likelihood with the introduction of a posterior mean. The performance of the modified tests was evaluated in comparison with the results of conventional tests obtained by Monte Carlo simulation using original specifications. Due to the conservative results, we concluded that the modification made by the inclusion of prior information Beta$(5, 5)$ in building the deviance test resulted in a promising test with satisfactory power values. The results of the modified Pearson's χ^2 test showed that, for some evaluated cases, the type I error values were not consistent with the specified nominal level, suggesting that the conventional form of this test is more appropriate to assess multinomial logit models goodness-of-fit.

Keywords: Correlated binomial · Deviance · False discovery rate · Monte Carlo simulation · Overdispersion · q-value.

Mathematics Subject Classification: Primary 62J12 · Secondary 62J15.

1. Introduction

The estimation of parameters in a multinomial logit model considers structures called contingency tables for their formalization (Jhun and Jeong, 2000). For the two-dimensional case, J categories (or classes) are represented in the vertical direction, while the response frequencies "success" and "failure" encoded by $i = 1, 2$, are described in the horizontal direction. Under this formalization, each cell is interpreted as a value given by y_{ij}, for $i = 1, 2$ and $j = 1, \ldots, J$. It is appropriate to evaluate multinomial logit model goodness-of-fit irrespective of the application and, for this purpose, the deviance and Pearson's χ^2 tests are used.

According to Dobson (2001), these tests are constructed considering the predicted probability distance of the proposed model in relation to the observed probability. However,

[*]Corresponding author. Email: macufla@ufla.br

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

these tests are sensitive to an overdispersion effect, which occurs when the sample variance exceeds the nominal variance assumed by the model. Bogutchi et al. (2006) explained that this effect results in incorrect standard deviations that become underestimated. Taking into account these evidences, Hinde and Demétrio (1998) claim that the predictions are inaccurate.

Several authors suggested alternative methods toreduce this effect. For example, we can mention Efron's (1986) point of view that considers a double family exponential, Smith (1989) who treats the generalized linear models as dispersion covariates, and Smith and Verbyla (1999) who consider an additional regression model to the dispersion parameter and also incorporate this parameter. Moreover, the authors show that the dispersion sub model is a gamma generalized linear model.

In the case of genuinely Bayesian tests involving multinomial distributions, Petri (2007) studied the relationship between the frequentist and Bayesian levels of significance evaluated by measures of evidence defined, respectively, by p-value and e-value, which is originated from the Full Bayesian Significance Test (FBST), presented by Pereira and Stern (1999). The authors concluded through simulation studies that the association between these measures is almost one to one. However, given this equivalence, the advantage of using the Bayesian alternative to significance (e-value) without major philosophical examination is considered only related to knowledge of the posterior distribution without the need to know the total sample space.

The motivation for this paper was inspired by the studies of Agresti and Min (2005), which deal with information from various articles (Good, 1956; Altham, 1969) discussing Bayesian inference for contingency tables. It is important to note that most of the articles cited deal with point estimation or significance testing and connections between Bayesian and frequentist results. We emphasize that the modification of the statistics of the deviance and chi-squared tests in this research was motivated by the fact that conventional tests Deviance e Pearson's χ^2 are founded on frequentist arguments. Thus, the incorporation of aprior information of the researcher in the evaluation of goodness-of-fit of multinomial models is made in order to reduce the uncertainty in the selection of models, since different models supposedly could be compared and adjusted. Thus, more information in the model selection can be added in the construction of these tests.

In summary, we propose an enhancement of deviance and Pearson's χ^2 tests using a Bayesian argument so that estimates of the proportions a replaced by estimates from the posterior distribution. Thus, by imposing a prior distribution the researcher has flexibility in adding information, e.g. the effects of asymmetry on the evaluation of goodness of fit for multinomial models. Thus, the interest in carrying out this study, by empirical simulation using the Monte Carlo method, is to verify if the inclusion of a Bayesian argument in the construction of each test can improve the control of type I errors and resulting power values.

This paper is organized as follows. In Section 2 we describe the statistical modeling and inference including a simulation study for the multinomial model, proposed test procedures, and the application of FDR criteria for determining an overall nominal level. In Section 3 we discuss the results of this study. Finally, in Section 4, some concluding remarks are given.

2. Statistical Modeling and Inference

This section is organized as follows. In Section 2.1 the multinomial model and the distributions used in the simulation are presented. In Section 2.2 a modification in the deviance and Pearson's χ^2 tests which consider the posterior mean of the binomial model is presented. In Section 2.3 the authors present an application of FDR (False Discovery Rate) for the purpose of detecting the occurrence of false positives, interpreted as the proportion of errors due to erroneous rejection of H_0 true, the deviance and Pearson's χ^2.

2.1 Multinomial model simulation

The multinomial model simulation used for this work considered J categories and the random variables vector $\boldsymbol{Y} = (Y_1, \ldots, Y_J)$, where each component represented the number of occurrences in the jth category, for $j = 1, \ldots, J$, associated with the vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$. Thus, Y_j follows a binomial correlated distribution with parameters n_j, π_j and ρ, denoted by $CB(n_j, \pi_j, \rho)$. It should be noted that each observation y_j from jth category, for $j = 1, \ldots, J$, was generated by a correlated binomial model developed by Luceño (1995). The probability distribution of $CB(n_j, \pi_j, \rho)$ is a mixture of two discrete distributions, the random variable Y_j stands for this mixture, which has a binomial distribution $B(n_j, \pi_j)$, with probability $(1 - \rho)$, and a modified Bernoulli represented by BernM(π_j) variable, assuming 0 or n_j values, with probability ρ (Fu and Sproule, 1995). The probability distribution of Y_j, given n_j, π_j and ρ, is:

$$\begin{aligned} P(Y_j | n_j, \pi_j, \rho) = & \binom{n_j}{y_j} \pi_j^{y_j} (1 - \pi_j)^{n_j - y_j} (1 - \rho) I_{A_1}(y_j) \\ & + \pi_j^{y_j/n_j} (1 - \pi_j)^{(n_j - y_j)/n_j} \rho I_{A_2}(y_j), \end{aligned} \tag{1}$$

where $A_1 = 0, 1, \ldots, n_j$, $A_2 = 0$, $n_j, y_j = 0, \ldots, n_j$ and $0 \leq \rho \leq 1$.

This can be verified using the probability generation function developed by Tallis (1962). The expectation and variance of Y_i are, respectively, $E(Y_j) = n_j \pi_j$ and $Var(Y_j) = \pi_j(1 - \pi_j)\{n_j + \rho n_j(n_j - 1)\}$, which means that for $\rho \neq 0$, the model includes extra-binomial variations. If ρ tends to 1, it creates an excess of n_j or zeros on the observed data. More details can be found in Tallis (1962). It should be noted that the $CB(n_j, \pi_j, \rho)$ model is equivalent to ordinary binomial model when $\rho = 0$.

In both situations the value of type I error and the power of the tests, which will be described in more detail in the next subsection, were computed. These values were derived from the proportion of times that these tests showed significance, given a 0.05 fixed nominal value, in a total of $10,000$ experiments simulated by the Monte Carlo method. The parametric values used are presented in Table 1. The Monte Carlo simulation algorithms was developed using software R, which is available at `www.R-project.org`; see R Development Core Team (2009).

Table 1. Parameters of the binomial distribution used in the Monte Carlo simulations.

J	$\pi_j (j = 1, \ldots, J)$	n_j	N
3	$(0.33, 0.33, 0.34)$	$(20; 20; 20)$	60
		$(80; 80; 80)$	240
5	$(0.20; 0.20; 0.20; 0.20; 0.20)$	$(20; \ldots; 20)$	100
		$(80; \ldots; 80)$	400
7	$(0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.10)$	$(20; \ldots; 20)$	140
		$(80; \ldots; 80)$	560

2.2 Proposed tests procedures

The null hypothesis evaluated by deviance and Pearson's χ^2 tests was described by

$$H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}^*, \tag{2}$$

where $\boldsymbol{\pi}^* = (\pi_1, \ldots, \pi_J)$ is the vector of probabilities estimated by (5). Possibly the most commonly used deviance and Pearson's χ^2 statistics were computed, respectively, according to

$$D = 2\sum_{j=1}^{J} \left\{ y_j \log\left(\frac{\tilde{\pi}_j}{\hat{\pi}_j}\right) + (n_j - y_j) \log\left(\frac{1-\tilde{\pi}_j}{1-\hat{\pi}_j}\right) \right\}, \tag{3}$$

$$Q = \sum_{j=1}^{J} n_j \frac{(\tilde{\pi}_j - \hat{\pi}_j)^2}{\hat{\pi}_j(1-\hat{\pi}_j)}, \tag{4}$$

where $\tilde{\pi}_j = \frac{y_j}{n_j}$ is the ML estimator of each component of π_j, for $j = 1, \ldots, J$, and $\hat{\pi}_j$ indicated the adjusted probability resulting from the logit model:

$$\hat{\pi}_j = \frac{e^{\hat{\beta}_0 + \hat{\beta}_j}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_j}}, j = 1, \ldots, J. \tag{5}$$

The linear term $\hat{\beta}_0 + \hat{\beta}_j$ was obtained through the logit transformation Dobson (2001) given by

$$\mathit{logit}\, \pi_j = \log\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_j, j = 1, \ldots, J, \tag{6}$$

where β_0 is the logit of the reference group, and β_j measures the difference in logits between level j of the category and the reference level.

The proposed modification in (3) and (4) statistics was made using Bayesian estimations instead of $\widetilde{\pi}_j$.The inclusion of a Bayesian argument with different informative priors aims to study the approximation of p-values obtained and modified in relation to the usual tests. Thus, by using posterior expectation instead of the maximum likelihood estimator (MLE) it will be possible to assess whether the tests mentioned in (3) and (4) showed some improvement in verification of multinomial models goodness-of-fit. To do this, we set the likelihood (7) as the function π_j, considering $\rho = 0$ for the equation (1),

$$p(y_j|\pi_j) \propto \pi_j^{\alpha}(1-\pi_j)^{\beta}. \tag{7}$$

Thus, if the prior density (8) is of the same form, with its own values α and β,

$$p(\pi_j) \propto \pi_j^{\alpha-1}(1-\pi_j)^{\beta-1}, \tag{8}$$

which is a beta distribution with parameters α and β: $\pi_j \sim \text{Beta}(\alpha, \beta)$. This prioris chosen because it is conjugate, thus the model becomes more flexible according to the researcher's interaction with the data, as inclusion of an non-informative prior can be obtained from a conjugate prior (Gammerman and Migon, 1993) by specifying the hyperparameter scale tending to zero and keeping the others constant. However, it should be emphasized that the main interest lies in posterior distribution, and since it is generally proper even when the prior distribution is improper, possible impropriety of prior distributions is not important.

The posterior density for π_j is (Gelman et al., 2004)

$$\begin{aligned} p(\pi_j|y_j) \propto & \pi_j^{y_j}(1-\pi_j)^{n_j-y_j}\pi_j^{\alpha-1}(1-\pi_j)^{\beta-1} = \\ & = \pi_j^{y_j+\alpha-1}(1-\pi_j)^{n_j-y_j+\beta-1} = \\ & = \text{Beta}(\pi_j|\alpha+y_j;\beta+n_j-y_j). \end{aligned}$$

Thus, the posterior mean represented the Bayesian estimator for π_j, for $j = 1, \ldots, J$, according to the expression (9), defining $\alpha^* = \alpha + y_j$ and $\beta^* = \beta + n_j - y_j$, is

$$E\left(\pi_j|y_j\right) = \frac{\alpha^*}{\alpha^* + \beta^*} = \frac{y_j + \alpha}{n_j + (\alpha + \beta)}. \tag{9}$$

The posterior variance is

$$Var\left(\pi_j|y_j\right) = \frac{(\alpha^* + y_j)(\beta^* + n_j - y_j)}{(\alpha^* + \beta^* + n_j)^2(\alpha^* + \beta^* + n_j + 1)} = \frac{E\left(\pi_j|y_j\right)\left[1 - E\left(\pi_j|y_j\right)\right]}{\alpha^* + \beta^* + n - 1}. \tag{10}$$

The relation between the sample proportion y_j/n_j (MLE estimator) and $E(\pi_j|y_j)$ is verified if y_j and $n_j - y_j$ become large with fixed α and β. Under these conditions:

$$E(\pi_j|y_j) \approx \frac{y_j}{n_j}, \tag{11}$$

$$Var\left(\pi_j|y_j\right) = \frac{1}{n_j}\frac{y_j}{n_j}\left(1 - \frac{y_j}{n_j}\right), \tag{12}$$

which approaches zero at rate $1/n_j$. According to Gelman et al. (2004), clearly, in the limit, the parameters of the prior distribution have no influence on the posterior distribution. By the above, if we assume the $\alpha = 0$ and $\beta = 0$ the approximation described in (11) resulted in the MLE estimator.

Thus, the values of the parameters that describe the posterior distribution assumed for the pair (α, β) were $(2, 8)$, $(5, 5)$, and $(8, 2)$ and were defined only for characterizing prior distributions with different degrees of symmetry. We assumed $\alpha = 0$ and $\beta = 0$, y_j/n_j is the maximum likelihood (ML) estimator of π_j to allow comparison between, on one hand, deviance and Pearson's χ^2 modified tests, and on the other hand, the conventional tests.

In terms of comparison, the performed tests were analyzed by the empirical probabilities concerning type I errors and power rates obtained via the Monte Carlo simulation and compared with the results of the original tests, where the values $(0, 0)$ were assumed for the prior information. As such, this paper will discuss the following results: type I error values and power. We emphasize that the choice of these prior distributions is justified by the fact that the inclusion of the Bayesian argument in the construction of Deviance and Pearson's χ^2 tests is very significant information to the researcher. Thus, the use of non-informative prior distributions has not been applied since, in general, these prior distributions are used in situations where data information is predominant when compared with the vague knowledge of the researcher.

2.3 Application of FDR (False Discovery Rate) criteria for determining an overall nominal level: A Monte Carlo Study

Considering the problem of detecting the occurrence of false positives, interpreted as the proportion of errors due to erroneous rejection of H_0 true, the deviance (3) and Pearson's χ^2 (4) tests were applied and their p-values computed, given a set of $m = 10,000$ null and independent hypotheses, for each sample generated under the hypothesis H_0 (2). At the end of the Monte Carlo simulation, it was possible to obtain the empirical distribution of p-values for each test.

The FDR criteria (false discovery rate) were used to determine an adjusted p-value, ie, a new measure of evidence defined as q-value (q) that measures the lowest false discovery rate, according to the methodology of Benjamini and Hochberg (1995). Therefore, the application of the FDR criteria was made assuming a set of $m = 10,000$ hypotheses H_0 and computing m_0 as the frequency of H_i, for $i = 1, \ldots, 10,000$, that were considered true. For each hypothesis H_i, the tests (3) and (4) were applied and the corresponding p-values were obtained.

With these specifications, we have assumed R as the number of rejected hypotheses, V as the number of true null hypotheses that were rejected and S as the number of false hypotheses rejected. Thus, FDR was defined as $E[V/R] = F$ if R represents a number other than zero and greater than V. In the situation of $F = 0$, the error of rejecting true null hypotheses is not committed.

The determination of the cut-off that controls the $FDR_{\delta\%} = n_c$ equivalent to an overall nominal level and interpreted as the value to reject all the hypotheses H_0 with p-values less than or equal to n_c were obtained according to the procedure described in the following steps:

(i) 1 - For each hypothesis H_{01}, H_{02}, ..., $H_{010,000}$ the corresponding value P_i, for $i = 1, \ldots, 10,000$, was obtained and put in a crescent order, that is, $P_{(1)}$, $P_{(2)}$, ..., $P_{(10.000)}$.

(ii) 2 - $q = (m \times P_i)/i$ was defined by determining the largest i for which FDR is controlled at a level q^* according to the relation $q* \geq (mP_i)/i$ and interpreted as the q-value cut-off.

3. Results and Discussion

According to the proposed objectives, the results presented below were obtained using the Monte Carlo method in order to investigate whether the deviance and chi-square tests, modified by replacing the MLE estimate for the posterior expectation (9) showed an improvement in the control of type I error and power.

The simulation studies were used to obtain empirical results that contribute to the evaluation of the proposed tests for two main reasons: (1) the tests listed in section 2 are asymptotically distributed as a χ^2 distribution and (2) obtaining analytical results through the construction of frequentist statistical tests with the incorporation of Bayesian arguments is complex. Thus, the correlated binomial distribution (section 2.1) was only a manner for generating samples with overdispersion effect in the simulation process, since one of the factors which cause this effect is associated with correlated responses.

Having been provided with this basic information, assuming different values for $\rho > 0$ $(0.20; 0.50$ and $0.70)$, it was possible to measure the overdispersion level present in the generated samples and consequently a favorable environment for evaluating the power of the tests was obtained. Assuming $\rho = 0$, samples are generated considering the binomial distribution (Luceño, 1995). Thus, it becomes possible to evaluate rejection rate sunder the null hypothesis defined in H_0 by using the correlated binomial model as a sample generating mechanism.

The functions implemented in the R software to generate contingency tables are described in the Appendix, keeping the procedure described in Section 2.1 with an illustrative example of the algorithms execution.

The results presented below refer to the performance of the statistics used in the deviance and Pearson's χ^2 tests to evaluate the multinomial logit model for goodness-of-fit. Considering that some prior information was incorporated to construct these tests, specifically assuming a Beta distribution (α, β) for the observed probability $(\widetilde{\pi})$ with the settings Beta$(2, 8)$, Beta$(5, 5)$ and Beta$(8, 2)$, the tests will now be called "modified deviance and Pearson's χ^2 tests", due to the inclusion of a posterior expectation.

Regarding the results of the value of type I error (Table 2), in the case of deviance and Pearson's χ^2 original tests, we observed that in all cases these tests controlled the type I error rates, with values equal to or below the 0.05 nominal level, except for the situation in which the binomial model was subjected to seven categories and larger samples $(N = 80)$, where the resultant rates were approximately 0.10. In comparison with the tests proposed by Sutradhar et al. (2008), built to assess binomial and/or multinomial models goodness-of-fit with overdispersion, based on Pearson's χ^2 statistic, our finding can be considered reasonable.

Park et al. (1996) studied the overdispersion effect considering three categories, underthe hypothesis $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}^*$, where $\boldsymbol{\pi}^* = (\pi_1, \ldots, \pi_J)$, fixed $J = 3$, in which the values of π were defined as, respectively, 0.4, 0.2, and 0.4. Under this setting, the authors assumed sample sizes for each π to be respectively, 10, 15, and 20 and they concluded that for all overdispersion levels $(\rho = 0.3, 0.5$ and $0.7)$, their test controlled the Type I error with rates close to nominal levels 0.10, 0.15, and 0.20. Another result was verified by Cressie and Read (1984). The authors examined a family of statistics defined by $\left\{I^\lambda; \lambda \in \Re\right\}$ where $\lambda = 1$ resulted in Pearson's χ^2 test. They followed an approach based on statistical moments of this statistic and they concluded that, under the hypothesis $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}^*$ where $\boldsymbol{\pi}^* = (\pi_1, \ldots, \pi_J)$, fixed $J = 4$, assuming $n_j = 20$, their test did not control type I error at a 0.10 nominal level.

Regarding the performance of modified tests, the results related to type I error control revealed that in almost all cases the deviance test controlled type I errors with a probability close to the 0.05 nominal level. Moreover, with the adoption of a prior distribution Beta$(8, 2)$, a right-skewed distribution, the deviance test has remained conservative. There is no discussion about the deviance performance in other bibliographies and, due to this fact we are able to highlight the relevance of this research.

For the Pearson's χ^2 test, we observed some inconsistent results influenced by the number of categories assigned to the binomial model. For J=3 categories, this test produced high Type I error rates, when the prior distribution Beta$(8, 2)$ was used. For higher J values (5 and 7), the test did not control type I errors, including a prior distribution Beta$(5, 5)$.

Table 3 shows the empirical power of original and modified deviance tests and Table 4 shows the empirical power of original and modified Pearson's χ^2 tests. According to Table 3, we observed that the deviance test resulted in expressive power rates in situations of large samples $(N = 80)$, when the number of categories was J=5 and J=7, given the

Table 2. Rejection rates obtained in deviance and Pearson's χ^2 used for validation of a multinomial logit model considering the parametric values π_j for $j = 1, \ldots, J$.

$J = 3(0.33; 0.33; 0.34)$								
	$n_j = 20$				$n_j = 80$			
	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
Deviance	0.0533	0.0275	0.0175	0.0010	0.0500	0.0455	0.0415	0.0140
Pearson	0.0483	0.0125	0.0665	0.9620	0.0481	0.0370	0.0395	0.3860
$J = 5(0.20; 0.20; 0.20; 0.20; 0.20)$								
	$n_j = 20$				$n_j = 80$			
	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
Deviance	0.0415	0.0180	0.0000	0.0000	0.0485	0.0515	0.0145	0.0000
Pearson	0.0380	0.0000	0.5370	1.0000	0.0455	0.0200	0.1135	1.0000
$J = 7(0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.10)$								
	$n_j = 20$				$n_j = 80$			
	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
Deviance	0.0224	0.0040	0.0000	0.0000	0.1045	0.0840	0.0040	0.0000
Pearson	0.0220	0.0000	0.9620	1.0000	0.1005	0.0385	0.5945	1.0000

overdispersion levels $\rho = 0.50$ and $\rho = 0.70$. Given the number of categories 5 and 7, we noted that the modified tests with the inclusion of the prior distributions Beta$(2, 8)$ and Beta$(5, 5)$ resulted in satisfactory power rates. However, this result was expected.

Table 3. Power values of deviance test obtained at different overdispersion levels of a multinomial logit model considering distinct number of populations and parametric values π_j, for $j = 1, \ldots, J$.

$J = 3(0.33; 0.33; 0.34)$								
	$n_j = 20$				$n_j = 80$			
ρ	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
0.20	0.3335	0.1785	0.1530	0.1060	0.4680	0.4650	0.4490	0.3525
0.50	0.4950	0.4010	0.2910	0.2330	0.6840	0.5760	0.5640	0.5294
0.70	0.5515	0.5070	0.2970	0.2470	0.6830	0.4685	0.4510	0.4525
$J = 5(0.20; 0.20; 0.20.0.20.0.20)$								
	$n_j = 20$				$n_j = 80$			
ρ	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
0.20	0.2670	0.2315	0.1620	0.1380	0.6680	0.6890	0.6025	0.3301
0.50	0.5130	0.4225	0.3400	0.2820	0.9250	0.8990	0.7965	0.5805
0.70	0.5730	0.5095	0.4325	0.3820	0.8990	0.8475	0.7635	0.5825
$J = 7(0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.10)$								
	$n_j = 20$				$n_j = 80$			
ρ	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
0.20	0.2130	0.2025	0.1680	0.1275	0.8215	0.7735	0.4420	0.1870
0.50	0.4390	0.4115	0.3620	0.2775	0.9675	0.9435	0.6680	0.4195
0.70	0.5200	0.5215	0.4735	0.3600	0.9320	0.8910	0.6690	0.5050

A question that naturally arises is how such test procedures perform if the Bayes-Laplace and Jeffreys priors are assumed. It is important to emphasize that in the case of the Beta$(5, 5)$ distribution, the modified deviance test proved to be a promising one because, in comparison to results of type I error control (Table 2), this test was conservative and, for this reason, we expected lower power.

Under the same settings, the results found in Table 4 revealed that the modified Pearson's χ^2 test was also a competitor test in comparison to the original version. However, it should be noted that the results that are discussed refer only to cases where the control of type I error was effective (Table 2). Thus, we proceed with the discussion of the results in Table 4.

Regarding the original Pearson's χ^2 test, for small samples ($n_j = 20$) we observed that the power rates (Table 4) were similar to those obtained with the deviance test (Table 3) in general. This fact was expected according to the similarity of type I error control for the two tests, expressed in Table 2. We observed the same behavior when considering the modified Pearson's χ^2 test with the prior distribution Beta$(2, 8)$. However, under the small samples situation ($n_j = 20$), with 3 categories and prior distribution Beta$(5, 5)$, where the Pearson's χ^2 test controlled the type I error (Table 2), we found that the modified test presented low power rates. This happened to all overdispersion levels (ρ) and larger samples ($n_j = 80$).

Increasing the number of categories J for 5 and 7, the results in bold in Table 3 were not interpreted because there was no control of the type I error by Pearson's χ^2 test, as reported on the discussion made on Table 2.

Table 4. Power values of deviance test obtained at different overdispersion levels of a multinomial logit model considering distinct number of populations and parametric values π_j for $j = 1, \ldots, J$.

			$J = 3(0.33; 0.33; 0.34)$					
		$n_j = 20$				$n_j = 80$		
ρ	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
0.20	0.3310	0.1750	0.5180	0.9810	0.4870	0.5140	0.5265	0.6725
0.50	0.5020	0.3940	0.8910	0.9900	0.7630	0.8700	0.8765	0.9320
0.70	0.5685	0.5230	0.9760	0.9840	0.7730	0.9660	0.9655	0.9810
			$J = 5(0.20; 0.20; 0.20.0.20.0.20)$					
		$n_j = 20$				$n_j = 80$		
ρ	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
0.20	0.2630	0.1990	0.8350	1.0000	0.6710	0.6750	0.7027	1.0000
0.50	0.5140	0.4145	0.9830	1.0000	0.9260	0.9610	0.9695	1.0000
0.70	0.5830	0.5560	0.9965	1.0000	0.9110	0.9950	0.9965	1.0000
			$J = 7(0.15; 0.15; 0.15; 0.15; 0.15; 0.15; 0.10)$					
		$n_j = 20$				$n_j = 80$		
ρ	MLE	(2,8)	(5,5)	(8,2)	MLE	(2,8)	(5,5)	(8,2)
0.20	0.2180	0.1930	0.9994	1.0000	0.7875	0.6800	0.9130	1.0000
0.50	0.4470	0.4120	0.9999	1.0000	0.9675	0.9670	0.9980	1.0000
0.70	0.5300	0.5340	1.0000	1.0000	0.9350	0.9950	1.0000	1.0000

Then, we discuss the application of FDR criteria (false discovery rate) to determine the overall nominal level. Note that the FDR procedure is based on the distribution of p-values, in cases of independent or multiple tests Benjamini and Yekutieli (2001). Thus, many situations can be exemplified. However, in this work we emphasize that the FDR procedure was applied in circumstances in which the estimates of MLE and $n_j = 80$ were considered. This was made because, in these cases, the tests were characterized in their original form, keeping their asymptotic properties in the sense that, for large samples, the tests statistics provide a better approximation to the chi-square distribution.

The results illustrated in Figure 1 correspond to application of the FDR procedurein the distribution of p-values of Pearson's χ^2 and deviance tests in order to determine an overall significance level q-value cut-off through the FDR criteria by setting the nominal level of significance of $\delta = 5\%$. This fact is expressive when considering a q-value close to 0.8, where a large number of tests were significant. Similarly, we observed the same behavior for $J = 5$ categories; see Figure 1 (second panel). When we considered $J = 7$ categories, this problem was not detected because assuming a q-value approximately equal to 0.07 for both tests resulted in a low number of significant tests; see Figure 1 (third panel).

4. Conclusion

According to the results obtained in the evaluations, the deviance test controlled type I errors under all evaluated situations with values close to or below the 0.05 nominal level. However, it should be noted that the modification given by the inclusion of a prior distribution Beta$(5, 5)$ caused high power rates. Therefore, we consider the deviance test more promising than conventional deviance tests under large sample situations. This prior distribution was efficient considering the simulated configurations in this research. However, it is not possible to ensure that this efficiency will be the same in different cases of probability and sample sizes.

Due to the Pearson's χ^2 test sensitivity for type I error control considering different prior distributions, the modified Pearson's χ^2 test showed no benefits to recommend its use when compared to the conventional test.

Based on the simulated results, the FDR criteria applied to Pearson's χ^2 and deviance tests in its original configuration for $J = 7$ categories and $n_j = 80$ samples provided an overall nominal level close to the nominal significance level fixed at 5%.

Acknowledgements

We would like to thank CAPES-PROCAD (DEX-UFLA /IME-USP) project for the financial assistance. The authors also would like to thank the anonymous referees for their valuable contribution to improve the quality of this article.

Appendix

R functions to generate contingency tables to obtain the rejection rates and power values via Monte Carlo method.

Table 5. R function to compute $y_j \sim CB(n_j, \pi_j, \rho)$

```
bc=function(n,phi,rho) {
    # Arguments: CB(nj,pij,rho) (Section 2.1) #
    x1=rbinom(1,n,phi)
    x2=rbinom(1,1,phi)*n
    u=rbinom(1,1,rho)
    y=(1-u)*x1+u*x2
        return (y)
 }
```

After the contingency tables generated, the adjusted probabilities were obtained using the lm command. Repeating this process $10,000$ times it was possible to estimate the rates by computing the empirical ratio given by the number of times that the p-value was greater than the 0.05 fixed nominal value.

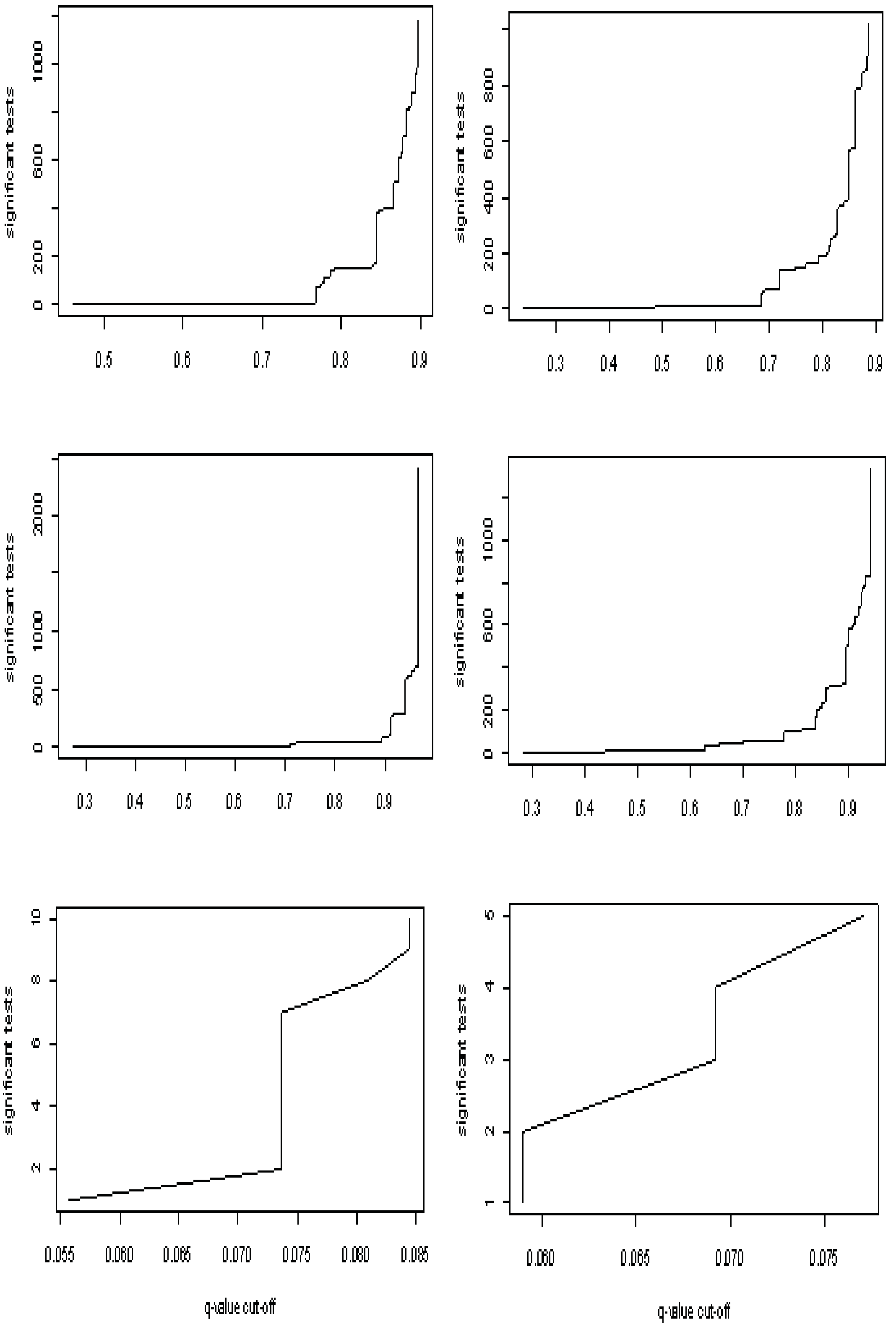

Figure 1. Number of significant results of the Pearson's χ^2 (on the left) and deviance (on the right) tests in function of the cut-off FDR values obtained at different nominal levels with $J = 3$ (first panel), $J = 5$ (second panel) and $J = 7$ (third panel).

Table 6. R Function to compute contingency tables used in the Monte Carlo simulation

```
table_cont=function(J,parj,nj,pho,alfa,beta) {
    # Arguments
    # J = total number of categories (Table 1 − Section 2.1)
             # parj = vetor pij (Table 1 − Section 2.1)
             # nj = vetor nj (Table 1 − Section 2.1)
             # pho = probability rho CB(nj,pij,rho)
             # (eq. 1, Section 2.1)
             # alfa and beta = hyperparameters of
             # beta prior distribution

          respns=matrix(0,J,6)

   for (j in 1:J)
   {
           sbc=bc(nj[j],parj[j],pho)    # number of occurences
                                        # (success) in jth category
           f=nj[j]−sbc                  # nj−yj
```

```
 #construction of contingency table
        respns[j,1]=j                  # jth category
        respns[j,2]=sbc                # yj
        respns[j,3]= f                 # nj−yj
        respns[j,4]=nj[j]              # nj
        respns[j,5]=(sbc+alfa)/(nj[j]+(alfa+beta))        # E(pij|yj)
        respns[j,6]=1−(sbc+alfa)/(nj[j]+(alfa+beta))      # 1 − E(pij|yj)
   }
        return (respns)
}
```

Table 7. Application of functions described in tables 5 and 6

```
        J=5;alfa=8;beta=2;pho=0.8
        pj<−as.matrix(c(0.20,0.20,0.20,0.20,0.20))
        nj<−as.matrix(c(50,50,50,50,50))
        table=table_cont(J,pj,nj,pho,alfa,beta)
```

References

Agresti, A., Min, Y., 2005. Frequentist performance of bayesian confidence intervals for comparing proportions in 2 × 2 contingency tables. Biometrics, 61, 515-523.

Altham, P.M.E., 1969. Exact bayesian analysis of 2×2 contingency table, and Fisher's exact significance test. Journal of the Royal Statistical Society, Series B (Methodological), 31, 261-269.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, series B, 57, 289-300.

Benjamini, Y., Yekutieli, D., 2001. On the control of discovery rate in multiple testing under dependency. The Annals of Statistics, 29, 1165-1188.

Bogutchi, T.F., Colosimo, E.A., Lamounier, J.A., 2006. Comparação entre testes para superdispersão em dados binários. Revista de Matemática e Estatística, 23, 55-70.

Cressie, N., Read, T.R.C., 1984. Multinomial Goodness-of-Fit Tests. Journal of the Royal Statistical Society, Series B (Methodological), 46, 440-464.

Dobson, A.J., 2001. Introduction to generalized linear models. 2nd edition. Chapman &

Hall, London.

Good, I.J., 1956. On the estimation of small frequencies in contingency tables. Journal of the Royal Statistical Society Series B, 18, 113-124.

Efron, B., 1986. Double exponential families and their use in generalized linear regression. Journal of the American Statistical Association, 81, 709-721.

Fu, J., Sproule, R.A., 1995. A generalization of the binomial distribution. Communications in Statistics: Theory and Methods, 24, 2645-2658.

Gammerman, D., Migon, H. S., 1993. Inferência Estatística: uma abordagem integrada. Instituto de Matemtica da Universidade do Rio de Janeiro, 207p.

Gelman, A., Carlin, J.B., Stern, H.S., 2004. Bayesian data analysis. Chapman and Hall/CRC, London.

Hinde, J., Demétrio, C.G.B., 1998. Overdispersion: models and estimation. Computational Statistics and Data Analysis, 27, 151-170.

Jhun, M., Jeong, H.C., 2000. Applications of bootstrap methods for categorical data analysis. Computational Statistics and Data Analysis, 35, 83-91.

Luceño, A., 1995. A family of partially correlated Poisson models for overdispersion. Computational Statistics and Data Analysis, 20, 511-520.

Park, C.G., Park, T.P., Shin, D.W., 1996. A simple method for generating correlated binary variates. American Statistician, 50, 306-310.

Petri, C., 2007. Relação entre níveis de significância Bayesiano e frequentista: e-value e p-value em tabelas de contingência. IME-USP - Instituto de Matemática e Estatística da Universidade de São Paulo Master thesis, 93p.

Pereira C.A.B., Stern, J.M., 1999. Evidence and Credibility: Full Bayesian Significance Test for precise Hypotheses. Entropy Journal, 1, 69-80.

R Development Core Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.r-project.org.

Smith, G.K., 1989. Generalized linear models with varying dispersion. Journal of the Royal Statistical Society, Series B (Methodological), 51, 47-60.

Smith, G.K., Verbyla, A.P., 1999. Adjusted likelihood methods for modelling dispersion in generalized linear models. Environmetrics, 10, 696-709.

Sutradhar, S.C., Nagaraj, K., Neerchal, Morel, J.G., 2008. A goodness-of-fit test for overdispersed binomial (or multinomial) models. Journal of Statistical Planning and Inference, 138, 1459-1471.

Tallis, G.M., 1962. The use of a generalized multinomial distribution in the estimation of correlation in discrete data. Journal of the Royal Statistical Society, Series B, 24, 530-534.

Tutia, M.H., Diniz, C.A.R., Leite, J.G., 2003. Bayesian inference for the parameter p of the correlated binomial distribution. Revista de Matemática e Estatística, 21, 85-94.