Pearson's Idea to test fitting in GLM

João F. Gomes

FCUL, Departamento de Estatística e Investigação Operacional, Edifício C6, Campo Grande, 1749-016 Lisboa Portugal

Ricardo T. Lemos

SIAM, Observatório Astronómico de Lisboa, Tapada da Ajuda, Edifício Este, 1349-018 Lisboa

Summary. In this manuscript we show that it is still possible, to make a classical Pearson Chi-Squared test when the observations are not identically distributed. We propose a methodology, such as the Pearson Chi-Squared Statistic, which compares the observed counts with their expected values under the multinomial setting. In this case, the mean value of the observed values changes in each observation. Simulation results show its good performance in all the cases analysed.

This methodology can be used to suppress the lack of goodness-of-fit test statistics in generalized linear models (GLM's), such as count data model first of all to test the model distribution and the link function when we make model assumptions. It can also be used, in binary models to test the quality of the covariates.

Keywords: Generalized Linear Models, Goodness-of-fit, Deviance, Pearson χ^2, Central Limit Theorem, Simulation

This work was partially supported by Fundação para a Ciência e Tecnologia, project POCTI-ISFL-1-209 (Centro de Matemática e Aplicações Fundamentais).

Adress for Correspondence: João Gomes, Faculdade de Ciências da Universidade de Lisboa, Departamento de Estatística e Investigação Operacional, Edifício C6, Campo Grande, 1749-016 Lisboa, Portugal
e-mail: jjgomes@fc.ul.pt

1. Introduction

Generalized linear models (GLM) have become a standard class of models for data analysis, being a valuable alternative to classical regression models in cases where the response variable is a count variable, i.e., a non-negative integer; examples of such models are the log-linear models and the binary models.

The models that we select to fit to data are usually chosen from a particular class of distributions, the exponential family with canonical parameter θ, taking the e.g. special form

$$f(y;\theta;\phi) = \exp\left((y\theta - b(\theta))/\phi) + c(y;\phi)\right),$$

for some specific functions b(.) and c(.), such that $\mu = E(Y) = b'(\theta)$ and $Var(Y) = b''(\theta)\phi$.

For the time being we will suppose that ϕ, a dispersion parameter, is known and will consist f(.) as function of μ.

Hence, our methodology will be used in the exponential family. This family includes common discrete distributions such as the Binomial and the Negative Binomial (if the nuisance parameters are known) and the Poisson and also continues distribution, e.g., the Normal, the Gamma and the Inverse Gaussian. The parameterization of each model can have seen in MacCullagh and Nelder (1989).

Pigeon and Heyse (1999) developed a test statistic suitable for binary and polytomous regression models and argued that it might offer some advantages over the Hosmer-Lemeshow statistic. Regarding other GLMs, such as count data models and models with censored or zero-inflated data, a robust and unbiased statistic seems to be lacking. In this paper, we attempted to develop such a statistic and, through simulation, to test its performance against alternative statistics.

This test is based in the Chi-squared Pearson test where were compare the cells counts with their expected values under the multinomial setting by applying a simple adjustment to solve the problem that has a distinct set of outcomes probabilities.

Our methodology will also be based in the Chi-squared statistic however other adjustment will be made to the exponential family.

A simulation study was made to show that our proposed statistic, JOR, has approximately a χ^2 distribution when we use a set of mean values and constant nuisance parameters in each distribution.

2. Presentation of the JOR test statistic

2.1. Fundamentals

We can define the Pearson Chi-Squared statistic in a few words. Let ($V_1,\dots,V_n$) be a random sample with distribution function F(.). Let be $I_1,\dots,I_k$ a partition of support of F(.) and $p_j = \int_{I_j} dF(x), j=1,...,k$. O_j is a number of observations in I_j, j=1,…,K. With $(O-np) \equiv (O_1 - np_1,...,O_{k-1} - np_{k-1})^T$,

$$X^2 = (O-np)^T \Sigma^{-1} (O-np) \sim \chi^2_{k-1} ,$$

where Σ is the non-singular variance-covariance matrix of (O-np). Computation shows that $X^2 = \sum_{i=1}^{k} \frac{(O_i - np_i)^2}{np_i}$. $O_j \sim Bi(n,p_j)$ and O_j-np_j, the difference between the observed and the expected cells, express lack of fit of the data to F() and Σ is non-singular if only K-1 of the cells were considered and the limit distribution is independent of F().

Let us consider now a random sample of *n* independent variables (V_i, i = 1, 2,…, n), belonging to the same family (exponential family in this case), all with the same support, with average value μ_i. Consider a partition of $\mathbb{R}$ into K intervals, $I_1 = (-\infty, a_1]$, $I_2 = (a_1, a_2]$, …, $I_{n-1} = (a_{K-2}, a_{K-1}]$, $I_n = (a_{K-1}, +\infty)$ so that $a_{j-1} \leq a_j$ for all j = 1, 2, …, K-1. Under the null hypothesis (H_0) of knowing the distribution function of the V_i ($F_i()$, i = 1, 2, …, n), it is possible to calculate $p_{ij} = P(V_i \in I_j)$. Hence, each of the *n* random variables is associated with *K* random variables with Bernoulli distribution, $B_{i,1},...,B_{i,K}$, with parameters $p_{i,1},...,p_{i,K}$ related to H_0. For a fixed *i*, the *K* Bernoulli random variables thus obtained are not independent: for $c \neq d$, $Cov\left[B_{i,c}, B_{i,d}\right] = -p_{i,c}\, p_{i,d}$. In contrast, for a fixed *j*, the *n* obtained Bernoulli random variables are independent.

The approach taken here to test the validity of H_0 is based on applying a version of the Central Limit Theorem in which the involved random variables need not have the same distribution. According to Karr (1993), if *n* random variables ($X_1, X_2, \dots, X_n$) are independent but not necessarily identically distributed, with $E\left[X_i\right] = 0$ and $E\left[X_i^2\right] < \infty$, a sufficient condition to satisfy the Central Limit Theorem is that the Lyapunov conditions are satisfied:

$$E\left[|X_i|^3\right] < \infty$$

for each *i* and

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E\left[|X_i|^3\right]}{\left(\sum_{i=1}^{n} Var\left[X_i\right]\right)^{3/2}} = 0 .$$

2.2. Construction of JOR

Admit *j* fixed. If the $X_{i,j}$ random variables are given by $X_{i,j} = B_{i,j} - p_{i,j}$, then

$$|X_{i,j}|^3 = \begin{cases} (1-p_{i,j})^3 \text{, with probability } p_{i,j} \\ p_{i,j}^3 \text{, with probability } 1-p_{i,j} \end{cases}$$

$$E\left[|X_{i,j}|^3\right] = p_{i,j}(1-p_{i,j})\left[(1-p_{i,j})^2 + p_{i,j}^2\right] < \infty$$

$$Var\left[X_{i,j}\right] = p_{i,j}(1-p_{i,j})$$

Thus

$$Q = \frac{\sum_{i=1}^{n} E\left[|X_i|^3\right]}{\left(\sum_{i=1}^{n} Var\left[X_i\right]\right)^{3/2}} = \frac{\sum_{i=1}^{n} p_{i,j}(1-p_{i,j})\left[(1-p_{i,j})^2 + p_{i,j}^2\right]}{\left(\sum_{i=1}^{n} p_{i,j}(1-p_{i,j})\right)^{3/2}}$$

Since $0 < p_{i,j} < 1$, then $1/2 < (1-p_{i,j})^2 + p_{i,j}^2 < 1$ and

$$\lim_{n\to\infty} \frac{1}{2\sqrt{\sum_{i=1}^{n} p_{i,j}(1-p_{i,j})}} < \lim_{n\to\infty} Q < \lim_{n\to\infty} \frac{1}{\sqrt{\sum_{i=1}^{n} p_{i,j}(1-p_{i,j})}} .$$

If the series $\sum_{i=1}^{n} p_{i,j}(1-p_{i,j})$ is convergent as $n \to \infty$, implying that $p_{i,j}$ converge to either 0 or 1 as $n \to \infty$, then the second Lyapunov condition is not verified and the Central Limit Theorem is not satisfied. However, this seems to be an uncommon situation in GLMs, since larger sample sizes are not usually associated with the convergence of $p_{i,j}$, and therefore the Lyapunov conditions should be met. For example, if the $p_{i,j}$ involved are random numbers between 0 and 1 then the convergence of Q to zero seems to occur (Figure 1).

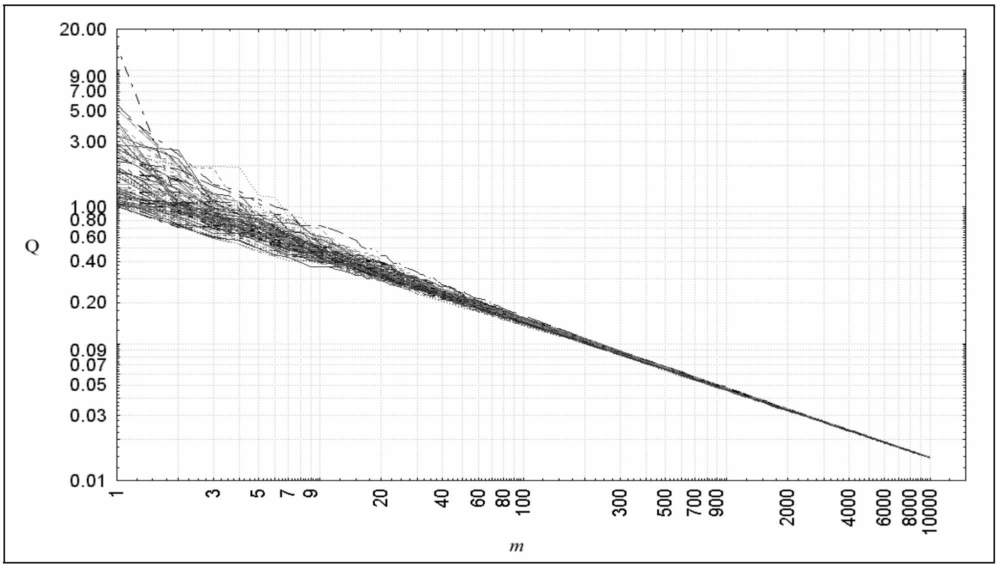


Figure 1. Convergence analysis of Q as $n \to \infty$, when $p_{i,j}$'s are pseudo-random numbers between 0 and 1. One hundred simulations are shown, with $\max(n) = 10^4$. For $n = 10^6$, $\overline{Q} = 4.6\times10^{-3}$, and for $n = 10^{10}$, $\overline{Q} = 1.5\times10^{-5}$.

Considering this and letting

$$Z_j = \frac{\sum_{i=1}^{n} B_{i,j} - \sum_{i=1}^{n} p_{i,j}}{\sqrt{\sum_{i=1}^{n} p_{i,j} - \sum_{i=1}^{n} p_{i,j}^2}}, \; j = 1, 2,\ldots, \mathrm{K},$$

then the Central Limit Theorem should be applicable, $Z_j \xrightarrow{d} N(0,1)$ and $(Z_1, Z_2, ..., Z_K) \xrightarrow{d} N\left(0, \boldsymbol{\Sigma}\right)$, where $\boldsymbol{\Sigma}$ denotes the variance-covariance matrix. In this matrix, all the diagonal elements are equal to 1, whereas the generic element (c, d) of $\boldsymbol{\Sigma}$, such that $c \neq d$, is given by the following expression:

$$Cov\left(Z_c, Z_d\right) = \frac{-\sum_{i=1}^{n} p_{i,c}\, p_{i,d}}{\sqrt{\sum_{i=1}^{n} p_{i,c}(1 - p_{i,c})}\sqrt{\sum_{i=1}^{n} p_{i,d}(1 - p_{i,d})}}.$$

It is also possible to demonstrate that $\boldsymbol{\Sigma}$ is singular, since the vector **r**, with j-th component given by $r_j = \sqrt{\sum_{i=1}^{n} p_{i,j}(1 - p_{i,j})}$, is an eigenvector of $\boldsymbol{\Sigma}$ with associated Eigen value 0.

Proof: the i-th element of the vector $\boldsymbol{\Sigma}.\mathbf{r}$ is given by

$$\left(\mathbf{\Sigma.r}\right)_i = \sum_{j=1}^{K} Cov\left[Z_i, Z_j\right] \times r_j = Cov\left[Z_i, \sum_{j=1}^{K} r_j Z_j\right] = Cov\left[Z_i, \sum_{j=1}^{K}\left(\sum_{i=1}^{n} B_{i,j}\right) - \sum_{j=1}^{K}\left(\sum_{i=1}^{n} p_{i,j}\right)\right]$$
$$= Cov\left[Z_i, n-n\right] = 0$$

hence, for all i, $(\mathbf{\Sigma.r})_i = 0$ and therefore $\mathbf{\Sigma.r} = 0.\mathbf{r}$.

The singularity of $\mathbf{\Sigma}$ results from considering all the random variables Z_i, associated to all the support of the V_i. By removing e.g. the last row and column of $\mathbf{\Sigma}$, a full rank matrix is generally obtained and the following test statistic may be constructed:

$$T = (Z_1, Z_2, ..., Z_{K-1})\mathbf{\Sigma}^{-1}(Z_1, Z_2, ..., Z_{K-1})^t,$$

which has, under H_0, asymptotic χ^2 distribution with $K - 1$ degrees of freedom. The proof that if $\mathbf{X} \cap N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ then $(\mathbf{X}-\boldsymbol{\mu})^t \mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \cap \chi_p^2$ can be found e.g. in Kshirsagar (1972) and Muirhead (1982).

Considering the purpose of this test statistic, two aspects are worth mentioning:

a) The partition intervals of $\mathbb{R}$ should be defined so that all the $p_{i,j}$ are neither very large (≈ 1) nor very small (≈ 0); this procedure is necessary to accelerate the convergence of the Z_j to normally distributed random variables.

b) To evaluate the goodness-of-fit of a model, it is convenient to assess its behaviour for small, intermediate and large expected average values (μ_i). This may be accomplished by sorting the observed sample by μ_i and then splitting it into g groups. The test statistic would then be

$$T = T_1 + ... + T_g = (Z_{1,1}, ..., Z_{1,K-1})\mathbf{\Sigma}_1^{-1}(Z_{1,1}, ..., Z_{1,K-1})^t + ... + (Z_{g,1}, ..., Z_{g,K-1})\mathbf{\Sigma}_g^{-1}(Z_{g,1}, ..., Z_{g,K-1})^t$$

which would have, under H_0, asymptotic χ^2 distribution with $g \times (K-1)$ degrees of freedom. Being the choice of g arbitrary, it should be made so as to ensure that in each group the asymptotic distribution is reached.

In view of these two aspects, the *JOR* test statistic is defined as

$$JOR = T_1 + T_2 ... + T_g = (Z_{1,1}, ..., Z_{1,K_1-1})\mathbf{\Sigma}_1^{-1}(Z_{1,1}, ..., Z_{1,K_1-1})^t$$
$$+ (Z_{2,1}, ..., Z_{2,K_2-1})\mathbf{\Sigma}_2^{-1}(Z_{2,1}, ..., Z_{2,K_2-1})^t + ... + (Z_{g,1}, ..., Z_{g,K_g-1})\mathbf{\Sigma}_g^{-1}(Z_{g,1}, ..., Z_{g,K_g-1})^t$$

where K_j denotes the number of partitions made to group j (j = 1, 2,…, g) and $\mathbf{\Sigma}_j$ is an $(\mathrm{K_j}-1) \times (\mathrm{K_j}-1)$ invertible variance-covariance matrix. Under H_0, *JOR* has asymptotic χ^2 distribution with $\sum_{j=1}^{g}\left(\mathrm{K_j}-1\right)$ degrees of freedom. The proposed method to partition $\mathbb{R}$ in the j-th group is the following:

Consider g fixed; the first interval of the j-th group is defined as $(-\infty, a_{j,1}]$, where $a_{j,1}$ is such that $\sum_{i=1}^{n} p_{i,1}\mathbb{I}_j \approx 5$, $\mathbb{I}_j$ taking the value 1 if the i-th observation belongs to group j and 0 otherwise; the second interval is defined as $(a_{j,1}, a_{j,2}]$, where $a_{j,2}$ is such that $\sum_{i=1}^{n} p_{i,2} I_j \approx 5$; etc.; this condition is relaxed only for the last interval, where the sum may be less than 5.

3. Performance analysis

In this work we simulated the behaviour of *JOR* for the Binomial, the Negative Binomial and the Poisson distributions, and also for the Gamma and the Normal distributions.

For all models, we split a sorted sample of n = 1000 mean values into g groups of nearly equal size (the last group could contain fewer observations than the others, which were equally sized), being the choice of g arbitrary. For group j (j = 1, 2,…, g), we partitioned $\mathbb{R}$ into K_j intervals, using the procedure mentioned in the previous section. With the probability distribution function (p.d.f.) associated with each regression model and the value of μ_i (nuisance parameters were fixed; see below), we also computed $p_{i,\ell}$ (ℓ = 1, 2,…, K_j) and generated a single observation for each V_i (i =1, 2,…, 1000). The generation of 1000 observations was repeated 1000 times; in each cycle, the *JOR* statistic was calculated. The distribution of *JOR* was compared to the χ^2 distribution with $d.f. = \sum_{j=1}^{g} \left(K_j - 1\right)$ by means of a smooth probability density function and a Kolmogorov-Smirnov goodness-of-fit test. Finally, we analysed if the type I error rate was correct, using α=0.1, 0.05 and 0.01. All these procedures were incorporated in Microsoft Visual Basic code interfaced with Excel. A collection of results is presented in Table 1 and in Graphic 1.

As indicated by the Kolmogorov-Smirnov test and by the smooth probability density function, *JOR* generally displayed the expected χ^2 distribution in all models, regardless of the grouping strategy. This first result points to the good performance of *JOR* even when groups are composed of less than 30 observations, suggesting rapid convergence to the asymptotic distribution.

As a corroborative result, *JOR*'s type I error rates were always close to expected values (α), irrespective of g.

Therefore, *JOR* seems to be an adequate statistic for testing goodness-of-fit in several distributions belonging to the exponential family when the observations have different mean values and nuisance parameters constant.

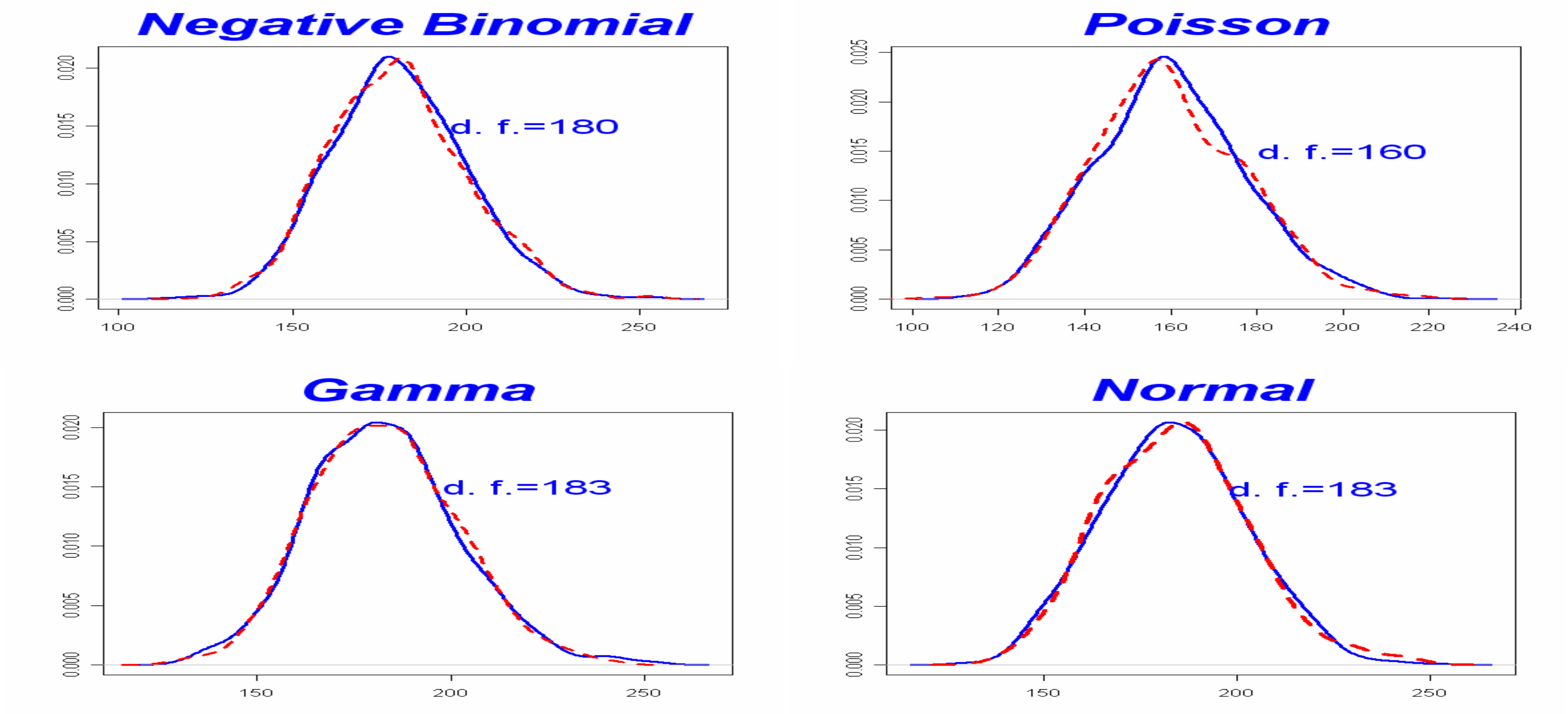
Also, we made some tests using alternative hypothesis, for example the test using Normal observations when H_0 was Gamma (the same dispersion parameter). Even though the distribution of the JOR is close to a Chi-squared, the number of degrees of freedom is very different from it would expected under H_0.

	Negative Binomial(h)												Poisson					
Probability Distribution Function	$\ln f(y;\mu)=\frac{y\ln\left(\mu/(\mu+h)\right)-h\ln\left(\mu/(\mu+h)\right)}{1}+\ln\binom{y+h-1}{y}$												$ln\ f(y;\mu)=\frac{y\ln(\mu)-\mu}{1}-\ln(y!)$, y,μ>0					
$\mu_i \in$	(0, 90)						(0, 450)						(0, 120)					
Nuisance Parameters	$h = 1$						$h = 5$						–					
Kolmogov-- Smirnov: # groups (g)	9	10	13	17	25	50	9	10	13	17	25	50	9	10	13	17	25	50
Kolmogov-- Smirnov: K-S p-value	*	*	****	****	*	*	*	***	*	*	**	*	*	*	*	*	*	*
Type I error rate: **α = 1%**	1.5%	1.2%	0.7%	0.8%	0.8%	0.6%	1.0%	1.2%	1.4%	0.8%	1.5%	1.0%	1.7%	1.0%	1.2%	0.8%	1.5%	0.4%
Type I error rate: **α = 5%**	5.5%	4.6%	3.9%	3.2%	4.7%	4.3%	4.7%	5.8%	6.8%	5.0%	4.3%	4.5%	5.0%	4.7%	4.7%	4.7%	4.5%	4.2%
Type I error rate: **α = 10%**	10.9%	9.3%	8.1%	8.2%	9.1%	8.0%	8.9%	9.9%	10.4%	10.8%	8.5%	8.0%	9.0%	9.7%	8.9%	8.7%	8.5%	7.8%

	Binomial(n)									Normal(ϕ)					
Probability Distribution Function	$\ln\ f(y;\mu)=\frac{y\ln\frac{\mu}{n-\mu}-n\ln(\frac{n}{n-\mu})}{1}+\ln\binom{n}{y}$									$\ln f(y;\mu,\phi)=\frac{y\mu-\mu^2/2}{\phi}-\frac{1}{2}\left(\frac{y^2}{\phi}+\ln(2\pi\phi)\right)$					
$\hat{\mu}_i \in$	(0, 0.9)			(0, 5)			(0, 10)			(0, 200)					
Nuisance Parameters	$n = 1$			$n = 5$			$n = 10$			ϕ = 1					
Kolmogov- -Smirnov: # groups (g)	7	10	20	7	10	20	7	10	20	9	10	13	17	25	50
Kolmogov- -Smirnov: K-S p-value	**	*	*	*	*	*	*	*	*	*	***	*	**	*	**
Type I error rate: α = 1%	1.6%	0.5%	1.0%	1.6%	1.0%	1.8%	0.7%	0.6%	1.0%	0.9%	1.6%	1.2%	1.7%	0.4%	1.2%
Type I error rate: α = 5%	5.7%	3.8%	4.9%	5.4%	5.5%	5.0%	5.3%	4.6%	5.7%	4.7%	6.0%	4.3%	5.1%	4.7%	4.1%
Type I error rate: α = 10%	10.5%	8.6%	10.9%	10.6%	10.5%	11.1%	10.1%	9.7%	10.0%	8.8%	10.4%	11.0%	10.2%	9.6%	8.4%

	Gamma(ϕ)
Probability Distribution Function	$\ln f(y,\mu)=\frac{y(-1/\mu)-\log(\mu)}{\phi}-(1/\phi)\log(\phi)+(1/\phi-1)\log y-\log(\Gamma(1/\phi))$

Table 1. Performance analysis of *JOR* in Negative Binomial, Poisson, Binomial and Normal regression models. Kolmogorov-Smirnov results: **** indicates p<0.05, *** 0.05<p<0.10, ** 0.10<p<0.20, * p>0.20.



Results: Estimates densities of JOR(- - - -) vs Chi-square(—).Negative Binomial(h=0.2), Poisson, Normal(ϕ=2), Gamma(ϕ=0.5). For all, mean values (μ_i, i=1 to 1000): Min=0.3, Max=200.

4. Concluding Remarks

The results show that, in these conditions, the performance of JOR is excellent. We can say that whenever there is a change in localization of the independent variables the classical test of Chi-squared is available.

If we have a generalized linear model (GLM) where we estimate μ_i after choose a probability distribution, link function, and estimate the nuisance parameters correctly (i.e., H_0), JOR can be a goodness-of-fit test.

If we reject H_0, in spite of the good qualities of explanatory variables, signifies that or probability distribution or link function were a bad choice. Through further simulations, we also found *JOR*'s type I errors not to be affected by reducing sample size (n = 100, 200, 300, with g = 2 or 5).

Nevertheless, our analysis was not exhaustive and additional tests must be performed before we can suggest that *JOR* may be an all-encompassing goodness-of-fit test statistic for GLMs.

For example, the irrelevance in the choice of g may not hold in cases where the μ_i form separate clusters instead of being more or less randomly distributed within their range. In these situations, we suggest selecting g so that each group contains roughly homogeneous μ_i, since this procedure maximises *JOR*'s degrees of freedom and is likely to optimize its performance.

When the null hypothesis of correct model specification is rejected, an additional aspect of *JOR* may be of interest: under H_0, *JOR* is composed of g terms, each with asymptotic $\chi^2_{n_g-1}$ distribution; hence, if H_0 is rejected, it is possible to analyse which term(s) contributed the most to a large value of the test statistic and to assess if the GLM performed poorer for small, intermediate or large μ_i (since the μ_i are sorted).

Apart from GLMs, *JOR* may also be suitable for censored regression models such as Tobit (Tobin, 1958) and Tobit-type (Terza, 1985; Brãnnãs, 1992; Caudill and Mixon, 1995), as well as for zero-inflated count data models (e.g., Lemos and Gomes, 2004). To facilitate its use, a program in R CRAN that computes *JOR* basing on a set of observations and model estimates is freely available on request to the authors.

We can make some questions yet. What is the power of JOR test, how much degrees of freedom we lost after estimate μ_i or how can we estimate the nuisance parameters including ϕ?

5. Acknowledgements

We thank Jorge Cadima for suggestions made at the stage of developing *JOR*. We would like to dedicate this manuscript to the memory of Orlando Oliveira, who gave valuable comments to this work, and to express our gratitude for his many years of friendship and support.

6. References

Brãnnãs, K. (1992) Limited Dependent Poisson Regression, *The Statistician* 41: 413-423.

Caudill, S.B. and Mixon, F.G. (1995) Modelling Household Fertility Decisions: Estimation and Testing of Censored Regression Models for Count Data. Empirical Economics 20: 183-196.

Christensen, R. (1997) *Log-linear Models and Logistic regression*, 2nd edn. New York, Springer.

Connover, W.J. (1999) *Practical Nonparametric Statistics*, 3rd edn. New York, John Wiley & Sons.

Cox, D.R. and Hinkley, D.V. (1974) *Theoretical statistics*, London, Chapman & Hall.

Fahrmeir, L. and Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York, Springer.

Goeman, J.J and Cessie, S. (2006) A goodness-of-Fit Test for Multinomial Logistic Regression, *Biometrics* (to appear)

Hosmer, D.W. and Lemeshow, S. 1989. *Applied Logistic Regression*, New York, John Wiley & Sons.

Karr, A.F. (1993) *Probability*, New York, Springer.

Kshirsagar, A.M. (1972) Multivariate Analysis, New York, Marcel-Dekker, Inc.

Lemos, R. e Gomes, J. (2004) Do local environmental factors induce daily and yearly variability in bluefin tuna (*Thunnus thynnus* L.) trap catches? *Ecological Modelling*, Volume 177, Issues 1-2 , 143-156.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman & Hall.

Muirhead, R.J. (1982) *Aspects of Multivariate Analysis*. New York: John Wiley & Sons.

Pigeon, J.P. and Heyse, J.S. (1999) An Improved Goodness of Fit Statistic for Probability Prediction Models. *Biometrical Journal*, 42, 71-82.

Terza, J. (1985) A Tobit-type Estimator for the Censored Poisson Regression Model. *Economics Letters*, 18, 361-365.

Tobin, J. (1958) Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26, 24-36.