A ROBUST SCORE TEST OF HOMOGENEITY FOR ZERO-INFLATED COUNT DATA

Nadeesha R. Mawella
Kansas State University, Manhattan, KS, USA - nadee@k-state.edu

Wei-Wen Hsu
Kansas State University, Manhattan, KS, USA - wwhsu@ksu.edu

David Todem
Michigan State University, East Lansing, MI, USA - todem@msu.edu

KyungMann Kim*
University of Wisconsin-Madison, Madison, WI, USA - kyungmann.kim@wisc.edu

Abstract

Score statistics are often used for the test of homogeneity in zero-inflated models for count data. The most cited justification is that it only requires the model fit under the null hypothesis. However, the true null models are often unknown in practice and these statistics can be invalid when the null models are not correctly specified. As an empirical evidence, an intensive simulation study indicates that the sizes of these score tests for zero-inflated Poisson or negative binomial models may behave extremely liberal and unstable when the mean function of baseline distribution or the baseline distribution itself is misspecified. In this paper, we propose a new score test of homogeneity for zero-inflated models which is robust to these misspecifications. Technically, the test is developed under the framework of Poisson-gamma mixture model which can provide a more general framework to incorporate various baseline distributions without specifying the mean function. The empirical performances of our test in finite samples are evaluated in simulation studies, and the dental caries data from the Detroit Dental Health Project (DDHP) and Girl Scout data from Scouting Nutrition and Activity Program (SNAP) are used to illustrate the proposed test.

Keywords: Poisson; negative binomial; latent class models; model misspecification

1. Introduction

Zero-inflated models provide a parametric framework to accommodate heterogeneity in a population, in which the heterogeneity is referred to zeros being generated from two different sources. In real applications of these models, the question of interest is typically whether the mixture probability or so-called mixing weight parameter of the model adequately represents the inherent heterogeneity in the population. In order to address this question of interest, the hypothesis of zero mixing weight is often evaluated using score test statistics under this class of models. Intuitively, a zero mixing weight indicates that the zeros are only generated from the non-degenerate distribution. Therefore, testing the hypothesis that mixing weight equals zero is equivalent to testing the homogeneous model against the heterogeneous model.

Score test has been suggested by many authors as a homogeneity test under this class of models due to the fact that it only requires the model fit under the null model (for example, see van den Broek, 1995; Deng and Paul, 2000; Jansakul and Hinde, 2002; Todem et al., 2012). However, the general limitation of the score test is that it relies on the assumption of a well-specified null model. The true null models are often unknown in practice, and as a result, it can be invalid when the null models are not correctly specified. For instance, the mean of the null model is often related to covariates through a link function, but it could be misspecified easily. The baseline distribution under the null model could also be misspecified. For instance, if the working baseline distribution is assumed to be Poisson but the true underlying distribution is negative binomial (NB). It is well known that count data often exhibit over-dispersion compared to the Poisson distribution. If such over-dispersion is not incorporated by the baseline model appropriately, the homogeneity test for zero-inflated data may provide unreliable conclusions about the population.

Here we propose a new score test of homogeneity which is robust to these misspecifications. Technically, the test is developed under the framework of Poisson-gamma mixture model which provides a general framework to incorporate the baseline distributions under the zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB) models. Our test can be performed without specifying the mean function and hence avoids the impact of any possible misspecification of mean under the null model. The proposed test assumes a more general model for the baseline model, and therefore it is robust to the misspecification of the baseline distribution. The rest of this paper is organized as follows. In Section 2, we describe the zero-inflated model in general and the score test for homogeneity. In Section 3, we describe misspecification in the homogeneity test for zero-inflated models. We introduce the new robust test for evaluating heterogeneity under zero-inflated models in Section 4. The performance of the proposed test is evaluated by simulation studies, under possible misspecification of the mean and the baseline distribution in Section 5. The proposed test is then applied to dental caries data from the Detroit Dental Health Project (DDHP) by Tellez et al. (2006) and Girl Scout data from the Scouting Nutrition and Activity Program (SNAP) by Rosenkranz et al. (2010), and its performance compared with the existing homogeneity tests in Section 6. We close with a brief discussion in Section 7.

2. Zero-inflated models and score test of homegeneity

Zero-inflated model is a two-component mixture model which combines a degenerate distribution at zero and a parametric non-degenerate distribution based on a fixed set of parameters, say ζ. Specifically, the zero-inflated model is defined as,

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)g_i(0;\zeta) & \text{if } y_i = 0 \\ (1-\omega_i)g_i(y_i;\zeta) & \text{if } y_i = 1, 2, \ldots \end{cases}$$

where Y_i is a count variable, with observed value y_i for $i = 1, \ldots, n$ and ω_i is the unknown mixture probability. For $\omega_i = 0$, the model reduces to the homogeneous model which corresponds to the non-degenerate distribution $g_i(y_i;\zeta)$.

To test for heterogeneity in population under the zero-inflated models, mixing weight is evaluated by the homogeneity tests. Assuming a constant mixing weight such that $\omega_i \equiv \omega$ for all i, the hypothesis $H_0 : \omega = 0$ is tested. Score test is often used to test for heterogeneity under this class of models as it requires the estimation only under the null model. The corresponding score test statistic is given by

$$S = S_\omega(\hat{\zeta}, 0)^T \widehat{V}_\omega^{-1} S_\omega(\hat{\zeta}, 0)$$

where $\hat{\zeta}$ is the maximum likelihood estimate of ζ and $\widehat{V}_\omega$ is the estimated variance of the score function $S_\omega(\hat{\zeta}, 0)$. With the assumption of constant mixing weight, this score test statistic will have an asymptotic χ_1^2 distribution under the null hypothesis.

3. Misspecification

Score test requires a well-specified model under the null hypothesis. However, in practice, this assumption may not be met which in turn may result in unreliable statistical inferences. In the score test of homogeneity for zero-inflated models, the null model corresponds to the non-degenerate distribution that could be misspecified in terms of the mean or the distribution itself. For instance, when the true mean is $\lambda^\star = \exp(\beta_0+\beta_1X_1+\beta_2X_2)$ with two independent covariates X_1 and X_2, the working mean of $\lambda = \exp(\beta_0 + \beta_1X_1)$ may be used under the test which would result in misspecification of the null model in terms of the mean function. Baseline distribution under the null model could also be misspecified. For instance, if the true baseline distribution under the null hypothesis is negative binomial, but the working baseline distribution is assumed as the Poisson distribution.

As an empirical evidence, an intensive simulation study indicated that size of the score test behaves extremely liberal and unstable when the mean function of baseline distribution is misspecified like this. Score test for zero-inflated Poisson or zero-inflated negative binomial model does not maintain the size at the nominal level when the baseline distribution itself is also misspecified as summarized below:

Table 1: Empirical sizes of the score test statistics at the nominal level 0.05 based on 1,000 samples generated with true mean $\lambda^* = \exp(0.6 + 0.45X_1)$.

Working mean function	Base line distribution: True	Base line distribution: Working	n = 50	100	200	500	1000
Misspecification of mean							
$\log(\lambda) = \beta_0$	Poisson	Poisson	0.064	0.078	0.082	0.091	0.144
$\log(\lambda) = \beta_0 + \beta_2 X_2$	Poisson	Poisson	0.071	0.087	0.092	0.095	0.137
Misspecification of baseline distribution							
$\log(\lambda) = \beta_0 + \beta_1 X_1$	NB	Poisson	0.931	0.997	1.000	1.000	1.000
$\log(\lambda) = \beta_0 + \beta_1 X_1$	Poisson	NB	0.029	0.027	0.028	0.025	0.029

$X_1 \sim$ U(0,1) and $X_2 \sim$ N(0,1) with $X_2 \in$ (-1,1).

4. Robust score test of homegeneity

Here we develop a robust homogeneity score test for zero-inflated data. The proposed test is developed under the framework of Poisson-gamma mixture model based on the model proposed by Kassahun et al. (2014).

4.1 Kassahun's model

A more general model for zero-inflated data which can accommodate zero-inflation, over-dispersion and correlation was proposed by Kassahun et al. (2014). Let Y_{ij} be the j^{th} outcome measured for subject i. Zero-inflated over-dispersed hierarchical Poisson model by Kassahun et al. (2014) is given by

$$Y_{ij}|\theta_{ij} \sim \begin{cases} 0 & \text{with prob } \pi_{ij} \\ \text{Pois}(\lambda_{ij} = \theta_{ij}K_{ij}) & \text{with prob } 1-\pi_{ij} \end{cases},$$

where $\theta_{ij} \sim \text{Gamma}(\alpha, \beta)$ with α and β as shape and scale parameters, $K_{ij} = \exp(X_{ij}^T\eta + Z_{ij}^T b_i)$ with X_{ij} and Z_{ij} as vectors of known covariates, and η is a vector of unknown coefficients and random effects b_i is Gaussian with mean 0 and covariance matrix D.

4.2 Proposed robust test

Here we consider the case for independent data. Based on the model above, we consider that Y_i $(i = 1, ..., n)$ are independent observations from a mixture of degenerate distribution at 0 and a Poisson distribution with a random mean Λ_i such that

$$Y_i|\Lambda_i \sim \begin{cases} 0 & \text{with prob } \omega_i \\ \text{Pois}(\Lambda_i) & \text{with prob } 1-\omega_i \end{cases}$$

where ω_i is the mixture probability and $\Lambda_i \sim \text{Gamma}(\alpha,\ \beta)$ with α and β as shape and scale parameters, respectively. The marginal distribution $f_Y(y)$ is given by

$$f_Y(y) = \int_0^\infty f_{Y,\Lambda}(y,\lambda)\, d\lambda = \int_0^\infty f_{Y|\Lambda}(y|\lambda)\, f_\Lambda(\lambda)\, d\lambda.$$

Hence, the zero-inflated distribution can be re-expressed as

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i) f_{Y_i}(0) & \text{if } y_i = 0 \\ (1-\omega_i) f_{Y_i}(y_i) & \text{if } y_i = 1, 2, \ldots \end{cases}$$

where

$$f_{Y_i}(y_i) = \frac{\Gamma(y_i+\alpha)}{y_i!\,\Gamma(\alpha)} \left(\frac{\beta}{1+\beta}\right)^{y_i} \left(\frac{1}{1+\beta}\right)^{\alpha},\ y_i = 0, 1, 2, \ldots,$$

a negative binomial distribution. Then, based on this zero-inflated distribution, standard likelihood approach can be used to derive the score vector, i.e. the first order derivative of the log likelihood function along with Fisher information matrix.

5. Simulation studies

Empirical performance of the proposed test is evaluated by investigating the type I error rates of the test under the misspecification of mean and baseline distribution. To evaluate the size of the test, data are generated with sample sizes 50, 100, 200, 400, 800 and 1500 under the mean functions of the baseline distribution; $\log(\lambda^\star) = 0.6$, $\log(\lambda^\star) = 0.6 + 0.45X_1$, and $\log(\lambda^\star) = 0.6 + 0.45X_1 - 0.2X_2$ where $X1$ and $X2$ are two independent covariates with X_1 as a continuos variable with uniformly distributed values on (0,1) and X_2 as a truncated normal random variable with values on (-1,1). Two simulation studies were conducted, the first with the data generated from negative binomial distribution with the dispersion parameter $\alpha = 0.8$ and various true mean functions and the second with the data generated from Poisson distribution under the same mean functions. Each simulation is repeated 1,000 times and empirical type I error rate of the proposed robust test is evaluated at the nominal level 0.05. Power of the test is evaluated under the constant mixing weights ($\omega^\star$) 0.1, 0.2.

With the first simulation study, the proposed robust test maintains the size well around the nominal level 0.05. The power of the test increases as the sample size increases or the mixing weight ($\omega^\star$) increases. Clearly, the test is robust to any mean function used in the data generating process and performs well in all cases. Withe second simulation study, the size of the test is conservative but remains stable as the sample size increases. This result is as expected because it is well known that Poisson is a special case of negative binomial. Therefore, the tests are conservative simply due to the efficiency issue. The power of the test increases as the sample size increases and as the mixing weight increases from 0.1 to 0.2. The test is robust to any mean function used in the data generation process. This result is not surprising as the proposed test has developed under the framework of Poisson-gamma mixture model which can provide a more general framework to incorporate Poisson distribution.

6. Examples

6.1 Dental caries data

To illustrate the proposed test we use dental caries data from the Detroit Dental Health Project (DDHP) by Tellez et al. (2006). The target population of this study is low-income African-American children under age six and their main caregivers who resided in Detroit, Michigan. Although the study is longitudinal in nature, we use cross-sectional data of 897 children surveyed in the first wave of examinations conducted between 2002 and 2003. The outcome variable DS represents the number of decayed surfaces in our analysis. We compare the results of proposed test with the results from existing testing procedures proposed by van den Broek (1995) and Jansakul and Hinde (2008), assuming a constant mixing weight.

Score test proposed by van den Broek (1995) is performed by specifying the distribution under the null model as Poisson distribution and the mean of the null model as a function of covariates Age (child's age in years), SI (the child's sugar intake), and interaction term Age*SI. The score test proposed by Jansakul and Hinde (2008) is conducted by specifying the distribution under the null model as negative binomial distribution and with the same mean function. Our proposed robust test does not require the mean of the null model as a function of covariates and the test can be performed without specifying the baseline distribution specifically as Poisson distribution or negative binomial distribution.

The proposed test statistic and the score test proposed by van den Broek (1995) reject the null hypothesis

Table 2: Comparison of score test statistic, degrees of freedom and associated p-values of homogeneity tests for dental caries data.

	van den Broek test	Jansakul and Hinde test	Robust test
df	1	1	1
Test statistic	30098.21	0.0106	19.0774
p-value	<0.001	0.9178	<0.001

at 5% significance level, supporting the hypothesis of heterogeneity. The score test proposed by Jansakul and Hinde (2008) fails to reject the null hypothesis. It is worth to note that rejecting the hypothesis of homogeneity under the van den Broek (1995) test does not give evidence that zero-inflated poisson model provides a best fit for the data. The test statistic of Jansakul and Hinde (2008) under the constant mixing weight fails to reject the homogeneity hypothesis suggesting that inflation and deflation at zero appear to be of the same magnitude so that the test is not powerful enough to capture the heterogeneity in data. This situation was pointed out by Todem et al. (2012). Our test, on the other hand, provides a stronger evidence for heterogeneity in the population in the view of p-values and consistent with Todem et al. (2012).

6.2 Girl Scouts data

The use of proposed score test is illustrated with girl scouts data from the study of Scouting Nutrition and Activity Program (SNAP) by Rosenkranz et al.(2010). The objective of the study is to evaluate the effectiveness of an intervention program for girl scout troops. In this study, seven Girl Scout troops were randomized to intervention (3 troops with 34 girls) and to standard-care control (4 troops with 42 girls). Intervention troop leaders trained to implement policies promoting physical activity (PA) and healthful eating opportunities at troop meetings. At each troop meeting during seven meetings from October 2007 to April 2008, a trained research assistant observed health and nutrition promotions by troop leaders using a customized SNAP Troop Observation Form. Research assistants were blind to the condition of each troop.

For our analysis we considered the nutrition promotion activities. Number of nutrition promotions implemented by troop leaders in every 5 minutes at the troop meeting was considered as the outcome variable. We conducted our proposed score test along with the score test proposed by van den Broek (1995) and the score test proposed by Jansakul and Hinde (2008) assuming a constant mixing weight. For van den Broek (1995) test and Jansakul and Hinde (2008) test, the mean was considered as $\lambda = \exp(\beta_0 + \beta_1 X_1)$ where X_1 is an indicator variable for the intervention.

Score test proposed by van den Broek (1995) rejects the null hypothesis at 5significance level. Our proposed robust test and the score test proposed by Jansakul and Hinde (2008) fail to reject the null hypothesis. To evaluate the test results, we consider several count models and the fitted models are compared by model fit index Akaike Information Criteria (AIC). Additionally, in those count data models, the mean function is considered as $\lambda = \exp(\beta_0 + \beta_1 X_1)$ where X_1 is an indicator variable for the intervention. From Table 6.3, it is clear that negative binomial (NB) model fits data well compared to other models as the NB model indicates the smallest AIC. The treatment effect is significant under each model indicating the intervention program has a significant effect on the outcome of nutrition promotions by girl scout leaders. Additionally, we compare the observed proportion of the nutritional promotions with the fitted proportions by NB model in Figure 1. It is clear that NB model fits data well and this result is consistent with the results of robust test.

Table 3: Comparison of score test statistic, degrees of freedom and associated p-values of homogeneity tests for Girl Scout data.

	van den Broek test	Jansakul and Hinde test	Robust test
df	1	1	1
Test statistic	56.8558	0.0302	0.057
p-value	<0.001	0.8620	0.8110

7. Discussion

The proposed robust score test can address the misspecification issue under the the framework of zero-inflated models for count data. The proposed test does not require specification of the mean function or the baseline distribution. The test can be performed without specifying the baseline distribution specifically as Poisson or negative binomial, and as a result it is robust to the misspecification of the baseline distribution under zero-inflated Poisson models or zero-inflated negative binomial models, which are mostly commonly used. The proposed test might not work well if the true baseline model is binomial. It would be of interest to study the extensions of the test when the true underlying distribution is binomial or any other count distribution. This test can be further extended to incorporate correlated count data. Since the true model is often unknown a priori, robust test approach would be an efficient approach to detect heterogeneity under zero-inflated models.

References

Deng, D. and S. R. Paul (2000). Score tests for zero inflation in generalized linear models. Canadian Journal of Statistics. 28(3):563–570.

Jansakul, N. and J. Hinde (2002). Score tests for zero-inflated poisson models. Computational Statistics & Data Analysis. 40 (15):75–96.

Jansakul, N. and J. P. Hinde (2008). Score tests for extra-zero models in zero-inflated negative binomial models. Communications in Statistics-Simulation and Computation 38 (16):92–108.

Kassahun, W., T. Neyens, C. Faes, G. Molenberghs, and G. Verbeke (2014). A zero-inflated overdispersed hierarchical poisson model. Statistical Modelling 14 (5):439–456.

Rosenkranz, R. R., T. K. Behrens, and D. A. Dzewaltowski (2010). A group-randomized controlled trial for health promotion in Girl Scouts: healthier troops in a SNAP (Scouting Nutrition & Activity Program). BMC Public Health. 10:81.

Tellez, M., W. Sohn, B. Burt, and A. Ismail (2006). Assessment of the relationship between neighborhood characteristics and dental caries severity among low-income African-Americans: A multilevel approach. Journal of Public Health Dentistry. 66:30-36.

Todem, D., W.-W. Hsu, and K. Kim (2012). On the efficiency of score tests for homogeneity in two-component parametric models for discrete data. Biometrics. 68(18):975–982.

van den Broek, J. (1995). A score test for zero inflation in a poisson distribution. Biometrics. (8):738–743.