Scientific Research

A Comparison of Two Test Statistics for Poisson Overdispersion/Underdispersion

Hongyue Wang[1], Changyong Feng[1*], Xinming Tu[1], Jeanne Kowalski[2]

[1]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, USA
[2]Department of Biostatistics & Bioinformatics, Rollins School of Public Health Winship Cancer Institute,
Emory University, Atlanta, USA
Email: *feng@bst.rochester.edu

ABSTRACT

Within the family of zero-inflated Poisson distributions, the data has Poisson distribution if any only if the mean equals the variance. In this paper we compare two closely related test statistics constructed based on this idea. Our results show that although these two tests are asymptotically equivalent under the null hypothesis and are equally efficient, one test is always more efficient than the other one for small and medium sample sizes.

Keywords: Poisson Distribution; Over-Dispersion; Asymptotic Relative Efficiency

1. Introduction

The Poisson distribution is the standard model for counting data, for example, the number of telephone calls within a specific time period [1]. One stringent condition for the Poisson distribution is that the mean equals variance. However, in practice, many counting data show some overdispersion, *i.e.* the variance is greater than the mean value. The Zero-inflated Poisson (ZIP) distribution [2] and the negative binomial distribution [3,4] have been proposed to catch this overdispersion in practical data. The ZIP and it's related regression methods have been developed and used in many different areas, such as substance use [5], microbiology [6,7], psychology [8], health information management [9], dentistry [10], transportation engineering [11], and manufacturing [2].

Many tests have been proposed to test the overdispersion in counting data [1,6,12-19]. El-Shaarawi [6] compares the properties of the likelihood ratio test, the Cochran test [13], and the Rao test [17]. His simulation result indicates that the likelihood ratio is always the best to keep the significance level in the cases of small or medium sample sizes. However, the Cochan and Rao tests are much more powerful than the likelihood ratio test in those cases.

In the Zero-inflated Poisson (ZIP) distribution an extra proportion of zeros is added to the probability of zeros in the Poisson distribution. Suppose that $X \overset{d}{=} ZV$, where Z and V are independent variables with

$\Pr\{Z = 1\} = 1 - \Pr\{Z = 0\} = p$ and $V \sim \text{Poisson}(\lambda)$. Then X has the *zero-inflated Poisson distribution*. Note that

$$\Pr\{X = k\} = \begin{cases} (1-p) + p\exp(-\lambda), & k = 0, \\ p\dfrac{\lambda^k \exp(-\lambda)}{k!}, & k = 1, 2, \cdots \end{cases} \quad (1)$$

The mean and variance of X are

$$\begin{aligned} \mu &= E(X) = p\lambda, \\ \sigma^2 &= \text{var}(X) = p\lambda + p(1-p)\lambda^2. \end{aligned} \quad (2)$$

Remark 1. Although in the definition of ZIP distribution, the parameter p in the Bernoulli distribution is required to be in $(0,1)$, formula (1) always define a valid probability distribution as long as

$$p \le \frac{1}{1 - \exp(-\lambda)}. \quad (3)$$

This means that p can be greater than 1. For example, when $\lambda = 0.1$, $p = 10.1$, formula (1) still defines a valid probability distribution. The formulas for mean and variances in (mean) still hold as long as (3) is true. The valid range of (λ, p) is the range below the curve in **Figure 1**.

Remark 2. The distribution in (1) can introduce both overdispersion and underdispersion.

1) if $0 < p < 1$, then $(1-p) + p\exp(-\lambda) > \exp(-\lambda)$ and $\sigma^2 < \mu$ (over dispersion).

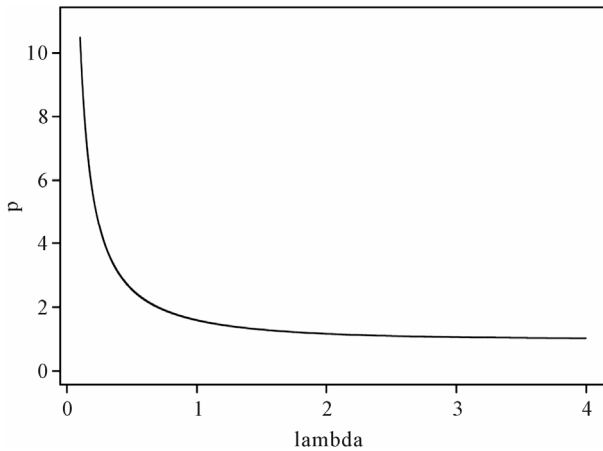2) if $1 < p < 1/(1 - \exp(-\lambda))$, then

*Corresponding author.

Figure 1. The range of *p* and *λ*.

$(1-p) + p \exp(-\lambda) < \exp(-\lambda)$ and $\sigma^2 < \mu$ (under dispersion).

Since the Poisson distribution is a special case of the distribution defined in (1), the likelihood ratio test (LRT) is a natural choice for testing the hypotheses

$$H_0 : \text{Data from Poisson distribution vs}$$
$$H_1 : \text{Data from distribution in (1),}$$ (4)

which is equivalent to

$$H_0 : p = 1 \text{ vs } H_1 : p \neq 1.$$

To construct LRT, we need to estimate two parameters λ and p. There are no closed form solutions for the score equation. Some iteration methods are called for the solution. Secondly, under the null hypothesis, the parameter p is on the boundary of the parameter region.

Many other methods have been proposed to test the hypotheses in (4). See, for example, [12] and [19]. Brown and Zhao [1] studied the hypotheses of the form

$$H_0 : X_i \sim \text{Poisson}(\lambda_i), \lambda_1 = \cdots = \lambda_n = \lambda \text{ vs}$$
$$H_1 : X_i \sim \text{Poisson}(\lambda_i), \lambda_i \text{ not all equal,}$$

and compared the behavior of their newly developed test with likelihood ratio test and several other tests. Feng *et al.* [14] derived the asymptotic distribution of the likelihood ratio test defined in [1] and corrected an error in that paper.

In this paper we construct two nonparametric test statistics and compare the efficiency, the empirical size, and power of two closely related tests, especially for the cases of small and medium sample sizes.

2. Test Statistics

From (1) we know that the data is from Poisson if and only if $\sigma^2 = \mu$. We can construct a test to study the difference between the sample mean and sample variance. Suppose $X_i, i = 1, \cdots n$, is a random sample. Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

be the sample mean and sample variances, respectively. Simple algebra shows that

$$\sqrt{n}S_n^2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 + o_p\left(n^{-1/2}\right).$$

Therefore

$$\sqrt{n}\left(S_n^2 - \bar{X}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(X_i - \mu\right)^2 - X_i\right] + o_p\left(n^{-1/2}\right).$$

Let $W_i = \left(X_i - \mu\right)^2 - X_i$. Under the null hypothesis

$$E\left(W_i\right) = 0, \text{Var}\left(W_i\right) = 2\lambda^2.$$

From central limit theorem,

$$\frac{\sqrt{n}\left(S_n^2 - \bar{X}\right)}{\sqrt{2}\lambda} \xrightarrow{d} N(0,1) \text{ as } n \to \infty. \quad (5)$$

The asymptotic result in (5) still holds with λ replaced by a consistent estimator of λ.

Under the null hypothesis, both $\bar{X}$ and S_n^2 are consistent estimators of λ. In fact, $\bar{X}$ is the MLE of λ under H_0. Based on this idea, we define two test statistics

$$T_1 \equiv \frac{\sqrt{n}\left(S_n^2 - \bar{X}\right)}{\sqrt{2}\bar{X}} = \sqrt{\frac{n}{2}}\left(\frac{S_n^2}{\bar{X}} - 1\right), \quad (6)$$

$$T_2 \equiv \frac{\sqrt{n}\left(S_n^2 - \bar{X}\right)}{\sqrt{2}S_n^2} = \sqrt{\frac{n}{2}}\left(1 - \frac{\bar{X}}{S_n^2}\right). \quad (7)$$

Since the exact variance of $S_n^2 - \bar{X}$ is $2\lambda^2/(n-1)$ [Bo], the test statistic proposed by Böhning [12] is

$$T_3 \equiv \frac{S_n^2 - \bar{X}}{\sqrt{2\bar{X}^2/(n-1)}} = \sqrt{\frac{n-1}{2}}\left(\frac{S_n^2}{\bar{X}} - 1\right),$$

which is asymptotically equivalent to T_1. This test is also called *Neyman-Scott test* in [1].

In the next section we study the relative efficiency of T_1 and T_2 and compare their empirical size and power for small and medium sample sizes.

3. Comparison of T_1 and T_2

Note that algebraically these two tests satisfy the relation

$$T_2 = \frac{T_1}{1 + T_1\sqrt{2/n}}.$$

If $T_1 > 0$ or $T_1 < -\sqrt{2n}$, then $|T_1| > |T_2|$. Otherwise $|T_1| \leq |T_2|$.

In this section we study the Pitman asymptotic relative efficiency (ARE) of T_1 with respect to T_2. The ARE is a large sample property of a test statistic. We also com-

pare the empirical sizes and powers of these two test statistics by simulations for the cases of small and medium sample sizes if the asymptotic distributions are used in those cases.

3.1. Relative Efficiency

In this subsection we study the Pitman efficiency of T_1 with respect to T_2. Note that

$$T_1 = \sqrt{\frac{n}{2}}\left(\frac{S_n^2}{\bar{X}} - \frac{\mu_1}{\mu_2}\right) + \sqrt{\frac{n}{2}}\left(\frac{\mu_1}{\mu_2} - 1\right),$$

$$T_2 = \sqrt{\frac{n}{2}}\left(\frac{\mu_2}{\mu_1} - \frac{\bar{X}}{S_n^2}\right) + \sqrt{\frac{n}{2}}\left(1 - \frac{\mu_2}{\mu_1}\right).$$

Let

$$\Delta(p) = \mu_2^2\sigma_1^2 + \mu_1^2\sigma_2^2 - 2\mu_1\mu_2\sigma_{12}$$

and

$$M_1(p) = \frac{\mu_1}{\mu_2} - 1 = (1-p)\lambda,$$

$$M_2(p) = 1 - \frac{\mu_2}{\mu_1} = \frac{(1-p)\lambda}{1+(1-p)\lambda}. \tag{8}$$

The asymptotic variances of T_1 and T_2 are

$$\Sigma_1(p) = \frac{\Delta(p)}{2\mu_2^4} \text{ and } \Sigma_2(p) = \frac{\Delta(p)}{2\mu_1^4},$$

respectively. The Pitman efficiency of T_1 with respect to T_2 is

$$e(1,2) = \frac{\left[M_1'(1)\right]^2}{\Sigma_1(1)} \cdot \frac{\Sigma_2(1)}{\left[M_2'(1)\right]^2} = 1.$$

This means that these two test statistics have the same efficiency in the large sample case.

3.2. Empirical Sizes and Powers

In this part we compare the empirical sizes and powers by simulations when the asymptotic distributions of T_1 and T_2 are used in the cases of small and medium sample sizes. The theoretical significance level is set at 0.05. We compare T_1 and T_2 for different p and λ. **Table 1** shows the simulation results (from 100,000 Monte Carlo repetitions).

Empirical size: Except for very small λ (for example, $\lambda = 0.1$), the empirical size of T_1 is very close to the theoretical significance level, even when $n = 20$. On the other hand, the empirical size of T_2 is far away from the theoretical value. For example, when $\lambda = 0.1$, the empirical size of T_2 is 0.03 even when $n = 500$. When $\lambda = 1$, 2 and 5, the empirical sizes of T_2 are well above the theoretical value for sample size $n = 50$.

Empirical Power: As p decreases, the distribution goes far and far away from the Poisson distribution. The powers of these two test statistics increase as p decreases. However, the power of T_1 increases much faster than that of T_2. For example, when $p = 0.7$, $\lambda = 1.0$ and $n = 100$, the powers are 0.519 and 0.297, respectively. When $p = 0.9$, $\lambda = 2.0$ and $n = 200$, the powers are 0.498 and 0.337, respectively.

The simulation results show that although these two test statistics are asymptotically equivalent under H_0, and have the same efficiency in the large sample case, T_1 is more efficient than T_2 in the small and medium sample cases.

Since

$$T_2 = T_1 - \sqrt{\frac{2}{n}}T_1^2 + \frac{2}{n}T_1^3 + O_p(1), \tag{9}$$

we also compare the behave of

$$T_4 = T_1 - \sqrt{\frac{2}{n}}T_1^2 + \frac{2}{n}T_1^3$$

with T_1 and T_2 (not reported here). We find that T_4 is very similar to T_1. This means the remainder term in (9) plays a very significant role in the cases of small and medium sample sizes.

4. Real Data Study

In this section, we apply these two test statistics to four real data sets. These data sets have sample sizes from relatively small $n = 44$ to relatively large $n = 539$. The results are summarized in **Table 2**.

Example 1: This data set is used in [1]. It contains the number of daily calls for standard services between 4:30 pm to 4:45 pm in an Israel call center within 44 consecutive days. More information about this data set can be found in Section 2 of [1]. We want to test if the data has a Poisson distribution. Here we assume that the number of calls from different days are independent. The sample mean and sample variances are 18.66 and 25.95. The p-value of T_1 is 0.07, which shows marginally significant overdispersion of the data. This is consistent with our impression. See Figure 2 in [1] for the histogram of the data. However, the p-value of T_2 is 0.19.

Example 2: This is another data set used in [1]. It contains the number of daily calls for internet services between 4:30 pm to 4:45 pm in an Israel call center within 107 consecutive days. Here we also assume that the number of calls from different days are independent. The sample mean and sample variances are 2.18 and 2.47. The p-values of T_1 and T_2 are 0.33 and 0.39, which show the Poisson distribution is a good approximation for the data. This is consistent with our impression. See Figure 1 in [1] for the histogram of the data.

Table 1. Power comparison of T_1 and T_2. Results were obtained from 100,000 MC replications.

n	Binomial parameter (p)							
	$p = 1.0$		$p = 0.9$		$p = 0.7$		$p = 0.5$	
	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
$\lambda = 0.1$								
20	0.021	0.001	0.023	0.001	0.028	0.001	0.032	0.001
50	0.044	0.007	0.052	0.008	0.067	0.012	0.077	0.017
100	0.045	0.016	0.056	0.021	0.090	0.031	0.123	0.051
200	0.044	0.023	0.058	0.032	0.090	0.056	0.136	0.095
500	0.049	0.032	0.057	0.042	0.109	0.084	0.177	0.139
$\lambda = 1.0$								
20	0.047	0.087	0.080	0.048	0.189	0.018	0.324	0.026
50	0.046	0.067	0.096	0.034	0.316	0.091	0.587	0.268
100	0.049	0.059	0.126	0.052	0.519	0.297	0.848	0.674
200	0.048	0.052	0.181	0.100	0.790	0.654	0.985	0.964
500	0.048	0.050	0.351	0.264	0.991	0.982	1.000	1.000
$\lambda = 2.0$								
20	0.046	0.092	0.122	0.027	0.438	0.028	0.746	0.126
50	0.048	0.068	0.196	0.044	0.765	0.408	0.979	0.853
100	0.050	0.049	0.303	0.129	0.961	0.869	1.000	0.999
200	0.049	0.054	0.498	0.337	1.000	0.998	1.000	1.000
500	0.049	0.051	0.854	0.789	1.000	1.000	1.000	1.000
$\lambda = 5.0$								
20	0.050	0.099	0.387	0.037	0.930	0.426	0.998	0.889
50	0.050	0.071	0.631	0.311	0.999	0.988	1.000	1.000
100	0.050	0.061	0.855	0.700	1.000	1.000	1.000	1.000
200	0.051	0.056	0.983	0.962	1.000	1.000	1.000	1.000
500	0.050	0.052	1.000	1.000	1.000	1.000	1.000	1.000

Table 2. Results from real data.

Example	n	$\overline{X}$	S_n^2	T_1		T_2	
				Statistic	p-value	Statistic	p-value
1	44	18.66	25.95	1.83	0.07	1.32	0.19
2	107	2.18	2.47	0.97	0.33	0.86	0.39
3	366	6.36	6.82	0.98	0.33	0.91	0.36
4	539	12.03	184.52	235.44	<0.0001	15.35	<0.0001

Example 3: This data set is reported in [12]. It contains the number of daily deaths of women with brain vessel disease during the year 1989 in West Germany. The sample mean and sample variances are 6.36 and 6.82. The *p*-values of T_1 and T_2 are 0.33 and 0.36, which shows the Poisson distribution is a good approximation for the data.

Example 4: This data is from a HIV prevention study finished at the University of Rochester School of Nursing. The study participants were 621 sexually active girls of ages 15 - 19 years. For more details about the study, please refer to [7]. One of the primary outcomes is the number of unprotected vaginal sex over the past 3 months. After 3 months of intervention, the number of unprotected vaginal sex wase available for 539 girls. The sample mean and sample variances are 12.03 and 184.52. Although both tests show very significant over-disper- sion for the data set (*p*-values < 0.0001), the values of the test statistics are very significantly different with $T_1 = 235.44$ and $T_2 = 15.35$. This phenomena can also be seen from (means) as $M_1(p)$ can be arbitrarily large, while $M_2(p)$ is bounded above by 1.

5. Discussion

In this paper we compare two test statistics which can be easily used to test the Poisson distribution versus the zero-inflated Poisson distributions. Both test statistics are asymptotically equivalent under null hypothesis and the relative Pitman efficiency is 1. However, they have very significantly different behaviors for small and medium sample sizes. While T_1 always has reasonable empirical size (under null hypothesis) and power (under alternative hypothesis) for small and medium sample sizes, T_2 shows some erratic behaviors even for medium sample sizes and may lead to wrong conclusion in practice (example 1). Therefore we should never use it in practice.

REFERENCES

[1] L. D. Brown and L. H. Zhao, "A Test for the Poisson Distribution," *Sankhyā*, Vol. 64, pp. 611-625

[2] D. Lambert, "Zero-Inflated Poisson Regression Models with an Application to Defects in Manufacturing," *Technometrics*, Vol. 34, No. 1, 1992, pp. 1-14. doi:10.2307/1269547

[3] R. A. Fisher, "The Negative Binomial Distribution," *Annals of Human Genetics*, Vol. 11, No. 1, 1941, pp. 182-187. doi:10.1111/j.1469-1809.1941.tb02284.x

[4] G. J. S. Ross and D. A. Preece, "The Negative Binomial Distribution," *The Statistician*, Vol. 34, No. 3, 1985, pp. 323-336. doi:10.2307/2987659

[5] S. M. DeSantis and D. Bandyopadhyay, "Hidden Markov Models for Zero-Inflated Poisson Counts with an Appli-

cation to Substance Use," *Statistics in Medicine*, Vol. 30, 2011, pp. 1678-1694. doi:10.1002/sim.4207

[6] A. El-Shaarawi, "Some Goodness-of-Fit Methods for the Poisson Plus Added Zeros Distribution," *Applied and Environmental Microbiology*, Vol. 49, 1985, pp. 1304-1306.

[7] Y. Xia, D. Morrison-Beedy, J. Ma, C. Feng, W. Cross and X. M. Tu, "Modeling Count Outcomes from HIV Risk Reduction Interventions: A Comparison of Competing Statistical Models for Count Responses," *AIDS Research and Treatment*, Vol. 2012, 2012, Article ID 593569. doi:10.1155/2012/593569

[8] T. Loeys, B. Moerkerke, O. De Smet, *et al.*, "The Analysis of Zero-Inflated Count Data: Beyond Zero-Inflated Poisson Regression," *British Journal of Mathematical and Statistical Psychology*, Vol. 65, No. 1, 2012, pp. 163-180. doi:10.1111/j.2044-8317.2011.02031.x

[9] A. Khan and M. Western, "Does Attitude Matter in Computer Use in Australian General Practice? A Zero-Inflated Poisson Regression Analysis," *Health Information Management Journal*, Vol. 40, 2011, pp. 23-29.

[10] B. T. Pahel, J. S. Preisser, S. C. Stearns, *et al.*, "Multiple Imputation of Dental Caries Data Using a Zero-Inflated Poisson Regression Model," *Journal of Public Health Dentistry*, Vol. 71, No. 1, 2011, pp. 71-78. doi:10.1111/j.1752-7325.2010.00197.x

[11] S. R. Hu, C. S. Li and C. K. Lee, "Assessing Casualty Risk of Railroad-Grade Crossing Crashes Using Zero-Inflated Poisson Models," *Journal of Transportation Engineering-ASCE*, Vol. 137, No. 8, 2011, pp. 527-536. doi:10.1061/(ASCE)TE.1943-5436.0000243

[12] D. Böhning, "A Note on a Test for Poisson Overdispersion," *Biometrika*, Vol. 81, No. 2, 1994, pp. 418-419. doi:10.1093/biomet/81.2.418

[13] W. Cochran, "Some Methods of Strengthening χ2 Tests," *Biometrics*, Vol. 10, No. 4, 1954, pp. 417-451. doi:10.2307/3001616

[14] C. Feng, H. Wang and X. M. Tu, "The Asymptotic Distribution of a Likelihood Ratio Test for the Poisson Distribution," *Sankhyā*, 2012 (in press)

[15] C. S. Li, "A Lack-of-Fit Test for Parametric Zero-Inflated Poisson Models," *Journal of Statistical Computation and Simulation,* Vol. 81, No. 9, 2011, pp. 1081-1098. doi:10.1080/00949651003677410

[16] C. S. Li, "Testing the Lack-of-Fit of Zero-Inflated Poisson Regression Models," *Communication in Statistics-Simulation and Computation*, Vol. 40, No. 4, 2011, pp. 497-510. doi:10.1080/03610918.2010.546541

[17] C. Rao and I. Chakravarti, "Some Small Sample Tests of Significance for a Poisson Distribution," *Biometrics*, Vol. 12, No. 3, 1956, pp. 264-282. doi:10.2307/3001466

[18] J. Van den Broeck, "A Score Test for Zero Inflation in a Poisson Distribution," *Biometrics*, Vol. 51, 1995, pp. 738-743.

[19] O. Thas and J. C. W. Rayner, "Smooth Tests for the Zero-Inflated Poisson Distribution," *Biometrics*, Vol. 61, 2005, pp. 808-815. doi:10.1111/j.1541-0420.2005.00351.x