In: Journal of Applied Statistical Science
Volume 18, Number 3, pp. 1–8
ISSN: 1067-5817

Conditional and Unconditional Tests for Comparing Several Poisson Means

***Jie Peng*[1*] *and K. Krishnamoorthy*[2†]**
[1]Department of Finance, Economics and Decision Science
St. Ambrose University
Davenport, Iowa 52803
[2]Department of Mathematics
University of Louisiana at Lafayette
Lafayette, LA 70504

Abstract

In this article, we consider the problem of testing equality of several Poisson means. Two tests, one based on the exact conditional distribution of the sample counts and another based on the parametric bootstrap (PB) approach, are considered. The conditional test is exact in the sense that its type I error rates never exceed the nominal level. These tests are compared with the Pearson chi-square test and Brown–Zhao test. Our comparison studies indicate that the conditional test and the PB test work satisfactorily even for small samples and/or small values of Poisson mean. The PB test offers more power than other tests in some situations. The tests are illustrated using an example.

Keywords and phrases: Exact conditional test; Power; Type I error rates; Variance stabilizing transformation.

1 Introduction

Poisson distribution is appropriate to describe the distribution of counts of rare events, and so it is sometimes referred to as the law of rare events. In general, this model is commonly used to study the number of random occurrences of an event over a time interval or a specified space. If the mean rate of occurrence of an event is λ, then the probability distribution of the number of occurrences X can be modeled by a Poisson distribution with mean λ, say, Poisson(λ). In this case, the probability mass function of X is given by

$$P(X = x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0.$$

Let $X_{i1}, \ldots, X_{in_i}$ be a sample from a Poisson(λ_i) distribution, $i = 1, \ldots, m$. Let $Y_i = \sum_{j=1}^{n_i} X_{ij}$ so that $Y_i \sim$ Poisson($n_i\lambda_i$), $i = 1, \ldots, m$. We consider testing

$$H_0 : \lambda_1 = \cdots = \lambda_m \ \text{ vs. } \ H_a : \lambda_i \neq \lambda_j \ \text{ for some } i \neq j \tag{1}$$

*E-mail address: PengJie@sau.edu
†E-mail address: krishna@louisiana.edu

based on the total counts $Y_1, \ldots, Y_m$.

The problem of estimating or testing the ratio of (or the difference between) two Poisson means commonly arises in clinical studies. For example, one may want to compare the incident rates of a disease in a treatment group and control group, where the incident rate is defined as the number of events (such as death) observed divided by the time at risk during the observed period. Several articles have been written on comparison of two Poisson means for the past seven decades; see Przyborowski and Wilenski (1940), Chapman (1952), Cox (1953), Jaech (1970), Krishnamoorthy and Thomson (2004) and Stamey and Katsis (2007), and the references therein.

The problem of comparing several Poisson rates arises in many practical situations. For example, in clinical trials one may want to compare the incidence rates (such as deaths or adverse symptoms) among several age groups of patients. Nelson et al. (2005, p. 30) provide an example where the arrival rates (per hour) of patients to six urgent clinics run by an HMO are to be compared based on samples of arrival counts from these clinics. Brown and Zhao (2002) provide a specific situation where it is of interest to compare the averages of number of service request calls per day at different call centers. Recently, Chiu and Wang (2009) provided an example where death rates of patients from four groups after heart valve replacement are compared.

By far the most common test for comparing several Poisson means is the usual χ^2-test based on the statistic

$$\chi^2 = \sum_{i=1}^{m} \frac{n_i(\widehat{\lambda}_i - \widehat{\lambda})^2}{\widehat{\lambda}}, \tag{2}$$

where $\widehat{\lambda}_i = \frac{Y_i}{n_i}$ and $\widehat{\lambda} = \frac{\sum_{i=1}^{m} Y_i}{\sum_{i=1}^{m} n_i}$. Under H_0, the above χ^2 statistic follows a χ^2_{m-1} distribution as $n_i \to \infty$. Brown and Zhao (2002) proposed a new test based on Anscombe's (1948) variance stabilizing transformation, and compared it with the χ^2-test, Neyman–Scott test and the likelihood ratio test. On the basis of their limited comparison study, Brown and Zhao have concluded that the Neyman–Scott test and the likelihood ratio test perform poorly with respect to type I error rates while their test is reasonably accurate if the means are not too small. Recently, Chiu and Wang (2009) compared a few asymptotic tests and their parametric bootstrap versions with respect to type I error rates and powers. On the basis of extensive simulation studies they concluded that none of the tests dominates others uniformly with respect to type I error rates or powers, and the Pearson χ^2-test is satisfactory for moderate to large values of mean and/or large sample sizes.

In this article, we consider two tests. One of the tests is based on the conditional distribution of $Y_1, \ldots, Y_m$ given the total $\sum_{i=1}^{m} Y_i$. As the conditional distribution $Y_1, \ldots, Y_m$ under H_0 in (1) is multinomial with the ith cell probability $n_i / \sum_{j=1}^{m} n_j$, p-values of the conditional test can be evaluated using multinomial probabilities, and so the conditional test is exact. Another test is based on the parametric bootstrap (PB) approach which utilizes the samples generated from Poisson distributions with estimated parameters, and this test is referred to as the PB test. In the following section, we describe all the tests and their null distributions. In Section 3, we estimate the type I error rates and powers of the tests using Monte Carlo simulation, and make some recommendations regarding the choice of the tests for applications. An example with real data is given in Section 4.

2 The Tests

Let $Y_1, \ldots, Y_m$ be independent Poisson random variables with mean $n_i\lambda_i$, $i = 1, \ldots, m$. In the following we shall outline the tests for (1) based on $Y_1, \ldots, Y_m$.

2.1 The Brown–Zhao Test

Brown and Zhao (2002) proposed the following test on the basis of Anscombe's (1948) variance stabilizing transformation. Let $Z_i = \sqrt{Y_i + 3/8}$, $i = 1, \ldots, m$, and $\bar{Z} = \frac{1}{m}\sum_{i=1}^{m} Z_i$. Anscombe has shown that $\mathrm{Var}(Z_i) = 1/4 + O(1/\lambda_i)$. Let $\nu(\lambda_i) = E(Z_i)$, $i = 1, \ldots, m$, so that $\nu(\lambda_i) = \nu(\lambda)$ under H_0 in (1), where λ is the unknown common mean under H_0. Thus, when λ_i's are large, we can regard $Z_1, \ldots, Z_m$ as independent random variables with $N(\nu(\lambda), 1/4)$ distribution, and as a result

$$4\sum_{i=1}^{m}(Z_i - \bar{Z})^2 \sim \chi^2_{m-1} \tag{3}$$

approximately under H_0. This test rejects H_0 if $4\sum_{i=1}^{m}(Z_i - \bar{Z})^2 > \chi^2_{m-1;1-\alpha}$. Brown and Zhao have noted that the above approximation is reasonably accurate even for fairy small λ's and m. The above test to the case of unequal sample sizes can not be applied, because in this case, equality of $\nu(\lambda_i)$'s does not necessarily imply the equality of the λ_i's.

2.2 Approximate and Exact Conditional Tests

It is easy to see that the conditional distribution of $Y_1, \ldots, Y_m$ given $\sum_{i=1}^{m} Y_i = T$ is multinomial with the probability mass function given by

$$P\left(Y_1 = y_1, \ldots, Y_m = y_m \,\middle|\, \sum_{i=1}^{m} Y_i = T\right) = \frac{T!}{y_1! \cdots y_m!} p_1^{y_1} \cdots p_m^{y_m}, \tag{4}$$

where $p_i = n_i\lambda_i / \left(\sum_{j=1}^{m} n_j\lambda_j\right)$, $i = 1, \ldots, m$. Letting $p_{i0} = n_i / \left(\sum_{j=1}^{m} n_j\right)$, we see that, conditionally given T, testing H_0 in (1) is equivalent to testing

$$H_0: \; p_1 = p_{10}, \ldots, p_{m-1} = p_{m-1,0} \;\text{ vs. }\; H_a: p_i \neq p_{i0} \;\text{ for some } i. \tag{5}$$

For testing (5), the popular chi-square statistic

$$\chi^2_c = \sum_{i=1}^{m} \frac{(Y_i - Tp_{i0})^2}{Tp_{i0}} \tag{6}$$

was introduced by Pearson (1900), which has the χ^2_{m-1} distribution asymptotically. It is interesting to note that the χ^2 statistic in (2) and the above χ^2_c statistic are the same, and so the asymptotic unconditional test and the asymptotic conditional test are the same. We simply refer to this asymptotic test as the χ^2-test.

Instead of using the chi-square approximation, we can compute the exact p-value of the conditional test using multinomial probabilities. In particular, the p-value can be computed using the expression

$$P(\chi^2 \geq \chi_0^2 | H_0) = \sum_{y_1=0}^{T} \sum_{y_2=0}^{T-y_1} \cdots \sum_{y_{m-1}=0}^{T-y_1-\cdots-y_{m-2}} \frac{T!}{y_1! \cdots y_m!} p_{10}^{y_1} \cdots p_{m0}^{y_m} I\left[\chi^2 \geq \chi_0^2\right], \tag{7}$$

where $p_{i0} = n_i / \sum_{j=1}^{m} n_j$, χ_0^2 is an observed value of the χ^2 statistic in (2) and $I[.]$ is the indicator function. This test rejects the H_0 in (1) whenever the above p-value is less than or equal to α. Note that this test is exact in the sense that the Type I error rates never exceed the nominal level α. We refer to the test based on the p-value in (7) as the conditional test.

Remark 1 The p-value in (7) is often difficult to compute when the number of means to be compared is moderate to large. In such cases, the conditional p-values can be estimated by simulating samples from a multinomial distribution with cell probabilities $p_{10}, \ldots, p_{m0}$ and the number of trials T. We shall describe this Monte Carlo approach in the following algorithm.

Algorithm 1

1. For a given m, sample sizes $n_1, \ldots, n_m$ and sample counts $y_1, \ldots, y_m$, compute the χ^2 statistic using (2) or (6); this is an observed value of the χ^2 statistic, and is denoted by χ_0^2.
2. Compute $p_{i0} = n_i / \sum_{j=1}^{m} n_j$, $i = 1, \ldots, m$.
3. Generate a large number (say, 10000) samples from the multinomial distribution with the above cell probabilities p_{i0}'s and the total number of trials T.
4. For each of the above generated samples, compute the χ^2 statistic using (6).
5. The proportion of the χ^2 statistics in step 4 that are greater than or equal to χ_0^2 in step 1 is a Monte Carlo estimate of the conditional p-value.

2.3 A Parametric Bootstrap Test

The PB test is based on the unconditional parametric bootstrap samples. This test utilizes the simulated samples from Poisson distributions with the estimated parameters $\widehat{\lambda}_i = \frac{Y_i}{n_i}$, $i = 1, \ldots, m$, where y_i is the total count from the ith sample. To handle zero count, we set $Y_i = 0.5$ when it is actually zero. The PB p-value can be obtained using the following algorithm.

Algorithm 2

1. For a given m and observed counts $y_1, \ldots, y_m$, compute the χ^2 statistic in (2), and denote it by χ_0^2.
2. Compute $\widehat{\lambda}_i = \frac{y_i}{n_i}$, $i = 1, \ldots, m$.
3. Generate $Y_i \sim$ Poisson$(n_i \widehat{\lambda}_i)$, $i = 1, \ldots, m$ and compute the chi-square statistic (6).
4. Repeat the step 3 for a large number of times, say, 10000.
5. The proportion of the χ^2 statistics that are greater than or equal to χ_0^2 in step 1 is the PB p-value.

3 Monte Carlo Estimates of Sizes and Powers

We estimated the type I error rates of the tests using Monte Carlo simulation when $n_1 = \cdots = n_m = 1$ and values of m considered in Table 1 of Brown and Zhao's (2002) paper. The estimated type I error rates are reported in Table 1a. We first observe from these table values that the Brown–Zhao (BZ) test is too conservative (the type I error rates are much smaller than the nominal levels) when λ's are unity. This implies that the BZ test could be very conservative for small values of λ's.

Table 1a. Monte Carlo Estimates of Type I Error Rates of the Tests $n_1 = \cdots = n_m = 1$; $\lambda_1 = \cdots = \lambda_m = \lambda$

			Tests			
m	λ	α	χ^2	BZ	Conditional	PB
20	1	.10	.0887	.0008	.0836	.0886
		.05	.0469	.0001	.0420	.0464
		.01	.0112	.0000	.0092	.0112
	5	.10	.1021	.1071	.0920	.0978
		.05	.0486	.0601	.0500	.0504
		.01	.0090	.0142	.0084	.0099
12	1	.10	.0893	.0024	.0900	.1040
		.05	.0491	.0005	.0372	.0456
		.01	.0105	.0001	.0092	.0104
	5	.10	.0990	.1131	.1000	.1060
		.05	.0505	.0601	.0428	.0432
		.01	.0111	.0146	.0100	.0115
5	1	.10	.1058	.0099	.0556	.0964
		.05	.0264	.0036	.0236	.0372
		.01	.0055	.0002	.0056	.0096
	5	.10	.0942	.1045	.0980	.1012
		.05	.0400	.0552	.0487	.0524
		.01	.0075	.0118	.0099	.0110

In Table 1b, we reported the type I error rates of the tests when the sample sizes are unequal. As noted earlier, Brown–Zhao's test is not available for the case of unequal sample sizes, and so it is not included in Table 1b. We observe from Tables 1a and 1b that the χ^2, conditional and PB tests exhibit similar performances in both Tables 1a and 1b. Among these three tests, the conditional test seems to be conservative when the values of common λ under H_0 are small. Over all, the χ^2, conditional and PB tests perform satisfactorily even for small values of λ.

Regarding power comparison among the tests, our preliminary power calculations (not reported here) showed that the χ^2-test, the conditional test and the PB test are equally powerful for large samples. The estimated powers of the χ^2, conditional and PB tests at the level 0.05 are given in Table 2. It is clear that the powers of the conditional test are always smaller than the PB test which is based on the unconditional joint sampling distribution. In some situations the PB test offers more power than the χ^2 test, and in other cases the powers of these two tests are not appreciably different.

On an overall basis, we can recommend the PB test followed by the χ^2 test; however, if simplicity is important, then the χ^2 test is preferable to the PB test. Finally, if controlling type I error rates is important, then the conditional test should be used.

Table 1b. Monte Carlo Estimates of Type I Error Rates of the Tests $m = 5$; $\lambda_1 = \cdots = \lambda_m = \lambda$; $\alpha = 0.05$

		Tests		
λ	$(n_1, \ldots, n_5)$	χ^2	Conditional	PB
.5	(5, 12, 5, 14, 22)	.050	.048	.050
	(10, 20, 30, 6, 40)	.050	.046	.049
	(30, 40, 30, 40, 30)	.048	.048	.049
	(4, 8, 12, 3, 20)	.053	.049	.049
1	(5 , 5, 5, 5, 5)	.049	.046	.051
	(8, 8, 8, 8, 8)	.050	.044	.046
	(10, 10, 10, 10, 10)	.048	.049	.050
	(10, 20, 30, 10, 15)	.049	.052	.053
	(20, 25, 20, 30, 20)	.049	.050	.050
4	(2, 2, 2, 2, 2)	.048	.050	.051
	(3, 3, 3, 3, 3)	.040	.049	.050
	(5, 4, 8, 3, 12)	.050	.046	.044
	(10, 14, 12, 22, 50)	.051	.050	.052

Table 2. Monte Carlo Estimates of Powers of the Tests when $\alpha = 0.05$

		Tests		
m	$(\lambda_1, \ldots, \lambda_m)$	χ^2	Conditional	PB
3	(1,2,3)	.1026	.0824	.1088
	(1,1,5)	.4396	.3900	.4496
	(1,4,8)	.5754	.5300	.5624
	(1,1,8)	.7962	.7568	.7968
	(1,4,15)	.9623	.9640	.9708
6	(1,2,3,4,5,6)	.3335	.3280	.3408
	(.5,.5,.5,.5,.5,3)	.2912	.2936	.3696
	(1,2,1,2,1,5)	.3653	.3776	.4024
	(2,5,1,3,2,6)	.3988	.4060	.4212
	(2,2,4,4,7,7)	.4062	.4092	.4188
	(3,6,2,3,5,1)	.3456	.3444	.3576
	(4,5,4,5,2,6)	.1447	.1508	.1536
	(.5,2,1,.4,3,5)	.4906	.4824	.5164

4 An Example

The data for this example are from Laird and Olivier (1981), which were also used by Chiu and Wang (2009) to illustrate test procedures for comparing Poisson means. To compare survival rates after heart valve replacement, a sample of 109 patients were observed, and they were classified by valve (aortic or mitral) and by age (< 55 or ≥ 55). For younger patients in the sample, 4 deaths in 1259 months of observation have occurred with aortic valve replacement, 1 death in 2082 months observation has occurred with mitral valve replacement; for older patients, 7 deaths in 1417 months of observation have occurred with aortic valve replacement, and 9 deaths in 1647 months of observation have occurred with mitral valve replacement. In this study, each patient at risk was observed until she/he died or the study ended. We shall use the χ^2-test, the conditional test and the PB test for comparing death rates of four groups. To apply the tests, we first note that $(y_1, n_1) = (4, 1259)$, $(y_2, n_2) = (1, 2082)$, $(y_3, n_3) = (7, 1417)$ and $(y_4, n_4) = (9, 1647)$. The χ^2-statistic in (2) is 8.57, and the p-value $P(\chi_3^2 > 8.57)$ is .036. The p-value of the conditional test is .034 and the p-value of the PB test is .032. The p-values of the conditional test and the PB test were estimated by Monte Carlo simulation using 10,000 runs. Note that, for this example, observed number of months for the four groups are large, and so the p-values of all three tests are practically the same. Thus, all the tests indicate that death rates among the four groups of patients are significantly different.

References

Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrics*, 35, 246–254.

Brown, L. D. and Zhao, L. H. (2002). A test for the Poisson distribution. *Shankya, Ser. A*, 64, 611–625.

Chapman, D. G. (1952). On tests and estimates for the ratio of Poisson means. *Annals of the Institute of Statistical Mathematics*, 4, 45–49.

Chiu, S. N. and Wang, L. (2009). Homogeneity tests for several Poisson populations. *Computational Statistics and Data Analysis*, 53, 4266–4278.

Cox D. R. (1953). Some simple approximate tests for Poisson variates. *Biometrika*, 40, 354-360.

Jaech J. L. (1970). Comparing two methods of obtaining a confidence interval for the ratio of Poisson parameters. *Technometrics*, 12, 383–387.

Krishnamoorthy, K. and Thomson, J. (2004). A more powerful test for comparing two Poisson means. *Journal of Statistical Planning and Inference*, 119, 23–35.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76, 231–240.

Nelson, P. R., Wludyka, P. S. and Copeland, K. A. F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*, Philadelphia: SIAM.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine*, 50, 157–175.

Przyborowski, J. and Wilenski, H. (1940). Homogeneity of results in testing samples from Poisson series. *Biometrika*, 31, 313–323.

Stamey, J. and Katsis, A. (2007). Sample size determination for comparing two Poisson rates with underreported counts. *Communications in Statistics-Simulation and Computation*, 36, 483–492.