Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference 119 (2004) 23–35

journal of statistical planning and inference

www.elsevier.com/locate/jspi

A more powerful test for comparing two Poisson means

K. Krishnamoorthy*, Jessica Thomson

Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA

Received 9 May 2002; received in revised form 15 July 2002

Abstract

The problem of hypothesis testing about two Poisson means is addressed. The usual conditional test (C-test) and a test based on estimated p-values (E-test) are considered. The exact properties of the tests are evaluated numerically. Numerical studies indicate that the E-test is almost exact because its size seldom exceeds the nominal level, and it is more powerful than the C-test. Power calculations for both tests are outlined. The test procedures are illustrated using two examples.

MSC: 62F03

Keywords: Binomial distribution; Exact power; Exact size; Parametric bootstrap; Poisson distribution; Sample size calculation; Standardized difference

1. Introduction

The Poisson distribution is well suited to model many processes in a broad variety of fields such as agriculture, ecology, biology, medicine, commerce, industrial quality control and particle counting to name just a few. Specific examples are given, among others, in Przyborowski and Wilenski (1940), Gail (1974), Shiue and Bain (1982), Nelson (1991), Sahai and Misra (1992), and Sahai and Khurshid (1993). A Poisson model is appropriate in a situation where we count the number of events X in a unit interval of time or on an object. If the mean rate of occurrence of events is λ, then the probability distribution of X can be modeled by a Poisson distribution with mean

* Corresponding author. Tel.: +1-337-482-5283; fax: +1-337-482-5346.
E-mail address: krishna@louisiana.edu (K. Krishnamoorthy).

PII: S0378-3758(02)00408-1

λ, say, Poisson(λ). In this case, the probability mass function of X is given by

$$P(X=x|\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}, \quad x=0,1,2,\ldots, \quad \lambda>0. \tag{1.1}$$

Let $X_{11},\ldots,X_{1n_1}$ and $X_{21},\ldots,X_{2n_2}$ be independent samples, respectively, from Poisson(λ_1) and Poisson(λ_2) distributions. It is well known that

$$X_1=\sum_{i=1}^{n_1} X_{1i} \sim \text{Poisson}(n_1\lambda_1)$$

independently of

$$X_2=\sum_{i=1}^{n_2} X_{2i} \sim \text{Poisson}(n_2\lambda_2). \tag{1.2}$$

Let k_1 and k_2 be the observed values of X_1 and X_2, respectively. The problem of interest here is to test

$$\text{H}_0: \lambda_1-\lambda_2 \leqslant d \quad \text{vs.} \quad \text{H}_a: \lambda_1-\lambda_2 > d, \tag{1.3}$$

where $d \geqslant 0$ is a given number, based on (n_1,k_1,n_2,k_2). Another way of testing the equality of two Poisson means is to test the hypotheses

$$\text{H}_0: \frac{\lambda_1}{\lambda_2} \leqslant c \quad \text{vs.} \quad \text{H}_a: \frac{\lambda_1}{\lambda_2} > c, \tag{1.4}$$

where $c>0$ is a specified number.

By far the most common method of testing the difference between two Poisson means is the conditional method that was first proposed by Przyborowski and Wilenski (1940). The conditional distribution of X_1 given X_1+X_2 follows a binomial distribution whose success probability is a function of the ratio λ_1/λ_2 (see Section 2). Therefore, hypothesis testing and interval estimation procedures can be readily developed from the exact methods for making inferences about the binomial success probability. In particular, Chapman (1952) proposed a confidence interval for the ratio λ_1/λ_2 which is deduced from the exact confidence interval for the binomial success probability due to Clopper and Pearsons (1934). Since then many papers have addressed these inferential procedures based on the conditional distribution; see Gail (1974), Shiue and Bain (1982), Nelson (1991), Sahai and Misra (1992), and Sahai and Khurshid (1993). These articles consider power computations when sample sizes $n_1=n_2=1$ and $n_1=n_2=n$. Specifically, Gail (1974) gives plots for determining the power for a given ratio λ_1/λ_2 and λ_1. Schwertman and Martinez (1994) give several binomial-normal based approximate methods for constructing confidence interval for $\lambda_1-\lambda_2$. Shiue and Bain (1982) pointed out situations where unequal sample sizes arise (see Example 2 in Section 5 of this paper), and suggested a normal based approximate method for computing sample sizes at a given level and power.

Although the conditional test is exact and simple to use, in the two-sample binomial case such a conditional test (Fisher's exact test) is known to be less powerful than some unconditional tests. For example, see Liddell (1978), Suissa and Schuster (1985), and Storer and Kim (1990). Furthermore, in the context of the present problem, Cousins

(1998) proposed a numerical method for computing a confidence interval for the ratio λ_1/λ_2 based on the joint distribution of X_1 and X_2. Even though Cousins' approach is numerically intensive, it provides shorter intervals than the conditional confidence interval due to Chapman (1952). In view of these results, in this article, we propose an unconditional test for testing the hypotheses in (1.3).

This article is organized as follows. In the following section, we describe the conditional test (C-test) due to Przyborowski and Wilenski (1940). In Section 3, we propose our new test which is obtained by suitably modifying the binomial test due to Storer and Kim (1990). Since the new test is essentially based on the estimated p-values of the standardized difference, we refer to this test as the E-test. In Section 4, we outline a numerical method for computing the exact sizes and powers of the proposed tests. Using this numerical method, we evaluated the sizes and powers of these two tests for various parameter configurations and nominal levels in Section 5. Our extensive numerical studies show that the performance of the E-test is very satisfactory in terms of size, and is more powerful than the C-test. In particular, there are situations (see Table 1(b)) where the E-test requires much smaller samples than those required for the C-test to attain a specified power. In Section 6, both tests are illustrated using two examples. Some concluding remarks are given in Section 7.

2. The conditional test

The conditional test (C-test) due to Przyborowski and Wilenski (1940) is based on the conditional distribution of X_1 given $X_1 + X_2 = k$. Note that the distribution of X_1 conditionally given $X_1 + X_2 = k$ is binomial with the number of trials k and success probability

$$p(\lambda_1/\lambda_2) = (n_1/n_2)(\lambda_1/\lambda_2)/(1 + (n_1/n_2)(\lambda_1/\lambda_2)). \tag{2.1}$$

This C-test rejects H_0 in (1.4), whenever the p-value

$$P(X_1 \geqslant k_1|k, p(c)) = \sum_{i=k_1}^{k} \binom{k}{i} (p(c))^i (1 - p(c))^{k-i} \leqslant \alpha, \tag{2.2}$$

where $p(c)$ is the expression in (2.1) with λ_1/λ_2 replaced by c. For a given c, the p-value in (2.2) can be easily computed using widely available softwares that compute the binomial distribution. Furthermore, the p-value for testing

$$\mathrm{H}_0 : \lambda_1 = \lambda_2 \quad \text{vs.} \quad \mathrm{H}_a : \lambda_1 \neq \lambda_2$$

is given by $2 \times \min\{P(X_1 \geqslant k_1|k, p(c)), P(X_1 \leqslant k_1|k, p(c))\}$.

3. The proposed test

The test we consider here is based on the standardized difference between X_1/n_1 and X_2/n_2. The variance of the difference $X_1/n_1 - X_2/n_2$ is given by

$$\mathrm{Var}(X_1/n_1 - X_2/n_2) = \frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}. \tag{3.1}$$

Note that X_i/n_i is an unbiased estimate of λ_i, $i = 1, 2 \ldots$. This leads to the unbiased variance estimate given by

$$\hat{V}_X = \frac{X_1/n_1}{n_1} + \frac{X_2/n_2}{n_2}. \tag{3.2}$$

Using this unbiased variance estimate, we consider the standardized difference

$$T_{X_1,X_2} = \frac{X_1/n_1 - X_2/n_2 - d}{\sqrt{\hat{V}_X}} \tag{3.3}$$

as a pivot statistic for our testing problem. For a given (n_1, k_1, n_2, k_2), the observed value of the pivot statistic T_{X_1,X_2} is given by

$$T_{k_1,k_2} = \frac{k_1/n_1 - k_2/n_2 - d}{\sqrt{\hat{V}_k}},$$

where $\hat{V}_k$ is defined similarly as $\hat{V}_X$ in (3.2) with X replaced by k.

The p-value for testing (1.3) is $P(T_{X_1,X_2} \geqslant T_{k_1,k_2}|\mathrm{H}_0)$ which involves the unknown parameter λ_2. However, following the approaches of Liddell (1978) and Storer and Kim (1990), we can estimate the p-value by replacing λ_2 by its estimate $\hat{\lambda}_{2k}$. Note that when $\lambda_1 = \lambda_2 + d$, we have

$$E\left(\frac{X_1 + X_2}{n_1 + n_2}\right) = \frac{n_1\lambda_1 + n_2\lambda_2}{n_1 + n_2} = \lambda_2 + \frac{dn_1}{n_1 + n_2}.$$

Therefore, for a given k_1 and k_2, an estimate of λ_2 is given by

$$\hat{\lambda}_{2k} = \frac{k_1 + k_2}{n_1 + n_2} - \frac{dn_1}{n_1 + n_2}. \tag{3.4}$$

Using this $\hat{\lambda}_{2k}$, we estimate the p-value $P(T_{X_1,X_2} \geqslant T_{k_1,k_2}|\mathrm{H}_0)$ by

$$\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} \frac{\mathrm{e}^{-n_1(\hat{\lambda}_{2k}+d)}(n_1(\hat{\lambda}_{2k} + d))^{x_1}}{x_1!} \frac{\mathrm{e}^{-n_2\hat{\lambda}_{2k}}(n_2\hat{\lambda}_{2k})^{x_2}}{x_2!} I[T_{x_1,x_2} \geqslant T_{k_1,k_2}], \tag{3.5}$$

where $I[.]$ denotes the indicator function. For a given nominal level α, the test rule is to reject H_0 in (1.3) whenever the estimated p-value in (3.5) is less than α. The p-value for testing

$$\mathrm{H}_0 : \lambda_1 = \lambda_2 + d \quad \text{vs.} \quad \mathrm{H}_a : \lambda_1 \neq \lambda_2 + d \tag{3.6}$$

can be computed using (3.5) with T_{X_1,X_2} and T_{k_1,k_2} replaced by their absolute values. The $\hat{\lambda}_{2k}$ in (3.4) may be less than or equal to zero. However, we note that $\hat{\lambda}_{2k} \leqslant 0$ implies that $k_1/n_1 - k_2/n_2 \leqslant d$, and in this case the null hypothesis in (1.3) cannot be rejected. In other words, it is not necessary to compute the p-value when $\hat{\lambda}_{2k} \leqslant 0$.

Remark 1. We see in Figs. 2a–f and in 4a–f, the size of the test attains the maximum at the boundary $\lambda_1 = \lambda_2 + d$. Therefore, it suffices to compute the p-value for testing (1.3) at the boundary $\lambda_1 = \lambda_2 + d$.

Remark 2. It should be noted that the p-value in (3.5) is essentially equal to the one based on the parametric bootstrap approach (see Efron, 1982, p. 29). Recall that in parametric bootstrap approach the p-value is obtained by using the distribution of the Monte Carlo samples generated from Poisson($n_1(\hat{\lambda}_{2k}+d)$) and Poisson($n_2\hat{\lambda}_{2k}$). In particular, in parametric bootstrap approach, the p-value is estimated by the proportion of simulated samples for which $T_{x_1,x_2} \geqslant T_{k_1,k_2}$ where as in (3.5) we use the exact sampling distribution of T_{X_1,X_2} based on Poisson($n_1(\hat{\lambda}_{2k}+d)$) and Poisson($n_2\hat{\lambda}_{2k}$). Therefore, the present approach is equivalent to the parametric bootstrap approach applied in an exact manner. Furthermore, it is known that the results of a pivot based parametric bootstrap approach (percentile t) are in general preferred to those based on non-pivot based approach (see Hall, 1992, p. 141). For this reason, we use T_{X_1,X_2} given in (3.3) instead of just the difference $X_1/n_1 - X_2/n_2$.

Remark 3. We consider the pivot statistic T_{X_1,X_2} instead of the one based on the pooled variance estimate $(1/n_1 + 1/n_2)((X_1 + X_2)/(n_1 + n_2))$ under $H_0 : \lambda_1 - \lambda_2 = 0$. In other contexts, such as two-sample binomial problem, pivot statistics based on a pooled variance estimate are commonly used. However, our preliminary numerical studies showed that the sizes of test based on the T_{X_1,X_2} and those of the one based on pooled variance estimate are almost the same but the former test has slightly better power properties than the latter.

Since our test is based on an estimated p-value, its properties should be evaluated. In the following section, we outline a method of computing the powers of the tests.

4. Power calculation

The power of a test is the probability that the null hypothesis is rejected when it is actually false. For hypotheses in (1.3), the power is the actual probability that the null hypothesis is rejected when λ_1 is indeed greater than $\lambda_2 + d$. For a given α, λ_i and n_i, $i = 1,2$, the exact power of the C-test is given by

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} \frac{e^{-n_1\lambda_1}(n_1\lambda_1)^{k_1}}{k_1!}\,\frac{e^{-n_2\lambda_2}(n_2\lambda_2)^{k_2}}{k_2!}\, I[P(X_1 \geqslant k_1 | k_1 + k_2, p(\lambda_1/\lambda_2)) \leqslant \alpha], \tag{4.1}$$

where X_1 follows a binomial distribution with number of trials $k_1 + k_2$ and success probability $p(\lambda_1/\lambda_2)$ is given in (2.1). Note that in (4.1), we are merely adding the probabilities over the set of values of (k_1, k_2) for which the null hypothesis is rejected. If $\lambda_1 \leqslant \lambda_2 + d$, then (4.1) gives the size of the test; otherwise it gives the power. The exact power of the E-test is given by

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} \frac{e^{-n_1\lambda_1}(n_1\lambda_1)^{k_1}}{k_1!}\,\frac{e^{-n_2\lambda_2}(n_2\lambda_2)^{k_2}}{k_2!}$$

$$\times I\left[\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} \frac{e^{-n_1(\hat{\lambda}_{2k}+d)}(n_1(\hat{\lambda}_{2k}+d))^{x_1}}{x_1!}\,\frac{e^{-n_2\hat{\lambda}_{2k}}(n_2\hat{\lambda}_{2k})^{x_2}}{x_2!}\right.$$

$$\left.\times\, I[T_{x_1,x_2} \geqslant T_{k_1,k_2}] \leqslant \alpha\right]. \tag{4.2}$$

As we already mentioned in Section 3, $\hat{\lambda}_{2k}$ could be negative if the observed difference $k_1/n_1 - k_2/n_2 \leqslant d$. In this case, we can not reject H_0 in (1.3), and so we set the p-value to one. In other words, we assign zero to the indicator function (by the multiplication sign) in (4.2) whenever $k_1/n_1 - k_2/n_2 \leqslant d$. Since the above power function involves four infinite series, we shall explain a computational method for evaluating the power. We note that the probabilities in (4.2) can be evaluated by first computing the probability at the mode of the Poisson distribution, and then computing the other terms using forward and backward recurrence relations for Poisson probability mass function. Toward this, we note that for $X \sim \text{Poisson}(\theta)$, the mode of X is the largest integer less than or equal to θ, and

$$P(X = k + 1) = \frac{\theta}{k + 1} P(X = k), \quad k = 0, 1, \ldots,$$

and

$$P(X = k - 1) = \frac{k}{\theta} P(X = k), \quad k = 1, 2, \ldots .$$

Evaluation of each sum in (4.2) may be terminated once the probability is less than a small number, say, 10^{-7}. Furthermore, the sum over x_1 inside the square bracket in (4.2) should be terminated once it reaches a value greater than α. Using these steps, we developed Fortran programs for computing p-values and powers of the E-test. The programs will be posted at the website http://lib.stat.cmu.edu.

Remark 4. Note that in the above method of computing powers, the tail probabilities (in each sum) which are less than 10^{-7} are omitted. In order to understand the magnitude of the sum of the neglected tail probabilities, we computed (4.2) using the above method and the direct method (each sum is evaluated from 0 to 100) when $(\lambda_1, \lambda_1, n_1, n_2) = (2, 2, 3, 4), (1, 1, 2, 2), (4, 4, 1, 1), (5, 5, 4, 4)$ and $(2, 4, 5, 7)$. The highest neglected probability in the direct method is $P(X = 101)$ for each case. This probability is less than 10^{-25} for all sample size and parameter configurations considered. Based on the computed values, we found that the sums of the neglected probabilities for the method discussed in the previous paragraph are less than 10^{-6} for all the five cases.

5. Power studies and sample size calculation

In view of (1.2), without loss of generality, we can take $n_1 = n_2 = 1$ when comparing the powers of the C-test and E-test. The powers of both tests are computed using the numerical method given in Section 4 when $H_0 : \lambda_1 - \lambda_2 \leqslant d$ vs. $H_a : \lambda_1 - \lambda_2 > d$. For the case of $d = 0$, the sizes and powers of both tests are plotted, respectively, in Figs. 1a–f and 2a–f. It is clear from Figs. 1a–f that the size of the E-test exceeds the nominal α by a negligible amount. The sizes of the C-test are always smaller than the nominal levels (for all the cases considered), which indicates that the C-test is too conservative. We observe from Figs. 2a–f that the power of the E-test is always

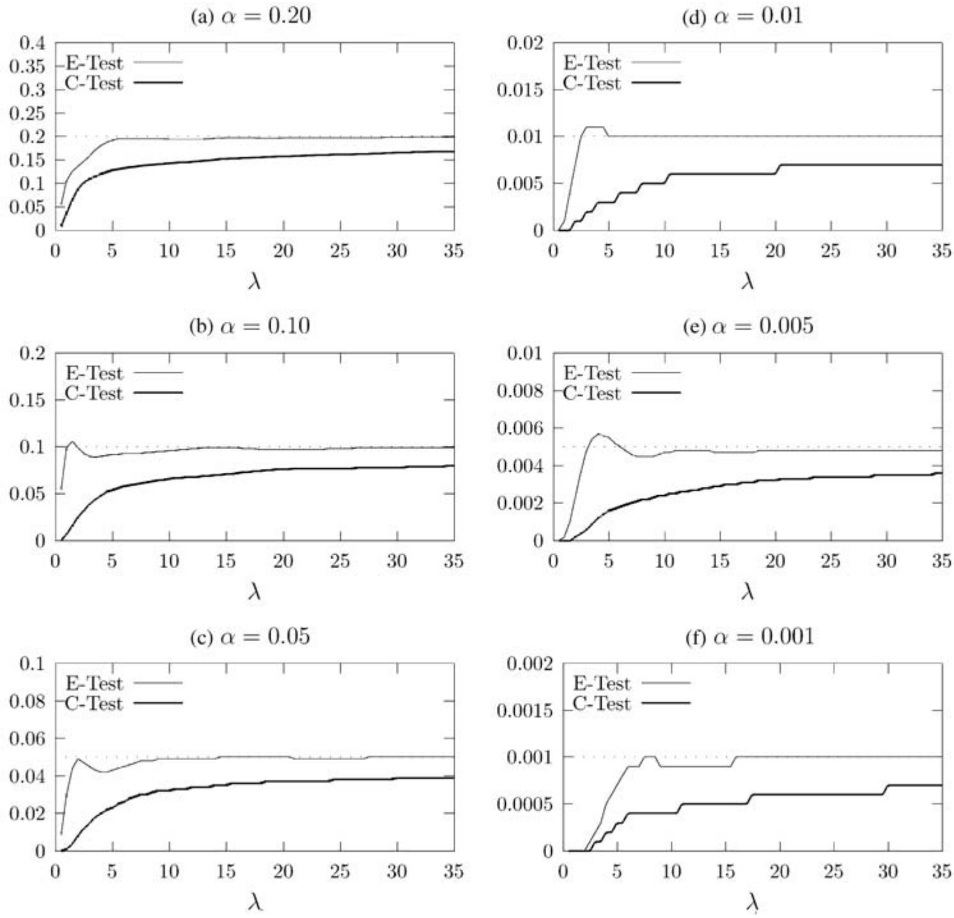

Fig. 1. Exact size of the E-test and the C-test as a function of λ at various nominal levels under the null hypothesis $H_0 : \lambda_1 - \lambda_2 = 0$: (a) $\alpha = 0.20$; (b) $\alpha = 0.10$; (c) $\alpha = 0.05$; (d) $\alpha = 0.01$; (e) $\alpha = 0.005$; (f) $\alpha = 0.001$.

larger than that of the C-test. The power curves of both tests are getting closer when λ_1 and λ_2 are large. In view of (1.2), we see that for fixed $\lambda_1 - \lambda_2 > 0$, the powers of both tests are increasing with respect to the sample size, and the difference between the powers are small for large sample sizes (see Figs. 2e and f). The sizes and powers of both tests are plotted, respectively, in Figs. 3a–f and Figs. 4a–f for a few values of $d > 0$. In Figs. 3a–f, we observe that the sizes of the E-test never exceed the nominal level, and the sizes of the C-test are always lower than those of the E-test. We observe again from Figs. 4a–f that the E-test is much more powerful than the C-test, and in some cases the differences between the powers are quite large (see Figs. 4a and d).

In order to understand the gain in power in terms of sampling cost, we computed the sample sizes for a given level and power. Although, expressions (4.1) and (4.2)

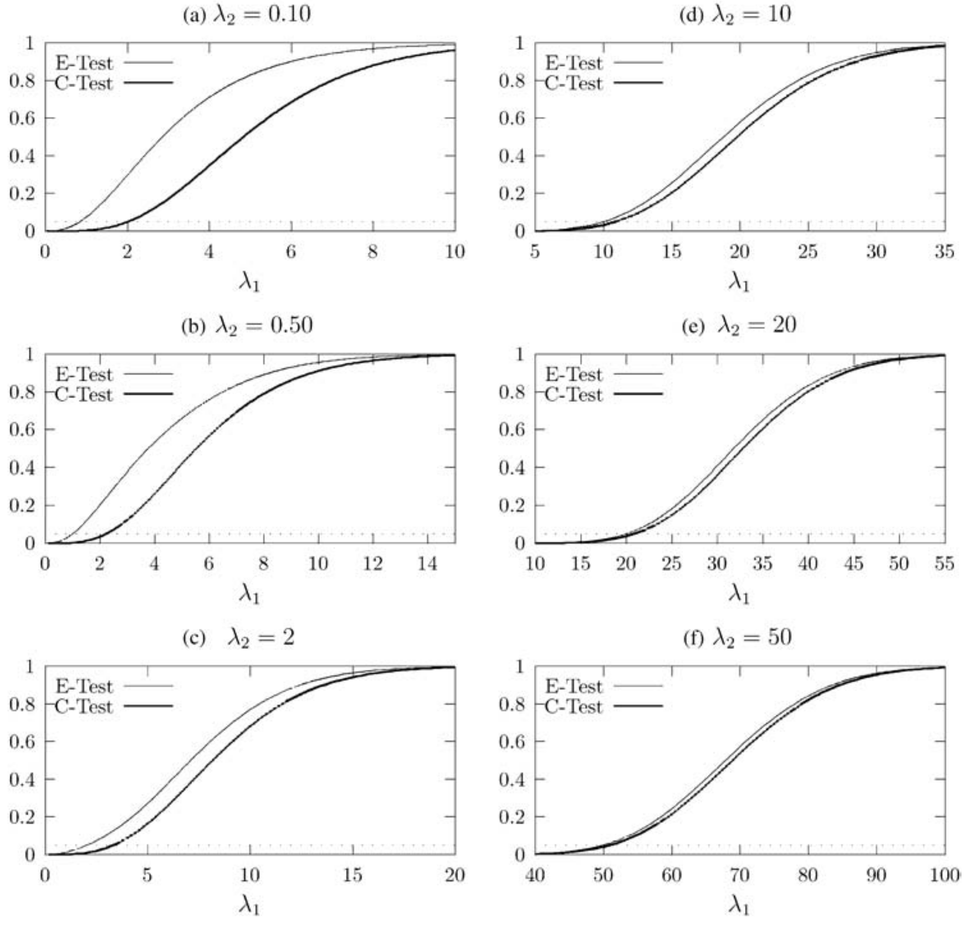

Fig. 2. Actual power of the E-test and the C-test as a function of λ_1 at the nominal level $\alpha = 0.05$ when $H_a : \lambda_1 - \lambda_2 > 0$: (a) $\lambda_2 = 0.10$; (b) $\lambda_2 = 0.50$; (c) $\lambda_2 = 2$; (d) $\lambda_2 = 10$; (e) $\lambda_2 = 20$; (f) $\lambda_2 = 50$.

can be used to find unequal sample sizes to attain a given power, for simplicity we computed equal sample sizes required to have powers of at least 0.80, 0.90 and 0.95 at the level of significance 0.05. We provide sample sizes in Table 1(a) when $d = 0$ and in Table 1(b) for some values of $d > 0$. The sizes of the tests are computed by choosing λ_1 such that $\lambda_1 - \lambda_2 = d$. We observe from these table values that the sizes of the E-test are less than or equal to the nominal level 0.05 for all the cases considered. To attain the same power, the sample size required by the E-test is always less than or equal to those required by the C-test. In particular, we see from Table 1(b) that the E-test requires much smaller sample sizes and thereby reduces the cost of sampling considerably.

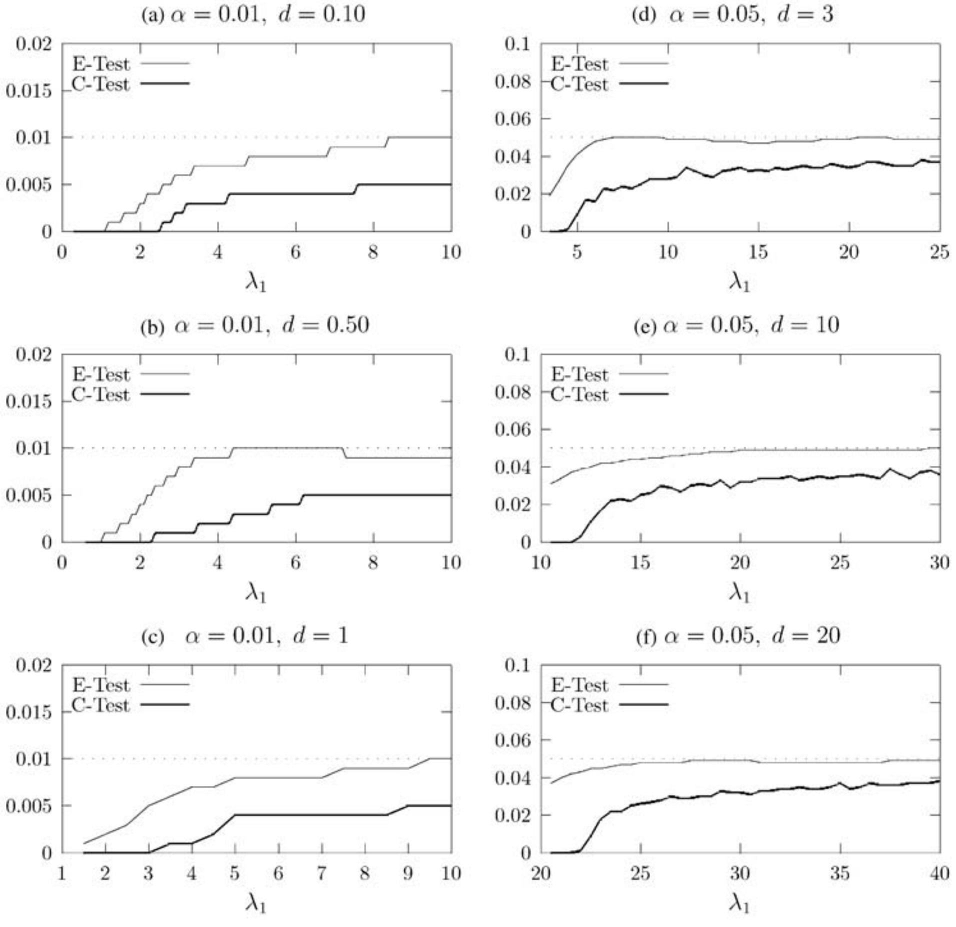

Fig. 3. Exact size of the E-test and C-test as a function of λ_1 at various nominal levels under the null hypothesis $H_0 : \lambda_1 - \lambda_2 = d$: (a) $\alpha = 0.01,\ d = 0.10$; (b) $\alpha = 0.01,\ d = 0.50$; (c) $\alpha = 0.01,\ d = 1$; (d) $\alpha = 0.05,\ d = 3$; (e) $\alpha = 0.05,\ d = 10$; (f) $\alpha = 0.05,\ d = 20$.

6. Illustrative examples

Example 1. Przyborowski and Wilenski (1940) considered this example for illustrating the C-test. Suppose that a purchaser wishes to test the number of dodder seeds (a weed) in a sack of clover seeds that he bought from a seed manufacturing company. A 100 g sample is drawn from a sack of clover seeds prior to being shipped to the purchaser. The sample is analyzed and found to contain no dodder seeds; that is, $k_1 = 0$. Upon arrival, the purchaser also draws a 100 g sample from the sack. This time, three dodder seeds are found in the sample; that is, $k_2 = 3$. The purchaser wishes to determine if the difference between the samples could not be due to chance. In this case $H_0 : \lambda_1 = \lambda_2$ vs. $H_a : \lambda_1 \neq \lambda_2$. Using the C-test, we computed the p-value as 0.2500. The E-test

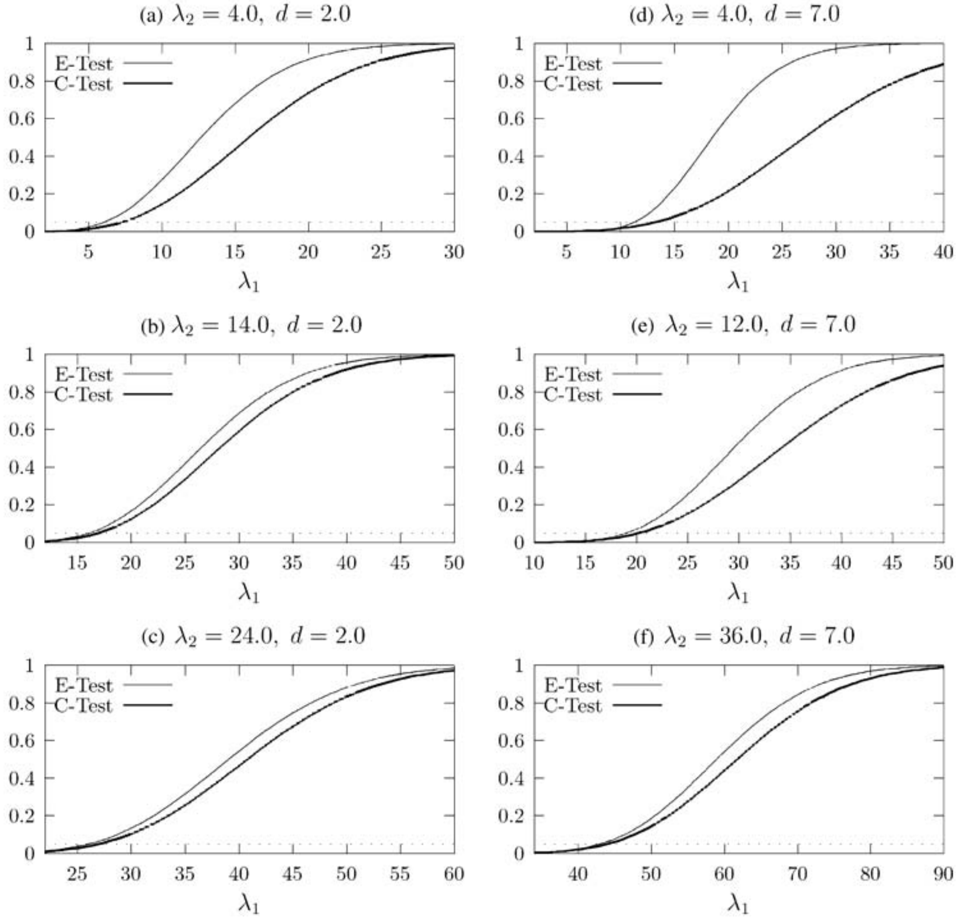

Fig. 4. Actual power of the E-test and the C-test as a function of λ_1 at the nominal level $\alpha = 0.05$ when $H_a : \lambda_1 - \lambda_2 > d$: (a) $\lambda_2 = 4.0,\ d = 2.0$; (b) $\lambda_2 = 14.0,\ d = 2.0$; (c) $\lambda_2 = 24.0,\ d = 2.0$; (e) $\lambda_2 = 4.0,\ d = 7.0$; (d) $\lambda_2 = 12.0,\ d = 7.0$; (f) $\lambda_2 = 36.0,\ d = 7.0$.

produced a p-value of 0.0884. Thus the E-test provides evidence to indicate that the difference between sample counts is not due to chance at the level of significance 0.10. Suppose that on another occasion it was found that $k_1 = 2$ and $k_2 = 6$. The C-test yielded a p-value of 0.2891, where as the E-test yielded a p-value of 0.1749. We again see that the p-value of the E-test is smaller than that of the C-test, although the conclusions of both tests are the same at any practical level of significance.

Example 2. This example is taken from Shiue and Bain (1982). Suppose in a fleet of planes a new type of component is being used. We wish to test whether this component is better than the current component being used in another fleet of planes. That is, we wish to test whether the failure rate of the new component is less than that of the

Table 1

(a) *Comparison of sample sizes required for a given level and power*
($H_a : \lambda_1 - \lambda_2 > 0$; $n_1 = n_2 = n$; $\alpha = 0.05$)

			Sample size n					Sample size n	
λ_1	λ_2	Power	C-test	E-test	λ_1	λ_2	Power	C-test	E-test
0.8	0.5	0.80	95 (0.040)	89 (0.050)	3.0	2.0	0.80	33 (0.042)	31 (0.050)
		0.90	129 (0.041)	123 (0.050)			0.90	45 (0.043)	43 (0.050)
		0.95	161 (0.042)	155 (0.050)			0.95	56 (0.044)	54 (0.050)
1.5	0.5	0.80	14 (0.029)	12 (0.045)	4.0	2.0	0.80	10 (0.037)	10 (0.050)
		0.90	18 (0.032)	17 (0.048)			0.90	14 (0.038)	13 (0.049)
		0.95	23 (0.033)	21 (0.049)			0.95	17 (0.039)	16 (0.050)
3.5	0.5	0.80	4 (0.008)	3 (0.044)	10.0	8.0	0.80	29 (0.046)	28 (0.050)
		0.90	4 (0.008)	4 (0.049)			0.90	40 (0.046)	39 (0.050)
		0.95	5 (0.012)	5 (0.047)			0.95	50 (0.047)	49 (0.050)

(b) *Comparison of sample sizes required for a given level and power*
($H_a : \lambda_1 - \lambda_2 > d$; $n_1 = n_2 = n$; $\alpha = 0.05$)

				Sample size n						Sample size n	
λ_1	λ_2	d	Power	C-test	E-test	λ_1	λ_2	d	Power	C-test	E-test
0.5	0.3	0.1	0.80	672 (0.045)	489 (0.050)	4.0	1.0	2.0	0.80	91 (0.044)	30 (0.049)
			0.90	921 (0.046)	678 (0.050)				0.90	124 (0.045)	41 (0.049)
			0.95	1156 (0.046)	856 (0.050)				0.95	155 (0.045)	52 (0.049)
0.7	0.3	0.1	0.80	95 (0.038)	67 (0.050)	7.0	1.0	2.0	0.80	9 (0.032)	3 (0.049)
			0.90	128 (0.041)	92 (0.049)				0.90	12 (0.033)	4 (0.045)
			0.95	159 (0.041)	116 (0.049)				0.95	14 (0.034)	5 (0.044)
1.2	0.3	0.1	0.80	21 (0.026)	14 (0.050)	11.0	1.0	2.0	0.80	4 (0.021)	1 (0.035)
			0.90	27 (0.029)	19 (0.050)				0.90	4 (0.021)	2 (0.035)
			0.95	33 (0.031)	24 (0.047)				0.95	5 (0.025)	2 (0.035)

current component. Let λ_1 be the failure rate of the current component with a fleet size of $n_1 = 20$ planes, and λ_2 be the failure rate of the new component with a fleet size of $n_2 = 10$ planes. We plan to observe both fleets for the same number of flying hours per plane. From laboratory testing, we know that the failure rate of the new component is approximately 2 failures per 100 flying hours. It is believed that the failure rate of the current component is twice that of the new component. Hence, $\lambda_1 = 0.04$ failures per flying hour and $\lambda_2 = 0.02$ failures per flying hour. We wish to determine how long the experiment should be conducted to test $H_0 : \lambda_1 = \lambda_2$ vs. $H_a : \lambda_1 > \lambda_2$ with $\alpha = 0.05$ and a power of 0.90.

Shiue and Bain (1982) applied the C-test, and reported that 97.5 flying hours per plane is adequate to attain a power of 0.90. This translates to 1950 flying hours for the first fleet and 975 flying hours for the second fleet. It should be noted that the power computed in their paper is based on an approximate method. The actual power of the

C-test (using the method in Section 5) is 0.8890. According to the exact method in Section 5, to attain a power of 0.90, 2026 flying hours are required for the first fleet and 1013 hours are required for the second fleet. This is considerably more hours than those given using an approximate power. If we use the E-test, then a total of 1886 flying hours for the first fleet and 943 flying hours for the second fleet are required to have a power of 0.90. Thus we can see that the E-test requires quite a bit less flying hours as a whole to achieve the same size and power of the test.

7. Concluding remarks

We studied the exact properties of the conditional C-test and the unconditional E-test based on the joint sampling distribution. Our extensive numerical studies in Section 5 clearly show that the E-test is much more powerful than the C-test, not withstanding its size occasionally exceeds the nominal level by a small amount. Considering the gain in power compared to the C-test and the reasonably good control of the size for the E-test, most practitioners would prefer the E-test. Furthermore, sample size calculation in Tables 1(a) and (b) demonstrates that the E-test is certainly preferable to the C-test in situations where the inspection of items is time consuming and/or expensive. Even though the application of E-test is numerically involved, modern computing technology and sophisticated softwares allow us to compute the p-values and powers of the E-test in a relatively easy manner. As already mentioned in Section 4, Fortran programs for implementing the E-test are posted at the website http://lib.stat.cmu.edu, and hence there will not be any technical difficulties in application of the E-test for practical purpose. We also evaluated the properties of both tests for two-sided alternative hypothesis. Since the performances of the tests are similar to those in Figs. 1–4, they are not reported here.

Acknowledgements

The authors are grateful to two reviewers for their valuable comments and suggestions.

References

Chapman, D.G., 1952. On tests and estimates for the ratio of Poisson means. Ann. Inst. Statist. Math. 4, 45–49.

Clopper, C.J., Pearsons, E.S., 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26, 404–413.

Cousins, R.D., 1998. Improved central confidence intervals for the ratio of Poisson means. Nucl. Instrum. Methods A 417, 391–399.

Efron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, Philadelphia, PA.

Gail, M., 1974. Power computations for designing comparative Poisson trials. Biometrics 30, 231–237.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.

Liddell, D., 1978. Practical tests of 2×2 contingency tables. Statistician 25, 295–304.

Nelson, L.S., 1991. Power in comparing Poisson means: II. Two-sample test. J. Qual. Technol. 23, 163–166.
Przyborowski, J., Wilenski, H., 1940. Homogeneity of results in testing samples from Poisson series. Biometrika 31, 313–323.
Sahai, H., Khurshid, A., 1993. Confidence intervals for the ratio of two Poisson means. Math. Sci. 18, 43–50.
Sahai, H., Misra, S.C., 1992. Comparing means of two Poisson distributions. Math. Sci. 17, 60–67.
Schwertman, N.C., Martinez, R., 1994. Approximate confidence intervals for the difference in two Poisson parameters. J. Statist. Comput. Simulation 50, 235–247.
Shiue, W., Bain, L.J., 1982. Experiment size and power comparisons for two-sample Poisson tests. Appl. Statist. 31, 130–134.
Storer, B.E., Kim, C., 1990. Exact properties of some exact test statistics comparing two binomial proportions. J. Amer. Statist. Assoc. 85, 146–155.
Suissa, S., Schuster, J.J., 1985. Exact unconditional sample sizes for the 2×2 binomial trials. J. Roy. Statist. Soc., Ser. A 148, 317–327.