Journal of the Chinese
Statistical Association
Vol. 48, (2010) 85–107

ASYMPTOTIC AND EXACT TWO-SAMPLE POISSON TESTS

Mingte Liu[1] and Huey-Miin Hsueh[2]

[1] **General Education Center, Tatung Institute of Commerce and Technology**

[2] **Department of Statistics, National Cheng-Chi University**

ABSTRACT

The Poisson distribution is popular in modeling a rare events in various fields such as biology, commerce, quality control, and so on. Many applications involve a comparison of two treatment groups based on two independent random samples drawn from Poisson distributions. In this study, the asymptotic power function and sample size formula of two types of Wald test are derived. Moreover, two exact testing procedures are introduced and investigated as well. The required sample size of the exact procedures can be found numerically. Through intensive numerical studies, the validity of these tests is justified. The performance of the two asymptotic tests are found to depend on the fraction of the two sample sizes and they tend to generate conclusions that are too liberal. In contrast, the two exact tests lead to more satisfactory statistical conclusions. Moreover, the asymptotic sample size formulae provide adequate approximations to the required sample size for the exact approach. A data set of breast cancer patients is analyzed for illustration.

Key words and phrases: Asymptotic test, exact test, Poisson, p-value, superiority.
JEL classification: C40.

1. Introduction

It is well known that the Poisson distribution is a suitable model for rare events in a variety of fields, such as biology, commerce, quality control, and so on. These applications are usually used to compare the means of two Poisson random variables. For example, to compare the rates of breast cancer of the group with/without X-ray fluoroscopy examination during treatment for tuberculosis, the equality of the mean numbers of cases in a given person-years at risk of the two groups are tested (Ng and Tang, 2005). Another study compared the failure rate of an airplane component during war time and the failure rate during peace time, see Shiue and Bain (1982).

Gail (1974) introduced two different experiments. In the first experiment, the total number of the two Poisson variables was predetermined. In the other experiment, the length of experiment duration was fixed instead. In the former experiment, the exact test is based on the conditional distribution given the fixed total number, which was proposed by Przyborowski and Wilenski (1940), and it appears to be an adequate testing method. This test is uniformly most powerful among unbiased tests. In the later experiment, which is more common in practice, an unconditional test is more suitable. In this study, all the testing procedures under investigation are unconditional.

The experimental durations of the two Poisson variables can be unequal due to a practical consideration. For example, in the comparison of the failure rate of an airplane component between war time and peace time, it is known that collecting data during war time is more difficult and expensive than during peace time. Hence the researcher may consider a shorter duration. Shiue and Bain (1982) generalized the conditional exact test and a normal approximated test to the case of unequal durations. Huffman (1984) proposed an approximate test based on the variance stabilizing transformation. Moreover, Thode (1997) provided an alternative normal approximated test. The two tests were shown to be superior to the test of Shiue and Bain (1982) by numerical results. Basically, these proposed methods were developed in terms of the difference of the two Poisson means in literatures. Alternatively, some authors expressed the comparison in terms of the ratio of the two positive means, see Ng and Tang (2005), Gu et al. (2008). Ng and Tang (2005) compared two normal approximated tests, which apply the logarithmic-transformed rate ratio in the numerator of the test statistic, and adapted two different estimations for the standard error. Gu et al. (2008) extended the numerical comparisons to more tests. However, all the existing procedures were studied and compared through numerical studies in most of the literatures. In this paper, we consider a comparison between two independent Poisson random samples within a fixed experiment duration. When the two sample sizes are unbalanced, the scenario is equivalent to the unequal duration case. The comparison would be expressed by the difference of the two means. In addition to the numerical results, more discussions on the theoretical performance would be provided.

As the sample sizes or the mean parameters are sufficiently large, asymptotic tests are often recommended in the literature. (See Shiue and Bain, 1982; Thode, 1997; Ng and Tang, 2005; and Gu el al., 2008). On the other hand, the exact testing procedures are more appropriate when both assumptions fail. In the problem of comparing two Poisson means, there involves a nuisance parameter, the common unknown mean value, under the null hypothesis. In the presence of nuisance parameters, Casella and Berger (1990) defined the standard p-value that considers the least favorable case under the principle of conservativeness. Specifically, the standard p-value requires a supremum search over the parameter space and more calculations are necessary. An infinite parameter space would even increase the difficulty of computation. Berger and Boos (1994) introduced a confidence-set p-value, which is defined as the maximal p-value over a confidence region of the nuisance parameter. The confidence-set p-value does not only have reduction in computation, but also is manifested to have a well-controlled type I error rate. On the other hand, Krishnamoorthy and Thomson (2004) inspired by Store and Kim (1990) developed a nearly exact testing method. The associated p-value is exact because it is evaluated under Poisson population distribution. The authors use an point estimate of the nuisance parameter in finding the p-value. The same test was studied in Gu et al. (2008). Although the estimated p-value was shown to perform well in selected scenarios in these papers, the testing procedure could not guarantee a well-controlled type I error rate theoretically.

Given a fixed unit experimental duration, two independent Poisson random samples are observed: $\{(Y_{11},\ldots,Y_{1n_1}),(Y_{21},\ldots,Y_{2n_2})\}$, where for $i=1,\ldots,n_1$, $j=1,\ldots,n_2$,

$$Y_{1i} \overset{\text{iid}}{\sim} \text{Poisson}(\lambda_1), \quad Y_{2j} \overset{\text{iid}}{\sim} \text{Poisson}(\lambda_2).$$

One is interested in testing the following statistical hypotheses,

$$H_0 : \lambda_1 = \lambda_2 \quad \text{vs.} \quad H_1 : \lambda_1 > \lambda_2.$$

Denote $\Delta = \lambda_1 - \lambda_2$. It is known that the two sums $Y_1 = \sum_{i=1}^{n_1} Y_{1i}$, $Y_2 = \sum_{j=1}^{n_2} Y_{2j}$ are sufficient statistics. The two sample means $\bar{Y}_1$, $\bar{Y}_2$ are the maximum likelihood estimator (MLE) of λ_1, λ_2, respectively. Further, the MLE of Δ is $\hat{\Delta} = \bar{Y}_1 - \bar{Y}_2$. In this study, the Wald statistic is applied,

$$Z = \frac{\hat{\Delta}}{\text{se}(\hat{\Delta})},$$

where the denominator can be any consistent estimate of the asymptotic standard error of $\hat{\Delta}$ under H_0. A large value of Z suggests the rejection of H_0. Two types of estimate of $\text{se}(\hat{\Delta})$ are commonly adopted. One applies the unrestricted MLE and the other one uses the restricted MLE under the null hypothesis. Three kinds of p-values will be investigated in the present study. While the asymptotic p-value is evaluated under a normal distribution, the confidence-set p-value and the estimated p-value are obtained under the Poisson distribution. To improve the inflation of type I error rate, the asymptotic testing procedure will be modified by adding an appropriate continuity correction factor. Pirie and Hamdan (1972) derived a continuity correction term when the two Poisson random samples are of equal size. In this paper, we will derive the adequate continuity correction term for general cases.

Note that the procedures can be easily extended to the case where every observation has a different experimental duration. Assume Y_{ij} be the Poisson random variable in the i-th group within m_{ij} units of duration, $i = 1, 2$, $j = 1, \ldots, n_i$. Then the MLEs become $\hat{\lambda}_1 = Y_1/n_1^*$, $\hat{\lambda}_2 = Y_2/n_2^*$, where $n_i^* = \sum_{j=1}^{n_i} m_{ij}$, $i = 1, 2$. Replacing n_i by n_i^* in the test statistic, one can employ the approach straightforward.

This articles is organized as follows. We will give the asymptotic properties and the sample size formula of the asymptotic testing procedures in Section 2. Next, the two exact tests, which include the confidence-set p-value and the estimated p-value, will be introduced in Section 3. Subsequently, numerical studies will be presented in Section 4. Section 5 will give an example for illustration. In this study, all numerical studies are conducted by MATLAB software and C++ language.

2. Asymptotic P-values

The asymptotic standard error of $\hat{\Delta}$ can be easily derived as $\{\lambda_1/n_1 + \lambda_2/n_2\}^{1/2}$. Hence, the unrestricted MLE of the standard error is $\mathrm{se}(\hat{\Delta})_U = \{\bar{Y}_1/n_1 + \bar{Y}_2/n_2\}^{1/2}$, and the corresponding Wald test statistic is

$$Z_U = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\bar{Y}_1}{n_1} + \frac{\bar{Y}_2}{n_2}}}.$$

On the other hand, under $H_0 : \lambda_1 = \lambda_2$, the restricted MLE of the standard error is $\mathrm{se}(\hat{\Delta})_R = \{\tilde{\lambda}_0/n_1 + \tilde{\lambda}_0/n_2\}^{1/2}$, where $\tilde{\lambda}_0 = (Y_1 + Y_2)/(n_1 + n_2)$. Consequently, the Wald test statistic is

$$Z_R = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\tilde{\lambda}_0}{n_1} + \frac{\tilde{\lambda}_0}{n_2}}}.$$

The two test statistics are functions of the sufficient statistics (Y_1, Y_2), which are independently Poisson distributed. It is known that as the sample sizes n_1, n_2 are sufficiently large and the mean values λ_1, λ_2 are not too small, a Wald statistic is approximately normally distributed. In the following, the asymptotic p-values of the observed z_U, z_R are

$$p_{A,U} = 1 - \Phi(z_U), \quad p_{A,R} = 1 - \Phi(z_R),$$

where $\Phi(\cdot)$ is the distribution function of $\mathrm{N}(0, 1)$. The null hypothesis is rejected if the p-value is not greater than the significance level α.

First, we explore the validity and asymptotic power function of the two asymptotic tests. Define $\rho = n_1/n_2$ as the sampling fraction of the first group to the second group. Given $\Delta = \Delta_0$, λ_2, ρ, n_2, it can be shown that as n_1, $n_2 \to \infty$,

$$Z_U - \mu \xrightarrow{d} \mathrm{N}(0,1) \quad \text{and} \quad Z_R\sigma - \mu \xrightarrow{d} \mathrm{N}(0,1).$$

In which,

$$\mu = \frac{\Delta_0}{\sqrt{\frac{(1+\rho)\lambda_2+\Delta_0}{n_2\rho}}}, \ \sigma = \sqrt{\frac{(1+\rho)\lambda_2+\rho\Delta_0}{(1+\rho)\lambda_2+\Delta_0}}.$$

At significance level α, H_0 is rejected if the test statistic exceeds z_α, where z_α is the $100(1-\alpha)$%-th percentile of $\mathrm{N}(0,1)$. Then the asymptotic power functions of Z_U, Z_R can be found as follows,

$$\bar{\beta}_{Z_U}(\Delta_0,\lambda_2,n_2,\rho) = 1-\Phi(z_\alpha-\mu), \quad \bar{\beta}_{Z_R}(\Delta_0,\lambda_2,n_2,\rho) = 1-\Phi(z_\alpha\sigma-\mu).$$

Under H_0, $\Delta_0 = 0$, then $\mu = 0$, $\sigma = 1$, and further $\bar{\beta}_{Z_U} = \bar{\beta}_{Z_R} = \alpha$. That is, both the two asymptotic tests successfully control their type I error rate at the significance level. The correspondent p-values are called valid. See Casella and Berger (1990).

When $\Delta_0 > 0$, $\mu > 0$, the asymptotic power $\bar{\beta}_{Z_U}$ can be shown always greater than α. It indicates that the testing procedure Z_U is an unbiased test approximately. Nevertheless, the unbiasedness of Z_R is not always true. When the first group has a smaller size than the second group, i.e. $\rho \le 1$, $\sigma \le 1$, the asymptotic power $\bar{\beta}_{Z_R}$ is always above the nominal level α and increases as Δ_0. On the contrary, if $\rho > 1$, the asymptotic power of Z_R can sometimes fall below the significance level, especially when the sample sizes are extremely unbalanced and the means are relatively small. For example, Figure 1 gives the plots of the asymptotic power function of Z_R for $\rho = 8, 20, 50$, $\lambda_2 = 0.3$ and $n_2 = 10$.

Once the asymptotic power function is derived, the sample size can be determined straightforward. Given α, ρ, λ_2, to achieve a prespecified power level $1-\beta$ at Δ_0, the sample size of the second group n_2 required for Z_U and Z_R should be at least

$$n_{2,Z_U} \ge \left(\frac{z_\alpha+z_\beta}{\Delta_0}\right)^2 \left\{\frac{\lambda_2(1+\rho)+\Delta_0}{\rho}\right\}, \tag{1}$$

and

$$n_{2,Z_R} \ge \left(\frac{z_\alpha\sigma+z_\beta}{\Delta_0}\right)^2 \left\{\frac{\lambda_2(1+\rho)+\Delta_0}{\rho}\right\}, \tag{2}$$

respectively. The size of the first group is found as $n_1 = [n_2\rho]+1$.

It can be seen that the powers and sample size formula of the two tests mainly differ in the multiple of z_α, σ. In the balanced case, i.e., $\rho = 1$, both tests are exactly of the same form. However, when the sample sizes are unbalanced, the performance of the two tests can be different and dependent on ρ. In general, $\bar{\beta}_{Z_U} < \bar{\beta}_{Z_R}$ if $\rho < 1$, and $\bar{\beta}_{Z_U} > \bar{\beta}_{Z_R}$, if $\rho > 1$. See Figure 2 for $n_2 = 10$, $\lambda_2 = 0.3$, $\Delta_0 \in (0,1)$, and $\rho = 3/5, 1, 5/3$. It indicates that Z_R is more powerful and requires less observations for a specified power than Z_U when the first group has a smaller size than the second group. The result is the opposite when the samples size of the first group is more than that of the second group. Hence, when the sampling cost for a subject from the second group is more expensive than from the first group, and one considers a study of $\rho \geq 1$, then Z_U is suggested.

When using the normal approximation in testing a parameter of a discrete distribution, a continuity correction is often added in the test statistic to reduce the inflation of the type I error rate. The continuity correction revised by Pirie and Hamdan (1972) is employed in the Poisson problem. If the support of the estimator $\hat{\Delta}$ has equal spacings with space b, the continuity corrected test statistic is

$$\frac{\hat{\Delta} - \frac{1}{2}b}{\mathrm{se}(\hat{\Delta})}.$$

Pirie and Hamdan (1972) indicated that in Poisson problem, the MLE $\hat{\Delta}$ has equal spacings if one of n_1, n_2 is an integer multiple of the other. Specifically, when $n_1 = n_2$, $b = 1$. We find that for any n_1, n_2, the sampling distribution of the MLE $\hat{\Delta}$ has equal spacings with space $b = 1/(2m)$, where m is the least common multiple of n_1, n_2. Consequently, the continuity-corrected $Z_{U,c}$ and $Z_{R,c}$ can be obtained accordingly.

3. Exact P-values

When the sample sizes are insufficient or the mean values are relatively small, exact testing procedures are more adequate than asymptotic ones. Given a realization of a test statistic, an exact p-value is defined and calculated under the exact null distribution. In many applications, the null distribution often involves an unknown nuisance parameter(s). In the following, several testing procedures to deal with unknown nuisance parameters in the literature are reviewed. Both the statistics Z_U, Z_R are functions of the sufficient statistics (Y_1, Y_2). Under the null hypothesis, $H_0 : \lambda_1 = \lambda_2 = \lambda > 0$,

Y_1, Y_2 independently follow a Poisson distribution with mean $n_1\lambda$, $n_2\lambda$, respectively. In which, the common mean value λ is an unknown nuisance parameter.

Casella and Berger (1990) defined the standard p-value as the supremum over the null parameter space. The standard p-value considers the most conservative scenario and hence can guarantee the validity. Let z_0 be the observed value of the Wald test. The standard p-value is given as

$$\sup_{\lambda>0} \mathrm{P}(Z \geq z_0 \mid \lambda) = \sup_{\lambda>0} \sum_{y_1' \geq 0} \sum_{y_2' \geq 0} \mathrm{poi}(y_1', n_1\lambda)\mathrm{poi}(y_2', n_2\lambda) I_{\{Z \geq z_0\}},$$

where Z is either Z_U or Z_R; $\mathrm{poi}(\cdot, v)$ is the probability function of Poisson distribution with mean v; and I is the indicator function. The null parameter space is unbounded in this problem, and the computation of the standard p-value is difficult in practice. In addition, without taking the data information into consideration, one may obtain an unnecessarily conservative conclusion.

To ease the computational burden brought by searching the supremum over an infinite interval, Berger and Boos (1994) proposed a confidence-set p-value that is the supremum over a confidence set of the nuisance parameter. Given an observation z_U of Z_U, the confidence-set p-value is defined as

$$p_{CI,U} = \sup_{\lambda \in C_\delta} \mathrm{P}(Z_U \geq z_U \mid \lambda) + \delta,$$

where C_δ is a $100(1-\delta)\%$ confidence interval of the nuisance parameter λ. On the other hand, given z_R,

$$p_{CI,R} = \sup_{\lambda \in C_\delta} \mathrm{P}(Z_R \geq z_R \mid \lambda) + \delta.$$

In which, δ is a positive real number and is far less than α for a nontrivial conclusion. In this study, we consider the following $100(1-\delta)\%$ exact confidence interval C_δ of λ,

$$\frac{1}{2(n_1+n_2)} \left(\chi^2_{1-\delta/2,\, 2(Y_1+Y_2)}, \chi^2_{\delta/2,\, 2(Y_1+Y_2+1)} \right),$$

where $\chi^2_{s,v}$ is the $100(1-s)$-th percentile of a chi-square distribution with degrees of freedom v. This exact confidence interval is constructed based on the exact relationships between Poisson and chi-square distribution random variables,

$$\mathrm{P}(Y_1 + Y_2 \leq y_0 \mid \lambda) = \mathrm{P}(\chi^2_{2(y_0+1)} > 2(n_1+n_2)\lambda),$$

and

$$\mathrm{P}(Y_1 + Y_2 \geq y_0 \,|\, \lambda) = \mathrm{P}(\chi^2_{2(y_0)} > 2(n_1 + n_2)\lambda).$$

See Casella and Berger (1990, p.434) and Sahai and Khurshid (1993). Although the exact confidence interval has been shown tend to be conservative, it is employed here for convenience sake. In the numerical study discussed in the next section, the confidence-set p-values are found by grid-search.

Krishnamoorthy and Thomson (2004) proposed an alternative exact p-value by using the restricted MLE of the nuisance parameter λ, $\tilde{\lambda}_0$. That is, given z_U, z_R, the estimated p-values are defined as

$$p_{E,U} = \mathrm{P}(Z_U \geq z_U \,|\, \tilde{\lambda}_0), \quad p_{E,R} = \mathrm{P}(Z_R \geq z_R \,|\, \tilde{\lambda}_0),$$

respectively. The estimated p-value has a great reduction in computation and is found to perform well empirically by the authors. However, the resultant p-value does not guarantee validity theoretically.

Given any $\lambda_1 > 0$, $\lambda_2 > 0$, the exact power of the test by using an exact p-value, p, can be calculated by

$$\sum_{y_1 \geq 0} \sum_{y_2 \geq 0} \mathrm{poi}(y_1, n_1\lambda_1)\mathrm{poi}(y_2, n_2\lambda_2) I_{\{p \leq \alpha\}},$$

where p can be a confidence-set p-value or an estimated p-value correspondent to the Wald statistic. Consequently, the required sample size of the second group to achieve a predetermined $1-\beta$ exact power level is found numerically as

$$\min\{n_2 : \sum_{y_1 \geq 0} \sum_{y_2 \geq 0} \mathrm{poi}(y_1, ([n_2\rho] + 1)\lambda_1)\mathrm{poi}(y_2, n_2\lambda_2) I_{\{p \leq \alpha\}} \geq 1 - \beta\}, \tag{3}$$

given ρ, λ_1, λ_2.

4. Numerical Study

In this numerical study, we investigate the performance of the asymptotic test by using Z_R, Z_U, respectively. The effect of a continuity correction is explored in both tests. The correspondent exact tests by using the estimated p-value and the confidence-set p-value are studied as well. As described above, the calculation of the exact power is straightforward when the test statistic depends on the data only through the two sufficient statistics Y_1, Y_2. Here, the exact type I error rate and the exact power of each test are calculated. We consider $\lambda_2 = 0.3, 0.4, 0.6, 1, 2, 3$, $\Delta_0 = 0, 1$ and $n_2 = 10, 30$, $\rho = 3/5, 1, 5/3$. The nominal significance level α is set as 0.05. The calculated type I error rate and power are presented in Table 1–4. The required samples sizes of the second group to achieve 80% power at $\Delta_0 = 0.6$ are provided in Table 5–7.

We first compare the two asymptotic tests in Table 1 to 4. Theoretically, at $\Delta_0 = 0$ the asymptotic sizes of the two tests are independent of ρ and equal to the nominal significance level α. However, the finite-sample results in Table 1 and Table 3 appear to be more consistent with the trend of the asymptotic power functions in Section 2. When $\rho = 3/5 < 1$, Z_R has more chance to reject the null hypothesis than Z_U. The trend becomes the opposite when $\rho > 1$. When ρ=1, both tests become equal and their sizes sometimes exceed the nominal level. When $\rho = 3/5$, the size of Z_R is not well-controlled at $\alpha = 5\%$. For $\rho = 5/3$, the inflation of the type I error rate of Z_U is even worse. In both cases, adding a continuity correction or increasing the sample size entail limited improvement. Overall speaking, Z_R is more robust to the choice of ρ than Z_U. Z_U is too liberal for $\rho > 1$ and too conservative for $\rho < 1$.

Next the two exact p-values are studied. Note that in finding the confidence-set p-value, the supremum is searched over 16 grids of the confidence interval of the common mean value. Table 1 and 3 show that the two exact approaches have their type I error rate well-controlled. The confidence-interval p-value is more conservative than the estimated p-value. The computations involved are greatly reduced for the estimated p-value. One should keep in mind that the estimated p-value is not a valid test theoretically. Although in these selected scenarios of our simulation, its type I error rate does not exceed the nominal level. It is possible that the estimated p-value has an inflated type I error rate in other cases.

Table 5–7 present the required sample size of the second group for 80% power at $\Delta_0 = 0.6$. The results for the two asymptotic tests are based on the asymptotic sample size formulae (1) and (2). For the two exact tests, the figures are the minimal integers such that the exact power achieves the level by (3). Between the two asymptotic tests, Z_R (Z_U) needs a slightly smaller sample than Z_U (Z_R) for $\rho < 1$ (> 1). However, with the smaller sample size, the exact type I error rate often exceeds the nominal level α. The inflation is more severe in the application of Z_U and showed limited improvement with the continuity correction.

Moreover, the sample sizes obtained for the two exact tests are near that of the asymptotic tests and the differences are within 3 units in all cases. With the calculated sample size, every exact test achieves the nominal power level and has a well-controlled type I error rate. In summary, although the exact tests are more time-consuming, they guarantee more adequate statistical conclusions. The asymptotic sample sizes (1) and (2) can be regarded as alternative quick solutions of (3) for the exact tests. The result is found to be close to the exact sample size.

5. A Real Example

In this section, the methods introduced are applied to the breast cancer study described in Ng and Tang (2005). Female subjects were classified according to whether they had been examined by using X-ray fluoroscopy during treatment for tuberculosis. The investigators suspect that the use of X-ray fluoroscopy will lead to a higher occurrence rate of breast cancer. Define λ_1 as the mean incidence number of breast cancer per person-year of the treatment group, in which patients had received X-ray; and λ_2 be the mean incidence number per person-year of the control group, in which patients were not examined by X-ray. Then we test the following hypothesis,

$$H_0 : \lambda_1 = \lambda_2 \quad \text{vs.} \quad H_1 : \lambda_1 > \lambda_2.$$

From Ng and Tang (2005), it was reported that the treatment group had $y_1 = 41$ cases of breast cancer in $n_1^* = 28010$ persons-year at risk and the control group had $y_2 = 15$ cases of breast cancer in $n_2^* = 19017$ person-years at risk. It was found that $\hat{\lambda}_1 = 1.464$, $\hat{\lambda}_2 = 0.789$ and $\tilde{\lambda}_0 = 1.191$ per 1000 person-year. Consequently, $z_U = 2.2047$, $z_R = 2.0818$ with asymptotic p-value 0.0137, 0.0187, respectively. The finding that the p-value of z_U is smaller than the p-value of z_R is consistent with our numerical results. When $\rho > 1$ (here, $28010/19017 = 1.47$), Z_U tends to have a more liberal conclusion than Z_R in an asymptotic test. The estimated p-value is evaluated at $\lambda_1 = \lambda_2 = \tilde{\lambda}_0 = 0.0011$. For the confidence-set p-value, the joint 99.9% (with $\delta = 0.001$) confidence set of (λ_1, λ_2) is $\{0.0008 \leq \lambda_1 = \lambda_2 \leq 0.00177\}$. And the supremum of the p-value of Z_R occurs at $\lambda_1 = \lambda_2 = 0.0014$, and the supremum of the p-value of Z_U occurs at $\lambda_1 = \lambda_2 = 0.0010$. The calculated p-value are reported in Table 8. All these p-values are less than $\alpha = 0.05$ and lead to the conclusion of rejecting the null hypothesis. The increase in the incidence rate of breast cancer by using the X-ray fluoroscopy achieves statistical significance.

6. Concluding Remarks

In this study, we investigate several asymptotic and exact statistical procedures for comparing two Poisson means. Two types of Wald test, Z_U, Z_R, are considered. The asymptotic power functions of the asymptotic procedures are derived and the correspondent asymptotic sample size formula are provided. Theoretically, the two asymptotic Wald tests are compared in terms of the power function and the required sample size. One concludes that the performance of the tests depend on the fraction ρ of the group sizes. Two exact procedures are introduced and the correspondent exact sample sizes can be found numerically. Based on the numerical studies, the asymptotic tests tend to have inflated type I error rates. On the contrary, the exact procedures have adequate performance overall, and dominate the asymptotic tests. Furthermore, in the numerical study, one observes that the result of comparison of the two test statistics depends on ρ as well. In summary, in application of an asymptotic procedure, Z_R is more robust than Z_U with respect to ρ and is hence suggested. In application of an exact procedure, the two test statistics both provide satisfactory performance and are comparable. Moreover, the quick solutions based on the asymptotic sample size formulae are found to provide good approximations to the exact sample sizes.

References

Berger, R. L. and Boos, D. D. (1994). P values maximized over a confidence set for the nuisance parameter. *Journal of the American Statistical Association*, **89**, 1012–1016.

Casella, G. and Berger, R. L. (1990). *Statistical Inference*, Pacific Grove, CA: Wadsworth.

Gail, M. (1974). Power computations for designing comparative Poisson trails. *Biometrics*, **30**, 231–237.

Gu, K., Ng, H. K., Tang, M. L., and Schucany, W. R. (2008). Testing the ratio of two Poisson rates. *Biometrical Journal*, **50**, 283–298.

Huffman, M. D. (1984). An improved approximate two-sample Poisson test. *Applied Statistics*, **33**, 224–226.

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

Ng, H. K. and Tang, M. L. (2005). Testing the equality of two Poisson means using the rate ratio. *Satistics in Medicine*, **24**, 955–965.

Pirie, W. R. and Hamdan, M. A. (1972). Some revised continuity corrections for discrete distributions. *Biometrics*, **28**(3), 693–701.

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

Sahai, H. and Khurshid, A. (1993). Confidence intervals for the mean of a Poisson distribution: A Review. *Biometrical Journal*, **35**, 857–867.

Shiue, W. and Bain, L. J. (1982). Experiment size and power comparisons for two-sample Poisson tests. *Applied Statistics*, **31**, 130–134.

Store, B. E. and Kim, C. (1990). Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association*, **85**(409), 146–155.

Thode, H. C. (1997). Power and sample size requirements for tests of differences between two Poisson rates. *The Statistican*, **46**, 227–230.

[Received April 2010; accepted July 2010.]

Table 1 The type I error rate ($\Delta_0 = 0$) of Z_R, Z_U for $n_2 = 10$.

			λ_2					
ρ	Test	p-value	0.3	0.4	0.6	1	2	3
3/5	Z_R	Asymptotic	0.0540	0.0528	0.0537	0.0519	0.0529	0.0524
		Asym+cont	0.0540	0.0528	0.0537	0.0505	0.0498	0.0496
		Confidence-set	0.0385	0.0412	0.0359	0.0375	0.0446	0.0483
		Estimated	0.0406	0.0476	0.0502	0.0466	0.0493	0.0483
	Z_U	Asymptotic	0.0219	0.0232	0.0266	0.0334	0.0410	0.0426
		Asym+cont	0.0219	0.0232	0.0265	0.0308	0.0378	0.0403
		Confidence-set	0.0385	0.0417	0.0402	0.0461	0.0482	0.0483
		Estimated	0.0389	0.0433	0.0459	0.0487	0.0482	0.0483
1	Z_R	Asymptotic	0.0497	0.0508	0.0515	0.0489	0.0496	0.0497
		Asym+cont	0.0331	0.0358	0.0371	0.0396	0.0414	0.0437
		Confidence-set	0.0448	0.0421	0.0454	0.0487	0.0475	0.0471
		Estimated	0.0448	0.0421	0.0454	0.0487	0.0496	0.0497
	Z_U	Asymptotic	0.0497	0.0508	0.0515	0.0489	0.0496	0.0497
		Asym+cont	0.0331	0.0358	0.0371	0.0397	0.0414	0.0437
		Confidence-set	0.0448	0.0421	0.0454	0.0487	0.0475	0.0471
		Estimated	0.0448	0.0421	0.0454	0.0487	0.0496	0.0497
5/3	Z_R	Asymptotic	0.0455	0.0474	0.0461	0.0484	0.0462	0.0467
		Asym+cont	0.0369	0.0388	0.0434	0.0447	0.0457	0.0460
		Confidence-set	0.0455	0.0474	0.0461	0.0482	0.0459	0.0467
		Estimated	0.0455	0.0474	0.0461	0.0484	0.0470	0.0491
	Z_U	Asymptotic	0.0799	0.0711	0.0644	0.0632	0.0563	0.0548
		Asym+cont	0.0724	0.0697	0.0629	0.0589	0.0562	0.0543
		Confidence-set	0.0353	0.0335	0.0324	0.0420	0.0457	0.0467
		Estimated	0.0455	0.0474	0.0461	0.0484	0.0470	0.0491

Table 2 The power at $\Delta_0 = 1$ of Z_R, Z_U for $n_2 = 10$.

			λ_2					
ρ	Test	p-value	0.3	0.4	0.6	1	2	3
3/5	Z_R	Asymptotic	0.7576	0.7037	0.6129	0.5024	0.3516	0.2834
		Asym+cont	0.7575	0.7034	0.6093	0.4873	0.3438	0.2723
		Confidence-set	0.6841	0.6209	0.5469	0.4524	0.3341	0.2705
		Estimated	0.7429	0.6816	0.5899	0.4848	0.3380	0.2705
	Z_U	Asymptotic	0.6505	0.5996	0.5275	0.4425	0.3139	0.2497
		Asym+cont	0.6500	0.5971	0.5162	0.4268	0.3005	0.2453
		Confidence-set	0.7042	0.6560	0.5878	0.4743	0.3374	0.2705
		Estimated	0.7282	0.6817	0.6013	0.4751	0.3374	0.2705
1	Z_R	Asymptotic	0.8387	0.7872	0.6996	0.5773	0.4073	0.3274
		Asym+cont	0.8044	0.7532	0.6641	0.5364	0.3818	0.3050
		Confidence-set	0.8323	0.7847	0.6992	0.5724	0.3988	0.3193
		Estimated	0.8323	0.7847	0.6994	0.5773	0.4073	0.3263
	Z_U	Asymptotic	0.8387	0.7872	0.6996	0.5773	0.4073	0.3274
		Asym+cont	0.8044	0.7532	0.6641	0.5364	0.3818	0.3050
		Confidence-set	0.8323	0.7847	0.6992	0.5724	0.3988	0.3193
		Estimated	0.8323	0.7847	0.6994	0.5773	0.4073	0.3263
5/3	Z_R	Asymptotic	0.8948	0.8522	0.7733	0.6419	0.4524	0.3657
		Asym+cont	0.8869	0.8422	0.7615	0.6275	0.4522	0.3589
		Confidence-set	0.8948	0.8521	0.7709	0.6339	0.4524	0.3657
		Estimated	0.8948	0.8523	0.7743	0.6499	0.4549	0.3729
	Z_U	Asymptotic	0.9208	0.8832	0.8086	0.6749	0.4876	0.3917
		Asym+cont	0.9140	0.8750	0.8002	0.6695	0.4875	0.3846
		Confidence-set	0.8693	0.8313	0.7595	0.6275	0.4524	0.3656
		Estimated	0.8948	0.8523	0.7743	0.6499	0.4549	0.3729

Table 3 The type I error rate ($\Delta_0 = 0$) of Z_R, Z_U for $n_2 = 30$.

			λ_2					
ρ	Test	p-value	0.3	0.4	0.6	1	2	3
3/5	Z_R	Asymptotic	0.0523	0.0525	0.0536	0.0524	0.0516	0.0510
		Asym+cont	0.0516	0.0493	0.0501	0.0496	0.0499	0.0502
		Confidence-set	0.0356	0.0405	0.0426	0.0483	0.0484	0.0489
		Estimated	0.0467	0.0476	0.0499	0.0483	0.0499	0.0496
	Z_U	Asymptotic	0.0314	0.0370	0.0404	0.0426	0.0447	0.0455
		Asym+cont	0.0297	0.0336	0.0374	0.0403	0.0431	0.0441
		Confidence-set	0.0450	0.0463	0.0477	0.0483	0.0482	0.0486
		Estimated	0.0490	0.0471	0.0477	0.0483	0.0499	0.0496
1	Z_R	Asymptotic	0.0492	0.0489	0.0498	0.0497	0.0497	0.0500
		Asym+cont	0.0394	0.0396	0.0408	0.0437	0.0453	0.0465
		Confidence-set	0.0486	0.0487	0.0475	0.0471	0.0486	0.0488
		Estimated	0.0486	0.0489	0.0498	0.0497	0.0497	0.0498
	Z_U	Asymptotic	0.0492	0.0489	0.0498	0.0497	0.0497	0.0500
		Asym+cont	0.0394	0.0396	0.0408	0.0437	0.0453	0.0465
		Confidence-set	0.0486	0.0487	0.0475	0.0471	0.0486	0.0488
		Estimated	0.0486	0.0489	0.0498	0.0497	0.0497	0.0498
5/3	Z_R	Asymptotic	0.0456	0.0467	0.0482	0.0470	0.0486	0.0491
		Asym+cont	0.0452	0.0451	0.0464	0.0469	0.0477	0.0480
		Confidence-set	0.0456	0.0467	0.0482	0.0470	0.0486	0.0490
		Estimated	0.0484	0.0497	0.0496	0.0475	0.0497	0.0499
	Z_U	Asymptotic	0.0639	0.0621	0.0593	0.0554	0.0547	0.0536
		Asym+cont	0.0598	0.0576	0.0570	0.0553	0.0537	0.0529
		Confidence-set	0.0393	0.0414	0.0459	0.0469	0.0479	0.0484
		Estimated	0.0453	0.0475	0.0496	0.0475	0.0497	0.0499

Table 4 The power at $\Delta_0 = 1$ of Z_R, Z_U for $n_2 = 30$.

			λ_2					
ρ	Test	p-value	0.3	0.4	0.6	1	2	3
3/5	Z_R	Asymptotic	0.9905	0.9805	0.9518	0.8765	0.6935	0.5676
		Asym+cont	0.9896	0.9791	0.9497	0.8716	0.6893	0.5640
		Confidence-set	0.9871	0.9768	0.9477	0.8697	0.6852	0.5595
		Estimated	0.9896	0.9788	0.9478	0.8697	0.6892	0.5618
	Z_U	Asymptotic	0.9865	0.9745	0.9415	0.8577	0.6709	0.5460
		Asym+cont	0.9853	0.9722	0.9376	0.8537	0.6658	0.5424
		Confidence-set	0.9889	0.9784	0.9478	0.8690	0.6852	0.5577
		Estimated	0.9889	0.9784	0.9478	0.8697	0.6892	0.5635
1	Z_R	Asymptotic	0.9984	0.9956	0.9842	0.9415	0.7918	0.6668
		Asym+cont	0.9979	0.9946	0.9816	0.9343	0.7814	0.6545
		Confidence-set	0.9983	0.9953	0.9832	0.9393	0.7885	0.6612
		Estimated	0.9984	0.9956	0.9842	0.9405	0.7918	0.6654
	Z_U	Asymptotic	0.9984	0.9956	0.9842	0.9415	0.7918	0.6668
		Asym+cont	0.9979	0.9946	0.9816	0.9343	0.7814	0.6545
		Confidence-set	0.9983	0.9953	0.9832	0.9393	0.7885	0.6612
		Estimated	0.9984	0.9956	0.9842	0.9405	0.7918	0.6654
5/3	Z_R	Asymptotic	0.9998	0.9991	0.9953	0.9730	0.8625	0.7457
		Asym+cont	0.9998	0.9991	0.9953	0.9720	0.8604	0.7425
		Confidence-set	0.9998	0.9991	0.9953	0.9726	0.8619	0.7447
		Estimated	0.9998	0.9992	0.9953	0.9746	0.8646	0.7479
	Z_U	Asymptotic	0.9999	0.9994	0.9963	0.9769	0.8730	0.7586
		Asym+cont	0.9998	0.9994	0.9963	0.9760	0.8716	0.7565
		Confidence-set	0.9998	0.9991	0.9953	0.9726	0.8612	0.7438
		Estimated	0.9998	0.9992	0.9953	0.9746	0.8646	0.7479

Table 5 To achieve 80% power at $\Delta_0 = 0.6$, the required sample size of the second group n_2 of Z_R, Z_U for $\rho = 3/5$. Based on the required samples n_2, the power and the type I error rate (in parentheses) are given.

Test	p-value		λ_2 0.3	0.4	0.6	1	2
Z_R	Asymptotic	n_2	27	31	41	59	105
		Power	0.8285	0.8088	0.8105	0.8045	0.8037
		(Size)	(0.0514)	(0.0543)	(0.0528)	(0.0516)	(0.0508)
	Asym+cont	Power	0.8179	0.8048	0.8067	0.8021	0.8017
		(Size)	(0.0506)	(0.0499)	(0.0509)	(0.0500)	(0.0499)
	Confidence-set	n_2	28	33	42	61	108
		Power	0.8147	0.8107	0.8118	0.8074	0.8055
		(Size)	(0.0449)	(0.0480)	(0.0477)	(0.0484)	(0.0488)
	Estimated	n_2	27	32	41	59	107
		Power	0.8078	0.8141	0.8018	0.8001	0.8068
		(Size)	(0.0451)	(0.0466)	(0.0485)	(0.0496)	(0.0498)
Z_U	Asymptotic	n_2	31	36	45	63	109
		Power	0.8301	0.8295	0.8216	0.8068	0.8053
		(Size)	(0.0345)	(0.0399)	(0.0426)	(0.0443)	(0.0469)
	Asym+cont	Power	0.8212	0.8188	0.8184	0.8039	0.8036
		(Size)	(0.0304)	(0.0368)	(0.0395)	(0.0428)	(0.0464)
	Confidence-set	n_2	28	33	42	61	108
		Power	0.8023	0.8097	0.8072	0.8056	0.8050
		(Size)	(0.0440)	(0.0442)	(0.0458)	(0.0483)	(0.0488)
	Estimated	n_2	27	32	41	60	107
		Power	0.8083	0.8141	0.8018	0.8086	0.8068
		(Size)	(0.0466)	(0.0469)	(0.0485)	(0.0499)	(0.0498)

Table 6 To achieve 80% power at $\Delta_0 = 0.6$, the required sample size of the second group n_2 of Z_R, Z_U for $\rho = 1$. Based on the required samples n_2, the power and the type I error rate (in parentheses) are given.

			λ_2				
Test	p-value		0.3	0.4	0.6	1	2
Z_R	Asymptotic	n_2	21	25	31	45	79
		Power	0.8244	0.8279	0.8084	0.8059	0.8017
		(Size)	(0.0512)	(0.0489)	(0.0498)	(0.0505)	(0.0499)
	Asym+cont	Power	0.7971	0.8012	0.7909	0.7931	0.7940
		(Size)	(0.0373)	(0.0396)	(0.0410)	(0.0448)	(0.0471)
	Confidence-set	n_2	20	24	31	45	80
		Power	0.8070	0.8088	0.8014	0.8032	0.8025
		(Size)	(0.0454)	(0.0487)	(0.0475)	(0.0488)	(0.0489)
	Estimated	n_2	20	24	31	45	79
		Power	0.8073	0.8140	0.8084	0.8058	0.8011
		(Size)	(0.0454)	(0.0487)	(0.0498)	(0.0497)	(0.0499)
Z_U	Asymptotic	n_2	21	25	31	45	79
		Power	0.8244	0.8279	0.8084	0.8059	0.8017
		(Size)	(0.0512)	(0.0489)	(0.0498)	(0.0505)	(0.0499)
	Asym+cont	Power	0.7971	0.8012	0.7909	0.7931	0.7940
		(Size)	(0.0373)	(0.0396)	(0.0410)	(0.0448)	(0.0471)
	Confidence-set	n_2	20	24	31	45	80
		Power	0.8070	0.8088	0.8014	0.8032	0.8025
		(Size)	(0.0454)	(0.0487)	(0.0475)	(0.0488)	(0.0489)
	Estimated	n_2	20	24	31	45	79
		Power	0.8073	0.8140	0.8084	0.8058	0.8011
		(Size)	(0.0454)	(0.0487)	(0.0498)	(0.0497)	(0.0499)

Table 7 To achieve 80% power at $\Delta_0 = 0.6$, the required sample size of the second group n_2 of Z_R, Z_U for $\rho = 5/3$. Based on the required samples n_2, the power and the type I error rate (in parentheses) are given.

			λ_2				
Test	p-value		0.3	0.4	0.6	1	2
Z_R	Asymptotic	n_2	18	20	26	37	64
		Power	0.8376	0.8151	0.8240	0.8090	0.8022
		(Size)	(0.0479)	(0.0457)	(0.0474)	(0.0496)	(0.0492)
	Asym+cont	Power	0.8339	0.8080	0.8131	0.8054	0.8003
		(Size)	(0.0415)	(0.0431)	(0.0452)	(0.0496)	(0.0484)
	Confidence-set	n_2	18	20	26	37	65
		Power	0.8376	0.8151	0.8165	0.8040	0.8066
		(Size)	(0.0479)	(0.0456)	(0.0474)	(0.0487)	(0.0489)
	Estimated	n_2	17	20	25	36	64
		Power	0.8188	0.8201	0.8004	0.8069	0.8030
		(Size)	(0.0467)	(0.0456)	(0.0495)	(0.0476)	(0.0499)
Z_U	Asymptotic	n_2	15	18	23	34	62
		Power	0.8182	0.8148	0.8030	0.7967	0.8008
		(Size)	(0.0753)	(0.0645)	(0.0598)	(0.0579)	(0.0532)
	Asym+cont	Power	0.8072	0.8058	0.8006	0.7927	0.7988
		(Size)	(0.0749)	(0.0630)	(0.0574)	(0.0579)	(0.0524)
	Confidence-set	n_2	18	20	26	37	65
		Power	0.8213	0.8078	0.8131	0.8038	0.8062
		(Size)	(0.0374)	(0.0374)	(0.0451)	(0.0449)	(0.0488)
	Estimated	n_2	17	20	25	36	64
		Power	0.8188	0.8201	0.8004	0.8069	0.8030
		(Size)	(0.0401)	(0.0446)	(0.0495)	(0.0476)	(0.0499)

Table 8 The asymptotic, estimated and confidence-set p-value of the Wald Z-test Z_R, Z_U.

p-value	$Z_U = 2.2047$	$Z_R = 2.0818$
Asymptotic	0.0137	0.0187
Estimated	0.0186	0.0177
Confidence-set	0.0188	0.0182

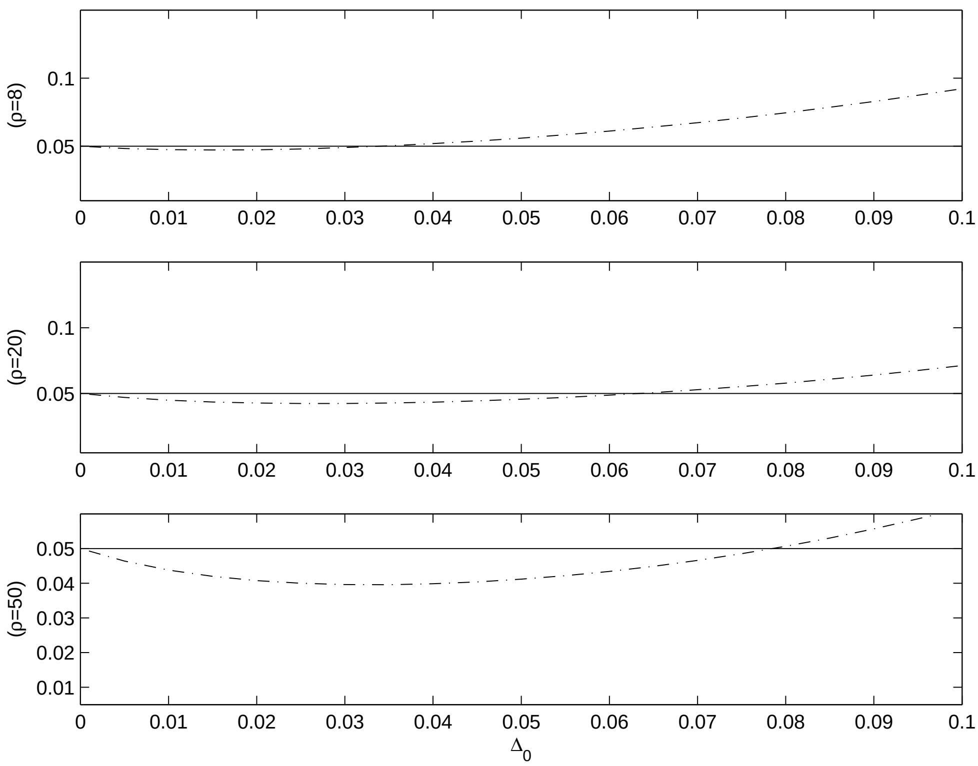

Figure 1 As $n_2 = 10$, $\lambda_2 = 0.3$, $\rho = 8, 20, 50$, the asymptotic power of Z_R (the dotted and dashed line) over $\delta_0 \in (0, 0.1)$.

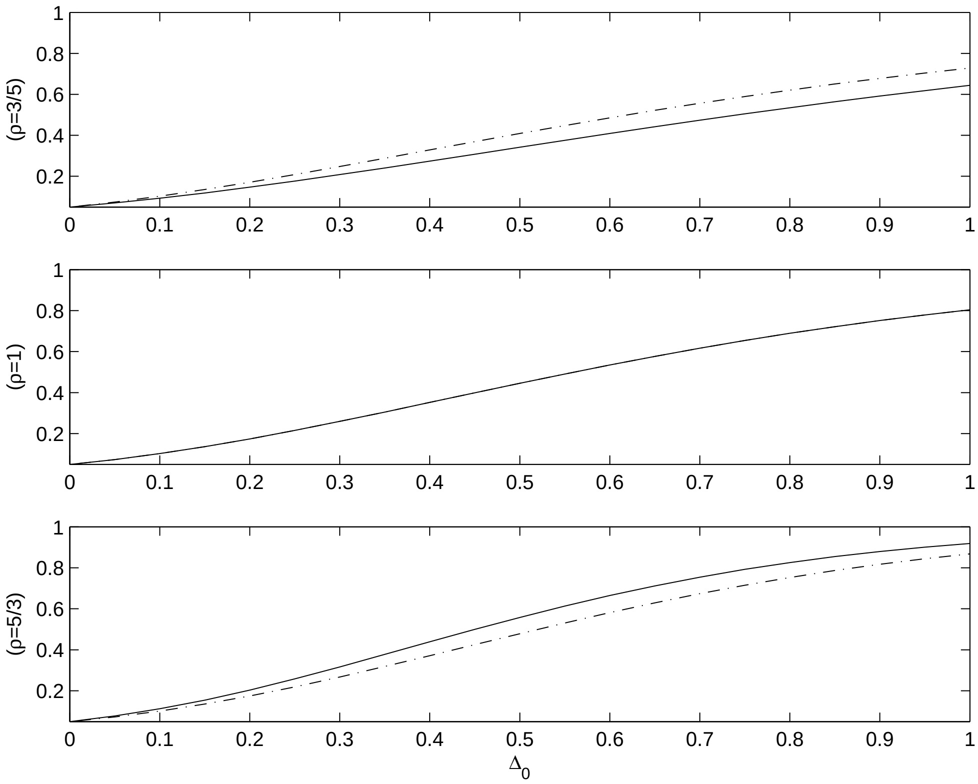

Figure 2 As $n_2 = 10$, $\lambda_2 = 0.3$, $\rho = 3/5, 1, 5/3$, the asymptotic powers of Z_R (the dotted and dashed line) and Z_U (the solid line) over $\delta_0 \in (0, 1)$.