Sankhyā : The Indian Journal of Statistics
Special issue in memory of D. Basu
2002, Volume 64, Series A, Pt. 3, pp 611-625

A TEST FOR THE POISSON DISTRIBUTION

By LAWRENCE D. BROWN
and
LINDA H. ZHAO
University of Pennsylvania, USA

SUMMARY. We consider the problem of testing whether a sample of observations comes from a single Poisson distribution. Of particular interest is the alternative that the observations come from Poisson distributions with different parameters. Such a situation would correspond to the frequently discussed situation of overdispersion.

We propose a new test for this problem that is based on Anscombe's variance stabilizing transformation. There are a number of tests commonly proposed, and we compare the performance of these tests under the null hypothesis with that of our new test. We find that the performance of our test is competitive with the two best of these. The asymptotic distribution of the new test is derived and discussed.

Use of these tests is illustrated through two examples of analysis of call-arrival times from a telephone call center. The example facilitates careful discussion of the performance of the tests for small parameter values and moderately large sample sizes.

1. Introduction

A variety of tests is available for testing whether a sample of observations comes from a Poisson distribution. This article proposes an additional test based on Anscombe's (1948) variance stabilizing transformation. We examine the performance of this test and compare it with three other tests in current use. We find this new test to be competitive in performance with the best of these alternatives. We recommend it on this basis, and also because the heuristic idea underlying it easily adapts for a variety of related applications.

We use call-arrival data gathered at an Israeli call center as motivation and illustration of the various problems and methodologies we discuss. We provide a very brief discussion in Section 2 of this application.

Paper received December 2001.
AMS (2000) *subject classification.* 62F03, 62F05.
Keywords and phrases. Poisson variables, Anscombe's transformation, likelihood ratio test, chi-squared test, overdispersion.

The three additional types of test statistics we examine are the likelihood ratio statistic, the corresponding chi-squared statistic sometimes called the "dispersion test", and a putatively normal version of this statistic sometimes attributed to Neyman and Scott. The performance of the Neyman-Scott test is shown to be inferior to those built from the remaining statistics. We favour the new test based on its ease of use, diagnostic ability and breadth of application.

Suppose the null hypothesis is true, i.e., the data come from a Poisson(λ) distribution. When λ is not small all three recommended tests (the new test, the dispersion χ^2, and the likelihood ratio) appear fully satisfactory for practical applications. When λ is small the nominal null distribution for the likelihood ratio test is quite inaccurate. The test should not then be used in the usual form as presented here.

In Section 5 we state the asymptotic distribution of our new test statistic as $n \to \infty$ with λ fixed. It is shown that this implies that the heuristic nominal null distribution is not fully accurate when λ is small, even if $n \to \infty$. Thus, when λ is small (say $\lambda \leqslant 5$), the new test we propose is slightly inaccurate. The source of that inaccuracy is explained in Section 4, and an easily implemented correction is proposed that is satisfactory for moderately large sample sizes (say 50 or more, depending partly on how small λ is).

In Section 2 we describe the call center data we use as an example of an application of our methodology. The various tests are described in Section 3, including the new test we propose based on Anscombe's variance stabilizing transformation. Section 4 presents some simulation results comparing our test and the various other tests. The asymptotic distribution of the new test is discussed in Section 5.

2. Call Center Arrival Data

The data accompanying our study was gathered at a relatively small Israeli bank telephone call center in 1999. The portion of data of interest to us here involves records of the arrival time of service-request calls to the center. These are calls in which the caller requests service from a call center representative. It is reasonable to conjecture that these arrival times are well modelled by an inhomogeneous Poisson process. The arrival rate for this process should depend only on the time of day, and perhaps other calendar related covariates such as month or day of the week. There are different categories of service that may be requested, and preliminary analysis clearly shows that this factor should also be considered since the arrival rate patterns

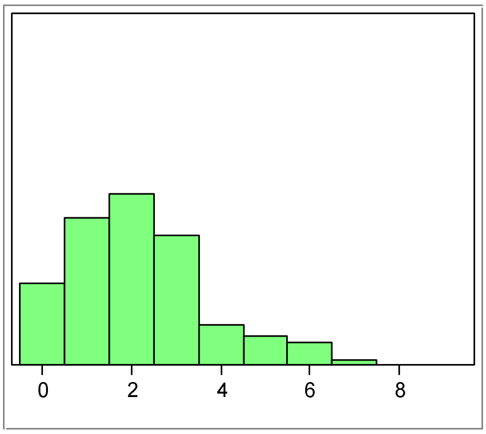

Figure 1: No. of daily calls for Internet Service arriving between 4:30pm and 4:45pm, Regular weekdays, Aug. – Dec. $n = 107$, $\bar{x} = 2.18$, $s^2 = 2.47$.

differ considerably. For more information about various aspects of this data see Brown et. al (2001a). Other features of the call arrival process are investigated in Brown, Mandelbaum, Sakov, Shen, Zeltyn and Zhao (2001b).

If the arrival process for a given call category is as above then the number of arrivals each day within any given interval of time should be independent Poisson variables with a parameter that depends only on the given time interval. If other covariates are involved (such as day of the week) then the Poisson parameter may also depend on these covariates.

The histograms in Figures 1 and 2 show the results from two typical samples. Figure 1 shows the number of standard calls arriving on each regular workday in Nov. and Dec., between 4:30pm and 4:45pm. Figure 2 is a similar histogram for the special category of calls requesting internet assistance arriving between 4:30pm and 4:45pm from Aug. through Dec. In each case it is of interest to test the null hypothesis that these data arise from Poisson populations with their own respective means. Note the different levels of calls/day in these two samples, as well as the different sample sizes.

One reason for considering standard calls only for Nov. and Dec. is that there is some evidence of an increased rate of standard calls in Nov. and Dec.

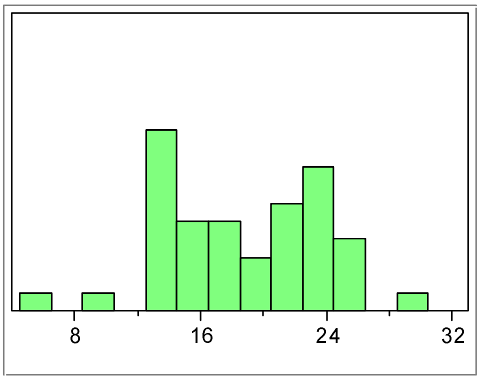

Figure 2: No. of daily calls for Standard Service arriving between 4:30pm and 4:45pm, Regular weekdays, Nov. – Dec. $n = 44$, $\bar{x} = 18.66$, $s^2 = 25.95$.

3. Tests for the Poisson Distribution

Let $X_1, \ldots, X_n$ be independent non-negative integer valued random variables with $P(X = x) = f(x)$. The basic null hypothesis of interest is that

$$H_0 : X_i \sim \text{Poiss}(\lambda_i), \quad \lambda_1 = \ldots = \lambda_n. \tag{1}$$

We consider the alternative hypothesis that

$$H_a : X_i \sim \text{Poiss}(\lambda_i), \ \sum (\lambda_i - \overline{\lambda})^2 > 0. \tag{2}$$

We propose a new test for this problem. We also briefly describe some other tests in common use for H_0. We will later focus our attention on properties of the new test in relation to the others.

3.1. *A new test based on Anscombe's statistic.* Anscombe (1948) derived the second order variance stabilizing transformation for a Poisson variable. Also see Bartlett (1947). If $N \sim \text{Poiss}(\lambda)$ Anscombe showed that

$$\text{Var}_\lambda \left(\sqrt{N + \frac{3}{8}} \right) = \frac{1}{4} + O\left(\frac{1}{\lambda}\right). \tag{3}$$

On this basis it is natural to define $Y_i = \sqrt{X_i + 3/8}$ and use the statistic

$$T_{new} = 4 \sum (Y_i - \overline{Y})^2$$

to provide a test for H_0.

Formula (3) suggests that Y_i is approximately normal with variance 1/4 and mean

$$\nu(\lambda_i) = E_{\lambda_i}(Y_i) = E_{\lambda_i}(\sqrt{N+3/8}). \tag{4}$$

Such an assertion is asymptotically correct as $\lambda_i \to \infty$. Under this approximation it would follow that when H_0 is true, T_{new} has approximately a Chi-squared distribution with $n-1$ df. We thus reject H_0 if $T_{new} > \chi^2_{n-1;1-\alpha}$. Further one may conclude that under H_a, T_{new} has approximately a non-central χ^2_{n-1} distribution. In summary it is reasonable to act as if

$$T_{new} \sim \chi^2_{n-1}\left(4\sum(\nu(\lambda_i) - \overline{\nu}_n)^2\right) \tag{5}$$

where

$$\overline{\nu}_n = \frac{1}{n}\sum_{i=1}^{n}\nu(\lambda_i).$$

The empirical results in Section 4 indicate that this approximation is reasonably accurate under H_0 even for fairly small λ and n. Further simulations we have carried out (not reported here) suggest that this approximation is also fairly good for a variety of choices of $\{\lambda_i\}$ in H_a, even for moderate n so long as none of the λ_i are small.

Section 5 presents some asymptotic theory concerning the distribution of T_{new}. This theory helps explain why (5) provides numerically satisfactory results even though it is not quite asymptotically valid as $n \to \infty$, even under H_0.

In the context of nonparametric density estimation Brown, Zhang and Zhao (2001) suggested using the transformation $\sqrt{N+1/4}$ instead of $\sqrt{N+3/8}$. This is because

$$E_\lambda(\sqrt{N+1/4}) = \sqrt{\lambda} + O(1/\lambda).$$

In the context of Brown, Zhang and Zhao (2001) accuracy in estimation of $\sqrt{\lambda}$ is of prime importance, rather than stability of the variance. However for the Poisson tests under investigation here validity of (3) is more important, and the transformation $\sqrt{X_i+3/8}$ performs slightly better than would $\sqrt{X_i+1/4}$.

Brown, Cai and DasGupta (2001) investigated confidence intervals for a Poisson mean. This is a related problem but techniques for best confidence intervals do not necessarily extend to best tests of H_0, and vice-versa. Some results about the confidence interval problem are also reported in Brown, Zhang and Zhao (2001).

The test statistic T_{new} appears to us a natural proposal given Anscombe's well known variance stabilizing transformation. We expect it has been used in the form (6) by some practitioners. But the only other reference we have found is Huffman (1984) that presents a sample size two ($n = 2$) version of this test, and also discusses testing a generalization of H_0 when $n = 2$.

3.2 *Likelihood ratio statistic.* The likelihood ratio statistic for testing H_0 versus H_a is

$$T_{LR} = 2\sum_{i=1}^{n} X_i \ln\left(\frac{X_i}{\overline{X}}\right). \tag{6}$$

Under the null hypothesis this statistic is asymptotically distributed as a Chi-squared variable with $n-1$ df. (asymptotically as $n \to \infty$ for fixed λ). Hence this test rejects H_0 when $T_{LR} > \chi^2_{n-1;1-\alpha}$.

Under alternatives in H_a this statistic has approximately a non-central Chi-squared distribution with $n-1$ df and non-centrality parameter $\psi^2 = \sum_{i=1}^{n}(\lambda_i - \overline{\lambda})^2/\overline{\lambda}$ where $\overline{\lambda} = \sum_{i=1}^{n}\lambda_i/n$. We write, $T_{LR} \sim \chi^2_{n-1}(\psi^2)$. This approximation is asymptotically valid as $\lambda \to \infty$ for fixed n with $\lambda_1, \ldots, \lambda_n$ chosen to depend on n in such a way that ψ^2 remains constant, or as $n \to \infty$ with $\lambda_1, \ldots, \lambda_n$ chosen so that $\liminf \overline{\lambda} > 0$ and $\psi^2 = O(\sqrt{n})$.

3.3 *Conditional Chi-squared statistic.* Under the null hypothesis the conditional distribution of $X_1, \ldots, X_n$ given $\sum X_i = n\overline{X}$ is multinomial $(n\overline{X}, (1/n, \ldots, 1/n))$. This motivates as a test statistic,

$$T_{CC} = \sum \frac{(X_i - \overline{X})^2}{\overline{X}} = \frac{(n-1)S^2}{\overline{X}} \tag{7}$$

where under H_0 has an (asymptotic) Chi-squared distribution with $n-1$ df. (Hence we reject H_0 if $T_{CC} > \chi^2_{n-1;1-\alpha}$.) This statistic can also be motivated as the asymptotic chi-squared approximation to the likelihood ratio test of Section 3.2. Some authors (e.g., Rice(1995)) call this the Poisson dispersion test or the variance test (Cochran, 1954). See also Agresti (1990, p. 479).

Under H_a, $T_{CC} \sim \chi^2_{n-1}(\psi^2)$, with this approximation being asymptotically valid under the same conditions as described for T_{LR}.

3.4 *Neyman-Scott statistic.* This statistic is directly motivated by the expression (7). It is often used as a test of H_0. See for example Lindsay (1995) and Jongbloed and Koole (2001) for application of this test to telephone call-center data. The statistic is

$$T_{NS} = \sqrt{\frac{n-1}{2}}\left(\frac{S^2}{\overline{X}} - 1\right).$$

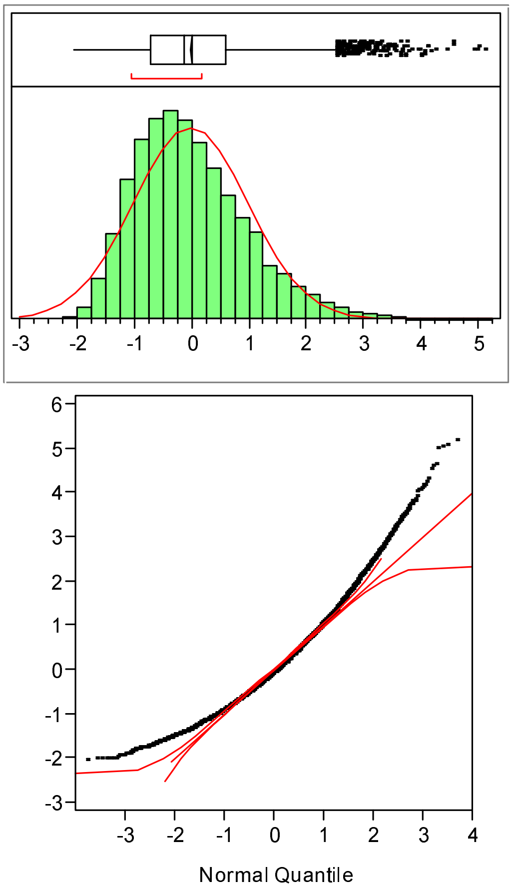

Figure 3: Histogram (with best fitting normal curve) and Normal Quantile plot for T_{NS}; $\lambda = 12$, $n = 12$, 10,000 Monte Carlo samples

This statistic is normalized so that asymptotically $T_{NS} \sim N(\psi^2/\sqrt{2n}, 1)$. (Hence this test rejects if $T_{NS} > \Phi^{-1}(1-\alpha)$.) The asymptotic assertion here is valid as $n \to \infty$ with $\lambda_1, \ldots, \lambda_n$ chosen so that $\psi^2 = O(\sqrt{n})$ and $\liminf \overline{\lambda} > 0$.

It can be seen that under H_0, T_{NS} is the standard normal approximation to the chi-squared statistic T_{CC}. The null distribution of T_{NS} is not close to its limiting normal distribution until n is moderately large. This is shown in Fig 3, which displays an empirical approximation to the null distribution of T_{NS} for the case $\lambda = 12$, $n = 12$.

The fact that the true null distribution of T_{NS} is not close to its nominal limiting distribution means that tests constructed using critical values from this will not have close to their nominal significance level. Correspondingly their nominal P-values based on the limiting distribution will also be considerably in error. For this reason we recommend against use of T_{NS}. (For comparative purposes we have nevertheless included T_{NS} in the numerical results in Section 4.)

4. Empirical Results Under H_0

This section reports selected empirical results about the null distribution of the statistics T_{new}, T_{LR}, T_{CC}, T_{NS}. These results are summarized in Table 1. This table gives information about the empirical type I error rates for tests computed using the nominal null distribution of various statistics. The table also contains an overall measure of how close is the empirical χ^2 or normal null distribution. The table also indirectly provides information about the accuracy of P-values calculated from the nominal distributions since accuracy of type I error rates and of P-values are linked concepts.

The general impression from the table is that the empirical type I error rates using any of T_{new}, T_{LR}, T_{CC} are reasonably accurate when $\lambda \geqslant 12$. Even when $\lambda = 5$ satisfactory accuracy is evident for T_{new} and T_{CC}. The results in Section 5 suggest a modified nominal null distribution be used when λ is even smaller to calculate critical values for T_{new}. The results in Section 5 also confirm that T_{LR} is a less desirable choice when $\lambda \leqslant 5$. Overall, the empirical type I errors using the T_{NS} are less accurate than those from the other three statistics, as one would also expect from the results reported in Fig 3.

Table 1. Empirical Type I errors (10,000 repetitions) and $\hat{D}^*_N$ defined in (8)

n	λ	Statistic	$\alpha = .1$ SE.=0.003	$\alpha = .05$ SE.=0.002	$\alpha = .01$ SE.=0.001	α=.005 SE.=0.001	$\hat{D}^* = \sup\lvert\hat{H} - G\rvert$ ESE.=0.007
20	5	T_{new}	0.1107	0.0585	0.0132	0.0070	0.0130
20	5	T_{LR}	0.1359	0.0724	0.0173	0.0089	0.0588
20	5	T_{CC}	0.0977	0.0495	0.0103	0.0059	0.0105
20	5	T_{NS}	0.1039	0.0620	0.0220	0.0148	0.0457
12	12	T_{new}	0.1050	0.0540	0.0122	0.0065	0.0094
12	12	T_{LR}	0.1102	0.0563	0.0120	0.0062	0.0130
12	12	T_{CC}	0.1007	0.0505	0.0104	0.0054	0.0057
12	12	T_{NS}	0.1082	0.0670	0.0260	0.0179	0.0611
5	25	T_{new}	0.1008	0.0510	0.0103	0.0053	0.0035
5	25	T_{LR}	0.1027	0.0517	0.0101	0.0051	0.0069
5	25	T_{CC}	0.0994	0.0490	0.0095	0.0046	0.0066
5	25	T_{NS}	0.1059	0.0696	0.0312	0.0231	0.0955

The quantities reported in Table 1 are defined as follows. Let G denote generally the nominal null cumulative distribution of a statistic T. (For T_{NS}, G is standard normal. For the other statistics G is χ^2_{n-1}.) Let κ_α denote the α critical values, $\kappa(\alpha) = G^{-1}(1-\alpha)$. Let H denote the true null distribution of the statistic. Then the true type I error is $1 - H(\kappa(\alpha))$. The table reports Monte-Carlo estimates based on 10,000 samples of these quantities for various statistics and values of n, λ. The standard errors reported in the table are the theoretical values $\sqrt{\alpha(1-\alpha)/10000}$.

Table 1 also reports a measure of the disparity between the nominal G and the true H as measured via the Kolmogorov-Smirnov distance

$$D^* = \sup_t |H(t) - G(t)|.$$

Again, the values reported derive from 10,000 simulations. To be more precise, each entry in the last column of the table reports the value of

$$\hat{D}^*_N = \sup_t |\hat{H}_N(t) - G(t)| \tag{8}$$

where $\hat{H}$ denotes the sample CDF from the N=10000 simulated values of T.

Simulated values of $\hat{D}^*_N$ have the Kolmogorov-Smirnov limiting distribution. This is not a normal distribution. In particular, a 95% confidence region for $H(t)$ is

$$\sup_t |H(t) - \hat{H}_n(t)| \leqslant 2\, ESE$$

where

$$ESE = \frac{1.96 \times 0.5}{1.36\sqrt{10000}} = 0.007.$$

For this reason we have chosen to report the effective standard error, ESE, as the measure of the precision of our Monte-Carlo simulation.

Note that for T_{new} and T_{CC}, D^*_N is acceptably small. Indeed, it is less than $2 \times ESE$, and hence using this we would not reject at level .05 the null hypothesis that $H = G$. This is also true for T_{LR} when $\lambda = 12$ and 25. But when $\lambda = 5$ the performance in this regard is less satisfactory, as is the performance of T_{NS} for all combinations of n, λ in the table.

5. Asymptotic Distribution of T_{new}

We have suggested approximating the null distribution of T_{new} as a Chi-squared with $n-1$ df. The empirical results in the previous section suggest that this approximation is satisfactory for practical applications. We now explore the asymptotic distribution of T_{new} as $n \to \infty$. We show that the limiting null distribution is not Chi-squared $(n-1)$ but is very close to Chi-squared $(n-1)$ so long as λ is not small. This closeness explains why the Chi-squared approximation is suitable for most practical applications. Finally, we also provide similar results about the distribution under H_a.

Note that $E_\lambda(T_{new}) = 4(n-1)\text{Var}_\lambda(Y)$. As noted at (3), Anscombe (1948) proved by an asymptotic expansion that

$$\xi(\lambda) \stackrel{\triangle}{=} 4\text{Var}_\lambda(Y) = 1 + O(1/\lambda). \tag{9}$$

This expression is not only asymptotically accurate — it is nearly the exact truth so long as $\lambda > 4$. Figs 4 and 5 show plots of $\xi(\lambda) = 4\text{Var}_\lambda(Y) = E_\lambda(T_{new})/(n-1)$ derived via direct calculation. In particular,

$$\xi(\lambda) = (n-1)^{-1}E_\lambda(T_{new}) \leqslant 1.0025. \tag{10}$$

(The maximum value of $E_\lambda(T_{new})$ occurs at approximately $\lambda = 5.5$.) This means that T_{new} is positively biased by at most a very small amount, and so suggests that a test based on T_{new} will not have significance levels much below their nominal value. That is, this suggests while the test based on T_{new} may be conservative, it will not be radical by very much.

The results in Figs 4 and 5 suggest that the distribution of Y may effectively be very close to normal. As further exploration of this possibility, note that if Y were exactly normal then we would have $\text{Var}((Y-\nu_\lambda)^2) = 2$. Fig 6 is a plot of

$$\rho(\lambda) = \left[\frac{\text{Var}((Y-\nu_\lambda)^2)}{2}\right]^{1/2}. \tag{11}$$

Note that $\rho(\lambda) \approx 1$ whenever $\lambda > 4$. In particular, $\rho(\lambda) \leqslant 1.054$ with the maximum occurring at $\lambda = 5.4$. Again, this suggests that the test based on T_{new} will be conservative for very small λ, but will not for any λ be "radical" by very much.

Here is a formal statement of the asymptotic result.

Theorem 5.1 *Assume H_0 is true, λ is fixed and $n \to \infty$. Then*

$$\frac{1}{\rho(\lambda)}\sqrt{\frac{n-1}{2}}\left(\frac{T_{new}}{n-1} - \xi(\lambda)\right) \to N(0,1) \tag{12}$$

in distribution where ξ, ρ are defined in (10) and (11).

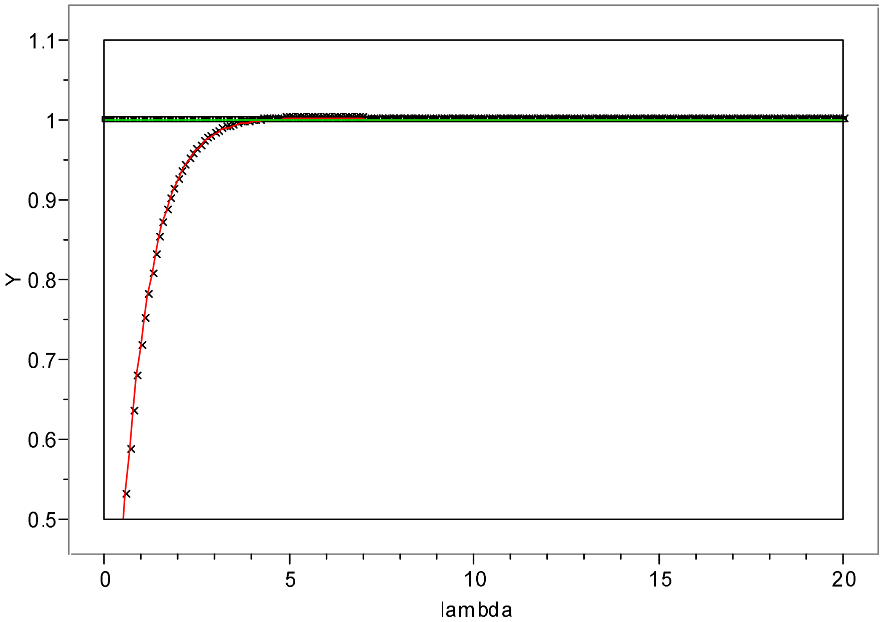

Figure 4: Plot of $E_\lambda(T_{new}/(n-1))$

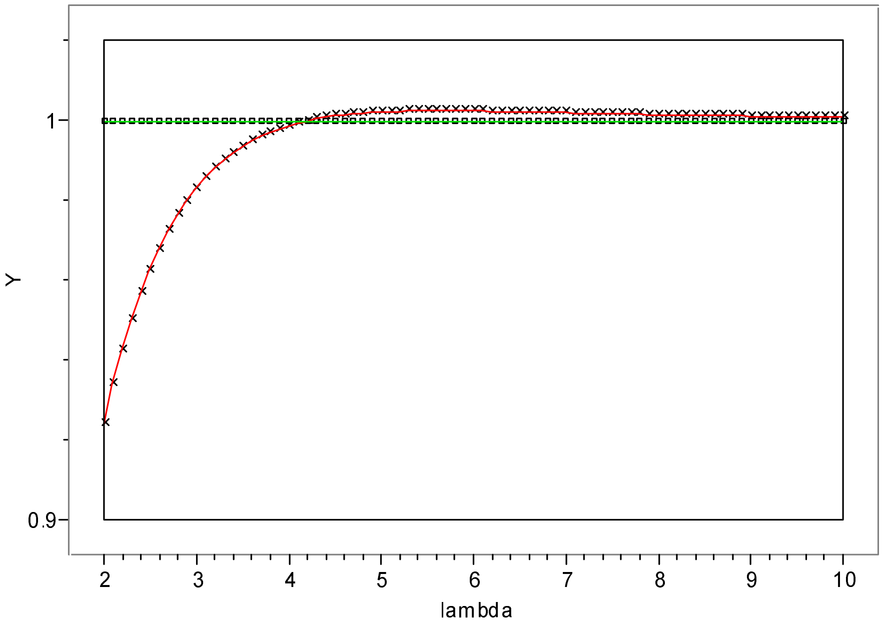

Figure 5: Detail of Fig 4

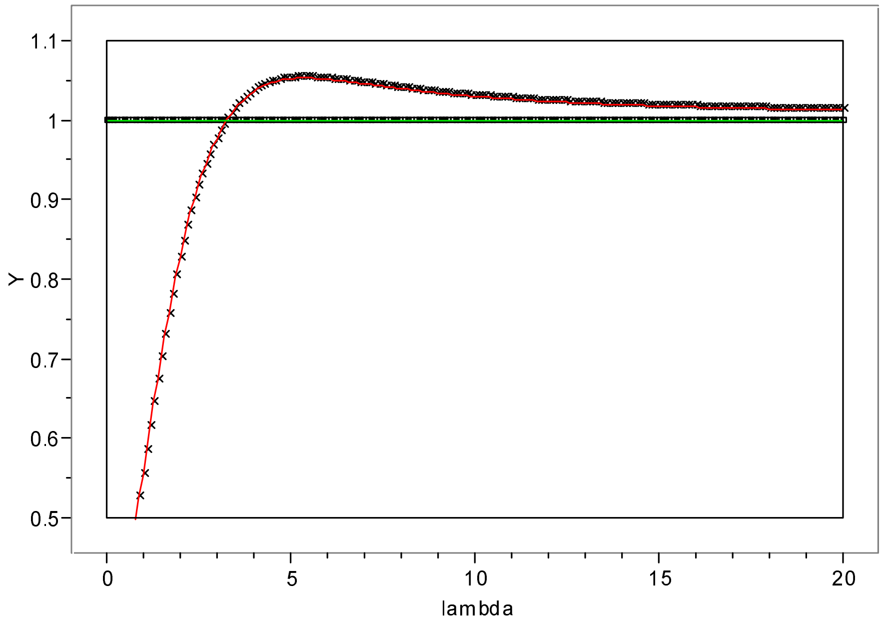

Figure 6: Plot of $\rho(\lambda)$ as defined in (11)

Remarks. Recall that if $Z \sim \chi^2_{n-1}$ then

$$\sqrt{\frac{n-1}{2}}\left(\frac{Z}{n-1}-1\right) \to N(0,1).$$

Note that both $\xi(\lambda) \approx 1$ and $\rho(\lambda) \approx 1$. It is thus clear that for large n, T_{new} is reasonably closely approximated as a χ^2_{n-1} variable, even though its asymptotic distribution is not exactly χ^2_{n-1} as $n \to \infty$ for fixed λ.

If one were in a situation where n is moderately large and λ is small then (12) suggests that the nominal χ^2 critical values (and P-values) can be slightly improved by calculating critical values (and P-values) from the normal distribution in (12) calculated at $\hat{\lambda} = \overline{X}$.

The formula for the approximate P-value thus becomes

$$P \approx 1 - \Phi^{-1}\left(\frac{1}{\rho(\overline{X})}\sqrt{\frac{n-1}{2}}\left(\frac{T_{new}}{n-1} - \xi(\overline{X})\right)\right). \qquad (13)$$

Proof: The result follows from the definition of ξ, ρ, the central limit theorem and Slutsky's theorem. We omit the straightforward details.

Similar reasoning using the central limit theorem for independent non-identically distributed random variables yields the following result.

THEOREM 5.2 *Let $\lambda_1, \ldots, \lambda_n$ depend on n. Let*

$$\begin{aligned} \overline{\xi}_n &= \frac{1}{n}\sum \xi(\lambda_i) \\ \overline{\rho}_n^2 &= \frac{1}{n}\sum \rho^2(\lambda_i) \\ \psi^2 &= 4\sum(\nu(\lambda_i) - \overline{\nu}_n)^2 \\ \overline{\nu}_n &= \frac{1}{n}\sum \nu(\lambda_i). \end{aligned} \tag{14}$$

Assume

$$\liminf_{n\to\infty} \overline{\rho}_n > 0 \;\; and \;\; \limsup_{n\to\infty} \overline{\rho}_n < \infty. \tag{15}$$

Then

$$P\left(\frac{1}{\overline{\rho}_n}\sqrt{\frac{n-1}{2}}\left(\frac{T_{new}}{n-1} - \overline{\xi}_n\right) > C\right) \to 1 - \Phi\left(C - \frac{4\sum(\nu(\lambda_i) - \overline{\nu}_n)^2}{\overline{\rho}_n\sqrt{2(n-1)}}\right) \tag{16}$$

as $n \to \infty$.

It is possible to effectively implement Theorem 5.2 to get values of the power of the test when more accuracy is desired than is provided by (5) and n is quite large. In order to best use (12) and (16) we suggest defining

$$\tilde{\overline{\xi}}_n = \frac{1}{n}\sum_{i=1}^{n} \xi(X_i)$$

$$\tilde{\overline{\rho}}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \rho^2(X_i), \tag{17}$$

since these are the obvious estimates of the corresponding quantities in (14). Then construct the test that rejects when

$$\frac{1}{\tilde{\overline{\rho}}_n}\sqrt{\frac{n-1}{2}}\left(\frac{T_{new}}{n-1} - \tilde{\overline{\xi}}_n\right) > \Phi^{-1}(1-\alpha). \tag{18}$$

Note for later use that under H_0

$$\xi(\overline{X}) \approx \tilde{\overline{\xi}}, \quad \rho^2(\overline{X}) \approx \tilde{\overline{\rho}}^2 \tag{19}$$

with asymptotic equality as $n \to \infty$. (19) should also be approximately valid when the alternative is not far from H_0. In such situations one could use the simpler values $\xi(\overline{X})$, $\rho^2(\overline{X})$ in place of $\tilde{\tilde{\xi}}$, $\tilde{\tilde{\rho}}^2$. Because of (19) the test in (18) is very similar to that described in (13).

Theorem 5.2 implies the power of the test given in (18) is

$$P_{\boldsymbol{\lambda}}(T_{new} \text{ satisfies (18)}) \to 1 \quad -E\left(\Phi\left(\frac{\tilde{\tilde{\rho}}_n}{\overline{\rho}_n}\Phi^{-1}(1-\alpha)\right)+\sqrt{\frac{n-1}{2}}\frac{\tilde{\tilde{\xi}}_n-\overline{\xi}_n}{\overline{\rho}_n}-\frac{4\sum(\nu(\lambda_i)-\overline{\nu}_n)^2}{\overline{\rho}_n\sqrt{2(n-1)}}\right), \tag{20}$$

where $\boldsymbol{\lambda} = \{\lambda_i\}$.

Now, $\tilde{\tilde{\rho}}_n \to \overline{\rho}_n$ in probability. Also, $\overline{\rho}_n \approx 1$ so long as $\min \lambda_i > 4$ as a consequence of the results plotted in Fig 6.

Let

$$\text{Var}(\sqrt{n-1}(\tilde{\tilde{\xi}}_n - \overline{\xi}_n)) = \epsilon(\boldsymbol{\lambda}).$$

Recall that $\xi(\lambda)$ is nearly constant for $\lambda > 4$. Hence ϵ is numerically quite small so long as $\min \lambda_i > 4$. It follows that then

$$P_{\boldsymbol{\lambda}}(T_{new} \text{ satisfies (18)}) = 1-E\left(\Phi(\Phi^{-1}(1-\alpha)-\frac{\psi^2}{\sqrt{2(n-1)}}+\epsilon(\boldsymbol{\lambda}))+O_p(1)\right)$$
$$\to 1-\Phi\left(\Phi^{-1}(1-\alpha)-\frac{\psi^2}{\sqrt{2(n-1)}}+\epsilon^*\right) \tag{21}$$

for some numerically small ϵ^*. (ϵ^* is numerically small because of its relation to the random variable $\epsilon(\boldsymbol{\lambda})$ which is also numerically small.)

If T_{new} were exactly noncentral χ^2 as assumed in (5) then we would have

$$P_{\boldsymbol{\lambda}}\left(\sqrt{\frac{n-1}{2}}\left(\frac{T_{new}}{n-1}-1\right) > \Phi^{-1}(1-\alpha)\right)$$
$$\to \quad 1-\Phi\left(\Phi^{-1}(1-\alpha)-\frac{\psi^2}{2\sqrt{n-1}}\right), \text{ under (5).} \tag{22}$$

Since ϵ^* is numerically small, these facts suggest that so long as all (or most) $\lambda_i > 4$, (5) is a very good approximation even though it is not asymptotically exact as $n \to \infty$ with $\overline{\lambda} = O(1)$.

Acknowledgments. The research of L.D. Brown was supported in part by NSF grant DMS-9971751. The research of L.H. Zhao was supported in part by NSF grant DMS-9971848.

References

AGRESTI, A. (1990). *Categorical Data Analysis.* Wiley, New York.

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

BARTLETT, M.S. (1947). The use of transformations, *Biometrics*, **3**, 39–52.

BROWN, L.D., CAI, T. and DASGUPTA, A. (2000). Interval estimation in exponential families, To appear in *Statist. Sinica.*

BROWN, L. D., GANS, N., MANDELBAUM, A., SAKOV, A., SHEN, H., ZELTYN, S. and ZHAO, L. (2001a). Empirical analysis of a telephone call center. Technical report. www-stat.wharton.upenn.edu/˜lbrown

BROWN, L.D., MANDELBAUM, A., SAKOV, A., SHEN, H., ZELTYN, S. and ZHAO, L. (2001b). Multifactor Poisson and Gamma-Poisson models for call center arrival times. Technical report. www-stat.wharton.upenn.edu/˜lbrown

BROWN, L.D., ZHANG, R. and ZHAO L. (2001). Root un-root methodology for nonparametric density estimation and Poisson random effects models. Technical report. www-stat.wharton.upenn.edu/˜lbrown

COCHRAN, W.G. (1954). Some methods of strengthening the common χ^2 tests, *Biometrics*, **10**, 417-451.

HUFFMAN, M.D. (1984). An improved approximate tow-sample Poisson test, *Appl. Statist.*, **33**, 224-226.

JONGBLOED, G. and KOOLE, G. (2001). Managing uncertainty in call centers using Poisson mixtures, *Appl. Stoch. Models Bus. Ind.*, **17**, 307-318.

LINDSAY, B.G. (1995). *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS regional conference series in probability and statistics, **5**.

RICE, J. (1995). *Mathematical Statistics and Data Analysis.* second edition. Duxbury press, CA.

LAWRENCE D. BROWN AND LINDA H. ZHAO
DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PA 19104, USA
E-mail: lbrown@wharton.upenn.edu
lzhao@wharton.upenn.edu