Accuracy of Power-Divergence Statistics for Testing Independence and Homogeneity in Two-Way Contingency Tables

Miguel A. García-Pérez, Vicente A. Núñez-Antón

Accuracy of Power-Divergence Statistics for Testing Independence and Homogeneity in Two-Way Contingency Tables

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2008-0137
Manuscript Type:	Original Paper
Date Submitted by the Author:	23-Jun-2008
Complete List of Authors:	García-Pérez, Miguel; Universidad Complutense, Facultad de Psicología Núñez-Antón, Vicente; Universidad del País Vasco, Fac. de CC. Económicas y Empresariales, Dpto. de Econometría y Estadística
Keywords:	poer-divergence statistics, contingency tables, independence, homogeneity
Abstract:	The small-sample accuracy of seven members of the family of power-divergence statistics for testing independence or homogeneity in contingency tables was studied via simulation at test sizes of .01 and .05 with marginal distributions that could be uniform or skewed and with sample sizes including sparseness conditions. The likelihood ratio statistic rejected the null hypothesis too often even with large table density, and none of the other five statistics outperformed Pearson's X2. A non-asymptotic variant of the latter was even more accurate with table densities of 1 observation/cell. These results advise against the use of the likelihood ratio statistic.
Note: The following files were submitted by the author for peer review, but cannot be converted	

contingency.zip

Accuracy of Power-Divergence Statistics for Testing Independence and Homogeneity in Two-Way Contingency Tables

MIGUEL A. GARCÍA-PÉREZ[†] AND VICENTE NÚÑEZ-ANTÓN[‡]

23rd June 2008

Abstract

The small-sample accuracy of seven members of the family of power-divergence statistics for testing independence or homogeneity in contingency tables was studied via simulation. The likelihood ratio statistic G^2 and Pearson's X^2 statistic are among these seven members, whose behavior was studied at nominal test sizes of .01 and .05 with marginal distributions that could be uniform or skewed and with a set of sample sizes that included sparseness conditions as measured through table density (i.e., the ratio of sample size to number of cells). The likelihood ratio statistic G^2 rejected the null hypothesis too often even with large table density, whereas Pearson's X^2 was sufficiently accurate and only presented a minor misbehavior when table density was less than 2 observations/cell. None of the other five statistics outperformed Pearson's X^2. A non-asymptotic variant of X^2 solved the minor inaccuracies of Pearson's X^2 and turned out to be the most accurate statistic for testing independence or homogeneity, even with table densities of 1 observation/cell. These results clearly advise against the use of the likelihood ratio statistic G^2.

Keywords: Power-Divergence Statistics; Independence; Homogeneity; Contingency Tables; Small-Sample Accuracy.
Mathematical Subject Classification 62F03.

[*]Departamento de Metodología, Universidad Complutense, Madrid, Spain
[†]Address for correspondence: Miguel A. García-Pérez, Departamento de Metodología, Facultad de Psicología, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain. Phone: +34 913 943 061; Fax: +34 913 943 189; E-mail: miguel@psi.ucm.es
[‡]Departamento de Econometría y Estadística, Universidad del País Vasco (E.A. III), Bilbao, Spain

1 Introduction

Consider a sample of N paired observations that are cross-classified in an $I \times J$ contingency table, and let (i) O_{ij} be the observed frequency in cell ij, (ii) $O_{i+} = \sum_{j=1}^{J} O_{ij}$ be the i-th row marginal, (iii) $O_{+j} = \sum_{i=1}^{I} O_{ij}$ be the j-th column marginal, and (iv) $E_{ij} = O_{i+}O_{+j}/N$ be the expected frequency in cell ij. Two-way contingency tables are the order of the day in many areas of the social and behavioral sciences for testing association between categorical variables (tests of independence) or for the comparison of distributions across populations (tests of homogeneity), although these tables are also used for fitting loglinear models to categorical data (Wickens, 1998). Pearson's X^2 statistic and the likelihood ratio statistic G^2 are most frequently used for testing independence or homogeneity in many areas of research (Cressie and Read, 1989), but they are only two members of a continuous family of power-divergence statistics (Cressie and Read, 1984).

The power-divergence statistic of index $\lambda \in \Re$ is given by

$$RC^{(\lambda)} = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{I} \sum_{j=1}^{J} O_{ij} \left\{ \left(\frac{O_{ij}}{E_{ij}} \right)^{\lambda} - 1 \right\} \tag{1}$$

with $RC^{(-1)}$ and $RC^{(0)}$ defined by continuity. This family includes Pearson's X^2 ($RC^{(1)}$) and the likelihood ratio statistic G^2 ($RC^{(0)}$) as well as other statistics such as Freeman-Tukey's T^2 ($RC^{(-1/2)}$), the modified likelihood ratio statistic GM^2 ($RC^{(-1)}$), and Neyman's modified X^2 statistic ($RC^{(-2)}$). For all λ, $RC^{(\lambda)}$ follows a discrete distribution that converges asymptotically to a continuous chi-square distribution on $(I-1)(J-1)$ degrees of freedom, although the rate of this convergence for different λ is unknown. Then, for any given sample size, asymptotic significance levels may mis-state actual levels.

A number of papers have investigated the conditions under which tests based on Pearson's X^2 statistic are accurate (Berry and Mielke, 1988; Bradley, Bradley, McGrath, and Cutcomb, 1979; Koehler, 1986; Lewis, Saunders, and Westcott, 1984; Martín Andrés and Herranz Tejedor, 2000). Wickens (1989, p. 30) summarized these conditions as follows:

1. For tests with 1 degree of freedom, all E_{ij} should exceed 2 or 3
2. With more degrees of freedom, $E_{ij} \simeq 1$ in a few cells is tolerable.
3. In large tables up to 20% of the cells can have E_{ij} appreciably less than 1.
4. The total sample should be at least 4 or 5 times the number of cells.
5. Samples should be appreciably larger when the marginal categories are not equally likely.

Yet, it is unclear whether these conditions generalize to all members of the power-divergence family, particularly to the widely used likelihood ratio statistic G^2. To the authors' knowledge, the only study to that effect was carried out by Rudas (1986), who determined empirical 90% and 95% points for testing independence with X^2, G^2, and $RC^{(2/3)}$ in 2×2, 2×3, 2×4, 2×6, 3×3, 3×6, and 5×6 tables with small sample sizes (15, 25, 35, 45, and 55 for tables with 5 or fewer degrees of freedom, 25, 35, 45, 55, 65, 75, 85, and 95 for 3×6 tables, and 50, 75, 100, and 125 for 5×6 tables). Rudas found that X^2 and $RC^{(2/3)}$ render similar results that match reasonably well the percentage points arising from the asymptotic χ^2 distribution, whereas G^2 appeared unsuitable because it rejected the null hypothesis too often.

Despite these results, and also despite the broader family of power-divergence statistics, the likelihood ratio statistic G^2 continues to be regarded as virtually the only alternative to Pearson's X^2 statistic for testing independence or homogeneity in contingency tables, and its use continues to be recommended with only non-specific warnings regarding its potential inaccuracies (Cressie and Read, 1989; Wickens, 1989, 1998). Everitt (1992, p. 72) disregarded all this evidence and expressed a clear preference for the likelihood ratio statistic G^2 over Pearson's X^2. Here we describe the results of a thorough analysis that addresses directly the small-sample accuracy (i.e., the actual Type-I error rate) of seven members of the family of power-divergence statistics when testing independence and homogeneity for a broad range of table dimensions and sample sizes. Our analysis includes either uniform marginal distributions or skewed marginal distributions in a way that allows addressing the relevance of the five conditions summarized by Wickens (1989) and listed above. Skewed marginal distributions are particularly revealing because they result in small expected frequencies in one or more cells for all tables, something that threatens the accuracy of the test (see Section 7.7.3 in Agresti, 1990). Our results indicate that a simple measure of sparseness (namely, table density defined as N/IJ) influences the accuracy of the tests and that Pearson's X^2 statistic is the most robust to sparseness conditions whereas the likelihood ratio statistic G^2 performs remarkably unsatisfactorily. A comparison with the non-asymptotic version of X^2 proposed by Berry and Mielke (1988) indicates that the latter clearly improves the performance of Pearson's X^2 statistic.

2 Method

Tables were generated under the independence and homogeneity sampling models with table dimensions between 2×2 and 8×12. Marginal distributions were either uniform or skewed (but of the same type for rows and columns in any given table).

Under the independence sampling model, uniform marginal distributions imply that the probability of an observation's falling in row i is $1/I$ and the probability of an observation's falling in column j is $1/J$, whereas skewed marginal distri-

butions imply that the probability of an observation's falling in row i is $2i/I(I+1)$ and the probability of an observation's falling in column j is $2j/J(J+1)$. Then, under the independence model the probability of an observation's falling in cell ij is just the product of the marginal probabilities of row i and column j. We used sample sizes N between 100 and 800 in steps of 100.

Under the homogeneity sampling model, column margins were fixed according to a uniform or skewed distribution as appropriate. These fixed numbers of observations were then randomly distributed across rows according to the probability functions for uniform or skewed row margins, as described in the preceding paragraph. The fixed number of observations at each column depended on the selected form of the marginal distribution for columns. With uniform marginal distributions, $O_{+j} = \text{nint}(N^*/J)$, where $\text{nint}(\cdot)$ represents the nearest integer; with skewed distributions, $O_{+j} = \text{nint}(2N^*j/J(J+1))$. We used here the same set of initial sample sizes N^* as described in the preceding paragraph for the independence sampling model, but the final set of sample sizes $N = \sum_{j=1}^{J} O_{+j}$ turned out to be slightly different on occasion as a result of the rounding operations involved here. Another consequence of the differences between independence and homogeneity sampling models is that tables in which the number of rows and columns are swapped (e.g., 2×5 and 5×2) yield different situations in the latter case but not in the former.

As an index of sparseness, we define the density of a table as N/IJ (in units of observations per cell, or obs/cell). Density thus indicates the number of observations that would fall in each cell in the case of a perfect uniform distribution of N (breakable) observations across IJ cells.

In each condition (table size, sample size, sampling model, and form of marginal distributions), 20,000 tables were generated none of whose rows or columns were empty of observations. Each of the N observations was randomly assigned to a cell in the table according to the corresponding probability distribution under the null hypothesis, and pseudo-random numbers required for this purpose were obtained with NAG subroutines G05DAF and G05DYF (Numerical Algorithms Group, 1999). For each table, power-divergence statistics were computed for $\lambda \in \{-1/2, 0, 1/3, 1/2, 2/3, 1, 3/2\}$. The empirical Type-I error rate for each power-divergence statistic was computed as the proportion of tables whose statistic value exceeded the critical value for a size-a test for $\alpha \in \{.05, .01\}$, that is, the value c_α satisfying $P(\chi^2_{(I-1)(J-1)} \geq c_\alpha) = \alpha$, where $\chi^2_{(I-1)(J-1)}$ is a chi-square random variable with $(I-1)(J-1)$ degrees of freedom. Figure 1 shows a graphical illustration of our approach, which also serves to present the motivation for our study by showing that the small-sample distribution of power-divergence statistics does not always resemble the asymptotic distribution.

3 Results

3.1 Tests Using the Asymptotic Distribution

Figure 2 shows the empirical Type-I error rate of the test of independence as a function of table density for each of the seven power-divergence statistics (rows), with either uniform marginal distributions (left column) or skewed marginal distributions (right column).

Ideally, all data points should fall on the horizontal lines at the nominal significance level, but there are apparent differences across power-divergence statistics in this respect. In one extreme, Pearson's X^2 statistic (the power-divergence statistic of index $\lambda = 1$; see the sixth row in Figure 2 yields the nominal significance level throughout the range of table densities when marginal distributions are uniform, and requires table densities above 2 obs/cell with skewed marginal distributions, although the size-.01 test is somewhat less accurate than the size-.05 test (compare the spread of data points around the applicable horizontal line in the upper and lower bundles). In the other extreme, Freeman-Tukey's T^2 (the power-divergence statistic of index $\lambda = -1/2$; see the first row in Figure 2) requires table densities above about 40 obs/cell (with uniform margins) or 60 obs/cell (with skewed margins) to behave properly. The widespread likelihood ratio statistic G^2 (the power-divergence statistic of index $\lambda = 0$; see the second row in Figure 2) also requires densities above about 20 obs/cell (with uniform margins) or 40 obs/cell (with skewed margins) to be accurate. Results for the test of homogeneity are omitted because they were indistinguishable by eye from those reported in Figure 2 for the test of independence.

3.2 Variants Using Concordant Estimates of Expected Frequencies

All of the results reported in Figure 2 were obtained with expected frequencies computed as usually, that is, $E_{ij} = O_{i+}O_{+j}/N$. It should be noted that these are maximum-likelihood estimates of the expected frequencies and, thus, they are consonant with the metric used by the likelihood ratio statistic. Read and Cressie (1988, p. 32) claimed that it would be reasonable to estimate expectations using the metric that is consonant with the power-divergence statistic that is to be used for testing the null hypothesis. Thus, one might surmise that the results would differ if this approach were taken and, for instance, Pearson's X^2 statistic were computed using minimum-X^2 estimates of expected frequencies. Clearly, the results would nevertheless stay the same for the likelihood ratio statistic G^2. Computing expected frequencies using the matching minimum-divergence estimate is difficult in the case of tests of independence for lack of a closed-form expression for the estimator, but Read and Cressie (1988, p. 32) showed that the minimum power-divergence estimate of index λ for expected frequencies in tests

of homogeneity is given by

$$E_{ij} = \frac{O_{+j}\left[\sum_{l=1}^{J} O_{il}^{\lambda+1}/O_{+l}^{\lambda}\right]^{1/(\lambda+1)}}{\sum_{k=1}^{I}\left[\sum_{l=1}^{J} O_{kl}^{\lambda+1}/O_{+l}^{\lambda}\right]^{1/(\lambda+1)}} \quad (2)$$

We thus repeated our analysis under the homogeneity sampling model by computing expected frequencies according to equation (2), but the results did not improve in any significant respect. The only differences occurred for table densities below 10 obs/cell, and the effect took the form of a general reduction of the empirical Type-I error rate in this range of small densities. As a consequence, the accuracy of Pearson's X^2 statistic deteriorated (compare the top row in Figure 3 with the sixth row in Figure 2). The most beneficial effect occurred for the power-divergence statistic of index $\lambda = 3/2$ and, although the results in these conditions (see the bottom row in Figure 3) represent a noticeable improvement over the corresponding results shown in the bottom row in Figure 2 for the same statistic computed from maximum-likelihood estimates of expected frequencies, these results do not represent an improvement over the conventional use of Pearson's X^2 statistic computed from maximum-likelihood estimates of expected frequencies (which were shown in the sixth row of Figure 2).

3.3 Tests Using the Non-Asymptotic Distribution

Read and Cressie (1988, ch. 5) showed that moment-correction terms for general λ can be derived that presumably increase the small-sample accuracy of power-divergence goodness-of-fit statistics in one-way multinomials, and García-Pérez and Núñez-Antón (2001) presented results indicating that the accuracy of the statistics improves when these correction terms are used. In the case of contingency tables, similar moment-correction terms can also be derived for Pearson's X^2 statistic, but the derivation for general λ does not appear to be feasible. The moment-corrected version of Pearson's X^2 statistic is the basis of the non-asymptotic test proposed by Berry and Mielke (1988), who claimed that this test is more accurate in sparse tables than the asymptotic test based on the original X^2 statistic.

Berry and Mielke (1988) also reported that the non-asymptotic test has some limitations (besides its formidable computational cost, which can be afforded with the help of the software provided by Berry and Mielke, 1986). Specifically, they claimed (i) that the test performs better for the independence than for the homogeneity model, (ii) that it does not perform well with 2×2 tables with $N = 20$, and (iii) that it performs poorly for tables with less than 3 degrees of freedom. These conclusions are a consequence of what we believe is an incorrect strategy for data analysis, as discussed next. Berry and Mielke judged the accuracy of the test by comparing the actual distribution of p-values with the theoretical

distribution. Consider the case of 2×2 tables with $N = 20$, equal marginals, and the independence model (i.e., the conditions for which data were reported in the top-left column of Table 1 in the paper of Berry and Mielke). The reported distribution of p-values yielded observed proportions of .8857, .0635, .0372, .0092, .0032, and .0012 respectively within the bins $(.1, 1]$, $(.05, .1]$, $(.01, .05]$, $(.005, .01]$, $(.001, .005]$, and $[0, .001]$, which implies expected proportions of .900, .050, .040, .005, .004, and .001, respectively. Berry and Mielke concluded that the non-asymptotic test does not perform well because of a significant goodness-of-fit test carried out on these data. But this omnibus goodness-of-fit test is assessing the similarity between the theoretical and empirical distributions of p-values throughout the entire range, when the truth is that only the agreement at the right tail matters when it comes to considering the accuracy of statistical tests (Read, 1984). From the data reported by Berry and Mielke and reproduced above, it follows that the proportion of p-values at or below .05 (i.e., the empirical Type-I error rate) is .0372 + .0092 + .0032 + .0012 = .0508 and that the proportion of p-values at or below .01 is .0092 + .0032 + .0012 = .0136. Looking at the results in this way, we see little justification for the conclusion that the non-asymptotic test performs poorly. More interestingly, the asymptotic test does not seem to perform poorly either when the results of Berry and Mielke are analyzed in this way: In the same conditions, the empirical Type-I error rates at the .05 and .01 levels are .0481 and .0109, respectively.

Rather than re-analyzing the results presented by Berry and Mielke (1988) as just discussed (and primarily because of the small range of conditions that they included in their study), we decided to run our own simulations under the conditions spanned by our previous analyses. The moment-corrected X^2 statistic has a Pearson Type III distribution with a parameter that depends on the row and column marginals and, then, the reference distribution varies across tables of the same size. For this reason, we varied slightly our strategy and computed the moment-corrected X^2 statistic and its p-value for each particular table using the subroutine of Berry and Mielke (1986) and then computed the proportion of tables (out of the 20,000 that we simulated per condition) whose p-value was at or below a = .05 or .01 (as applicable). Note that this approach is thoroughly analogous to our previous approach of comparing the conventional power-divergence test statistic for a given table with the critical value for a size-a test from the asymptotic distribution, which can only be used when the reference distribution does not change from table to table.

Figure 4 shows the results. Moment correction did improve the accuracy of Pearson's X^2 statistic particularly at low table densities, which thus makes this statistic exquisitely accurate regardless of table density, test size, and form of marginal distributions (compare the top row in Figure 4 with the sixth row in Figure 2). Interestingly, when its outcomes are analyzed as we did, this statistic does not seem to have any of the limitations reported by Berry and Mielke (1988), namely, different performance in the homogeneity and independence sampling

models (compare the top and bottom rows in Figure 4) and poor performance in tables with less than 3 degrees of freedom or in 2×2 tables with small sample sizes.

4 Conclusion

Although all the members of the family of power-divergence statistics converge asymptotically to a chi-square distribution, they do so at different rates and, hence, their small-sample accuracy is not guaranteed. Our results indicate that Pearson's X^2 statistic (the power-divergence statistic of index $\lambda = 1$) yields the most accurate tests of independence and homogeneity even with sparse tables, only failing to reach the nominal rejection rates when table density is below 2 obs/cell. The superior accuracy of Pearson's X^2 statistic compared to the remaining members of the family of power-divergence statistics was also found in studies on goodness-of-fit statistics in one-way multinomials (García-Pérez and Núñez-Antón, 2001; Read, 1984).

The minor inaccuracies of Pearson's X^2 statistic at low table densities can be remedied with recourse to moment-correction terms that render a moment-corrected statistic with a Pearson Type III distribution. In these conditions, the corrected X^2 statistic is extremely accurate with table densities as low as 1 obs/cell. García-Pérez and Núñez-Antón (2001) reported the same result when similar moment-correction terms were used to modify Pearson's X^2 statistic for use in goodness-of-fit tests in one-way multinomials.

In sum, our results advise against the use of the likelihood ratio statistic G^2 for testing independence or homogeneity in contingency table analysis. Among the other members of the family of power-divergence statistics, only the use of Pearson's X^2 statistic is advisable either in its original form (when table density is at or above 2 obs/cell) or with moment correction (when table density is lower). In the latter case, the software provided by Berry and Mielke (1986) greatly facilitates the computation of the moment-correction terms and the determination of a p-value from the reference Pearson Type III distribution. An interesting consequence of our results is that the set of conditions summarized by Wickens and stated in the Introduction for the accuracy of Pearson's X^2 statistic in contingency table analysis can be replaced by a simple rule that looks into table density to determine whether or not moment-correction terms are required to guarantee accurate tests.

Acknowledgements

This research was supported by grants SEJ2005-00485 (Ministerio de Educación y Ciencia), MTM2004-00341 (Ministerio de Educación y Ciencia and FEDER), MTM2007-60112 (Ministerio de Educación y Ciencia and FEDER), and IT-334-07 (Departamento de Educación del Gobierno Vasco - UPV/EHU Econometrics Research Group).

References

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

Berry, K.J., Mielke, P. W., Jr. (1986). R by C chi-square analyses with small expected cell frequencies. *Educational and Psychological Measurement* 46:169-173.

Berry, K.J., Mielke, P.W., Jr. (1988). Monte Carlo comparisons of the asymptotic chi-square and likelihood-ratio tests with the nonasymptotic chi-square test for sparse $r \times c$ tables. *Psychological Bulletin* 103:256-264.

Bradley, D.R., Bradley, T.D., McGrath, S.G., Cutcomb, S.D. (1979). Type I error rate of the chi-square test of independence in $R \times C$ tables that have small expected frequencies. *Psychological Bulletin* 86:1290-1297.

Cressie, N., Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society - Series B* 46:440-464.

Cressie, N., Read, T.R.C. (1989). Pearson's X^2 and the loglikelihood ratio statistic G^2: A comparative review. *International Statistical Review* 57:19-43.

Everitt, B.S. (1992). *The Analysis of Contingency Tables*. 2nd ed. London: Chapman & Hall.

García-Pérez, M.A., Núñez-Antón, V. (2001). Small-sample comparisons for power-divergence goodness-of-fit statistics for symmetric and skewed simple null hypotheses. *Journal of Applied Statistics* 28:855-874.

Koehler, K.J. (1986). Goodness-of-fit tests for log-linear models in sparse contingency tables. *Journal of the American Statistical Association* 81:483-493.

Lewis, T., Saunders, I.W., Westcott, M. (1984). The moments of the Pearson chi-squared statistic and the minimum expected value in two-way tables. *Biometrika* 71:515-522. [Correction published in *Biometrika*, 1989, 76:407.]

Martín Andrés, A., Herranz Tejedor, I. (2000). On the minimum expected quantity for the validity of the chi-squared test in 2×2 tables. *Journal of Applied Statistics* 27:807-820.

Numerical Algorithms Group (1999). *NAG Fortran library manual*. Mark 19. Oxford: Author.

Read, T.R.C. (1984). Small-sample comparisons for the power-divergence goodness-of-fit statistics. *Journal of the American Statistical Association* 79:929-935.

Read, T.R.C., Cressie, N.A.C. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data*. New York: Springer.

Rudas, T. (1986). A Monte Carlo comparison of the small sample behaviour of the Pearson, the likelihood ratio and the Cressie-Read statistics. *Journal of Statistical Computation and Simulation* 24:107-120.

Wickens, T.D. (1989). *Multiway Contingency Tables Analysis for the Social Sciences*. Hillsdale, N.J.: Erlbaum.

Wickens, T.D. (1998). Categorical data analysis. *Annual Review of Psychology* 48:537-557.

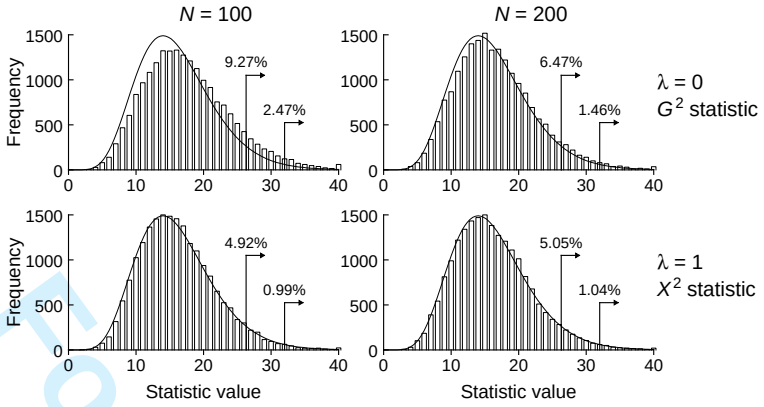

Figure 1: Illustration of the approach taken in this paper. Each panel shows a histogram of power-divergence statistics (the likelihood ratio statistic G^2 in the top row and Pearson's X^2 statistic in the bottom row) for testing independence. Data come from 20,000 5×5 tables with sample sizes of $N = 100$ (left column) and $N = 200$ (right column). Table density was thus 4 obs/cell in the left column and 8 obs/cell in the right column. Marginal distributions were uniform. The continuous curve depicts the asymptotic χ^2 distribution on 16 degrees of freedom. The vertical lines in each panel indicate the critical limits for a size-α test based on the asymptotic distribution, with the leftmost line corresponding to $\alpha = .05$ (i.e., $c_\alpha = 26.296$) and the rightmost line corresponding to $\alpha = .01$ (i.e., $c_\alpha = 32.000$). Percentages printed above these lines indicate the empirical Type-I error rate that those critical limits produce, whose agreement with the nominal Type-I error rates (5% and 1%, respectively) varies with sample size for the likelihood ratio statistic G^2 (top row) but not so much for Pearson's X^2 statistic (bottom row).

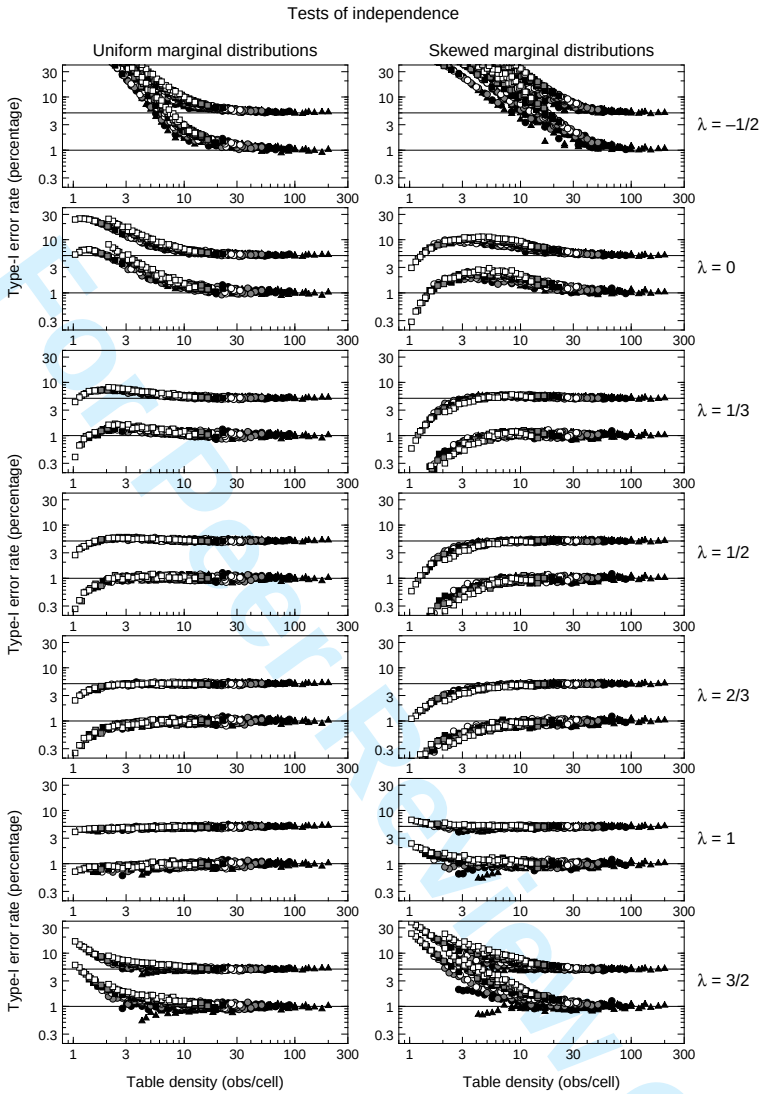

Figure 2: Empirical Type-I error rates of the omnibus size-.05 (upper bundle in each panel) and size-.01 (lower bundle in each panel) test of independence as a function of table density for each member of the power-divergence family of statistics (rows) and for either uniform marginal distributions (left column) or skewed marginal distributions (right column). Results are reported for tables of dimensions between 2×2 and 8×12 with sample sizes between 100 and 800 in steps of 100. Different symbols indicate results for tables with different numbers of rows (solid triangles: 2 rows; solid circles: 3 rows; gray circles: 4 rows; open circles: 5 rows; solid squares: 6 rows; gray squares: 7 rows; open squares: 8 rows). The horizontal lines across each panel, at ordinates of 5% and 1%, indicate the nominal Type-I error rates.

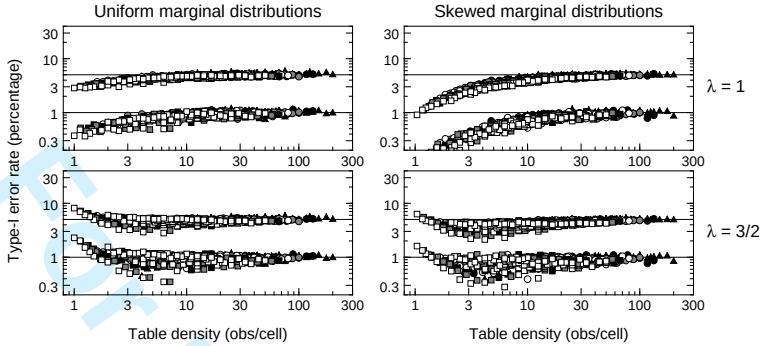

Figure 3: Empirical Type-I error rate of the power-divergence statistics of indices $\lambda = 1$ (top row) and $\lambda = 3/2$ (bottom row) in tests of homogeneity when expected frequencies are computed by minimizing the power-divergence measure of the same index. Graphical conventions as in Figure 2.

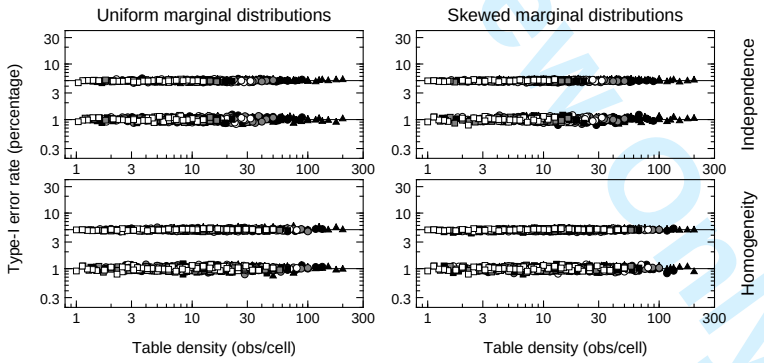

Figure 4: Empirical Type-I error rate of the moment-corrected, non-asymptotic X^2 statistic of Berry and Mielke (1988) in tests of independence (top row) and homogeneity (bottom row). Graphical conventions as in Figure 2.