IMS Collections
Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman
Vol. 8 (2012) 181–189

DOI: 10.1214/11-IMSCOLL812

Accurate approximations to the distribution of a statistic testing symmetry in contingency tables

John E. Kolassa* **and Hema Gayat Bhagavatula**

Rutgers University

Abstract: This manuscript examines this task of approximating significance levels for a test of symmetry in square contingency tables. The null sampling distribution of this test statistic is the same as that of the sum of squared independent centered binomial random variables, weighted by their separate sample size; each of these variables may be taken to have success probability half. This manuscript applies an existing asymptotic correction to the standard chi-squared approximation to the distribution of the quadratic form of a random vector confined to a multivariate lattice, when the quadratic form is formed from the inverse variance matrix of the random vector. This manuscript also investigates non-asymptotic corrections to approximations to this distribution, when the separate binomial sample sizes are small.

Contents

1. Introduction

Consider a square table of random counts X_{jk}, for $j, k \in \{1, \ldots, d\}$, such that

$$(X_{11}, X_{12}, \ldots, X_{1d}, X_{21}, X_{22}, \ldots, X_{2d}, \ldots, X_{d1}, X_{d2}, \ldots, X_{dd})$$

has a multinomial distribution, and let π_{jk} denote the cell probability corresponding to X_{jk}. Bowker (1948) introduces statistic for testing the null hypothesis that

Department of Statistics and Biostatistics, Hill Center, Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854 USA e-mail: kolassa@stat.rutgers.edu; bhagavat@eden.rutgers.edu

*This material is based in part upon work supported by the National Science Foundation under Grant Number DMS 0906569. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors thank the editor and an anonymous reviewer for helpful comments.

AMS 2000 subject classifications: Primary 62E17, 60K35; secondary 62H17

Keywords and phrases: conditional inference, Bowker's test of symmetry, Yarnold approximation

$\pi_{jk} = \pi_{kj}$ for all j and k, vs. the unrestricted alternative, of the form $W = \sum_{j=1}^{d-1}\sum_{k=2}^{d}(X_{jk} - X_{k,j})^2/(X_{jk} + X_{kj})$. As long as $\pi_{jk} > 0$ for all $j \neq k$, then, as $X_{..} = \sum_{j=1}^{d}\sum_{k=1}^{d} X_{jk}$ increases,

$$P[W \leq w] \doteq G_m(w) \tag{1}$$

for $m = d(d-1)/2$, and G_m the χ^2_m distribution function, where the subscript refers to the degrees of freedom. Here $\doteq$ represents an order of accuracy generally adequate for most practical problems, subject to sufficiently large $X_{jk} + X_{kj}$.

When some of the off-diagonal probabilities π_{jk} are small, and $X_{..}$ only moderate, the above χ^2 approximation may not be sufficiently adequate. This manuscript investigates two improvements to this approximation. The first of these adjusts for discreteness of the distribution of W, and the second exploits simplifications arising when the denominator of some summands of W are small.

Let $\mathbf{T} = (X_{12}, X_{13}, \cdots, X_{d-2d-1}, X_{d-2d}, X_{d-1d})$ and let $\mathbf{N}$ denote

$$(X_{12} + X_{21}, X_{13} + X_{31}, \cdots, X_{d-2d-1} + X_{d-1d-2}, X_{d-2d} + X_{dd-2}, X_{d-1d} + X_{dd-1}).$$

Let $m = (d-1)d/2$. Then for each $l \in \{1, \ldots, m\}$, $T_l \sim \text{Bin}(1/2, N_l)$, and

$$W = \sum_{l=1}^{m}(T_l - N_l/2)^2/(N_l/4). \tag{2}$$

Hence Bowker's test of symmetry can be recast as a multivariate binomial testing task; this more general problem will be addressed in the remainder of the paper. Krampe and Kuhnt (2007) perform this test conditionally on the value of $\mathbf{N}$; we test conditionally as well. Specifically, if we define $H(w;\mathbf{N}) = P[\sum_{l=1}^{m}(T_l - N_l/2)^2/(N_l/4) \leq w]$, then the balance of this paper concerns evaluating $H(w;\mathbf{N})$. This test might be performed exactly (Krampe and Kuhnt, 2007), but some very popular statistical packages lack this functionality (Oster, 2003). Ludbrook (2008), Oster and Hilbe (2008a), and Oster and Hilbe (2008b) review abilities of various commercial software packages to perform exact inference in a similar, but not identical, setting.

2. An approximation to the distribution function of a score test for binomial testing

Let $\mathcal{E} = \{\mathbf{t} | \mathbf{t}^\top \boldsymbol{\Sigma}^{-1} \mathbf{t} \leq w\}$ be the elliptical set of $\mathbf{t}$ giving rise to $W \leq w$. Kolassa (2006) reviews the multivariate Edgeworth series, which allows the approximation of probabilities of sets like $\mathcal{E}$, as long as $\mathbf{T}$ satisfies regularity conditions.

When $\mathbf{T}$ has a lattice distribution, approximating probabilities of elliptical regions becomes tricky. Bhattacharya and Rao (1976) review approximations to probabilities for events described in terms of random vectors confined to a multivariate lattice, and Yarnold (1972) addresses the problem of evaluating this approximation for convex sets, and in particular for standardized ellipses. The Yarnold approximation is the χ^2 approximation plus the difference between the actual number of points in the ellipse and the volume of the ellipse divided by the volume of a unit cube of the lattice, times the normal approximation to the density at each point on the ellipse boundary. Specifically, suppose $\mathbf{T}$ arises as the mean of n independent and identically distributed random vectors having a finite sixth moment and confined to a unit lattice, standardized so that the variance matrix of $\mathbf{T}$ is fixed as n

increases. Then

$$H(w;\mathbf{N}) \approx G_m(w) + (N(w) - V(w))\,\frac{\exp(-w/2)}{(2\pi)^{m/2}\det\boldsymbol{\Sigma}^{1/2}}, \tag{3}$$

where $N(w)$ is the number of vectors of integers $\mathbf{m}$ such that $\mathbf{m}+\mathbf{y}\in\mathcal{E}$,

$$V(w) = \frac{(\pi w)^{m/2}\det\boldsymbol{\Sigma}^{1/2}}{\Gamma(m/2+1)}$$

is the volume of $\mathcal{E}$, and $\approx$ represents a uniform absolute error of $O(n^{-1})$ as n increases. Kolassa (2003) reviews this approximation, as applied to score tests for discrete regression models.

In the present context, when applying Bowker's test, (3) may be simplified. The matrix $\boldsymbol{\Sigma}$ is a diagonal matrix with $\mathbf{N}/4$ on its diagonal. Hence $\det\boldsymbol{\Sigma}^{1/2} = \prod_{l=1}^{m}(N_l/4)$, and

$$\mathcal{E} = \{\mathbf{t}|\sum_{l=1}^{m}(2T_l - N_l)^2/N_l \le w\}.$$

Let M be the least common multiple of $\{N_l|l=1,\ldots,m\}$, and let $R_l = M/N_l$. Then $\mathcal{E} = \{\mathbf{t}|\sum_{l=1}^{m} R_l(2T_l - N_l)^2 \le Mw\}$. Hence the discontinuities in W can be located exactly via integer arithmetic.

3. Adjustments for small denominators

I first introduce some notation. For any set $\mathcal{S}\subset\{1,\ldots,m\}$, let $\mathbf{T}_{\mathcal{S}}$ and $\mathbf{N}_{\mathcal{S}}$ represent $\mathbf{T}$ and $\mathbf{N}$, with only those components retained with indices in $\mathcal{S}$. Furthermore, let $\#(\mathcal{S})$ represent the cardinality of $\mathcal{S}$.

Note that when $N_l = 0$, hypotheses concerning the associated binomial probabilities are not testable, and the contribution from T_l to W is undefined. Documentation for standard statistical software, (including SAS Institute Inc. (2010)) indicates that the programmers set such terms to zero in (2), but counts them when calculating m in (1). I recommend omitting these observations from $\mathbf{T}$, and from the calculated degrees of freedom. That is, I define

$$H(w;\mathbf{N}) = H(w;\mathbf{N}_{\mathcal{A}^c}), \tag{4}$$

for $\mathcal{A} = \{l|N_l = 0\}$, to remove the indeterminant terms arising from zero numerators and denominators, and if (1) is used to approximate the distribution of W, then entries with zero denominators are removed before calculating degrees of freedom. Furthermore, for those l such that $N_l = 1$, the contribution to W is always 1. Let $\mathcal{B} = \{l|N_l = 1\}$. Note that

$$H(w;\mathbf{N}) = H(w - \#(\mathcal{B});\mathbf{N}_{(\mathcal{A}\cup\mathcal{B})^c}) \tag{5}$$

Furthermore, for indices l such that $N_l = 2$, then the contribution to W is 0 with probability 1/2, and 2 with probability 1/2. and for indices l such that $N_l = 3$, then the contribution to W is 1/3 with probability 3/4, and 3 with probability 1/4. Let $\mathcal{C} = \{l|N_l = 2\}$, and $\mathcal{D} = \{l|N_l = 3\}$. Then

$$\begin{aligned} H(w;\mathbf{N}) \quad = \quad & \sum_{j=0}^{\#(\mathcal{C})}\sum_{k=0}^{\#(\mathcal{D})}\left(\frac{1}{2}\right)^{\#(\mathcal{C})}\left(\frac{3}{4}\right)^{k}\left(\frac{1}{4}\right)^{\#(\mathcal{D})-k}\times \\ & H(w - \#(\mathcal{B}) - 2j - 3k - \frac{1}{3}(\#(\mathcal{D}) - k);\mathbf{N}_{(\mathcal{A}\cup\mathcal{B}\cup\mathcal{C}\cup\mathcal{D})^c}). \end{aligned} \tag{6}$$

It may happen that $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D} = \{1, \ldots, m\}$; in this case, $\mathbf{N}_{(\mathcal{A}\cup\mathcal{B}\cup\mathcal{C}\cup\mathcal{D})^c}$ is a vector of zero length, and when H has a zero-component second argument we define

$$H(w;) = \begin{cases} 0 & \text{if } w < 0 \\ 1 & \text{if } w \geq 0 \end{cases} \tag{7}$$

4. Results

In this section we compare the result of our approximation in four cases. The first two of these cases involve binomial sample sizes small enough to merit the use of Yarnold's approximation of Section 2, but large enough not to require the refinements of Section 3. The last two of these cases involve both the Yarnold refinement and the exact convolution approach of Section 3.

4.1. Yarnold's approximation

Yarnold approximations for two sets of binomial sample sizes are presented in Figures 1–4.

Figure 1 presents the Chi-square and Yarnold approximations with binomial sample sizes 4,4,5, and Figure 2 shows the error of these two approximations. The Yarnold approximation is far more accurate than is the standard Chi-square approximation.

Figure 3 presents the Chi-square and Yarnold approximations with binomial sample sizes 8,9,10, and Figure 4 shows the error of these two approximations. Again, the Yarnold approximation is far more accurate than is the standard Chi-square approximation.

4.2. The addition of exact convolution

Yarnold approximations for two sets of binomial sample sizes are presented in Figures 5–8; additionally, these approximations also involve the correction for small sample sizes. The first set of binomial observations exhibit the effect of dropping the binomials with sample size one, and the second shows the effect of exactly convolving the approximation for observations with larger samples with those for sample sizes two and three.

Figure 5 presents the Chi-square and Yarnold approximations with binomial sample sizes 1,4,4,5, both with and without the adjustment for the sample size 1 category, and Figure 6 shows the error of these approximations. The Yarnold and small sample size adjustments each improve accuracy, and the combination is better than either of the two approximations separately.

Figure 7 presents the Chi-square and Yarnold approximations with binomial sample sizes 2,3,4,5, both with and without the adjustment for the sample sizes 2 and 3 category, and Figure 8 shows the error of these approximations. The Yarnold and small sample size adjustments each improve accuracy, and the combination is better than either of the two approximations separately.

5. Conclusion

We have investigated the problem of evaluation of symmetry in two-dimensional tables, recast this problem as a more general question of tests of multiple binomial

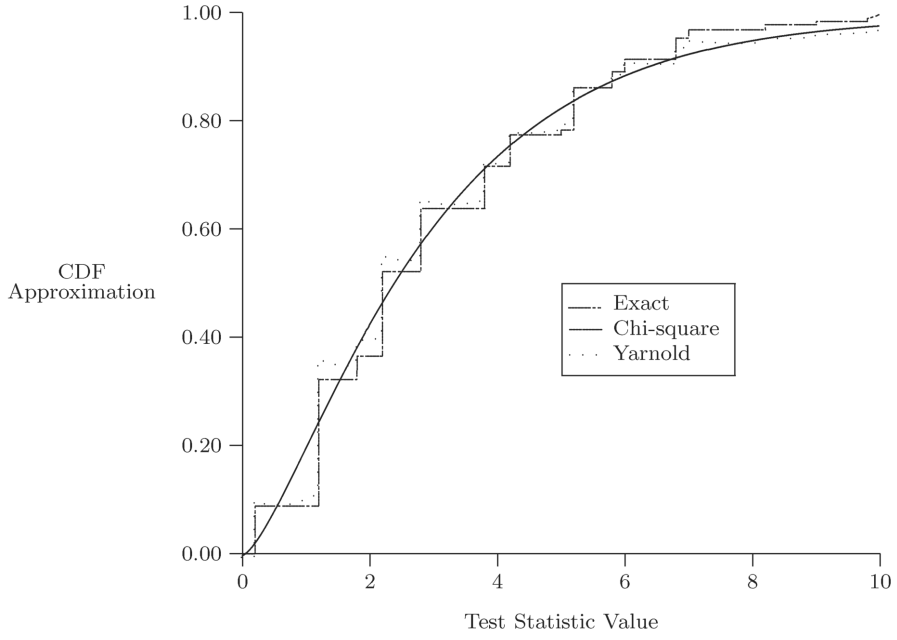

FIG 1. *Approximate and Exact Distribution Functions for Tests of Multiple Binomial Parameters.*[1]

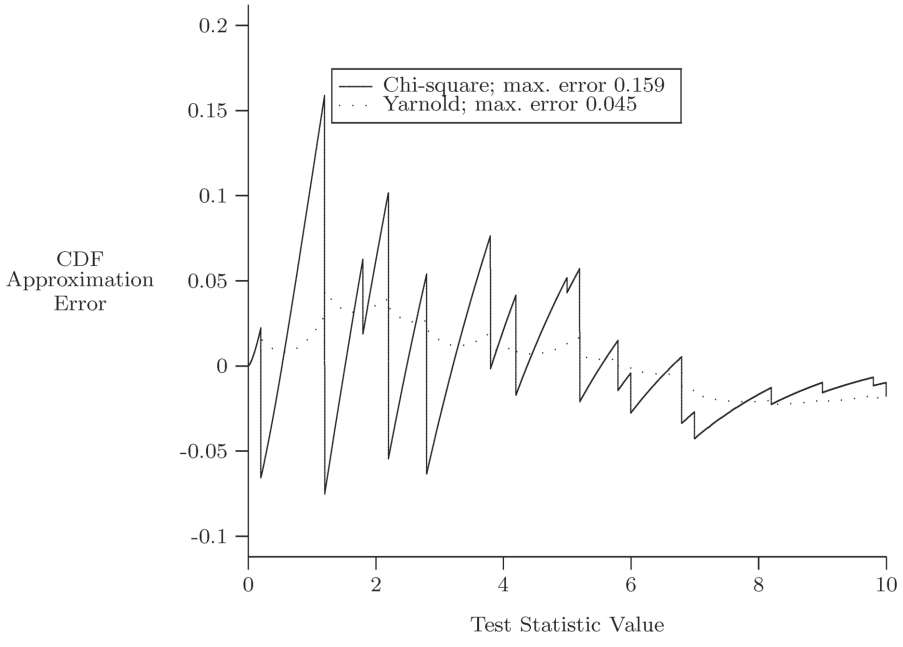

FIG 2. *Errors in Distribution Function Approximations for Tests of Multiple Binomial Parameters.*

[1]The approximate and exact distribution functions are based on sample sizes of 4, 4, and 5.

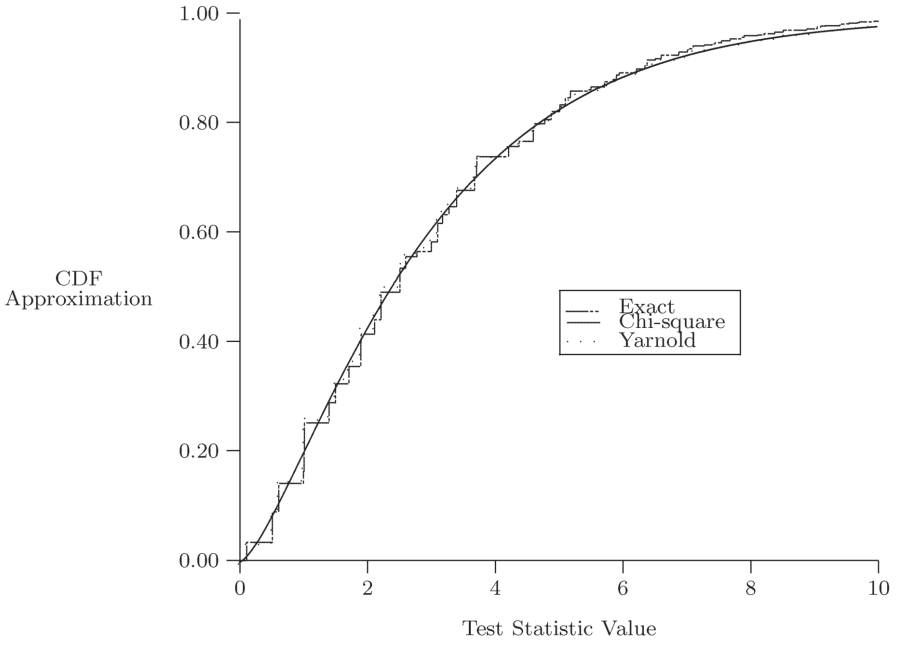

FIG 3. *Approximate and Exact Distribution Functions for Tests of Multiple Binomial Parameters.*[1]

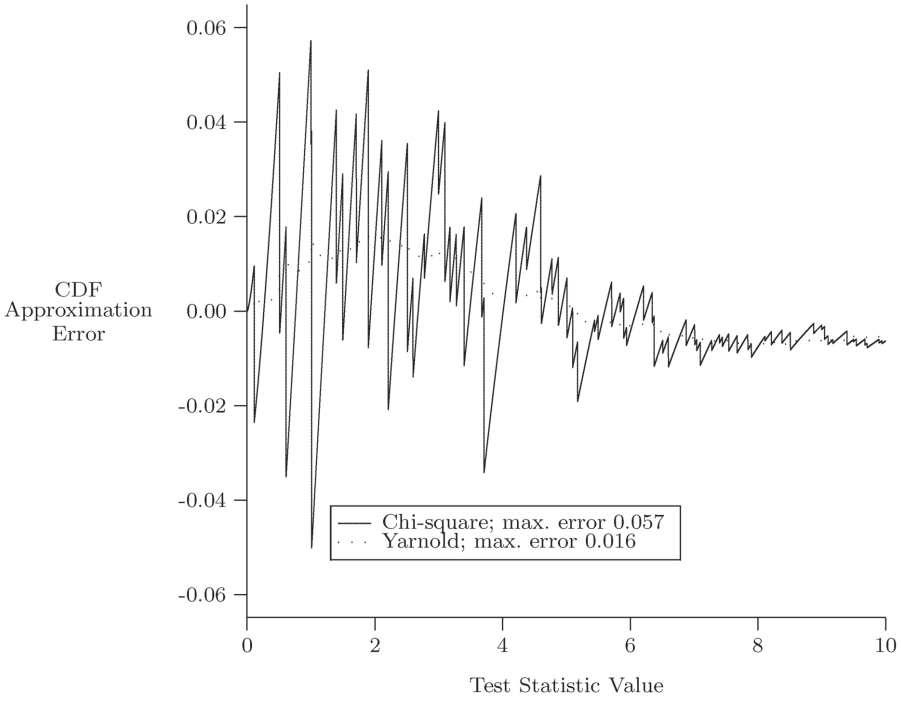

FIG 4. *Errors in Distribution Function Approximations for Tests of Multiple Binomial Parameters.*

[1]The approximate and exact distribution functions are based on sample sizes of 8, 9, and 10.

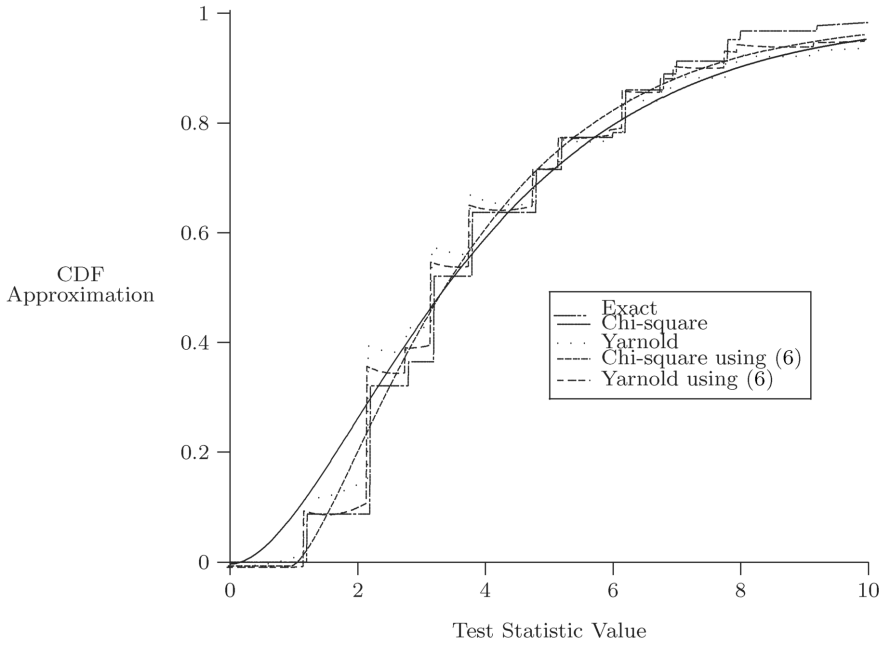

FIG 5. *Approximate and Exact Distribution Functions for Tests of Multiple Binomial Parameters.*[1]

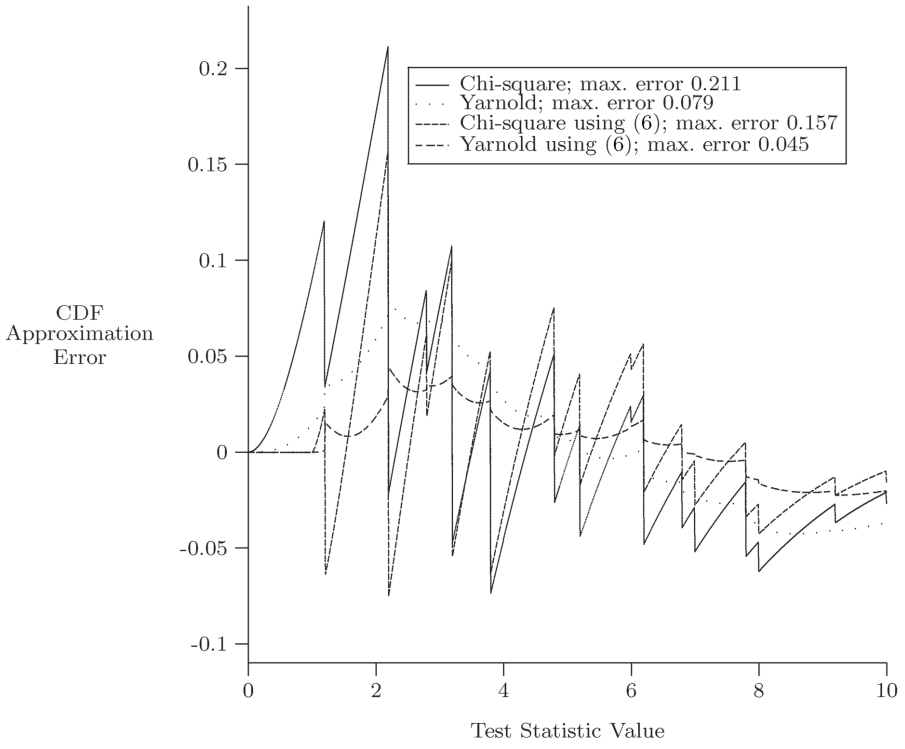

FIG 6. *Errors in Distribution Function Approximations for Tests of Multiple Binomial Parameters.*

[1]The approximate and exact distribution functions are based on sample sizes of 1, 4, 4, and 5.

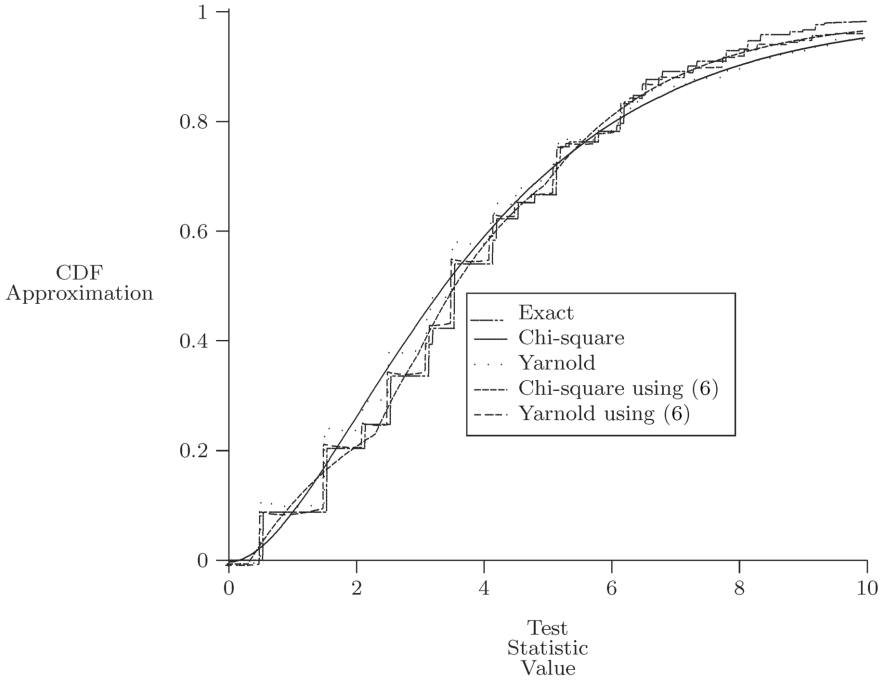

FIG 7. *Approximate and Exact Distribution Functions for Tests of Multiple Binomial Parameters.*[1]

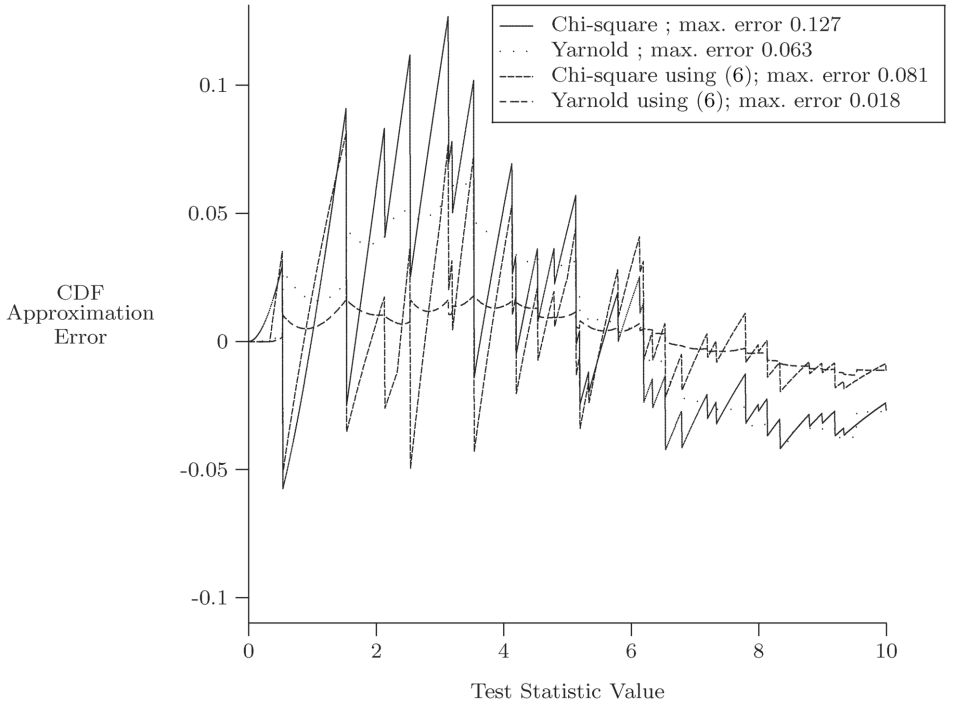

FIG 8. *Errors in Distribution Function Approximations for Tests of Multiple Binomial Parameters.*

[1]The approximate and exact distribution functions are based on sample sizes of 2, 3, 4, and 5.

parameters, and provided approximations to p-values that are more accurate than those commonly employed. This approximation combines an accurate approximation for the probabilities that lattice random vectors lie in an ellipse, and exact convolution for binomial cells for which this exact convolution is easy. The resulting approximation is easy to apply, and far more accurate than are conventional approximations.

References

Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley.

Bowker, A. H. (1948). A Test for Symmetry in Contingency Tables. *Journal of the American Statistical Association* **43** 572–574.

Kolassa, J. E. (2003). Continuity Correction for the Score Statistic in Discrete Regression Models. In *Crossing Boundaries: Statistical Essays in Honor of Jack Hall* (J. E. Kolassa and D. Oakes, eds.) Institute of Mathematical Statistics, Hayward, CA.

Kolassa, J. E. (2006). *Series Approximation Methods in Statistics, 3rd Edn.* Springer – Verlag.

Krampe, A. and Kuhnt, S. (2007). Bowker's test for symmetry and modifications within the algebraic framework. *Computational Statistics and Data Analysis* **51** 4124–4142.

Ludbrook, J. (2008). Analysis of 2×2 tables of frequencies: matching test to experimental design. *Int. J. Epidemiol.* **37** 1430–1435.

Oster, R. A. (2003). An Examination of Statistical Software Packages for Categorical Data Analysis Using Exact Methods-Part II. *The American Statistician* **57** 201–213.

Oster, R. A. and Hilbe, J. M. (2008a). An Examination of Statistical Software Packages for Parametric and Nonparametric Data Analyses Using Exact Methods. *The American Statistician* **62** 74–84.

Oster, R. A. and Hilbe, J. M. (2008b). Rejoinder to "An Examination of Statistical Software Packages for Parametric and Nonparametric Data Analyses Using Exact Methods". *The American Statistician* **62** 173–176.

SAS Institute Inc. (2010). *SAS OnlineDoc 9.2: PDF Files, Second Edition* The FREQ Procedure. SAS Institute Inc., Cary, NC.

Yarnold, J. K. (1972). Asymptotic Approximations for the Probability that a Sum of Lattice Random Vectors Lies in a Convex Set. *The Annals of Mathematical Statistics* **43** 1566–1580.