Computing the confidence levels for a root-mean-square test of goodness-of-fit, II

William Perkins[*]

School of Mathematics
Georgia Institute of Technology
686 Cherry St.
Atlanta, GA 30332-0160
e-mail: `wperkins3@math.gatech.edu`

Mark Tygert[†]

Courant Institute of Mathematical Sciences
NYU
251 Mercer St.
New York, NY 10012
e-mail: `tygert@aya.yale.edu`

and

Rachel Ward[‡]

Department of Mathematics
University of Texas
1 University Station, C1200
Austin, TX 78712
e-mail: `rward@math.utexas.edu`

Abstract: This paper is an extension of our earlier article, "Computing the confidence levels for a root-mean-square test of goodness-of-fit;" unlike in the earlier article, the models in the present paper involve parameter estimation — both the null and alternative hypotheses in the associated tests are composite. We provide efficient black-box algorithms for calculating the asymptotic confidence levels of a variant on the classic χ^2 test. In some circumstances, it is also feasible to compute confidence levels via Monte-Carlo simulations.

AMS 2010 subject classifications: Primary 62G10, 62F03; secondary 65C60.

Keywords and phrases: chi-square, significance, Euclidean, quadratic.

[*] Supported in part by NSF Grant OISE-0730136 and an NSF Postdoctoral Research Fellowship.

[†] Supported in part by an Alfred P. Sloan Research Fellowship.

[‡] Supported in part by an NSF Postdoctoral Research Fellowship and a Donald D. Harrington Faculty Fellowship.

Contents

1. Introduction

A basic task in statistics is to ascertain whether a given set of independent and identically distributed (i.i.d.) draws does not come from any member of a specified family of probability distributions (the specified family is known as the "model"). The present paper considers the case in which the draws are discrete random variables, taking values in a finite set. In accordance with the standard terminology, we will refer to the possible values of the discrete random variables as "bins" ("categories," "cells," and "classes" are common synonyms for "bins"). In earlier work, Perkins, Tygert, and Ward (2011b) treated the special case in which the "family" of distributions constituting the model in fact consists of a single, fully specified probability distribution. The present article focuses on models parameterized with a single scalar; our techniques extend straightforwardly to any parameterization with multiple scalars (or, equivalently, to any parameterization with a vector).

A natural approach to ascertaining whether a given set of i.i.d. draws does not come from the model uses a root-mean-square statistic. To construct this statistic, we estimate both the parameter and the probability distribution over the bins using the given i.i.d. draws, and then measure the root-mean-square difference between this empirical distribution and the model distribution corresponding to the estimated parameter (for details, see, for example, Rao, 2002; Varadhan, Levandowsky, and Rubin, 1974, page 123; or Section 3 below). If the draws do in fact arise from the specified model, then with high probability this root-mean-square is not large. Thus, if the root-mean-square statistic is large, then we can be confident that the draws did not arise from the model.

To quantify "large" and "confident," let us denote by x the value of the root-mean-square for the given i.i.d. draws; let us denote by X the root-mean-square statistic constructed for different i.i.d. draws that definitely do in fact come from the model. The P-value P is then defined to be the probability that $X \geq x$ (viewing X — but not x — as a random variable). The confidence level that the given i.i.d. draws do not arise from the model is the complement of the P-value, namely $1 - P$.

Unfortunately, the confidence levels for the simple root-mean-square are different for different models. In order to avoid this seeming inconvenience (at least asymptotically), one may weight the average in the root-mean-square by the reciprocals of the model probabilities associated with the various bins, obtaining the classic χ^2 statistic of Pearson (1900); see Remark 3.1 below. However, with the now widespread availability of computers, direct use of the simple root-mean-square statistic has become feasible (and actually turns out to be very convenient). The present paper provides efficient black-box algorithms for computing the confidence levels for any model with a smooth parameterization, in the limit of large numbers of draws. Calculating confidence levels for small numbers of draws via Monte-Carlo simulations can also be practical. The many advantages to using the root-mean-square have been discussed at length by Perkins, Tygert, and Ward (2011a).

The remainder of the present article has the following structure: Section 2 reviews previously developed techniques utilized in the following sections. Section 3 details the simple statistic discussed above, expressing the asymptotic confidence levels for the associated goodness-of-fit test in a form suitable for rapid computation. Section 4 applies the algorithms of the present paper to several examples. Section 5 draws some conclusions.

2. Preliminaries

This section summarizes a previously introduced numerical method.

The following theorem (proven in Section 3 of Perkins, Tygert, and Ward, 2011b) expresses the cumulative distribution function of the sum of the squares of independent centered Gaussian random variables as an integral suitable for evaluation via quadratures.

Theorem 2.1. *Suppose that n is a positive integer, X_1, X_2, …, X_{n-1}, X_n are i.i.d. Gaussian random variables of zero mean and unit variance, and σ_1, σ_2, …, σ_{n-1}, σ_n are positive real numbers. Suppose in addition that X is the random variable*

$$X = \sum_{k=1}^{n} |\sigma_k \, X_k|^2. \tag{1}$$

Then, the cumulative distribution function F of X is

$$F(x) = \int_0^\infty \operatorname{Im}\left(\frac{e^{1-t}\, e^{it\sqrt{n}}}{\pi\left(t - \frac{1}{1-i\sqrt{n}}\right) \prod_{k=1}^{n} \sqrt{1 - 2(t-1)\sigma_k^2/x + 2it\sigma_k^2\sqrt{n}/x}} \right) dt \tag{2}$$

for any positive real number x, and $F(x) = 0$ for any nonpositive real number x. The square roots in (2) denote the principal branch, and Im *takes the imaginary part.*

Remark 2.2. The absolute value of the expression under the square root in (2) is always greater than $\sqrt{n/(n+1)}$. Therefore,

$$\left| \prod_{k=1}^{n} \sqrt{1 - 2(t-1)\sigma_k^2/x + 2it\sigma_k^2\sqrt{n}/x} \right| > \left(\frac{n}{n+1} \right)^{n/4} > \frac{1}{e^{1/4}} \tag{3}$$

for any $t \in (0, \infty)$ and any $x \in (0, \infty)$. Thus, the integrand in (2) is never large for $t \in (0, \infty)$.

Remark 2.3. An efficient means of evaluating (2) numerically is to use adaptive Gaussian quadratures (see, for example, Section 4.7 of Press et al., 2007). To attain double-precision accuracy (roughly 15-digit precision), the domain of integration for t in (2) need be only $(0, 40)$ rather than the whole $(0, \infty)$. Good choices for the lowest orders of the quadratures used in the adaptive Gaussian quadratures are 10 and 21, for double-precision accuracy. (See Section 3 of Perkins, Tygert, and Ward, 2011b, for the details.)

3. The simple statistic

This section details the simple root-mean-square statistic discussed briefly in Section 1, determining its probability distribution in the limit of large numbers of draws, assuming that the draws do in fact come from the specified model. The distribution determined in this section yields the confidence levels (in the limit of large numbers of draws): Given a value x for the root-mean-square statistic constructed from i.i.d. draws coming from an unknown distribution, and given the value of the maximum-likelihood estimate $\hat{\theta}$ for the parameter of the distribution, the confidence level that the draws do not come from the specified model is the probability that the root-mean-square statistic is less than x when constructed from i.i.d. draws that do come from the model distribution associated with the parameter $\hat{\theta}$. (Please note that the definition in (5) and (6) below of the simple statistic involves the maximum-likelihood estimate $\hat{\theta}$. Maximum likelihood is the canonical method for parameter estimation, and is the focus of the present paper. See formulae (16) and (17) below regarding likelihood and maximum-likelihood estimation.)

3.1. The distribution of the goodness-of-fit statistic

To begin, we set notation and form the goodness-of-fit statistic X to be analyzed. Given n bins, numbered 1, 2, ..., $n-1$, n, we denote by $p_1(\theta)$, $p_2(\theta)$, ..., $p_{n-1}(\theta)$, $p_n(\theta)$ the probabilities associated with the respective bins under the specified model, where θ is a real number parameterizing the model; of course,

$$\sum_{k=1}^{n} p_k(\theta) = 1 \tag{4}$$

for any parameter θ. In order to obtain a draw conforming to the model for a particular value of θ, we select at random one of the n bins, with probabilities $p_1(\theta)$, $p_2(\theta)$, …, $p_{n-1}(\theta)$, $p_n(\theta)$. We perform this selection independently m times. For $k = 1, 2, \dots, n-1, n$, we denote by Y_k the fraction of times that we choose bin k (that is, Y_k is the number of times that we choose bin k, divided by m); obviously, $\sum_{k=1}^n Y_k = 1$. We define X_k to be $\sqrt{m}$ times the difference of Y_k from its expected value using the maximum-likelihood estimate $\hat{\theta}$ of the actual parameter θ, that is,

$$X_k = \sqrt{m}\,(Y_k - p_k(\hat{\theta})) \tag{5}$$

for $k = 1, 2, \dots, n-1, n$. Finally, we form the statistic

$$X = \sum_{k=1}^{n} X_k^2, \tag{6}$$

and now determine its distribution in the limit that the number m of draws is large. (The root-mean-square statistic $\sqrt{\sum_{k=1}^n (mY_k - mp_k(\hat{\theta}))^2/m}$ is the square root of X. As the square root is a monotonically increasing function, the confidence levels are the same whether determined via X or via $\sqrt{X}$; for convenience, we focus on X below.)

Remark 3.1. The classic χ^2 test for goodness-of-fit of Pearson (1900) replaces (6) with the statistic

$$\chi^2 = \sum_{k=1}^{n} \frac{X_k^2}{p_k(\hat{\theta})}, \tag{7}$$

where X_1, X_2, …, X_{n-1}, X_n are the same as in (5) and (6), and $\hat{\theta}$ is the maximum-likelihood estimate of the parameter.

For definiteness, we will be assuming that p_1, p_2, …, p_{n-1}, p_n are differentiable as functions of the parameter θ, that the maximum of the likelihood occurs in the interior of the domain for θ, that the maximum-likelihood estimate $\hat{\theta}$ is almost surely the correct value for the actual parameter θ as $m \to \infty$, and that the variance of $\hat{\theta}$ tends to zero as $m \to \infty$ (thus $\hat{\theta}$ is not "random" in the limit of large numbers of draws). As detailed, for example, by Moore and Spruill (1975) and by Kendall et al. (2009) in a chapter on goodness-of-fit (see also Remark 3.3 below), the multivariate central limit theorem then shows that the joint distribution of X_1, X_2, …, X_{n-1}, X_n converges in distribution as $m \to \infty$, with the limiting generalized probability density proportional to

$$\exp\left(-\sum_{k=1}^{n} \frac{x_k^2}{2p_k(\hat{\theta})}\right) \cdot \delta\left(\sum_{k=1}^{n} x_k\right) \cdot \delta\left(\sum_{k=1}^{n} x_k \frac{d}{d\theta} \ln(p_k(\theta))\bigg|_{\theta=\hat{\theta}}\right), \tag{8}$$

where δ is the Dirac delta, and $\hat{\theta}$ is the maximum-likelihood estimate of the parameter.

The generalized probability density in (8) is a centered multivariate Gaussian distribution concentrated on the intersection of two hyperplanes that both pass through the origin (the intersection of the hyperplanes consists of all the points such that $\sum_{k=1}^{n} x_k = 0$ and $\sum_{k=1}^{n} x_k \frac{d}{d\theta} \ln(p_k(\theta))\big|_{\theta=\hat{\theta}} = 0$); the restriction of the generalized probability density (8) to the intersection of the hyperplanes is also a centered multivariate Gaussian. Thus, the distribution of X defined in (6) converges as $m \to \infty$ to the distribution of the sum of the squares of $n-2$ independent Gaussian random variables of mean zero whose variances are the variances of the restricted multivariate Gaussian distribution along its principal axes (see, for example, Chapter 25 of Kendall et al., 2009). Given these variances, Remark 2.3 describes an efficient algorithm for computing the probability that the associated sum of squares is less than any particular value; this probability is the desired confidence level, in the limit of large numbers of draws. For a detailed discussion, see Section 3.2 below.

To compute the variances of the restricted multivariate Gaussian distribution along its principal axes, we perform the following four steps:

1. Form an $n \times 2$ matrix H whose columns both include a vector that is normal to the hyperplane consisting of the points $(x_1, x_2, \ldots, x_{n-1}, x_n)$ such that
$$\sum_{k=1}^{n} x_k = 0, \tag{9}$$
and also include a vector that is normal to the hyperplane consisting of the points $(x_1, x_2, \ldots, x_{n-1}, x_n)$ such that
$$\sum_{k=1}^{n} x_k \frac{d}{d\theta} \ln(p_k(\theta))\bigg|_{\theta=\hat{\theta}} = 0, \tag{10}$$
where $\hat{\theta}$ is the maximum-likelihood estimate of the parameter. For example, we can take the entries of H to be
$$H_{k,j} = \begin{cases} 1, & j = 1 \\ \frac{d}{d\theta} \ln(p_k(\theta))\big|_{\theta=\hat{\theta}}, & j = 2 \end{cases} \tag{11}$$
for $k = 1, 2, \ldots, n-1, n$ and $j = 1, 2$, where again $\hat{\theta}$ is the maximum-likelihood estimate of the parameter.
2. Form an orthonormal basis for the column space of H, by constructing a pivoted QR decomposition
$$H_{n\times 2} = Q_{n\times 2} \cdot R_{2\times 2} \cdot \Pi_{2\times 2}, \tag{12}$$
where the columns of Q are orthonormal, R is upper-triangular, and Π is a permutation matrix. (See, for example, Chapter 5 of Golub and Van Loan, 1996, for details on the construction of such a pivoted QR decomposition.)
3. Form the $n \times n$ diagonal matrix D with the entries
$$D_{j,k} = \begin{cases} 1/p_k(\hat{\theta}), & j = k \\ 0, & j \neq k \end{cases} \tag{13}$$

for $j, k = 1, 2, \dots, n-1, n$, where $\hat{\theta}$ is the maximum-likelihood estimate of the parameter. Then, multiply D from both the left and the right by the orthogonal projection $(\mathbf{1} - QQ^\top)$ onto the intersection of the hyperplanes consisting of the points satisfying (9) and (10), obtaining the $n \times n$ matrix

$$B = (\mathbf{1} - QQ^\top)\, D\, (\mathbf{1} - QQ^\top), \tag{14}$$

where $\mathbf{1}$ is the $n \times n$ identity matrix.

4. Find the eigenvalues of the self-adjoint matrix B defined in (14). By construction, exactly two of the eigenvalues of B are zeros. The other eigenvalues of B are the reciprocals of the desired variances of the restricted multivariate Gaussian distribution along its principal axes.

Remark 3.2. The $n \times n$ matrix B defined in (14) is the sum of a diagonal matrix and a low-rank matrix. The methods of Gu and Eisenstat (1994, 1995) for computing the eigenvalues of such a matrix B require only either $\mathcal{O}(n^2)$ or $\mathcal{O}(n)$ floating-point operations. Note that the $\mathcal{O}(n^2)$ methods of Gu and Eisenstat (1994, 1995) are more efficient than the $\mathcal{O}(n)$ procedure of Gu and Eisenstat (1995), unless n is impractically large.

Remark 3.3. Under appropriate regularity conditions, it is easy to derive the homogeneous linear constraint — analogous to (10) — that

$$\sum_{k=1}^{n} X_k \, \frac{d}{d\theta} \ln(p_k(\theta)) \bigg|_{\theta=\hat{\theta}} = 0, \tag{15}$$

where $\hat{\theta}$ is the maximum-likelihood estimator. The following is a sketch of the proof of (15).

To determine the maximum-likelihood estimate $\hat{\theta}$, we consider the likelihood, namely the multinomial distribution

$$L(y_1, y_2, \dots, y_{n-1}, y_n, \theta) = m! \prod_{k=1}^{n} \frac{(p_k(\theta))^{my_k}}{(my_k)!}. \tag{16}$$

Maximizing (16) defines $\hat{\theta}$ via the formula

$$0 = \frac{\partial}{\partial\theta} \ln(L(Y_1, Y_2, \dots, Y_{n-1}, Y_n, \theta)) \bigg|_{\theta=\hat{\theta}} = \sum_{k=1}^{n} mY_k \, \frac{d}{d\theta} \ln(p_k(\theta)) \bigg|_{\theta=\hat{\theta}}. \tag{17}$$

It follows from (4) that

$$\sum_{k=1}^{n} \frac{d}{d\theta} \, p_k(\theta) = 0 \tag{18}$$

for any parameter θ, in particular for $\theta = \hat{\theta}$. Combining (17) and (18) yields that

$$\sum_{k=1}^{n} (Y_k - p_k(\hat{\theta})) \, \frac{d}{d\theta} \ln(p_k(\theta)) \bigg|_{\theta=\hat{\theta}} = 0. \tag{19}$$

Combining (19) and (5) yields (15), as desired.

3.2. A procedure for computing the confidence levels

An efficient method for calculating the confidence levels in the limit of large numbers of draws proceeds as follows. Given i.i.d. draws from any distribution — not necessarily from the model — we can form the associated statistic X defined in (6) and (5); in the limit of large numbers of draws, the confidence level that the draws do not arise from the model is then just the cumulative distribution function $F(x)$ in (2) evaluated at $x = X$, with σ_k^2 in (2) obtained via Step 4 of the algorithm of Section 3.1 (after all, $F(x)$ is the probability that x is greater than the sum of the squares of independent centered Gaussian random variables whose variances are given by Step 4 above). Remark 2.3 describes an efficient means of evaluating $F(x)$ numerically.

4. Numerical examples

This section illustrates the performance of the algorithm of Section 3.2 via several numerical examples.

Figure 1 and Table 1 correspond to the first example. The model distribution for the first example has 4 bins, with the probabilities indicated in Table 4. We will detail the interpretation of the figures and tables shortly.

Figure 2 and Table 2 correspond to the second example. The model for the second example is the Zipf distribution on 100 bins. The row for Figure/Table 2 in Table 4 provides a definition of the Zipf distribution.

Figure 3 and Table 3 correspond to the third example. The model for the third example is the standard Poisson distribution. The row for Figure/Table 3 in Table 4 provides a definition of the Poisson distribution.

To test our algorithms, we conduct computational simulations. In every simulation, we choose the number m of draws to be a very large number, namely $m = 100{,}000$. (The algorithms of the present paper concern the limit as $m \to \infty$.) Part (a) of the examples uses $j = 1{,}000$ simulations; part (b) of the examples uses $j = 10{,}000$ simulations. The convergence (as j increases) of the plotted points to the straight line of unit slope through the origin provides numerical validation of our algorithms, for the following reasons.

To create the plots, we run j simulations, each taking $m = 100{,}000$ i.i.d. draws from the model distribution with the specified parameter θ. For each simulation, we compute the statistic X defined in (6), forming Y_1, Y_2, …, Y_{n-1}, Y_n and $\hat{\theta}$ needed in (5) and (6) using the generated draws. We then compute the asymptotic confidence level associated with each of these values for X, as described in Section 3.2, and sort the resulting confidence levels. These sorted results are the vertical coordinates of the points in the plots; the horizontal coordinates are the equispaced numbers $1/(2j)$, $3/(2j)$, …, $(2j-3)/(2j)$, $(2j-1)/(2j)$.

As the number j of simulations increases, and insofar as the number m of draws is very large, the plotted points should converge to the straight line through the origin of slope 1 (and, indeed, our experiments demonstrate this).

The dotted line in each plot is the straight line through the origin of slope 1. The trials converge correctly: The root-mean-square statistics for about $\alpha\%$ of the simulations should have P-values of $\alpha\%$ or less, for every $\alpha \in (0, 100)$; in the limit that both the number m of draws and the number j of simulations are large, the computed P-values for exactly $\alpha\%$ of the simulations should be less than or equal to $\alpha\%$ (this follows from the definition of P-values; it also follows from the fact that the confidence levels for the statistic X are given by its cumulative distribution function F, and from the fact that $F(X)$ is uniformly distributed over $(0, 1)$ for any random variable X distributed according to a continuous cumulative distribution function F).

The following list describes the headings of the tables:

- j is the number of simulations conducted in generating the associated plot.
- θ is the parameter for the model distribution used in generating the i.i.d. draws.
- n is the number of bins/categories/cells/classes in the model (see Remark 4.2 regarding the Poisson distribution of the third example).
- l is the maximum number of quadrature nodes required to evaluate the confidence level for any of the j root-mean-square statistics produced by the simulations.
- t is the total number of seconds required to perform the quadratures for evaluating the confidence levels for all j of the root-mean-square statistics produced by the simulations.
- s is the total number of seconds required to perform all j simulations.
- $p_k(\theta)$ is the probability associated with bin k ($k = 1, 2, \dots, n-1, n$), as a function of the parameter θ.
- $\hat{\theta}(Y_1, Y_2, \dots, Y_{n-1}, Y_n)$ is the maximum-likelihood estimate of the parameter θ, as a function of the fractions Y_1, Y_2, …, Y_{n-1}, Y_n of the draws in the respective bins (Section 3 provides a detailed definition of Y_1, Y_2, …, Y_{n-1}, Y_n).

We used Fortran 77 and ran all examples on one core of a 2.2 GHz Intel Core 2 Duo microprocessor with 2 MB of L2 cache. Our code is compliant with the IEEE double-precision standard (so that the mantissas of variables have approximately one bit of precision less than 16 digits, yielding a relative precision of about 2E–16). We diagonalized the matrix B defined in (14) using the Jacobi algorithm (see, for example, Chapter 8 of Golub and Van Loan, 1996), not taking advantage of Remark 3.2. We generated the pseudorandom numbers used in the simulations via (Mitchell-Moore-Brent-Knuth) lagged Fibonacci sequences (see, for example, Section 7.1.5 of Press et al., 2007).

Observation 1. It is easy to compute the confidence levels (in the limit of large numbers of draws) for a distribution having infinitely many bins, but only to any arbitrary accuracy that is greater than the machine precision. Specifically, given a fully specified model distribution and an extremely small positive real number ε, we would retain the smallest possible number of bins whose associated probabilities $p_1, p_2, \dots, p_{n-1}, p_n$ satisfy $p_1+p_2+\cdots+p_{n-1}+p_n \geq 1-\varepsilon$, and then proceed with the computation as if these finitely many were the only bins. When there is a parameter θ being estimated, we observe that the maximum-likelihood estimate $\hat{\theta}$ typically has variance zero and is almost surely correct in the limit of large numbers of draws; thus, as before, we may retain the smallest possible number of bins whose associated probabilities $p_1(\hat{\theta})$, $p_2(\hat{\theta})$, …, $p_{n-1}(\hat{\theta})$, $p_n(\hat{\theta})$ satisfy $p_1(\hat{\theta}) + p_2(\hat{\theta}) + \cdots + p_{n-1}(\hat{\theta}) + p_n(\hat{\theta}) \geq 1 - \varepsilon$, and then proceed with the computation as if these finitely many were the only bins. (Needless to say, if the fraction of the experimental draws falling outside the finitely many retained bins is significantly greater than ε, then we can be highly confident that the draws did not arise from the model.)

Remark 4.1. For the second example (the Zipf distribution), we computed the maximum-likelihood estimate $\hat{\theta}$ from the data Y_1, Y_2, …, Y_{n-1}, Y_n by finding the zero of the function $g(\hat{\theta}) = f(\hat{\theta}) - \sum_{k=1}^{n} Y_k \ln(k) = 0$, where f is the same as in Table 4, namely $f(\hat{\theta}) = \Big(\sum_{k=1}^{n} k^{-\hat{\theta}} \ln(k)\Big) \Big/ \Big(\sum_{k=1}^{n} k^{-\hat{\theta}}\Big)$. We evaluated the zero $\hat{\theta}$ numerically, via bisection (see, for example, Chapter 9 of Press et al., 2007).

Remark 4.2. For the third example (the Poisson distribution), we employed Observation 1, with $\varepsilon = 10^{-8}$.

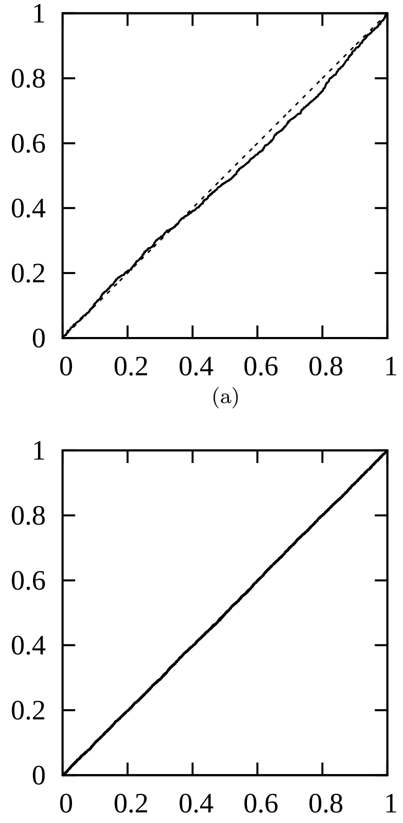
5. Conclusion

This paper provides efficient black-box algorithms for computing the confidence levels for one of the simplest, most natural goodness-of-fit statistics, in the limit of large numbers of draws. Although the present paper focuses on families of probability distributions parameterized with a single scalar (and the predecessor to this article focuses on fully specified distributions), our methods extend straightforwardly to any parameterization with multiple scalars (or, equivalently, to any parameterization with a vector). Furthermore, our methods can handle arbitrarily weighted means in the root-mean-square, in addition to the usual, uniformly weighted average considered above.

There are many advantages to using the simple root-mean-square, as shown by Perkins, Tygert, and Ward (2011a). With the now widespread availability of computers, calculating the relevant P-values via Monte-Carlo simulations can be feasible; the algorithms of the present paper can also be suitable, and are efficient and easy-to-use.

Acknowledgements

We would like to thank Tony Cai, Jianqing Fan, Peter W. Jones, Ron Peled, and Vladimir Rokhlin for many helpful discussions.



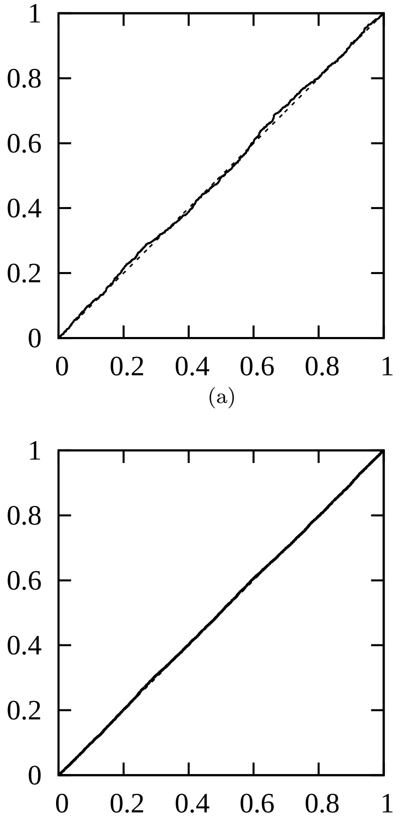
(a)

(b)

Fig. 1: 2 × 2 contingency-table/cross-tabulation of Table 4

Table 1
Values for Figure 1

	j	θ	n	l	t	s
(a)	10^3	.03	4	190	.43E0	.44E1
(b)	10^4	.03	4	190	.43E1	.45E2



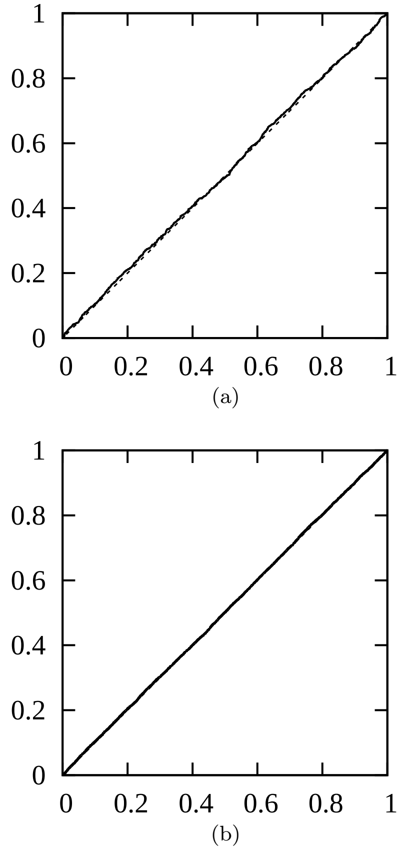
(a)

(b)

Fig. 2: Zipf distribution of Table 4

Table 2
Values for Figure 2

	j	θ	n	l	t	s
(a)	10^3	1	100	350	.92E1	.13E2
(b)	10^4	1	100	390	.11E3	.13E3



(a)

(b)

Fig. 3: Poisson distribution of Table 4

Table 3

Values for Figure 3

	j	θ	n	l	t	s
(a)	10^3	10.3	36	290	.37E1	.86E1
(b)	10^4	10.3	36	330	.37E2	.86E2

Table 4
Values for Figures 1–3 and Tables 1–3

Fig./Table #	$p_k(\theta)$	$\hat{\theta}(Y_1, Y_2, \dots, Y_{n-1}, Y_n)$
Fig./Table 1	$p_1 = .04{\cdot}\theta, p_2 = .04(1{-}\theta)$ $p_3 = .96{\cdot}\theta, p_4 = .96(1{-}\theta)$	$\hat{\theta} = Y_1 + Y_3$
Fig./Table 2	$p_k = k^{-\theta} \Big/ \sum_{i=1}^{n} i^{-\theta}$	$\hat{\theta} = f^{-1}\big(\sum_{k=1}^{n} Y_k \, \ln(k)\big)$, $f(\hat{\theta}) = \big(\sum_{k=1}^{n} k^{-\hat{\theta}} \, \ln(k)\big) \Big/ \big(\sum_{k=1}^{n} k^{-\hat{\theta}}\big)$
Fig./Table 3	$p_k = e^{-\theta}\theta^{k-1}/(k-1)!$	$\hat{\theta} = \sum_{k=1}^{\infty}(k-1)\, Y_k$

References

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*, 3rd ed. Johns Hopkins University Press, Baltimore, Maryland.

Gu, M. and Eisenstat, S. C. (1994). A stable and efficient algorithm for the rank-one modification of the symmetric eigenproblem. *SIAM J. Matrix Anal. Appl.* **15** 1266–1276.

Gu, M. and Eisenstat, S. C. (1995). A divide-and-conquer algorithm for the symmetric tridiagonal eigenproblem. *SIAM J. Matrix Anal. Appl.* **16** 172–191.

Kendall, M. G., Stuart, A., Ord, K. and Arnold, S. (2009). *Kendall's Advanced Theory of Statistics* **2A**, 6th ed. Wiley.

Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* **3** 599–616.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, Series 5,* **50** 157–175.

Perkins, W., Tygert, M. and Ward, R. (2011a). χ^2 and classical exact tests often wildly misreport significance; the remedy lies in computers. Technical Report No. 1108.4126, arXiv. http://cims.nyu.edu/∼tygert/abbreviated.pdf.

Perkins, W., Tygert, M. and Ward, R. (2011b). Computing the confidence levels for a root-mean-square test of goodness-of-fit. *Appl. Math. Comput.* **217** 9072–9084.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (2007). *Numerical Recipes*, 3rd ed. Cambridge University Press, Cambridge, UK.

Rao, C. R. (2002). Karl Pearson chi-square test: The dawn of statistical inference. In *Goodness-of-Fit Tests and Model Validity* (C. Huber-Carol, N. Balakrishnan, M. S. Nikulin and M. Mesbah, eds.) 9–24. Birkhäuser, Boston.

Varadhan, S. R. S., Levandowsky, M. and Rubin, N. (1974). *Mathematical Statistics. Lecture Notes Series.* Courant Institute, NYU, New York.