The Canadian Journal of Statistics
Vol. 35, No. 1, 2007, Pages 125–133
La revue canadienne de statistique

Cramér–von Mises statistics for discrete distributions with unknown parameters

Richard A. LOCKHART, John J. SPINELLI and Michael A. STEPHENS

Key words and phrases: Component; empirical distribution function; goodness-of-fit.

MSC 2000: Primary 62G30; secondary 62G10.

Abstract: Choulakian, Lockhart & Stephens (1994) proposed Cramér–von Mises statistics for testing fit to a fully specified discrete distribution. The authors give slightly modified definitions for these statistics and determine their asymptotic behaviour in the case when unknown parameters in the distribution must be estimated from the sample data. They also present two examples of applications.

Statistiques de Cramér–von Mises pour des lois discrètes dont les paramètres sont inconnus

Résumé : Choulakian, Lockhart & Stephens (1994) ont proposé des statistiques de Cramér–von Mises permettant de tester l'adéquation d'une loi discrète complètement spécifiée. Les auteurs donnent des définitions légèrement modifiées de ces statistiques et en déterminent le comportement asymptotique dans le cas où certains paramètres de la loi doivent être estimés à partir de données. Ils présentent en outre deux exemples d'application.

1. INTRODUCTION

In Choulakian, Lockhart & Stephens (1994), Cramér–von Mises statistics W^2, U^2 and A^2 were defined for testing a discrete distribution. Asymptotic theory was given for the case where the distribution tested was fully specified, and the tests for a discrete uniform distribution with k cells were discussed in detail. In this paper we modify the definitions slightly to allow for the same values of the statistics to be obtained if the cell order is completely reversed. This does not affect the results in Choulakian, Lockhart & Stephens (1994).

More importantly, we also add the asymptotic theory for the case when parameters of the tested distribution must be estimated from the sample data. The theory has in fact been used for several special cases: testing the Poisson distribution (Spinelli & Stephens 1997) and testing the exponential distribution with grouped data (Spinelli 2001).

Consider a discrete distribution with k cells labelled $1, \ldots, k$, and with probability p_i of falling into cell i. Suppose N independent observations are given; let o_i be the observed number of observations and $e_i = Np_i$ be the expected number in cell i. Let $S_j = \sum_{i=1}^{j} o_i$ and $T_j = \sum_{i=1}^{j} e_i$. Then S_j/N and $H_j = T_j/N$ are the cumulated histograms of observed and expected values and correspond to the empirical distribution function $F_N(x)$ and the cumulative distribution function $F(x)$ for continuous distributions. Suppose $Z_j = S_j - T_j, j = 1, \ldots, k$; the weighted mean of the Z_i is $\bar{Z} = \sum_{j=1}^{k} Z_j t_j$, where $t_j = (p_j + p_{j+1})/2$, with $p_{k+1} = p_1$. The modified Cramér–von Mises statistics are then defined as follows:

$$W_d^2 = N^{-1} \sum_{j=1}^{k} Z_j^2 t_j;$$

$$U_d^2 = N^{-1} \sum_{j=1}^{k} (Z_j - \bar{Z})^2 t_j;$$

$$A_d^2 = N^{-1} \sum_{j=1}^{k} Z_j^2 t_j / \{H_j(1 - H_j)\}.$$

Note that $Z_k = 0$ in these summations, so that the last term in W_d^2 is zero. The last term in A_d^2 is of the form $0/0$, and is set equal to zero.

The well-known Pearson chi-squared statistic is

$$X^2 = \sum_{i=1}^{k} (o_i - e_i)^2 / e_i .$$

Statistics corresponding to the Kolmogorov–Smirnov statistics for continuous observations are

$$\begin{aligned} D_d^+ &= \max_j (Z_j)/\sqrt{N}\,, \\ D_d^- &= \max_j (-Z_j)/\sqrt{N}\,, \\ D_d &= \max_j |Z_j|/\sqrt{N}\,. \end{aligned}$$

A feature of these statistics is that the Cramér–von Mises and Kolmogorov–Smirnov statistics take into account the order of the cells in contrast to the Pearson X^2 statistic. In Choulakian, Lockhart & Stephens (1994), p_j was used instead of t_j in these definitions; as stated above, the results in that paper still hold with the change in weights, since the concern was with the uniform distribution where all $t_j = p_j = 1/k$. However, the new definitions ensure that in a more general pattern of cell probabilities, if the cells are completely reversed in order, the values of the statistics are unaltered. This would seem to be a desirable quality; for instance, in testing the binomial distribution, where one statistician might cumulate the successes, and another the failures, or in a test involving categorical data such as the tones of a photograph, light to dark, or vice versa.

The statistic U_d^2 is intended for use with a discrete distribution around a circle; the other statistics will change their values with different choices of origin, but U_d^2 is unchanged; this is why p_{k+1} is defined to be p_1.

The above definitions can be put into matrix notation. Let a superscript $\top$, e. g., $\mathbf{Z}^\top$, denote the transpose of a vector or matrix. Let $\mathbf{I}$ be the $k \times k$ identity matrix, and let $\mathbf{p}^\top$ be the $1 \times k$ vector $(p_1, \ldots, p_k)$. Suppose $\mathbf{D}$ is the $k \times k$ diagonal matrix whose jth diagonal entry is p_j, $j = 1, \ldots, k$ and let $\mathbf{E}$ be the diagonal matrix with diagonal entries t_j, and $\mathbf{K}$ be the diagonal matrix whose (j, j)th element is $1/\{H_j(1 - H_j)\}$, $j = 1, \ldots, k-1$ and $\mathbf{K}_{kk} = 0$. Let o_i and e_i be arranged into column vectors $\mathbf{o}$, $\mathbf{e}$ (so that, for example, the jth component of $\mathbf{o}$ is o_j, $j = 1, \ldots, k$). Then $\mathbf{Z} = \mathbf{A}\mathbf{d}$, where $\mathbf{d} = \mathbf{o} - \mathbf{e}$ and $\mathbf{A}$ is the $k \times k$ partial-sum matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{pmatrix}.$$

The definitions become

$$\begin{aligned} W_d^2 &= \mathbf{Z}^\top \mathbf{E} \mathbf{Z}/N, & (1)\\ U_d^2 &= \mathbf{Z}^\top (\mathbf{I} - \mathbf{E}\mathbf{1}\mathbf{1}^\top)\mathbf{E}(\mathbf{I} - \mathbf{1}\mathbf{1}'\mathbf{E})\mathbf{Z}/N, & (2)\\ A_d^2 &= \mathbf{Z}^\top \mathbf{E}\mathbf{K}\mathbf{Z}/N, & (3)\\ X^2 &= (\mathbf{d}^\top \mathbf{D}^{-1}\mathbf{d})/N = \mathbf{Z}^\top {\mathbf{A}^{-1}}^\top \mathbf{D}^{-1}\mathbf{A}^{-1}\mathbf{Z}/N. & (4) \end{aligned}$$

These matrix forms have been introduced for the asymptotic theory, but it is convenient also to use them to calculate the statistics.

2. ASYMPTOTIC THEORY

2.1. All parameters known.

All four statistics above are of the general form $S = \mathbf{Y}^\top \mathbf{M}\mathbf{Y}$, where $\mathbf{Y} = \mathbf{Z}/\sqrt{N}$ and $\mathbf{M}$ is symmetric. For W_d^2, $\mathbf{M} = \mathbf{E}$, for U_d^2, $\mathbf{M} = (\mathbf{I} - \mathbf{E}\mathbf{1}\mathbf{1}^\top)\mathbf{E}(\mathbf{I} - \mathbf{1}\mathbf{1}^\top\mathbf{E})$, and for A_d^2, $\mathbf{M} = \mathbf{E}\mathbf{K}$. Also $\mathbf{Y}$ has mean $\mathbf{0}$. Suppose its covariance matrix is $\boldsymbol{\Sigma}_y$; then S may be written

$$S = \mathbf{Y}^\top \mathbf{M}\mathbf{Y} = \sum_{i=1}^{k-1} \lambda_i (\mathbf{w}_i^\top \mathbf{Y})^2, \tag{5}$$

where λ_i are the $k-1$ non-zero eigenvalues of $\mathbf{M}\boldsymbol{\Sigma}_y$ and $\mathbf{w}_i$ are the corresponding eigenvectors, normalized so that $\mathbf{w}_i^\top \boldsymbol{\Sigma}_y \mathbf{w}_j = \delta_{ij}$ where δ_{ij} is 1 if $i = j$ and 0 otherwise. In (5), the term $s_i^2 = (\mathbf{w}_i'\mathbf{Y})^2$ is called the ith component of the statistic. As $N \to \infty$, the distribution of $\mathbf{Y}$ tends to the multivariate normal with mean 0 and variance $\boldsymbol{\Sigma}_y$. The distribution of a typical s_i tends to univariate normal, mean 0, variance 1 and in the limit the s_i are independent. The limiting distribution of S is thus that of S_∞ where

$$S_\infty = \sum_{i=1}^{k-1} \lambda_i u_i^2, \tag{6}$$

and where the u_i are independent weighted χ_1^2 variables.

In order to find the λ_i we need $\boldsymbol{\Sigma}_y$. This is found as follows. Calculate the $k \times k$ matrix

$$\boldsymbol{\Sigma}_0 = \mathbf{D} - \mathbf{p}\mathbf{p}^\top; \tag{7}$$

this is the covariance matrix of $(\mathbf{o} - \mathbf{e})/\sqrt{N}$. Then $\boldsymbol{\Sigma}_y = \mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}^\top$, with entries $\Sigma_{ij} = \min(H_i, H_j) - H_i H_j$. This is the covariance matrix of $\mathbf{Y} = \mathbf{Z}/\sqrt{N} = \mathbf{A}\mathbf{d}/\sqrt{N}$.

For the appropriate $\mathbf{M}$ for the statistic required, the eigenvalues λ_i, $i = 1, \ldots, k$ of $\mathbf{M}\boldsymbol{\Sigma}_y$ are then used in (6) to obtain the limiting distribution of the statistic.

It was pointed out in Choulakian, Lockhart & Stephens (1994) that one can work with $\mathbf{X} = \mathbf{M}^{1/2}\mathbf{Y}$; the covariance of $\mathbf{X}$ is then $\boldsymbol{\Sigma}_X = \mathbf{M}^{1/2}\boldsymbol{\Sigma}_y\mathbf{M}^{1/2}$. It may be shown that the eigenvalues λ_i of $\boldsymbol{\Sigma}_X$ are the same as those of $\mathbf{M}\boldsymbol{\Sigma}_y$; the advantage of using $\boldsymbol{\Sigma}_X$ is that it is symmetric, which may be useful when using certain programmes to find eigenvalues.

2.2. Parameters unknown.

In this section the above theory will be extended to the case where the tested distribution contains unknown parameters θ_i. Let $\theta = (\theta_1, \ldots, \theta_m)^\top$ be the vector of m parameters.

In much of what follows we use results given by Bishop, Fienberg & Holland (1975). The parameters must be estimated efficiently, for example by maximizing the multinomial likelihood (MML) obtained from the multinomial distribution of the o_i, $i = 1, \ldots, k$. The log-likelihood is (omitting irrelevant constants)

$$L^* = \sum_{i=1}^{k} o_i \log p_i,$$

and p_i contains the unknown parameters. The MML estimation consists of solving the m equations

$$\frac{\partial L^*}{\partial \theta_j} = \sum_{i=1}^{k} \frac{o_i}{p_i} \frac{\partial p_i}{\partial \theta_j} = 0,$$

for $j = 1, \ldots, m$.

Let $\hat{\theta}$ be the MML estimate of θ, let $\hat{\mathbf{p}}$ be the estimate of $\mathbf{p}$, evaluated using $\hat{\theta}$, and let $\hat{\mathbf{e}}$ be the estimated vector of expected values in the cells, with components $\hat{e}_j = N\hat{p}_j$. Then let $\hat{\mathbf{d}} = \mathbf{o} - \hat{\mathbf{e}}$ and $\widehat{\mathbf{Z}} = \mathbf{A}\hat{\mathbf{d}}$.

Define a k by m matrix $\mathbf{B}$ with entries

$$\mathbf{B}_{i,j} = \partial p_i / \partial \theta_j$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, m$. The matrix $\mathbf{B}^\top \mathbf{D}^{-1} \mathbf{B}$ is the Fisher information matrix. Define $\mathbf{V} = (\mathbf{B}^\top \mathbf{D}^{-1} \mathbf{B})^{-1}$. The asymptotic covariance of $\hat{\theta}$ is then $\mathbf{V}/N$, the covariance of $\hat{\mathbf{d}}/\sqrt{N}$ is $\mathbf{\Sigma}_d = \mathbf{\Sigma}_0 - \mathbf{B}\mathbf{V}\mathbf{B}^\top$, where $\mathbf{\Sigma}_0$ is defined in (7), and the covariance of $\widehat{\mathbf{Z}}/\sqrt{N} = \mathbf{A}\hat{\mathbf{d}}/\sqrt{N} = \widehat{\mathbf{Y}}$ is

$$\mathbf{\Sigma}_u = \mathbf{A}\mathbf{\Sigma}_d \mathbf{A}^\top.$$

Then, as in the previous section, where parameters were known, the weights λ_i in the asymptotic distribution (6) are the k eigenvalues of $\mathbf{M}\mathbf{\Sigma}_u$ for the appropriate $\mathbf{M}$ for the statistic required. Again the transformation $\mathbf{X} = \mathbf{M}^{1/2}\mathbf{Y}$ may be made and the eigenvalues will be those of the symmetric matrix $\mathbf{M}^{1/2}\mathbf{\Sigma}_u\mathbf{M}^{1/2}$.

In practice, in order to calculate the statistics, using (1)–(4), the various vectors and matrices must be replaced by their estimates where necessary. For example, let matrix $\widehat{\mathbf{D}}$ be $\mathbf{D}$ with $\mathbf{p}$ replaced by $\hat{\mathbf{p}}$ and similarly obtain $\widehat{\mathbf{B}}, \widehat{\mathbf{E}}, \widehat{\mathbf{V}}, \widehat{\mathbf{K}}$ and $\widehat{\mathbf{\Sigma}}_0$ using estimates in an obvious way. The eigenvalues will also be found using the estimated matrices $\widehat{\mathbf{\Sigma}}_u$ and $\widehat{\mathbf{M}}$. Consistent estimates of the λ_i will be obtained and (6) used to find the estimated asymptotic distribution.

Thus the steps are:

1. Calculate $\widehat{\mathbf{V}} = (\widehat{\mathbf{B}}^\top \widehat{\mathbf{D}}^{-1} \widehat{\mathbf{B}})^{-1}$.
2. Calculate $\widehat{\mathbf{\Sigma}}_d = \widehat{\mathbf{\Sigma}}_0 - \widehat{\mathbf{B}}\widehat{\mathbf{V}}\widehat{\mathbf{B}}^\top$ and $\widehat{\mathbf{\Sigma}}_u = \mathbf{A}\widehat{\mathbf{\Sigma}}_d\mathbf{A}^\top$.
3. For the statistic required, let $\widehat{\mathbf{M}}$ be the estimate of the appropriate $\mathbf{M}$. Find the k eigenvalues of $\widehat{\mathbf{M}}\widehat{\mathbf{\Sigma}}_u$, or those of the symmetric matrix $\widehat{\mathbf{M}}^{1/2}\widehat{\mathbf{\Sigma}}_u\widehat{\mathbf{M}}^{1/2}$ and use them in (6) to obtain the asymptotic distribution.

2.3. Percentage points.

Percentage points of S_∞, using exact or estimated λs, can be found to high accuracy by the method of Imhof (1961). However, for practical purposes, they can be well approximated in the upper tail by the percentage points of $S1$, where $S1$ has the distribution $a + b\chi^2_p$, and the a, b, p are chosen so that the first three cumulants of $S1$ match those of the asymptotic distribution of S in (6). The cumulants of the distribution in (6) are $\kappa_j = 2^{j-1}(j-1)! \sum_{i=1}^{k-1} \lambda_i^j$. In particular, the mean κ_1 is $\sum_{i=1}^{k-1} \lambda_i$, the variance κ_2 is $\sum_{i=1}^{k-1} 2\lambda_i^2$ and κ_3 is $8\sum_{i=1}^{k-1} \lambda_i^3$. Then for the $S1$ approximation, $b = \kappa_3/(4\kappa_2)$, $p = 8\kappa_2^3/\kappa_3^2$, and $a = \kappa_1 - bp$. We have found this approximation to be accurate in the upper tail, at levels $\alpha < 0.15$, but the accuracy falls off at the 0.25 and 0.50 levels. More accurate approximations have been given by Solomon & Stephens (1977).

2.4. Long-tailed distributions.

Some distributions may have long tails with an infinite set of cells; for example, a long tail to the right, with probabilities diminishing as the cell index i increases. Of course, in practice, there is a finite set of cells with data. Suppose $i = k^*$ is the largest cell index with data in the cell. A practical procedure is then to do the above analysis with, say, $k = k^* + 20$ cells, and repeat with, say, $k = k^* + 50$ cells; these numbers should be increased until the percentage points agree to the desired accuracy. For such a distribution one would be unlikely to accumulate from the right, and a statistician may prefer the original definitions (see Choulakian, Lockhart & Stephens 1994) with p_j replacing t_j. The asymptotic theory goes through as above with p_j replacing t_j in all the definitions. See, for example, Spinelli & Stephens (1997).

3. NUMERICAL CALCULATIONS

3.1. Checks on the λ calculations.

For the X^2 statistic, when parameters are known, the eigenvalues of $\mathbf{D}^{-1}\boldsymbol{\Sigma}_0$ should be $k-1$ values equal to one, and one zero. This will give the well-known result that the asymptotic distribution is χ^2_{k-1}. When parameters are estimated by maximizing the multinomial likelihood, the asymptotic distribution is χ^2_{k-m-1} so there should be $k-m-1$ ones and $m+1$ zeros as eigenvalues of $\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{\Sigma}}_d$.

For the Cramér–von Mises statistics, the sum of the eigenvalues in (6) will be the limiting expected value of the appropriate statistic. When parameters are known the expected values of the statistics can be calculated exactly using the multinomial distribution of $\mathbf{o}$. They are as follows:

$$\begin{aligned}
\mathrm{E}(W_d^2) &= \sum_{j=1}^{k} t_j H_j(1-H_j), \\
\mathrm{E}(U_d^2) &= \mathrm{E}(W_d^2) - \sum_{i=1}^{k}\sum_{j=1}^{k} t_i t_j \{\min(H_i, H_j) - H_i H_j\}, \\
\mathrm{E}(A_d^2) &= 1 - t_k.
\end{aligned}$$

These provide a good check on the eigenvalues. When parameters have to be estimated, the limiting expected values are given by these formulas minus the trace of $\mathbf{MBVB}^{\top}$. This quantity depends on the distribution tested and cannot usually be put in a simple form.

4. EXAMPLES

We illustrate the difference in the calculations for the two cases (parameters known or unknown) by a simple example. Suppose there are $k = 10$ cells, and it is desired to test $\mathcal{H}_0$: the (linear) probabilities are $p_i = 0.1 + b(i - 5.5)$, $-1/45 < b < 1/45$. This form ensures that the sum is 1 for all b.

Example 1 (Parameter known): Suppose first that b is known to be 0.02, giving cell probabilities $p_i = 0.01$, 0.03, 0.05, $\ldots$, 0.19, and suppose the observed 50 values give cell counts $1, 3, 6, 2, 9, 3, 4, 6, 7, 9$. The Kolmogorov–Smirnov statistics are $D_d^+ = 1.202$, $D_d^- = 0.000$, $D_d = 1.202$. The Cramér–von Mises statistics are $W_d^2 = 0.344$, $U_d^2 = 0.138$, $A_d^2 = 2.071$, and Pearson's $X^2 = 14.732$.

Since the tested probabilities are known, we need the eigenvalues discussed in Section 2.1, for the appropriate statistic. For W^2, U^2 and A^2, the eigenvalues are given in Table 1 and upper tail percentage points of (6) are in Table 2.

The P-value for statistic W_d^2 is 0.11, for U_d^2 is 0.16, and for A_d^2 is 0.07; for Pearson's X^2, the P-value is 0.10.

Example 2 (Parameter estimated): Now suppose the given data (observed values in the cells) are the same as above, but the linear probabilities model will be fitted, with the value of b estimated by MML. This value is $\hat{b} = 0.0128444$ and the corresponding probabilities are

$$0.0422, 0.0550, 0.0679, 0.0807, 0.0936,$$

$$0.1064, 0.1193, 0.1321, 0.1450, 0.1578.$$

Then Pearson's $X^2 = 9.499$, to be compared with the χ^2_8 distribution; the P-value is greater than 0.5. The Cramér–von Mises statistics are $W_d^2 = 0.052$, $U_d^2 = 0.050$ and $A_d^2 = 0.284$, and the Kolmogorov–Smirnov statistics are $D_d^+ = 0.570$, $D_d^- = 0.157$ and $D_d = 0.570$.

TABLE 1: Eigenvalues for Cramér–von Mises statistics.

Example 1: W_d^2	0.1030	0.0271	0.0132	0.0083	0.0056
	0.0035	0.0019	0.0008	0.0001	0.0000
	sum	0.1634			
Example 1: U_d^2	0.0280	0.0262	0.0093	0.0076	0.0049
	0.0031	0.0017	0.0007	0.0001	0.0000
	sum	0.0817			
Example 1: A_d^2	0.5000	0.1667	0.0833	0.0500	0.0333
	0.0238	0.0179	0.0139	0.0111	0.0000
	sum	0.9000			
Example 2: W_d^2	0.0456	0.0173	0.0092	0.0062	0.0045
	0.0032	0.0020	0.0011	0.0000	0.0000
Example 2: U_d^2	0.0271	0.0169	0.0080	0.0061	0.0043
	0.0030	0.0019	0.0010	0.0000	0.0000
Example 2: A_d^2	0.2403	0.0959	0.0531	0.0342	0.0240
	0.0179	0.0139	0.0111	0.0000	0.0000

Then we follow the steps in Section 2.2. For W_d^2, U_d^2 and A_d^2 the eigenvalues are given in Table 1 and percentage points in Table 2.

The significance levels of all four statistics W_d^2, U_d^2, A_d^2, and X^2 are now greater than 0.5. These higher P-values demonstrate the common phenomenon in testing fit, that estimation of the parameters generally gives a better fit when a model with fixed parameters is marginal as in Example 1.

TABLE 2: Asymptotic percentage points for Examples 1 and 2.

	α					
	0.500	0.250	0.100	0.050	0.025	0.010
Example 1: W_d^2	0.1155	0.2083	0.3483	0.4642	0.5853	0.7514
Example 1: U_d^2	0.0671	0.1063	0.1564	0.1941	0.2317	0.2815
Example 1: A_d^2	0.6737	1.1473	1.8325	2.3919	2.9771	3.7778
Example 2: W_d^2	0.0689	0.1139	0.1770	0.2279	0.2811	0.3538
Example 2: U_d^2	0.0562	0.0883	0.1299	0.1617	0.1940	0.2375
Example 2: A_d^2	0.3850	0.6267	0.9614	1.2301	1.5101	1.8933

5. CONVERGENCE TO ASYMPTOTIC POINTS

In Section 2 we have given the calculations to obtain asymptotic points for the Cramér–von Mises statistics W_d^2, U_d^2 and A_d^2. It is known that in the continuous case the points for finite n converge rapidly to the asymptotic, so that these may be used for n as low as 20. However, the Kolmogorov–Smirnov D_d does not converge so quickly. For the discrete analogues, we have examined the convergence by taking 10000 Monte Carlo (MC) samples from various $\mathbf{p}$ vectors and for sample sizes 25, 50, 100, 200, and 500. Tables showing these studies based on probabilities in Examples 1 and 2 are included in a research report, obtainable from the first author. Comments on these studies are as follows.

TABLE 3: Eigenvalues and percentage points for Cramér–von Mises statistics: Uniform case.

	Eigenvalues				
W_d^2	0.1022	0.0262	0.0121	0.0072	0.0050
	0.0038	0.0031	0.0027	0.0026	0.0000
	sum	0.1650			
U_d^2	0.0262	0.0262	0.0072	0.0072	0.0038
	0.0038	0.0028	0.0028	0.0025	0.0000
	sum	0.0825			
A_d^2	0.5000	0.1667	0.0833	0.0500	0.0333
	0.0238	0.0179	0.0139	0.0111	0.0000
	sum	0.9000			

	Percentage Points α				
	0.250	0.100	0.050	0.025	0.010
W_d^2	0.2090	0.3480	0.4629	0.5830	0.7484
U_d^2	0.1060	0.1542	0.1905	0.2268	0.2748
A_d^2	1.1473	1.8325	2.3919	2.9771	3.7778

The percentage points for the Cramér–von Mises again converge to the asymptotic points very quickly and the asymptotic points can certainly be used with good accuracy for sample sizes greater than 25.

The statistic D_d was studied by Pettitt & Stephens (1977) for the case when the cell probabilities are completely specified as in Example 1. Our Monte Carlo studies confirm that D_d then has a very discrete distribution with few distinct values in the upper tail. This is because for large values of D_d, many configurations of o_i can give the same statistic. Thus it is difficult to achieve a test of exact size α. When parameters must be estimated, there will be many patterns of p_i and so the distribution takes many more values. However, statistic D_d is known not to have as good power as the Cramér–von Mises statistics, so we shall not consider this further.

Finally, although for small sample sizes such as 25 and 50, the expected numbers in the cells do not conform to the generally assumed rules (e.g., that e_i should be nearly always greater than 5) necessary to obtain convergence of Pearson's X^2 to the χ^2 distribution, statistic X^2 also converges very well.

6. POWER

In this section we give a small power study. The null hypothesis is that the distribution is the discrete uniform with 10 cells so that $p_i = 0.1$, $i = 1, \ldots, 10$. On the alternative, the cell probabilities are $p_i = 0.1 + b(i - 5.5)$ as in Example 1. The test size is $\alpha = 0.10$.

For the power study, the asymptotic percentage points for the uniform distribution with 10 cells are given in Table 3. The eigenvalues λ_i are also given, for completeness. An interesting result from Choulakian, Lockhart & Stephens (1994) is that, for A^2, the $k-1$ eigenvalues are exactly the first $k-1$ values in the continuous case.

Figure 1 shows the power of W_d^2, U_d^2, A_d^2, and X^2 for sample size 25, as b moves from 0.00 to 0.02. Figure 2 gives similar plots for sample size 100.

The figures demonstrate that the Cramér–von Mises statistics are more powerful than X^2 when the probabilities in the cells are in a steadily increasing pattern. The results are similar for $b < 0$ when the probabilities decrease. These patterns of probability, compared with the null, are quite common so that the Cramér–von Mises statistics, especially W_d^2 and A_d^2, can be recommended for testing fit.

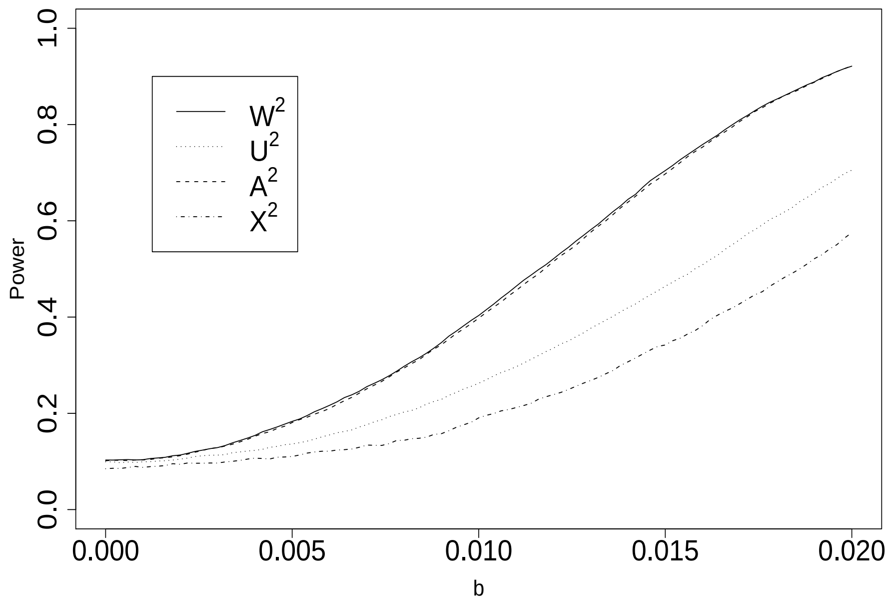

FIGURE 1: Power results for W_d^2, U_d^2, A_d^2, and X^2: test size 0.1, sample size 25.

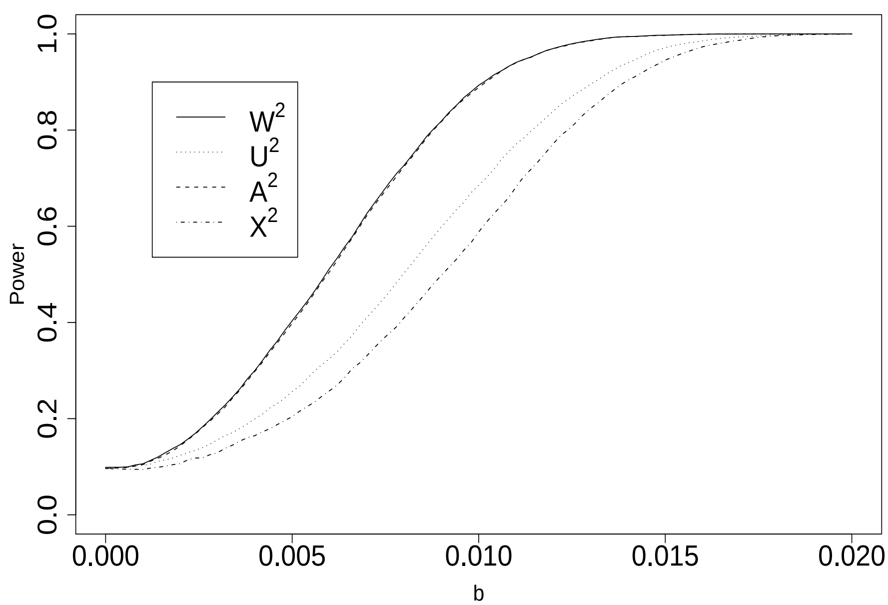

FIGURE 2: Power results for W_d^2, U_d^2, A_d^2, and X^2: test size 0.1, sample size 100.

7. SUMMARY

In this article we have defined statistics of the Cramér–von Mises type for testing fit to a discrete distribution. It is shown how to obtain asymptotic percentage points for the statistics, both when the distribution is completely specified, or when unknown parameters must be estimated from the data. An example is discussed, for both cases. Monte Carlo studies suggest that the asymptotic distributions may be used in practice for finite samples of reasonable size. A small power study (testing the discrete uniform distribution) is included.

ACKNOWLEDGEMENTS

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to this agency, and for helpful comments of the Editor, Associate Editor and referees.

REFERENCES

Y. M. Bishop, S. E. Fienberg & P. W. Holland (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Massachusetts.

V. Choulakian, R. A. Lockhart & M. A. Stephens (1994). Cramér–von Mises statistics for discrete distributions. *The Canadian Journal of Statistics*, 22, 125–137.

J. P. Imhof (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419–426.

A. N. Pettitt & M. A. Stephens (1977). The Kolmogorov–Smirnov goodness-of-fit statistic with discrete and grouped data. *Technometrics*, 19, 205–210.

H. Solomon & M. A. Stephens (1977). The distribution of a sum of weighted chi-square variables. *Journal of the American Statistical Association*, 72, 881–885.

J. J. Spinelli (2001). Testing fit for the grouped exponential distribution. *The Canadian Journal of Statistics*, 29, 451–458.

J. J. Spinelli & M. A. Stephens (1997). Cramér–von Mises tests of fit for the Poisson distribution. *The Canadian Journal of Statistics*, 25, 257–268.

Received 29 November 2005
Accepted 31 July 2006

Richard A. LOCKHART: lockhart@stat.sfu.ca
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, British Columbia
Canada V5A 1S6

John J. SPINELLI: jspinelli@bccrc.ca
British Columbia Cancer Research Centre
Vancouver, British Columbia
Canada V5Z 1L3

Michael A. STEPHENS: stephens@stat.sfu.ca
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, British Columbia
Canada V5A 1S6

Statistical Society of Canada
Société Statistique du Canada