Use of the Kolmogorov–Smirnov, Cramér–Von Mises and Related Statistics without Extensive Tables

By M. A. Stephens

McGill University, Montreal, Canada

[Received March 1969]

Summary

This paper gives modifications of eleven statistics, usually used for goodness of fit, so as to dispense with the usual tables of percentage points. Some test situations are illustrated, and formulae given for calculating significance levels.

1. Introduction

Various statistical problems reduce to testing the null hypothesis H_0 that a random sample $y_1, y_2, \ldots, y_n$ comes from a uniform distribution with limits 0, 1, written $U(0,1)$. Natural test statistics are the Kolmogorov–Smirnov (KS) or Cramér–von Mises (CvM) types, or the statistics, here called "periodogram statistics", based on the deviation of the ith order statistic of the y set from its expected value. In the literature, extensive tables exist giving the distributions, on H_0, or the percentage points, of most of these statistics. The purpose of this paper is to show how these tables may be replaced, for most practical test purposes, by a short table of percentage points. For each test statistic, T say, a simple modification T^* is given, and T^* is compared with the given percentage points. In effect, formulae are given from which the true null percentage points of T, for a sample of size n, may be very closely approximated; if the approximate point is calculated for significance level α, and if α' is the true level attained, the error $|\alpha' - \alpha|$ is negligibly small. We also give expressions to calculate the significance level of a given value of T, on H_0, provided the value lies in one of the tails. The techniques should be useful if a computer is to be used to give either the result of a test or the significance level of T. The statistics discussed are those usually known by the symbols D_n^+, D_n^-, D_n, V_n, W_n^2, U_n^2, C_n^+, C_n^-, C_n, K_n, and A_n; these are defined below, but from now on we shall omit the subscript n. When a statistic has been modified on the lines to be given in the paper, we add an asterisk; for example D becomes D^*, W^2 becomes W^{2*}, etc.

The test of H_0 often arises from the use of a random sample of x_i to test a null hypothesis H_1 concerning a random variable x; by suitable transformations, the y_i are found from the x_i, and, if H_1 is true, the y_i will be $U(0,1)$. Then H_0 is tested concerning the y_i.

In the next two sections we give four examples of this situation: the examples are used to introduce and illustrate the various statistics considered. The first two illustrations, in Section 2, are essentially goodness-of-fit tests, and introduce the KS and CvM statistics and their extensions for points on a circle; the second pair of examples, in Section 3, which arise with time series periodogram analysis and regression analysis, lead more naturally to the C^+, C, and K statistics, which we have called the "periodogram statistics". For each of these two groups, we give (a) the computational forms of the usual test statistics; (b) the modified statistics, in Tables 1 and 3, together with the set of significance points to be used with the modified form; (c) simple expressions to calculate the significance level of a given value of T, in the upper tail (Table 2).

After the computational forms, references are given. These have been restricted, for the older statistics, primarily to tables of exact percentage points which have been used to check the accuracy of the approximations, but for more recent statistics, references have also been included to give explanatory material or the original definitions. For the older statistics, the references given should be used as sources for the original definitions and the distribution theory. In Section 4 we give some observations on the relative properties of the various test statistics; some of those which were originally designed for observations on a circle will be found to have greater power on a line for some alternatives than, perhaps, the well-established statistic D. The accuracy of the approximations given by using the modified forms in this paper is discussed in Section 5.

2. Test Statistics and Modifications: KS and CvM Types

2.1. *Testing Goodness of Fit*

We start with two illustrations of H_0 arising from an original null hypothesis H_1, both connected with goodness-of-fit testing.

Example 1

Suppose H_1 is that a random sample $x_1, x_2, ..., x_n$ comes from a given distribution $F(x)$; the y_i are calculated from $y_i = F(x_i)$, $i = 1, 2, ..., n$, and the original H_1 now reduces to testing H_0 concerning the y_i.

Example 2

A second example is in testing H_1, that a random sample $x_1, x_2, ..., x_n$ comes from an exponential distribution, density $f(x) = \theta e^{-\theta x}$, $x \geqslant 0$, but θ not known. The goodness-of-fit test just above cannot therefore be applied; but two transformations may be used to produce uniform observations. To show these, first let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be the order statistics of the sample, and let $x_{(0)}$ be 0. Let $d_r = (n+1-r)(x_{(r)} - x_{(r-1)})$, for $r = 1, 2, ..., n$. The two transformations to y are then, each for $i = 1, 2, ..., n-1$,

$$J: \quad y_i = \sum_{r=1}^{i} x_r / S,$$

and

$$K: \quad y_i = \sum_{r=1}^{i} d_r / S,$$

where

$$S = \sum_{r=1}^{n} x_r = \sum_{r=1}^{n} d_r.$$

Then either J or K produce a set y_i ($i = 1, 2, ..., n-1$), which are $U(0, 1)$. Thus H_0 is tested using the y_i.

For tests of goodness of fit, the usual statistics measure in some way the discrepancy between the empirical distribution function of the x_i and the theoretical $F(x)$, and are chosen not to depend on the particular $F(x)$; these are the KS and CvM statistics. Example 2 above has been chosen to emphasize that the lower tail of the test statistic might have to be used in the test; this is because it is quite possible, using transformation J, to produce, if the x_i are *not* exponential, a set y_i which is *superuniform*, that is, more evenly spaced that one would expect on H_0 (Seshadri *et al.*, 1969). When this occurs the test statistics take very small values. Another example of superuniform observations, occurring in a time sequence, is given by Pearson (1963).

2.2. *Computational Formulae*

Computational formulae for the KS and CvM statistics are given below, together with those for U^2, V and A, statistics introduced by Watson, Kuiper and Ajne. They were designed for points on a circle, but may also be used for points on a line. A comment on their expected powers is in Section 4. For all the statistics, suppose y_i $(i = 1, 2, ..., n)$ are in ascending order, and let $\bar{y}$ be the mean of the y_i.

Kolmogorov–Smirnov statistics

These are calculated as follows:

$$D^+ = \max_{1 \leqslant i \leqslant n} (in^{-1} - y_i), \quad D^- = \max_{1 \leqslant i \leqslant n} \{y_i - n^{-1}(i-1)\},$$

$$D = \max(D^+, D^-), \quad V = D^+ + D^-.$$

References: D^+, D^-, D: Owen (1962).

V: Kuiper (1962), Stephens (1965).

Cramer–von Mises statistics

These are calculated from

$$W^2 = \sum_{i=1}^{n} \left(y_i - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}, \quad U^2 = W^2 - n(\bar{y} - \tfrac{1}{2})^2,$$

$$A = \frac{n}{4} - \frac{2}{n} \sum_{j=2}^{n} \sum_{i=1}^{j-1} m_{ij},$$

with $m_{ij} = y_j - y_i$ if $y_j - y_i \leqslant \frac{1}{2}$ and $m_{ij} = 1 - (y_j - y_i)$ if $y_j - y_i > \frac{1}{2}$.

References: W^2: Pearson and Stephens (1962), Stephens and Maag (1968).

U^2: Watson (1961), Stephens (1964).

A: Ajne (1968), Stephens (1969a).

2.3. *Modifications and Approximations*

Table 1 gives the proposed modifications for the KS and CvM group. The steps in a test of H_0, using any typical statistic T, are then as follows:

(a) Calculate the original statistic T using the formulae above.

(b) Corresponding to T in column 1 of Table 1, calculate the modified form T^* in Column 2.

(c) If T^* exceeds a value in column 3, for level α, reject H_0 at significance level α. This assumes the usual upper tail test; obvious modifications are made for the lower tail or for a two-tailed test.

The points in column 3 are the asymptotic percentage points for $D\sqrt{n}$, $V\sqrt{n}$, W^2, U^2 and A.

For a given T, calculate T^*; then the significance level *in the upper tail* is calculated as in Table 2 for the various statistics. Each formula (except for W^2) is the first term of the relevant asymptotic distribution of T^*; values of z and α are given for which this first term gives two-decimal or three-decimal accuracy. The asymptotic distribution of W^2 (Anderson and Darling, 1952) is too difficult even to use the first term; the expression given is empirically derived, and gives an error less than 0·002 for α between 0·15 and 0·05, and less than 0·001 for α between 0·05 and 0·01; for $\alpha = 0{\cdot}01$ the formula gives 0·0094 and for $\alpha = 0{\cdot}001$ it gives 0·0007.

TABLE 1

Modified Kolmogorov–Smirnov, Cramér–von Mises and related statistics

Statistic T	*Modified form T**	*Percentage points for T**				
		15·0%	10·0%	5·0%	2·5%	1·0%
	Upper tail					
D^+	$D^+(\sqrt{n}+0{\cdot}12+0{\cdot}11/\sqrt{n})$	0·973	1·073	1·224	1·358	1·518
D^-	$D^-(\sqrt{n}+0{\cdot}12+0{\cdot}11/\sqrt{n})$	0·973	1·073	1·224	1·358	1·518
D	$D(\sqrt{n}+0{\cdot}12+0{\cdot}11/\sqrt{n})$	1·138	1·224	1·358	1·480	1·628
V	$V(\sqrt{n}+0{\cdot}155+0{\cdot}24/\sqrt{n})$	1·537	1·620	1·747	1·862	2·001
W^2	$(W^2-0{\cdot}4/n+0{\cdot}6/n^2)(1{\cdot}0+1{\cdot}0/n)$	0·284	0·347	0·461	0·581	0·743
U^2	$(U^2-0{\cdot}1/n+0{\cdot}1/n^2)(1{\cdot}0+0{\cdot}8/n)$	0·131	0·152	0·187	0·221	0·267
A	$(A-0{\cdot}7/n+0{\cdot}9/n^2)(1{\cdot}0+1{\cdot}23/n)$	0·431	0·516	0·656	0·797	0·982
	Lower tail					
D	$D(\sqrt{n}+0{\cdot}275-0{\cdot}04/\sqrt{n})$	0·610	0·571	0·520	0·481	0·441
V	$V(\sqrt{n}+0{\cdot}41-0{\cdot}26/\sqrt{n})$	0·976	0·928	0·861	0·810	0·755
W^2	$(W^2-0{\cdot}03/n)(1{\cdot}0+0{\cdot}5/n)$	0·054	0·046	0·037	0·030	0·025
U^2	$(U^2-0{\cdot}02/n)(1+0{\cdot}35/n)$	0·038	0·033	0·028	0·024	0·020
A	$(A-0{\cdot}01/n)(1{\cdot}0-0{\cdot}2/n)$	0·079	0·065	0·050	0·040	0·032

TABLE 2

Formulae for the upper tail significance level

Test statistic	$\alpha(z)$	z_2	α_2	z_3	α_3
D^+, D^-	$\exp(-2z^2)$	0·82	0·26	1·00	0·135
D	$2\exp(-2z^2)$	0·91	0·38	1·08	0·194
V	$(8z^2-2)\exp(-2z^2)$	1·06	0·74	1·26	0·447
W^2	$0{\cdot}05\exp(2{\cdot}79-6z)$		See Section 2		
U^2	$2\exp(-2z^2\pi^2)$	0·29	0·26	0·34	0·135
A	$1{\cdot}273\exp(-\pi^2 z/2)$	0·11	0·74	0·21	0·452

For a given value z of the modified test statistic T^*, the significance level α in the upper tail (i.e. $P(T^*>z)=\alpha$) is given approximately by $\alpha(z)$; z_2 and z_3 are respectively the approximate values beyond which the formula is accurate to two decimal places and three decimal places; the corresponding values of α are then α_2 and α_3.

3. PERIODOGRAM STATISTICS

Example 3. *Time-series analysis*

Suppose a test is to be made of H_1: a given time series consists of x – values which are serially independent. Let the periodogram ordinates, based on a sample of

S observations, be p_r, $r = 1,2,...,m$; $m = [\frac{1}{2}S]$. Let

$$P = \sum_{r=1}^{m} p_r \quad \text{and} \quad y_i = \sum_{r=1}^{i} p_r/P,$$

$i = 1,2,...,m$. Put $n = m-1$. Then, if H_1 is true, the y_i $(i = 1,2,...,n)$ are $U(0,1)$. (y_m, that is, y_{n+1}, is identically 1.) For such a problem, it is considered more natural to consider the ith order statistic of the y set, (this will be y_i itself) and to base a test on the maximum deviation, for all i, of y_i from its expected value $i/(n+1)$ (Durbin, 1969).

Thus test statistics are

$$C^+ = \max_{1 \leqslant i \leqslant n+1} (im^{-1} - y_i), \quad C^- = \max_{1 \leqslant i \leqslant n+1} (y_i - im^{-1}),$$

$$C = \max(C^+, C^-), \quad K = C^+ + C^-.$$

C^+, C^-, C (usually written with subscript n) are also called respectively c^-, c^+, c, with no subscripts.

References: C^+, C^-, C: Durbin (1969).

K: Brunk (1962), Stephens (1969b).

Example 4. *Regression analysis*

Suppose a hyperplane has been fitted to the first $r-1$ observations in a usual regression model, and let the residual of the rth observation from the hyperplane be called W_r; the standardized cumulated sum of squares of the W_r will yield uniform observations, assuming the null hypothesis H_1, that the usual model holds with normal independent errors of constant variance. For further details of this application, for which the authors also recommend the statistics C, and for earlier references to these statistics, see Brown and Durbin (1968).

TABLE 3

Modified periodogram statistics

Statistic T	*Modified form* T^*	*Percentage points for* T^*				
		15·0%	0·0%	5·0%	2·5%	1·0%
	Upper tail					
C^+	$(C^+ + 0{\cdot}4/n)(\sqrt{n} + 0{\cdot}2 + 0{\cdot}68/\sqrt{n})$	0·973	1·073	1·224	1·358	1·518
C^-	$(C^- + 0{\cdot}4/n)(\sqrt{n} + 0{\cdot}2 + 0{\cdot}68/\sqrt{n})$	0·973	1·073	1·224	1·358	1·518
C	$(C + 0{\cdot}4/n)(\sqrt{n} + 0{\cdot}2 + 0{\cdot}68/\sqrt{n})$	1·138	1·224	1·358	1·480	1·628
K	$\{K - 1/(n+1)\}\{\sqrt{(n+1)} + 0{\cdot}1555 + 0{\cdot}24/\sqrt{(n+1)}\}$	1·537	1·620	1·747	1·862	2·001
	Lower tail					
C	$(C + 0{\cdot}5/n)(\sqrt{n} + 0{\cdot}44 - 0{\cdot}32/\sqrt{n})$	0·610	0·571	0·520	0·481	0·441
K	$\{K - 1/(n+1)\}\{\sqrt{(n+1)} + 0{\cdot}41 - 0{\cdot}26/\sqrt{(n+1)}\}$	0·976	0·928	0·861	0·810	0·755

Modifications for C and K are in Table 3; the table is to be used in the same way as Table 1. For an approximation to the upper tail significance level for a given value of C, calculate C^* and insert in the formula for D, Table 2; for K, calculate K^* and use the formula for V, Table 2.

4. Properties of the test Statistics

Large values of the test statistics will usually be regarded as significant. D^+ will tend to detect a movement of the y_i towards 0, D^- towards 1; D and W^2 will detect a shift of the mean in either direction. V and U^2, when used for points on a line, roughly detect a change in variance rather than in mean, either because the y_i have clustered together or moved into two groups towards 0 and 1. In goodness-of-fit tests of the type in Example 1, these properties may be readily interpreted in terms of changes of mean or of variance of $F(x)$; this is more difficult if the y_i have resulted from a complicated transformation of the original x_i. In many situations, D and W^2 seem to find almost the same samples significant, and similarly V and U^2 (Seshadri *et al.*, 1969; Stephens, 1969a).

For points on a circle, V, U^2 or A must be calculated; the other statistics take different values with different choices of origin of x (and hence of y).

Small values of the test statistics will suggest superuniformity, as discussed in Example 2. This could occur unexpectedly whenever the original random variables x_i undergo a complicated transformation to produce the final y_i. The need to test for superuniform observations has not been generally emphasized; for example, significance points in the lower tail of D have only just been calculated by this author.

5. Comments on the Modifications

The modifications give, in effect, approximate values for the percentage points of a typical statistic T, by solving $T^* = z(\alpha)$, where $z(\alpha)$ is the asymptotic percentage point at level α. These approximations came about because it was noted for D, and then for V, that the percentage points for $\alpha = 0{\cdot}10$, $0{\cdot}05$, $0{\cdot}025$, and $0{\cdot}01$ were in proportions which remained remarkably constant for all n. Thus if $D_n(\alpha)$ is the α-level point for D at sample size n, it could be possible to derive an approximation for $D_n(\alpha)$, calculated from its asymptotic point by means of a factor dependent only on n: thereafter the best expression was found by trial.

For K, the ratios of percentage points were not constant and the relation $P\{K_n < z\} = P\{V_{n+1} < z + 1/(n+1)\}$, given by Stephens (1969b), both explains why this is to be expected (if the points for V are in constant ratios) and also shows how to find the modification K^* in Table 3. It also suggested the approach followed for several of the other statistics, when the ratios of percentage points are not stable; first a correction factor is added to give more stable ratios, and then the multiplication factor is provided. Thus the addition of $-0{\cdot}7/n + 0{\cdot}9/n^2$ to all the upper tail points for A, for a given n, will give ratios of new points close to those of the asymptotic percentage points; then the factor $1 + 1{\cdot}23/n$ is found to give the necessary approximation.

The ratios of the exact points are not, of course, perfectly steady even after modification on the lines above; variations of the constants in the expressions for T^* will therefore give greater accuracy for $\alpha = 0{\cdot}01$, say, at the expense of accuracy for $\alpha = 0{\cdot}10$. The constants were chosen to give best results for α near $0{\cdot}05$, and for $n \geqslant 8$. It should be noted that the "exact" points of several statistics, for example W^2, U^2 and A, have been themselves derived by various approximate methods, so that there is no point in searching for too great apparent accuracy.

The accuracy of the final choice is shown in Table 4; this gives values of the true significance level α' attained if the modified form is used at level α. The α' values are found by first solving $T^* = z(\alpha)$, where $z(\alpha)$ is the asymptotic point at level α, to give the approximate value T^* of T; the α' for this value is then estimated from tables of exact points or probabilities. The values $n = 4$ are included to show that the approximations are surprisingly good for very low values of n, though such small samples are not realistic in practice. The formulae given should certainly make the proposed tests of practical value for $n \geq 8$.

TABLE 4

Values of α' using modified statistics at nominal level α

U.T. = upper tail L.T. = lower tail

Statistic		n	α			
			0·10	0·05	0·025	0·01
D	U.T.	4	0·102	0·050	0·023	0·008
		10	0·100	0·050	0·025	0·009
		50	0·100	0·050	0·025	0·010
D	L.T.	4	0·100	0·047	0·024	0·010
		10	0·100	0·050	0·025	0·012
V	U.T.	4	0·102	0·050	0·025	0·007
		10	0·100	0·049	0·024	0·009
		50	0·100	0·050	0·025	0·010
V	L.T.	4	0·094	0·050	0·028	0·013
		10	0·098	0·050	0·026	0·012
W^2	U.T.	4	0·100	0·052	0·025	0·009
		10	0·100	0·050	0·025	0·010
		20	0·100	0·050	0·025	0·010
W^2	L.T.	4	0·085	0·047	0·022	0·01
		10	0·100	0·054	0·028	0·014
		20	0·096	0·053	0·025	0·01
U^2	U.T.	4	0·101	0·053	0·024	0·008
		10	0·100	0·050	0·025	0·010
		20	0·100	0·050	0·025	0·010
U^2	L.T.	4	0·091	0·045	0·022	0·006
		10	0·103	0·051	0·024	0·009
C	U.T.	4	0·109	0·059	0·029	0·010
		10	0·100	0·051	0·025	0·010
		50	0·099	0·050	0·025	0·010
A	U.T.	4	0·108	0·052	0·021	0·003
		10	0·098	0·050	0·025	0·009
		50	0·010	0·050	0·025	0·010
A	L.T.	4	0·101	0·063	0·043	0·029
		10	0·100	0·050	0·025	0·011

ACKNOWLEDGEMENT

This work was partly supported by the National Research Council of Canada.

REFERENCES

AJNE, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika*, **55**, 343–354.

ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Ann. Math. Statist.*, **23**, 193–212.

BROWN, R. L. and DURBIN, J. (1968). Methods of investigating whether a regression relationship is constant over time. Paper given at European meeting, IMS, Amsterdam.

BRUNK, D. H. (1962). On the range of the difference between hypothetical distribution function and Pyke's modified empirical distribution function. *Ann. Math. Statist.*, **33**, 525–532.

DURBIN, J. (1969). Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika*, **56**, 1–15.

KUIPER, N. H. (1962). Tests concerning random points on a circle. *Proc. Koninkl. Nederl. Akad. van Wetenschappen*, Ser. A, **63**, 38–47.

OWEN, D. B. (1962). *A Handbook of Statistical Tables*. Addison-Wesley.

PEARSON, E. S. (1963). Tests for randomness of points on a line. *Biometrika*, **50**, 77–87.

PEARSON, E. S. and STEPHENS, M. A. (1962). The goodness-of-fit tests based on W_N^2 and U_N^2. *Biometrika*, **49**, 397–402.

SESHADRI, V., CSORGO, M. and STEPHENS, M. A. (1969). Tests for the exponential distribution using Kolmogorov-type statistics. *J. R. Statist. Soc.* B, **31**, 499–509.

STEPHENS, M. A. (1964). The distribution of the goodness-of-fit statistic U_N^2. II. *Biometrika*, **51**, 393–397.

—— (1965). The goodness-of-fit statistic V_N: distribution and significance points. *Biometrika*, **52**, 309–321.

—— (1969a). A goodness-of-fit statistic for the circle, with some comparisons. *Biometrika*, **56**, 161–168.

—— (1969b). Results from the relation between two statistics of the Kolmogorov–Smirnov type. *Ann. Math. Statist.*, **40**, 1833–1837.

STEPHENS, M. A. and MAAG, U. R. (1968). Further significance points for W_N^2. *Biometrika*, **55**, 428–430.

WATSON, G. S. (1961). Goodness-of-fit tests on a circle. I. *Biometrika*, **48**, 109–114.