ETH Library

Tests of the Kolmogorov-Smirnov type for exponential data with unknown scale, and related problems

Journal Article

Author(s):
Margolin, Barry H.; Maurer, Willi

Publication date:
1976

Permanent link:
https://doi.org/10.3929/ethz-b-000422555

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Biometrika 63(1), https://doi.org/10.1093/biomet/63.1.149

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Biometrika (1976), 63, 1, *pp.* 149–60
Printed in Great Britain

Tests of the Kolmogorov–Smirnov type for exponential data with unknown scale, and related problems

By BARRY H. MARGOLIN
Department of Statistics, Yale University, New Haven, Connecticut

AND WILLI MAURER
Wirtschafts-Mathematik AG and Swiss Federal Institute of Technology, Zürich

Summary

Let $\hat{D}$, $\hat{D}^+$ and $\hat{D}^-$ denote Kolmogorov–Smirnov type one-sample statistics to test goodness of fit in the presence of unknown nuisance parameters; then the distributions of $\hat{D}$, $\hat{D}^+$ and $\hat{D}^-$ depend on the population sampled and the estimator used. Simulation has been the primary tool for studying these statistics. Recently, Durbin obtained the distributions of $\hat{D}$, $\hat{D}^+$ and $\hat{D}^-$ in terms of a Fourier transform for a wide class of underlying populations, and produced explicit results for the exponential case. In this paper, the distribution functions of $\hat{D}$, $\hat{D}^+$ and $\hat{D}^-$ for the exponential case are derived from general results for order statistics, and computationally efficient approximations to these distribution functions are obtained. In the course of this derivation, Bonferroni inequalities of Kounias, and Sobel & Uppuluri are generalized. Certain problems of goodness-of-fit testing in the presence of nuisance parameters, whose solutions make use of existing tables, are also discussed. These problems include the Pareto, Rayleigh, power function, and uniform distributions.

Some key words: Bonferroni inequality; Composite hypothesis; Dirichlet distribution; Distance statistic; Goodness of fit; Kolmogorov–Smirnov tests; Nuisance parameter; Order statistic.

1. Introduction

A voluminous literature exists for tests of goodness of fit based on the sample distribution function, but, as Durbin (1973, p. 47) has commented, it is surprising how little theoretical work has been done on tests based on the sample distribution function when the population distribution function is postulated only up to a set of nuisance parameters. Few realistic goodness-of-fit problems involve a fully specified population distribution function; more typically, the form of the null hypothesis of interest in a goodness-of-fit problem is that, for θ a vector,

$$H_0 : F(x) = F_0(x; \theta) \quad (\theta \in \Omega),$$

that is, a composite rather than a simple hypothesis.

In this situation, it is tempting to replace the nuisance parameters in $F_0(x;\theta)$ by efficient estimators $\hat{\theta}$, and then to form test statistics based on $F_0(x;\hat{\theta})$ as if it were a fully specified distribution function. Thus, one might construct tests for a random sample of size n based on the maximum deviation between the sample distribution function, say $F_n(x)$, and

$F_0(x;\hat{\theta})$, that is tests of the form: reject H_0 if

$$\begin{aligned}&\text{(i)}\quad D_n(\hat{\theta}) = \sup_x |F_n(x) - F_0(x;\hat{\theta})| > C_{\alpha n},\\ &\text{(ii)}\quad D_n^+(\hat{\theta}) = \sup_x \{F_n(x) - F_0(x;\hat{\theta})\} > C_{\alpha n}^+,\\ &\text{(iii)}\quad D_n^-(\hat{\theta}) = \sup_x \{F_0(x;\hat{\theta}) - F_n(x)\} > C_{\alpha n}^-.\end{aligned} \tag{1·1}$$

The test statistics in (1·1) mirror the Kolmogorov (1933) statistic, although the sampling distributions under H_0 are no longer distribution-free when $F_0(x;\theta)$ is continuous, but depend on both $F_0(x;\theta)$ and the choice of estimator for θ. Nevertheless, the class of tests in (1·1) retains certain attractive features that commend the Kolmogorov test when no nuisance parameter is involved. First, it avoids arbitrarily discretizing the hypothesized distribution function and transforming the sample observations to counts, a process that is troublesome in small samples, but is required by the natural competitor to $D_n(\hat{\theta})$, the chi-squared goodness-of-fit test. Secondly, as Durbin (1975) has shown, the distribution functions of the statistics in (1·1) can be obtained for certain $F_0(x;\theta)$ and maximum likelihood estimation of θ. Thirdly, simulation studies by Lilliefors (1967, 1969) and Stephens (1974) support the thesis that the tests in (1·1) will generally be more sensitive to departures from H_0 than is the chi-squared test. Thus, the class of tests represented in (1·1) warrants further exploration.

To date, research on the behaviour of the tests in (1·1) has proceeded along two lines. Simulation has been the primary path followed. For example, Lilliefors (1967), via simulation, has produced tables of critical values for test (i) in (1·1) for a normal distribution with unknown mean and variance, together with estimation of these two nuisance parameters by the corresponding sample mean and variance. Lilliefors (1969) has also studied via simulation the case of an exponential distribution with unknown scale parameter estimated by the sample mean. Both papers include brief power studies. Stephens (1974), also using simulation, has investigated the cases Lilliefors studied, but under a wider variety of alternative distributions for $F(x)$. In an analytic direction, Durbin (1975) appears to be the only author to date who has an exact, albeit complex, numerical method for evaluating tail-area probabilities of $D_n(\hat{\theta})$, $D_n^+(\hat{\theta})$ and $D_n^-(\hat{\theta})$ for a wide class of cases; Durbin demonstrates the computational feasibility of his technique by producing tables of exact critical values of the statistics in (1·1) for the exponential case.

In this paper, the case of exponentiality is considered in detail. In §2 this case is formulated explicitly and distribution functions for the statistics in (1·1) are derived analytically in closed form expressions. In §3, for purposes of deriving computationally efficient approximations to these distribution functions, new Bonferroni inequalities are produced. These new inequalities generalize earlier results of Kounias (1968) and Sobel & Uppuluri (1972). In §4, the Bonferroni inequalities of §3 are used to determine critical values for the statistics in (1·1); in the process, one obtains new insight into the behaviour of the corresponding tests. Finally in §5, attention is turned to a class of distributions $F_0(x;\theta)$ for which existing tables may be employed. These include the Pareto, Rayleigh, power function, and uniform distributions.

2. AN EXACT SOLUTION FOR THE EXPONENTIAL CASE

After observing a random sample $X_1, \ldots, X_n$, one may wish to determine whether these data are reasonably consistent with an assumption of exponentiality for the sampled population. Specifically, the hypothesis is:

$$H_0 : F_0(x;\theta) = 1 - e^{-x/\theta} \quad (\theta > 0, x > 0). \tag{2·1}$$

The alternative hypothesis is that H_0 is not tenable. Were a specific alternative in mind, one presumably would employ a testing procedure tailored to that alternative.

Many tests of fit of the hypothesis in (2·1) exist; Shapiro & Wilk (1972) proposed an analysis of variance test for exponentiality and also referenced an extensive literature of other tests of fit for this case. Attention in this paper will focus on the sampling behaviour of the test statistic (i) in (1·1) for the hypothesis in (2·1) together with $\hat{\theta} = (X_1 + \ldots + X_n)/n$, the maximum likelihood estimate of θ. This test statistic is (Lilliefors, 1969)

$$D_n(\hat{\theta}) = \sup_x |F_n(x) - \{1 - \exp(-nx/\Sigma X_i)\}|. \tag{2·2}$$

If $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ denote the order statistics corresponding to the original sample, then it can easily be verified that

$$D_n(\hat{\theta}) = \max\{D_n^+(\hat{\theta}), D_n^-(\hat{\theta})\}, \tag{2·3}$$

where

$$\begin{aligned} D_n^+(\hat{\theta}) &= \max_{1 \leqslant i \leqslant n}\left[\frac{i}{n} - \{1 - \exp(-nX_{(i)}/\Sigma_j X_{(j)})\}\right], \\ D_n^-(\hat{\theta}) &= \max_{1 \leqslant i \leqslant n}\left[\{1 - \exp(-nX_{(i)}/\Sigma_j X_{(j)})\} - \frac{(i-1)}{n}\right]. \end{aligned} \tag{2·4}$$

From (2·3) and (2·4) it follows that, for $0 \leqslant \gamma \leqslant 1$,

$$\begin{aligned} \operatorname{pr}\{D_n(\hat{\theta}) \leqslant \gamma\} &= \operatorname{pr}\{D_n^+(\hat{\theta}) \leqslant \gamma, D_n^-(\hat{\theta}) \leqslant \gamma\} \\ &= \operatorname{pr}\left\{\frac{i}{n} - \gamma \leqslant 1 - \exp(-nX_{(i)}/\Sigma_j X_{(j)}) \leqslant \frac{(i-1)}{n} + \gamma;\ i = 1, \ldots, n\right\} \qquad (2·5) \\ &= \operatorname{pr}\left\{\frac{1}{n}\log\left(\frac{n}{n-i+n\gamma}\right) \leqslant X_{(i)}/\Sigma_j X_{(j)} \leqslant \frac{1}{n}\log^*\left(\frac{n}{n-i+1-n\gamma}\right);\ i = 1, \ldots, n\right\}, \end{aligned}$$

where $\log^*(x) = \log x\ (x > 1)$, $\log^*(x) = n\ (x \leqslant 1)$.

If $Y_i = X_i/\Sigma X_j\ (i = 1, \ldots, n)$, then $\Sigma Y_i = 1$; from Wilks (1962, p. 179), one learns that the joint distribution of $Y_1, \ldots, Y_n$ is Dirichlet $D_n(1, \ldots, 1)$, that is

$$f(Y_1, \ldots, Y_n) = (n-1)!\, I\left\{\left(\sum_{i=1}^{n} Y_i = 1\right) \cap \bigcap_{i=1}^{n} (Y_i > 0)\right\}, \tag{2·6}$$

where $I(S)$ is the indicator function for set S.

If $Y_{(1)} \leqslant \ldots \leqslant Y_{(n)}$ denote the order statistics corresponding to the $Y_1, \ldots, Y_n$, then $Y_{(i)} = X_{(i)}/\Sigma_j X_{(j)}$ and

$$f(Y_{(1)}, \ldots, Y_{(n)}) = n!\,(n-1)!\, I\left\{\left(\sum_{j=1}^{n} Y_{(j)} = 1\right) \cap (0 \leqslant Y_{(1)} \leqslant \ldots \leqslant Y_{(n)})\right\}. \tag{2·7}$$

Note (David & Johnson, 1948) that the distributions in (2·6) and (2·7) do not depend on θ.

As a consequence of the constraint on $Y_{(1)}, \ldots, Y_{(n)}$, the following inequalities hold with probability one:

$$0 \leqslant Y_{(i)} \leqslant 1/(n-i+1) \quad (i = 1, \ldots, n-1); \quad 1/n \leqslant Y_{(n)} \leqslant 1.$$

It then follows that (2·5) is equivalent to

$$\operatorname{pr}\{D_n(\hat{\theta}) \leqslant \gamma) = \operatorname{pr}\{L(i,n,\gamma) \leqslant Y_{(i)} \leqslant U(i,n,\gamma);\, i = 1, \ldots, n\}, \tag{2·8}$$

where

$$\begin{aligned} L(i,n,\gamma) &= \max\left\{0, \frac{1}{n}\log\left(\frac{n}{n-i+n\gamma}\right)\right\} \quad (i = 1, \ldots, n-1), \\ L(n,n,\gamma) &= \max\left\{\frac{1}{n}, \frac{1}{n}\log\left(\frac{1}{\gamma}\right)\right\}, \\ U(i,n,\gamma) &= \min\left\{\frac{1}{n-i+1}, \frac{1}{n}\log^*\left(\frac{n}{n-i+1-n\gamma}\right)\right\} \quad (i = 1, \ldots, n). \end{aligned} \tag{2·9}$$

This attention to bounds will avoid computational problems later.

In principle, the probability in (2·8) could be obtained directly by integrating the probability density in (2·7) over the interval in R^n defined by the expressions in (2·8) and (2·9); the integration, however, is best done in a two-stage operation.

First, define

$$\bar{F}(r_1, r_2, \ldots, r_n) = \operatorname{pr}\left\{\bigcap_{i=1}^{n} (Y_{(i)} \geqslant r_i)\right\}$$

for all values of $(r_1, \ldots, r_n)$. Then, because $0 < Y_{(1)} \leqslant \ldots \leqslant Y_{(n)}$, it follows that

$$\bar{F}(r_1, \ldots, r_n) = \bar{F}(C_1, \ldots, C_n) \tag{2·10}$$

for $C_i = \max\{0, \max(r_1, \ldots, r_i)\}$. Note that $0 \leqslant C_1 \leqslant \ldots \leqslant C_n$.

Secondly, it follows from (2·8) that

$$\begin{aligned} \operatorname{pr}\{D_n(\hat{\theta}) \leqslant \gamma\} = \bar{F}\{&L(1,n,\gamma), \ldots, L(n,n,\gamma)\} \\ &-[\bar{F}\{U(1,n,\gamma), L(2,n,\gamma), \ldots, L(n,n,\gamma)\} + \ldots \\ &+\bar{F}\{L(1,n,\gamma), \ldots, L(n-1,n,\gamma), U(n,n,\gamma)\}] + \ldots \\ &+(-1)^n \bar{F}\{U(1,n,\gamma), \ldots, U(n,n,\gamma)\}. \end{aligned} \tag{2·11}$$

Equation (2·11) is justified by an argument identical to that which justifies expressing the probability of an n-dimensional interval in terms of an n-dimensional distribution function (Wilks, 1962, p. 49).

An expression for $\bar{F}(C_1, \ldots, C_n)$ has been obtained by Maurer & Margolin (1976) as a special case of results for distribution functions of subsets of order statistics, namely

$$\begin{aligned} \bar{F}(C_1, \ldots, C_n) &= \operatorname{pr}\left\{\bigcap_{i=1}^{n} (Y_{(i)} > C_i)\right\} \\ &= (-1)^{\frac{1}{2}n(n+1)} \sum_{a_{n-1}=n-1}^{n} \cdots \sum_{a_1=1}^{a_2} (-1)^{a_1+\cdots+a_n}\, n! \left\{\prod_{j=1}^{n+1} (a_j - a_{j-1})!\right\}^{-1} \\ &\quad \times \prod_{j=1}^{n} \binom{a_j - a_{j-1} - 1}{a_j - j} \left\{1 - \sum_{i=1}^{n} C_i(a_i - a_{i-1})\right\}^{n-1} I\left\{\sum_{i=1}^{n} C_i(a_i - a_{i-1}) \leqslant 1\right\}, \end{aligned} \tag{2·12}$$

where $a_0 = 0$, $a_n = n$.

The result in (2·12) can be restated more compactly as

$$\operatorname{pr}\left\{\bigcap_{i=1}^{n}(Y_{(i)} > C_i)\right\} = n!\sum_{m=1}^{n}(-1)^{n-m}\Sigma_r\left(\prod_{i=1}^{m} r_i!\right)^{-1}\left(1-\sum_{i=1}^{m} C_i r_i\right)^{n-1} I\left(\sum_{i=1}^{m} C_i r_i \leqslant 1\right), \qquad (2{\cdot}13)$$

where Σ_r denotes summation over all positive integers $(r_1, \ldots, r_m)$ such that $\Sigma_i r_i = n$.

Formula (2·13) can be derived from (2·12) essentially by deleting the terms with combinatorial coefficients equal to zero. The proof will be omitted. Durbin (1975) obtained formula (2·13) via Fourier inversion; he observed that (2·13) yielded the distribution function of $D_n^+(\hat\theta)$ for the exponential case if one sets $C_i = L(i, n, \alpha)$ $(i = 1, \ldots, n)$. Upon combining the expressions in (2·9), (2·10), (2·11) and (2·13), one arrives at a closed analytic expression for the distribution function of $D_n(\hat\theta)$.

The statistics $D_n^+(\hat\theta)$ and $D_n^-(\hat\theta)$ were first studied by Durbin (1975); he proved that the two statistics are not identically distributed and tabulated their critical values. It is worth noting that if the alternative to (2·1) is one-sided, then $D_n^+(\hat\theta)$ and $D_n^-(\hat\theta)$ are intuitively appealing test statistics. For alternatives that are one-sided, and for alternative test procedures, see Hollander & Proschan (1972).

The distribution in (2·6) arises in many other seemingly unrelated contexts (David, 1970, §5·4). Thus, for example, the statistics in (1·1) may be used to test whether data observed are consistent with the hypothesis that the observations are distances between successive points dropped at random in the interval [0, 1], as Durbin (1975) has indicated. Similarly, one is also able to test whether data observed are consistent with the hypothesis that the $Y_1, \ldots, Y_n$ are distances between midpoints of successive arcs of equal length that have been randomly placed on the perimeter of a circle of unit circumference.

In §4 computationally efficient approximations to the distribution function of $D_n(\hat\theta)$ will be derived; these approximations are based on new Bonferroni inequalities to be presented in the next section.

3. IMPROVED BONFERRONI BOUNDS FOR THE PROBABILITY OF A UNION

For a finite set of events $\{A_1, \ldots, A_n\}$ associated with a probability space $(\Omega, \mathscr{F}, P)$, Kounias (1968) has proved that

$$\operatorname{pr}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n}\operatorname{pr}(A_i) - \max_{j}\sum_{i \neq j}\operatorname{pr}(A_i A_j), \qquad (3{\cdot}1)$$

which clearly improves on the simple Bonferroni upper bound of $\Sigma\operatorname{pr}(A_i)$. Following Sobel & Uppuluri (1972), one can define the degree of a bound, such as that in (3·1), as the maximum number of events in any intersection whose probability is needed to evaluate the bound; the bound in (3·1) is then of degree two. Sobel & Uppuluri produce upper bounds of even degree, and lower bounds of odd degree, that are clear improvements over the standard Bonferroni bounds of degree one less, for degrees at least 2, but their results apply only if the events are exchangeable. The theorem that follows generalizes their results to non-exchangeable events.

THEOREM 3·1. Let

$$S_1 = \sum_{i=1}^{n}\operatorname{pr}(A_i), \quad S_2 = \sum_{i<j}\operatorname{pr}(A_i A_j), \quad S_3 = \sum_{i<j<k}\operatorname{pr}(A_i A_j A_k), \quad \ldots, \quad S_n = \operatorname{pr}\left(\bigcap_{i=1}^{n} A_i\right);$$

similarly, for a fixed integer r *such that* $1 \leqslant r \leqslant n$, *set*

$$S_2^{(r)} = \sum_{i \neq r} \mathrm{pr}\,(A_i A_r), \quad S_3^{(r)} = \sum_{\substack{i<j \\ i \neq r \neq j}} \mathrm{pr}\,(A_i A_j A_r), \quad \dots .$$

Then, for any odd degree ν $(3 \leqslant \nu \leqslant n)$,

$$\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} S_\alpha + \max_r S_\nu^{(r)} \leqslant \mathrm{pr}\left(\bigcup_{i=1}^{n} A_i\right), \tag{3·2}$$

and for any even degree ν $(2 \leqslant \nu \leqslant n)$,

$$\mathrm{pr}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} S_\alpha - \max_r S_\nu^{(r)}. \tag{3·3}$$

Proof. Let $I_i(\omega)$ be the indicator random variable of A_i. Here $\max\{I_1(\omega), \dots, I_n(\omega)\}$ is the indicator random variable for $A_1 \cup \dots \cup A_n$.

For a fixed integer r such that $1 \leqslant r \leqslant n$, define

$$T_2(\omega) = \sum_{i<j} I_i(\omega) I_j(\omega), \quad T_3(\omega) = \sum_{i<j<k} I_i(\omega) I_j(\omega) I_k(\omega), \quad \dots,$$

$$T_2^{(r)}(\omega) = \sum_{i \neq r} I_i(\omega) I_r(\omega), \quad T_3^{(r)}(\omega) = \sum_{\substack{i<j \\ i \neq r \neq j}} I_i(\omega) I_j(\omega) I_r(\omega), \quad \dots .$$

The claim is made that for all $\omega \in \Omega$ and odd ν such that $3 \leqslant \nu \leqslant n$,

$$\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} T_\alpha + T_\nu^{(r)} \leqslant \max\{I_1(\omega), \dots, I_n(\omega)\}. \tag{3·4}$$

To prove (3·4), one must consider three possible cases.

(i) First, if $I_i(\omega_0) = 0$ for all i, then both the left-hand side and the right-hand side of (3·4) are zero.

(ii) Secondly, if $I_r(\omega_0) = 0$ and exactly $n-t-1$ other indicator variables are zero for $\omega = \omega_0$ $(t = 1, \dots, n-1)$, then t indicator variables are one for $\omega = \omega_0$, and the left-hand side of (3·4) is

$$\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} \binom{t}{\alpha} = -\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha} \binom{t}{\alpha}. \tag{3·5}$$

$$= 1 - \binom{t-1}{\nu-1}, \tag{3·6}$$

which is less than or equal to 1, the right-hand side of (3·4).

(iii) Thirdly, if $I_r(\omega_0) = 1$ and exactly t other indicator variables are one for $\omega = \omega_0$ $(t = 0, 1, \dots, n-1)$, then $n-t-1$ indicator variables are zero, and the left-hand side is

$$\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} \binom{t+1}{\alpha} + \binom{t}{\nu-1} = 1 - \binom{t}{\nu-1} + \binom{t}{\nu-1} = 1,$$

as is the right-hand side.

Thus (3·4) is proved. Taking expectations in (3·4), one proves that

$$\sum_{\alpha=1}^{\nu-1} (-1)^{\alpha-1} S_\alpha + S_\nu^{(r)} \leqslant \mathrm{pr}\left(\bigcup_{i=1}^{n} A_i\right). \tag{3·7}$$

Equation (3·7) clearly implies (3·2), since (3·7) holds for all r $(1 \leqslant r \leqslant n)$.

To prove (3·3), one first establishes the inequality

$$\max\{I_1(\omega),\ldots,I_n(\omega)\} \leqslant \sum_{\alpha=1}^{\nu-1}(-1)^{\alpha-1}T_\alpha - T_\nu^{(r)} \tag{3·8}$$

for ν even $(2 \leqslant \nu \leqslant n)$. Again this can be done directly by considering the cases (i)–(iii) above and using (3·6).

The inequality in (3·8), in turn, proves that

$$\operatorname{pr}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{\alpha=1}^{\nu-1}(-1)^{\alpha-1}S_\alpha - S_\nu^{(r)}. \tag{3·9}$$

Equation (3·9) holds for all r, so that (3·3) is also established.

Four comments concerning Theorem 3·1 are in order.

(I) If $\nu = 2$, then (3·3) reduces to Kounias's result in (3·1).

(II) For exchangeable events, (3·2) and (3·3) reduce to the bounds of Sobel & Uppuluri (1972, equations (1·1), (1·2)).

(III) Various other inequalities generalizing the results of Sobel & Uppuluri to nonexchangeable events, and generalizing the results of Kounias beyond $\nu = 3$, can be derived in a fashion similar to Theorem 3·1. Only one will be presented here.

(IV) Sobel & Uppuluri (1972, p. 1558) indicate that the 'simple-minded' replacement of $P_\alpha n!/\{\alpha!(n-\alpha)!\}$ in their results by S_α for nonexchangeable events does not follow from their work. Such replacement would yield an upper bound for ν even, and a lower bound for ν odd, of

$$\sum_{\alpha=1}^{\nu-1}(-1)^{\alpha-1}S_\alpha + (-1)^{\nu-1}\frac{\nu}{n}S_\nu. \tag{3·10}$$

Because $n^{-1}\Sigma_r S_\nu^{(r)} = n^{-1}\nu S_\nu$, it follows by averaging the bounds in either (3·2) or (3·3) that (3·10) is valid but is weaker than (3·2) and (3·3).

For reasons that will be clearer in § 4, one further generalization of Theorem 3·1 is needed. Consider any disjoint subsets $J_1,\ldots,J_m$ with union $(1,\ldots,n)$, and with $m \leqslant n$; extend the notation of Theorem 3·1 by letting

$$S_2^{(r_k)} = \sum_{\substack{i\in J_k\\ i\neq r_k}} \operatorname{pr}(A_i A_{r_k}), \quad S_3^{(r_k)} = \Sigma \operatorname{pr}(A_i A_j A_{r_k}), \quad \ldots,$$

for $r_k \in J_k$ $(k = 1,\ldots,m)$ the second sum being for $i < j \in J_k$, $i \neq r_k \neq j$. Then the following theorem can be proved.

THEOREM 3·2. *For any odd degree* ν $(3 \leqslant \nu \leqslant n)$,

$$\sum_{\alpha=1}^{\nu-1}(-1)^{\alpha-1}S_\alpha + \sum_{k=1}^{m}\max_{r_k\in J_k} S_\nu^{(r_k)} \leqslant \operatorname{pr}\left(\bigcup_{i=1}^{n} A_i\right), \tag{3·11}$$

and for any even degree ν $(2 \leqslant \nu \leqslant n)$,

$$\operatorname{pr}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{\alpha=1}^{\nu-1}(-1)^{\alpha-1}S_\alpha - \sum_{k=1}^{m}\max_{r_k\in J_k} S_\nu^{(r_k)}. \tag{3·12}$$

Proof. A proof similar to that of Theorem 3·1 is easily constructed and will be omitted.

Note that Theorem 3·1 is the special case of Theorem 3·2 with $m = 1$. For the case $\nu = 2$, see Kounias (1968). Finally, for exchangeable events, (3·11) and (3·12) are weaker than (3·2) and (3·3), respectively.

In § 4 these inequalities will be applied to the approximation of the distribution function of $D_n(\hat{\theta})$.

4. APPROXIMATIONS TO THE DISTRIBUTION FUNCTION OF $D_n(\hat\theta)$

The key to approximating the distribution function of $D_n(\hat\theta)$ via the results of § 3 is the recognition that (2·8) implies that

$$\operatorname{pr}\{D_n(\hat\theta) \geqslant \gamma\} = \operatorname{pr}\left[\bigcup_{i=1}^{n}\{Y_{(i)} \notin [L(i,n,\gamma), U(i,n,\gamma)]\}\right]. \qquad (4\cdot1)$$

If one defines the events

$$A_i = (Y_{(i)} \notin [L(i,n,\gamma), U(i,n,\gamma)]), \quad B_i = \{Y_{(i)} < L(i,n,\gamma)\},$$
$$B_{i+n} = \{Y_{(i)} > U(i,n,\gamma)\} \quad (i = 1, \ldots, n),$$

then $A_i = B_i \cup B_{i+n}$, $B_i \cap B_{i+n} = \phi$ and (4·1) may be written as

$$\operatorname{pr}\{D_n(\hat\theta) \geqslant \gamma\} = \operatorname{pr}\left(\bigcup_{i=1}^{n} A_i\right) = \operatorname{pr}\left\{\left(\bigcup_{i=1}^{n} B_i\right) \cup \left(\bigcup_{i=1}^{n} B_{i+n}\right)\right\}.$$

By the principle of inclusion-exclusion, an alternative expression can be derived for the tail area probability of $D_n(\hat\theta)$:

$$\operatorname{pr}\{D_n(\hat\theta) \geqslant \gamma\} = \sum_{i=1}^{n} \operatorname{pr}(A_i) - \sum_{i<j} \operatorname{pr}(A_iA_j) + \sum_{i<j<k} \operatorname{pr}(A_iA_jA_k) - \ldots. \qquad (4\cdot2)$$

Equation (4·2) together with the results of § 3 then gives rise to various inequalities that may be used to approximate (4·2) to any desired degree of accuracy. Samples of size 3 to 10 were considered. Work was not carried beyond $n = 10$ because Durbin's (1975) Table 3 had been verified to be correct in all cases for $n \leqslant 10$, and there seemed little point in proceeding further, given that the feasibility of the approximation approach had been established.

The bounds programmed are specified in the footnote to Table 4·3.

It is important to note that all the probabilities involved can be expressed simply and linearly in various terms of the form

$$\operatorname{pr}\left\{\left(\bigcap_{j=1}^{s} \bar B_{i_j}\right) \cap \left(\bigcap_{j=s+1}^{t} B_{n+i_j}\right)\right\} = \operatorname{pr}\left\{\bigcap_{j=1}^{t} (Y_{(i_j)} > \delta_{i_j})\right\}, \qquad (4\cdot3)$$

where

$$\delta_{i_j} = \begin{cases} L(i_j, n, \gamma) & (j = 1, \ldots, s), \\ U(i_j, n, \gamma) & (j = s+1, \ldots, t). \end{cases}$$

The terms in (4·3) may then be evaluated via (2·10) and (2·12), or more easily, via formula (4·5) of Maurer & Margolin (1976).

It is instructive to examine in detail a representative example of the computations involved in these approximations to the distribution function of $D_n(\hat\theta)$; $n = 7$ and $\gamma = 0\cdot309$ are chosen for this purpose, the γ value being the 0·20 critical value reported by Lilliefors (1969).

The marginal behaviour of each order statistic is contained in Table 4·1.

Table 4·2 contains the values of the second-degree terms, $\{\operatorname{pr}(B_iB_j), \operatorname{pr}(B_iB_{n+j}), \operatorname{pr}(B_{n+i}B_j), \operatorname{pr}(B_{n+i}B_{n+j})\}$ that are greater than zero to four decimals; all omitted terms and terms replaced by a dash are zero to four decimals. Double precision accuracy was

Table 4·1. $\operatorname{pr}(B_i)$ *and* $\operatorname{pr}(B_{i+n})$, $n = 7$, $\gamma = 0\cdot309$

i	1	2	3	4	5	6	7
Tail probabilities							
$\operatorname{pr}\{Y_{(i)} < L(i, 7, 0\cdot309)\}$	0·0000	0·0000	0·0196	0·0295	0·0235	0·0108	0·0000
$\operatorname{pr}\{Y_{(i)} > U(i, 7, 0\cdot309)\}$	0·0628	0·0666	0·0373	0·0060	0·0000	0·0000	0·0000

Table 4·2. *Second-degree terms*, $n = 7$, $\gamma = 0{\cdot}309$

(i, j)	$\mathrm{pr}\,(B_iB_j)$	$\mathrm{pr}\,(B_iB_{n+j})$	$\mathrm{pr}\,(B_{n+i}B_j)$	$\mathrm{pr}\,(B_{n+i}B_{n+j})$
1, 2	—	—	—	0·0213
1, 3	—	—	—	0·0041
1, 6	—	—	0·0003	—
2, 3	—	—	—	0·0105
3, 4	0·0067	0·0002	—	0·0002
3, 5	0·0026	—	—	—
3, 6	0·0006	—	—	—
4, 5	0·0081	—	—	—
4, 6	0·0017	—	—	—
5, 6	0·0035	—	—	—
Column sum	0·0232	0·0003	0·0004	0·0360

maintained, however, in all later computations, including the column sums. Two points concerning Table 4·2 are noteworthy. First,

$$\sum_{1 \leqslant i < j \leqslant 7} \mathrm{pr}\,(B_i B_{7+j}), \qquad \sum_{1 \leqslant i < j \leqslant 7} \mathrm{pr}\,(B_{7+i} B_j)$$

are both negligible; since

$$\mathrm{pr}\,\{D_7^+(\hat{\theta}) \geqslant \gamma, D_7^-(\hat{\theta}) \geqslant \gamma\} \leqslant \sum_{1 \leqslant i < j \leqslant 7} \{\mathrm{pr}\,(B_i B_{7+j}) + \mathrm{pr}\,(B_{7+i} B_j)\},$$

the events $\{D_7^+(\hat{\theta}) \geqslant \gamma\}$, $\{D_7^-(\hat{\theta}) \geqslant \gamma\}$ are nearly exclusive. In all cases examined, near exclusivity was found; this is consistent with the behaviour of the usual Kolmogorov statistics reported by Vandewiele & Noé (1967), namely,

$$\mathrm{pr}\,(D_n^+ \geqslant \gamma, D_n^- \geqslant \gamma) \leqslant \mathrm{pr}\,(D_n^+ \geqslant \gamma)\,\mathrm{pr}\,(D_n^- \geqslant \gamma).$$

Secondly, for fixed i, both $\mathrm{pr}\,(B_iB_k)$ and $\mathrm{pr}\,(B_{7+i}B_{7+k})$, for $k \neq i$, $1 \leqslant k \leqslant 7$ and $1 \leqslant i \leqslant 7$, decrease monotonically either as k takes the values $i+1, \ldots, 7$, or as k takes the values $i-1, \ldots, 1$.

Third-order terms have not been given to save space. Based on these computations and others for $\gamma = 0{\cdot}310$, one finds the results in Table 4·3. Thus, the tail area probability for $\gamma = 0{\cdot}310$ is closer to 0·20 than is the tail area probability for $\gamma = 0{\cdot}309$; to be sure, the difference is negligible for all intents and purposes, but $\gamma = 0{\cdot}310$ is, to three decimals, the value obtained by Durbin (1975).

When $n = 7$ and $\gamma = 0{\cdot}310$, it was determined that, to three decimal places,

$$\mathrm{pr}\{D_n^+(\hat{\theta}) \geqslant \gamma\} = 0{\cdot}137, \quad \mathrm{pr}\{D_n^-(\hat{\theta}) \geqslant \gamma\} = 0{\cdot}063;$$

in all cases studied it was observed that $\mathrm{pr}\{D_n^+(\hat{\theta}) \geqslant \gamma\} > \mathrm{pr}\{D_n^-(\hat{\theta}) \geqslant \gamma\}$. This marked asymmetry, also evident in Table 4·1, is to be contrasted with the known symmetry in the behaviour of the usual Kolmogorov statistic. It is precisely this asymmetry that is exploited by the bounds in Theorem 3·2.

Table 4·1 also illustrates another feature of the test, namely that

$$\mathrm{pr}\{Y_{(1)} \leqslant L(1, 7, 0{\cdot}309)\} = 0, \quad \mathrm{pr}\{Y_{(7)} \geqslant U(7, 7, 0{\cdot}309)\} = 0.$$

In Durbin's (1975) Table 3 one sees that $\gamma > n^{-1}$ for all critical values of interest; this implies that $U(n, n, \gamma) = 1$ and $L(1, n, \gamma) = 0$. Therefore, if $\gamma > n^{-1}$, then

$$\mathrm{pr}\{Y_{(n)} \geqslant U(n, n, \gamma)\} = 0 = \mathrm{pr}\{Y_{(1)} \leqslant L(1, n, \gamma)\}. \tag{4·4}$$

Statement (4·4) will be true for this test irrespective of the true model. The test as constructed should then exhibit insufficient sensitivity to the behaviour of the extreme order statistics when the alternative hypothesis is true.

Table 4·3. *Probability bounds for* $\mathrm{pr}\{D_7(\hat{\theta}) \geqslant \gamma\}$

γ	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
0·309	0·2562	0·2243	0·2079	0·2030	0·2025	0·2017	0·1963
0·310	0·2514	0·2204	0·2044	0·1996	0·1992	0·1984	0·1932

The bounds are: (i) standard Bonferroni upper bound, degree 1; (ii) Theorem 3·1 upper bound, degree 2; (iii) Theorem 3·2 upper bound, degree 2, $m = 2$, $J_1 = \{1, \ldots, n\}$, $J_2 = \{n+1, \ldots, 2n\}$; (iv) standard Bonferroni upper bound, degree 3; (v) Theorem 3·2 lower bound, degree 3, m, J_1, J_2 as in (iii); (vi) Theorem 3·1 lower bound, degree 3; (vii) standard Bonferroni lower bound, degree 2.

Certain modifications of the test may counterbalance the insensitivity and asymmetry noted. A simple modification, for example, is the introduction of a weighting function, in the same spirit that Renyi (1953) and others have proposed weighting functions for the standard Kolmogorov statistic. Consider the general class of tests that follows:

Reject the hypothesis of exponentiality if

$$W_n(\hat{\theta}) = \max\{W_n^+(\hat{\theta}),\, W_n^-(\hat{\theta})\} \geqslant C_{\alpha n}, \tag{4·5}$$

where

$$W_n^+(\hat{\theta}) = \max_{1 \leqslant i \leqslant n}\left(w_i^+\left[\frac{i}{n} - \{1 - \exp(-nX_{(i)}/\Sigma_j X_{(j)})\}\right]\right),$$

$$W_n^-(\hat{\theta}) = \max_{1 \leqslant i \leqslant n}\left[w_i^-\left\{1 - \exp(-nX_{(i)}/\Sigma_j X_{(j)}) - \left(\frac{i-1}{n}\right)\right\}\right],$$

and $\{w_i^+\}$, $\{w_i^-\}$ are sets of specified weights; Durbin (1975) has also briefly discussed this class of tests.

Tests of the class in (4·5) require evaluation of an expression of the form

$$\mathrm{pr}(a_i < Y_{(i)} < b_i;\ i = 1, \ldots, n),$$

which can be evaluated exactly by the method of Durbin (1975) and can be approximated by the bounds of the present section. One possible set of weights that will give greater import to the extreme order statistics is $w_i^+ = w_i^- = \{i(n-i+1)/(n+1)^2\}^{-1}$. A more extreme set of weights, taking all w^+ and w^- to be zero except for $w_n^- = 1$, produces a test equivalent to Fisher's (1929) maximum harmonic test with critical region $Y_{(n)} > k_{n,\alpha}$, a test that is not thought of as a global goodness-of-fit test for exponentiality, but rather as a test for slippage.

5. PROBLEMS FOR WHICH EXISTING TABLES MAY BE EMPLOYED

The tables produced by Durbin (1975) for the exponential goodness-of-fit test under discussion have wider applicability than has been indicated previously. The following lemma extends the applicability in one direction.

LEMMA 5·1. *If* $X_1, \ldots, X_n$ *are a random sample from a population whose distribution function* $F_X(x;\theta)$ *has a maximum likelihood estimate* $\hat{\theta}(X)$, *and if* $Y_i = g(X_i)$ $(i = 1, \ldots, n)$, *where* g *is a monotonic transformation involving no parameters with inverse* g^{-1}, *then the value of the statistic* $D_n^*\{\hat{\theta}(Y)\}$ *in* (1·1) *based on* $Y_1, \ldots, Y_n$ *together with maximum likelihood estimation for*

θ has the same value as $D_n\{\hat{\theta}(X)\}$ in (1·1) based on $X_1, \ldots, X_n$. Therefore, the two statistics have the same distribution.

Proof. We have

$$D_n^*\{\hat{\theta}(Y)\} = \max_{1 \leqslant i \leqslant n} \max\left(\left[\frac{i}{n} - F_Y\{Y_{(i)};\hat{\theta}(Y)\}\right], \left[F_Y\{Y_{(i)};\hat{\theta}(Y)\} - \frac{i-1}{n}\right]\right).$$

It suffices then and is easy to show that for all i, $F_Y(Y_{(i)}, \theta)$ equals $F_X(X_{(i)};\theta)$ if g is monotone increasing, and equals $1 - F_X(X_{(n-i+1)};\theta)$ if g is monotone decreasing. This, together with the fact that $\hat{\theta}(Y) = \hat{\theta}(X)$ under the conditions of the theorem, yields the desired result.

The invariance in Lemma 5·1 extends to many statistics based on $F_X\{X_{(i)}, \hat{\theta}(X)\}$, such as the various goodness-of-fit test statistics considered by Stephens (1974).

Some examples of the use of Lemma 5·1 to extend the applicability of Durbin's table (1975) follow; all involve testing for goodness of fit via (1·1) together with maximum likelihood estimation for θ:

(i) *Power function distribution*, $\theta > 0: f_Y(y) = \theta^{-1}y^{(1-\theta)/\theta}$ for $0 \leqslant y \leqslant 1$. Here $X = g^{-1}(Y) = -\log Y$ has the distribution function in (2·1).

(ii) *Pareto distribution*, $\theta > 0: f_Y(y) = \theta^{-1}y^{-(\theta+1)/\theta}$ for $y > 1$. Here, $X = g^{-1}(Y) = \log Y$ has the distribution function in (2·1).

(iii) *Rayleigh distribution*, $\theta > 0: f_Y(y) = 2y\theta^{-1}e^{-y^2/\theta}$ for $0 < y$. Here, $X = g^{-1}(Y) = Y^2$ has the distribution function in (2·1).

Similarly, a table of the distribution of the test statistic in (1·1) for the two-parameter normal model together with maximum likelihood estimation (Lilliefors, 1967) may be used for the two-parameter log normal model with maximum likelihood estimation.

An extension of the usefulness of existing tables in another direction is indicated by the following lemma, which is stated without proof:

LEMMA 5·2. *Let $M(x)$ be a Borel-measurable function on the real line such that the integral*

$$N(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} M(x)\,dx$$

exists for $-\infty < \theta_1 < \theta_2 < \infty$. The null hypothesis of interest is

$$H_0: F(x;\theta_1,\theta_2) = \begin{cases} 0 & (x < \theta_1), \\ \displaystyle\int_{\theta_1}^{x} \{M(x)/N(\theta_1,\theta_2)\}\,dx & (\theta_1 \leqslant x \leqslant \theta_2), \\ 1 & (x > \theta_2). \end{cases} \tag{5·1}$$

Consider a goodness-of-fit statistic to test (5·1) *based on a random sample $X_1, \ldots, X_n$ $(n \geqslant 3)$ of the form*:

$$\mathscr{D}_n(\hat{\theta}_1, \hat{\theta}_2) = \max_{2 \leqslant i \leqslant n-1} \max\left[\left\{\frac{i-1}{n-2} - F(X_{(i)};\hat{\theta}_1,\hat{\theta}_2)\right\}, \left\{F(X_{(i)};\hat{\theta}_1,\hat{\theta}_2) - \frac{i-2}{n-2}\right\}\right], \tag{5·2}$$

where $(\hat{\theta}_1, \hat{\theta}_2)$ are maximum likelihood estimates. Then under H_0, the distribution of $\mathscr{D}_n(\hat{\theta}_1, \hat{\theta}_2)$ is identical to the distribution of D_{n-2}, the standard Kolmogorov statistic based on a sample of size $n-2$ from a population with fully specified distribution function.

As an example, one could employ this result in testing goodness of fit of the uniform distribution $U[\theta_1, \theta_2]$. Note also that for the case of only one truncation point, the statistic in (5·2) would be altered in an obvious way; $\mathscr{D}_n(\hat{\theta})$ would then be distributed as D_{n-1}.

Finally, although §5 has focused on extending the applicability of the table for $D_n(\hat{\theta})$, analogous results could have been stated for the one-sided statistics $D_n^+(\hat{\theta})$ and $D_n^-(\hat{\theta})$.

The work was supported by U.S. Army, Navy, Air Force and N.A.S.A. under a contract administered by the Office of Naval Research, and by the Swiss National Foundation.

REFERENCES

DAVID, F. N. & JOHNSON, N. L. (1948). The probability integral transformation when parameters are estimated from the sample. *Biometrika* **35**, 182–90.

DAVID, H. A. (1970). *Order Statistics*. New York: Wiley.

DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. Philadelphia: S.I.A.M.

DURBIN, J. (1975). Kolmogorov–Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. *Biometrika* **62**, 5–22.

FISHER, R. A. (1929). Tests of significance in harmonic analysis. *Proc. R. Soc.* A **125**, 54–9.

HOLLANDER, M. & PROSCHAN, F. (1972). Testing whether new is better than used. *Ann. Math. Statist.* **43**, 1136–46.

KOLMOGOROV, A. (1933). Sulla determinazione empirica di una legga di distribuzione. *Giorn. Ist. Ital. Attuari* **4**, 83–91.

KOUNIAS, E. G. (1968). Bounds for the probability of a union of events, with applications. *Ann. Math. Statist.* **39**, 2154–8.

LILLIEFORS, H. W. (1967). On the Kolmogorov–Smirnov tests for normality with mean and variance unknown. *J. Am. Statist. Assoc.* **62**, 399–402.

LILLIEFORS, H. W. (1969). On the Kolmogorov–Smirnov tests for the exponential distribution with mean unknown. *J. Am. Statist. Assoc.* **64**, 387–9.

MAURER, W. & MARGOLIN, B. H. (1976). The multivariate inclusion-exclusion formula and order statistics from dependent variates. *Ann. Statist.* **4**. To appear.

RENYI, A. (1953). On the theory of order statistics. *Acta Math. Hung.* **4**, 191–231.

SHAPIRO, S. S. & WILK, M. B. (1972). An analysis of variance test for the exponential distribution (complete samples). *Technometrics* **14**, 355–70.

SOBEL, M. & UPPULURI, V. R. R. (1972). On Bonferroni-type inequalities of the same degree for the probability of unions and intersections. *Ann. Math. Statist.* **43**, 1549–58

STEPHENS, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *J. Am. Statist. Assoc.* **69**, 730–7.

VANDEWIELE, G. & NOÉ, M. (1967). An inequality concerning tests of fit of the Kolmogorov–Smirnov type. *Ann. Math. Statist.* **38**, 1240–4.

WILKS, S. S. (1962). *Mathematical Statistics*. New York: Wiley.

[*Received December* 1974. *Revised July* 1975]