The Annals of Statistics
1980, Vol. 8, No. 5, 989–1001

A TEST FOR GOODNESS-OF-FIT BASED ON AN EMPIRICAL PROBABILITY MEASURE

BY ROBERT V. FOUTZ

Virginia Polytechnic Institute and State University

A goodness-of-fit test is proposed for the simple hypothesis specifying a continuous p-variate distribution. For a suitably defined empirical probability measure, the proposed test is based on the supremum of the absolute differences between hypothesized and empirical probabilities, the supremum being taken over all possible events. This test statistic is shown to be distribution free in the general p-variate case, its exact null distribution is indicated, and its asymptotic null distribution is obtained.

1. Introduction. Let $X_1, X_2, \cdots, X_{n-1}$ be independent indentically distributed, p-dimensional random vectors defined on a probability space $(\Omega, \mathcal{Q}, P)$. Denote Euclidean p-space by R^p, and denote the corresponding class of Borel sets by $\mathfrak{B}^p$. The probability distribution of X_1 is the probability measure, $\hat{P}$, on $(R^p, \mathfrak{B}^p)$ that is defined for all B in $\mathfrak{B}^p$ by $\hat{P}(B) = P(X_1 \in B)$. Throughout, $\hat{P}$ is assumed to be absolutely continuous with respect to Lebesgue measure on $\mathfrak{B}^p$. For a specified absolutely continuous probability distribution $\hat{P}_H$, this paper proposes a test for the simple goodness-of-fit hypothesis $H: \hat{P} = \hat{P}_H$.

Perhaps a more conventional statement of the goodness-of-fit hypothesis is $H': G = G_H$, where $G(x)$ and $G_H(x)$ are the continuous distribution functions corresponding to the probability distributions $\hat{P}$ and $\hat{P}_H$. Statements H and H' are equivalent due to the unique correspondence between a probability distribution and its distribution function; however, the statements reflect different approaches to testing goodness-of-fit. A natural test for $H': G = G_H$ would seem to involve a comparison of the hypothesized distribution function G_H with an empirical distribution function, say G_n. The Kolmogorov-Smirnov and the Cramér-von Mises tests are well-known examples of such empirical distribution function (EDF) tests. The EDF tests reject H' when a measure of the difference between G_H and G_n is large. In general, EDF tests are distribution free for testing goodness-of-fit in the case of a univariate distribution, but the tests do not extend readily to distribution free tests in the multivariate case. To emphasize another inherent property of EDF tests, let $\hat{P}_n$ be the empirical probability distribution corresponding to an empirical distribution function G_n, and define B_x in $\mathfrak{B}^p$ by

$$B_x = \{y = (y_1, y_2, \cdots, y_p)' : y_j \leqslant x_j, \quad j = 1, 2, \cdots, p\}.$$

A comparison of $G_H(x)$ and $G_n(x)$ is a comparison of hypothesized probabilities, $\hat{P}_H(B_x)$, and empirical probabilities, $\hat{P}_n(B_x)$, of *only* events of the restricted form B_x

Received January 1978; revised July 1979.

AMS 1970 *subject classifications.* Primary 62G10; secondary 62E15.

Key words and phrases. Goodness-of-fit, empirical probability measure, nonparametric test, statistically equivalent blocks, spacings, multivariate distribution.

for x in R^p. This paper takes the view that the test should compare hypothesized and empirical probabilities of *all* events B in $\mathfrak{B}^p$. Accordingly, the proposed test for $H : \hat{P} = \hat{P}_H$ is based on a suitable empirical probability distribution $\hat{P}_n$, and the test statistic takes the form

$$F_n = \sup\{|\hat{P}_H(B) - \hat{P}_n(B)|;\ B \in \mathfrak{B}^p\}. \tag{1.1}$$

It compares the hypothesized and empirical probabilities of all possible events. Unlike EDF test statistics, F_n is distribution free in the general p-variate case.

In Section 2 a suitable empirical probability distribution $\hat{P}_n$ is specified for the test statistic in (1.1). Since the direct computation of F_n by (1.1) is impossible, a result is given in Section 3 that permits F_n to be computed indirectly. In Section 4, the null distribution of F_n is shown to be independent of both the hypothesized $\hat{P}_H$ and the dimension p; the exact null distribution is indicated; and the asymptotic null distribution is obtained. Finally, some empirical results are given in Section 5 for the univariate case ($p = 1$).

2. An empirical probability distribution. A standard notion of an empirical distribution function from the sample $X_1, X_2, \cdots, X_{n-1}$ is

$$G_n(x) = \frac{1}{n-1}\Sigma_{i=1}^{n-1}\left[\Pi_{j=1}^{p} I_{(-\infty, X_{ij}]}(x_j)\right],$$

where x_j and X_{ij} are the jth components of x and X_i respectively and where $I_A(x)$ is the indicator of the set A. The corresponding empirical probability distribution is discrete, giving mass $1/(n-1)$ to the points $X_1, X_2, \cdots, X_{n-1}$. This discrete empirical distribution would clearly not be appropriate for the test statistic (1.1) since, with $\hat{P}_H$ a continuous distribution, F_n would equal 1 almost surely. For the statistic F_n to be meaningful when $\hat{P}_H$ is a continuous distribution, a continuous empirical probability distribution, $\hat{P}_n$, must be used in (1.1). We will define such a continuous empirical distribution by "spreading" the mass over "statistically equivalent blocks" that are defined from the sample. The following prescription for using the sample to partition R^p into statistically equivalent blocks is taken from Anderson (1966):

Let $h_1(x), h_2(x), \cdots, h_{n-1}(x)$ be $n-1$ real valued functions, not necessarily different, such that the distribution of $h_\alpha(X)$ is a continuous distribution when X is distributed according to $\hat{P}, \alpha = 1, \cdots, n-1$. Let $k_1, k_2, \cdots, k_{n-1}$ be a permutation of $1, 2, \cdots, n-1$. Use $h_{k_1}(x)$ to order $X_1, X_2, \cdots, X_{n-1}$; and define $X^{(k_1)}$ as the k_1th in this ordering. Then the "cut"

$$h_{k_1}(x) = h_{k_1}(X^{(k_1)})$$

defines two "blocks"

$$B_{1\cdots k_1} = \{x : h_{k_1}(x) \leqslant h_{k_1}(X^{(k_1)})\},$$
$$B_{k_1+1\cdots n} = \{x : h_{k_1}(X^{(k_1)}) < h_{k_1}(x)\}.$$

The procedure is continued. Suppose $0 < k_2 < k_1$. Use $h_{k_2}(\boldsymbol{x})$ to order the $k_1 - 1$ $\boldsymbol{X}_\alpha$'s in $B_{1\cdots k_1}$, and define $\boldsymbol{X}^{(k_2)}$ as the k_2th in this ordering. Let

$$B_{1\cdots k_2} = B_{1\cdots k_1} \cap \{\boldsymbol{x} : h_{k_2}(\boldsymbol{x}) \leqslant h_{k_2}(\boldsymbol{X}^{(k_2)})\},$$

$$B_{k_2+1\cdots k_1} = B_{1\cdots k_1} \cap \{\boldsymbol{x} : h_{k_2}(\boldsymbol{X}^{(k_2)}) < h_{k_2}(\boldsymbol{x})\}.$$

If $k_1 < k_2$, rank the $n - k_1$ $\boldsymbol{X}_\alpha$'s in $B_{k_1+1\cdots n}$ according to $h_{k_2}(\boldsymbol{x})$, and let $\boldsymbol{X}^{(k_2)}$ be the $(k_2 - k_1)$th in the ranking. Then

$$B_{k_1+1\cdots k_2} = B_{k_1+1\cdots n} \cap \{\boldsymbol{x} : h_{k_2}(\boldsymbol{x}) \leqslant h_{k_2}(\boldsymbol{X}^{(k_2)})\},$$

$$B_{k_2+1\cdots n} = B_{k_1+1\cdots n} \cap \{\boldsymbol{x} : h_{k_2}(\boldsymbol{X}^{(k_2)}) < h_{k_2}(\boldsymbol{x})\}.$$

At the end of the mth stage there will be $m + 1$ blocks: $B_{1\cdots j_1}, B_{j_1+1\cdots j_2}, \cdots, B_{j_m+1\cdots n}$, where $j_1 < \cdots < j_m$ are $k_1, \cdots, k_m$ arranged in ascending order; and, finally, after $n - 1$ stages there will be n blocks $B_1, B_2, \cdots, B_n$ in $\mathfrak{B}^p$.

The blocks are statistically equivalent in the sense of

THEOREM 2.1. *Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{n-1}$ be independent with common probability distribution $\hat{P}$ on $(R^p, \mathfrak{B}^p)$. Let $B_1, B_2, \cdots, B_n$ be the statistically equivalent blocks that are constructed from the sample using cutting functions $h_\alpha(\boldsymbol{x})$, $\alpha = 1, 2, \cdots, n - 1$. If $h_\alpha(\boldsymbol{X}_1)$ has a continuous distribution for $\alpha = 1, 2, \cdots, n-1$, then $\hat{P}(B_1), \hat{P}(B_2), \cdots, \hat{P}(B_n)$ are distributed as the n spacings (coverages) determined from a random sample of size $n - 1$ from the uniform distribution on* (0, 1).

The proof is given in Anderson (1966).

A consequence of Theorem 2.1 is that $E\hat{P}(B_\alpha) = 1/n$ for $\alpha = 1, 2, \cdots, n$. It is therefore natural to specify an empirical distribution, $\hat{P}_n$, on $(R^p, \mathfrak{B}^p)$ by

DEFINITION 2.1. The probability measure $\hat{P}_n$ is an empirical distribution on $(R^p, \mathfrak{B}^p)$ if it satisfies

$$\hat{P}_n(B_\alpha) = 1/n \tag{2.1}$$

for $\alpha = 1, 2, \cdots, n$.

The specifications (2.1) defines a unique probability measure on the algebra generated by the statistically equivalent blocks $B_1, B_2, \cdots, B_n$. Since this algebra is contained in $\mathfrak{B}^p$, $\hat{P}_n$ may be extended to a distribution on $\mathfrak{B}^p$. The intent of Definition 2.1 is to call any such extension an empirical distribution on $\mathfrak{B}^p$.

It has already been pointed out that a continuous empirical distribution, $\hat{P}_n$, is required for the test statistic F_n of (1.1): $\hat{P}_n$ should spread the mass $1/n$ continuously over each statistically equivalent block. In Definition 2.2 we let the hypothesized $\hat{P}_H$ dictate the manner in which this mass is distributed over each block. For a subset B of the block B_α, the definition requires $\hat{P}_n(B)$ to be proportional to $\hat{P}_H(B)$, the constant of proportionality being $1/(n\hat{P}_H(B_\alpha))$.

DEFINITION 2.2. Let $B_1, B_2, \cdots, B_n$ be the statistically equivalent blocks that are constructed from the sample $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{n-1}$; and let $\hat{P}_H$ be the hypothesized distribution used in the definition of F_n in (1.1). The empirical distribution in (1.1), $\hat{P}_n$, is defined for each B in $\mathcal{B}^p$ by

$$\hat{P}_n(B) = \Sigma_{\alpha=1}^{n} \frac{1}{n} \frac{\hat{P}_H(B \cap B_\alpha)}{\hat{P}_H(B_\alpha)}.$$

3. A computational result.

The computation of the statistic

$$F_n = \sup\{|\hat{P}_H(B) - \hat{P}_n(B)|;\ B \in \mathcal{B}^p\}$$

would seem to require the impossible task of directly comparing hypothesized and empirical probabilities of all events in $\mathcal{B}^p$. That F_n may be computed indirectly is a consequence of

THEOREM 3.1. *Let $B_1, B_2, \cdots, B_n$ be the statistically equivalent blocks constructed from the sample $X_1, X_2, \cdots, X_{n-1}$; and define $D_{n1} < D_{n2} < \cdots < D_{nn}$ to be the ordered values of $\hat{P}_H(B_1), \hat{P}_H(B_2), \cdots, \hat{P}_H(B_n)$. Then*

$$F_n = \max\{\alpha/n - (D_{n1} + D_{n2} + \cdots + D_{n\alpha});\ \alpha = 1, 2, \cdots, n-1\}. \tag{3.1}$$

The proof requires

LEMMA 3.1. *For $\hat{P}_n$ given by Definition 2.2 and with B a Borel subset of a statistically equivalent block B_α, we have*

$$0 \leqslant \frac{\hat{P}_H(B) - \hat{P}_n(B)}{\hat{P}_H(B_\alpha) - \hat{P}_n(B_\alpha)} \leqslant 1.$$

PROOF. With $\hat{P}_n(B) = (\hat{P}_H(B)/\hat{P}_H(B_\alpha))/n$ and with $\hat{P}_n(B_\alpha) = 1/n$, straightforward algebraic manipulations show

$$\frac{\hat{P}_H(B) - \hat{P}_n(B)}{\hat{P}_H(B_\alpha) - \hat{P}_n(B_\alpha)} = \frac{\hat{P}_H(B)}{\hat{P}_H(B_\alpha)}$$

the right side of which is easily seen to be between 0 and 1 when B is a subset of B_α.

PROOF OF THEOREM 3.1. Define $n + 1$ classes of events in $\mathcal{B}^p$ by

$$\begin{aligned}
\mathcal{B}_{(0)} &= \{\phi\} \\
\mathcal{B}_{(1)} &= \{B_1, B_2, \cdots, B_n\} \\
\mathcal{B}_{(2)} &= \{B_{j_1} \cup B_{j_2};\ 1 \leqslant j_1 < j_2 \leqslant n\} \\
\mathcal{B}_{(n-1)} &= \{B_{j_1} \cup \cdots \cup B_{j_{n-1}};\ 1 \leqslant j_1 < \cdots < j_{n-1} \leqslant n\} \\
\mathcal{B}_{(n)} &= \{R^p\},
\end{aligned}$$

and write

$$\mathfrak{B}_X = \mathfrak{B}_{(0)} \cup \mathfrak{B}_{(1)} \cup \cdots \cup \mathfrak{B}_{(n)}$$

as the algebra generated by the statistically equivalent blocks. The first step in proving Theorem 3.1 is to show that F_n may be constructed solely from events in $\mathfrak{B}_X$ by showing

$$\text{(3.2)}\quad \sup\{|\hat{P}_H(B) = \hat{P}_n(B)|;\ B \in \mathfrak{B}^p\} = \max\{|\hat{P}_H(B) - \hat{P}_n(B)|;\ B \in \mathfrak{B}_X\}.$$

Select an arbitrary B in $\mathfrak{B}^p$. Expression (3.2) will follow upon producing a corresponding event B^* in $\mathfrak{B}_X$ for which

$$\text{(3.3)}\qquad |\hat{P}_H(B^*) - \hat{P}_n(B^*)| \geqslant |\hat{P}_H(B) - \hat{P}_n(B)|.$$

In the case where $\hat{P}_H(B) - \hat{P}_n(B) > 0$, define a subclass $\mathcal{C}$ of statistically equivalent blocks by

$$\mathcal{C} = \{B_\alpha : \hat{P}_H(B \cap B_\alpha) - \hat{P}_n(B \cap B_\alpha) > 0\},$$

and select the required

$$B^* = \bigcup\nolimits_{B_\alpha \in \mathcal{C}} B_\alpha.$$

To see that B^* satisfies (3.3) use the definition of $\mathcal{C}$ and Lemma 3.1 to argue the two inequalities below:

$$\begin{aligned}
\hat{P}_H(B) - \hat{P}_n(B) &= \Sigma_{\alpha=1}^{n} \hat{P}_H(B \cap B_\alpha) - \hat{P}_n(B \cap B_\alpha) \\
&\leqslant \Sigma_{B_\alpha \in \mathcal{C}} \hat{P}_H(B \cap B_\alpha) - \hat{P}_n(B \cap B_\alpha) \\
&\leqslant \Sigma_{B_\alpha \in \mathcal{C}} \hat{P}_H(B_\alpha) - \hat{P}_n(B_\alpha) \\
&= \hat{P}_H(B^*) - \hat{P}_n(B^*).
\end{aligned}$$

The selection of B^* when $\hat{P}_H(B) - \hat{P}_n(B) < 0$ may be made similarly to complete the justifications of (3.3) and (3.2).

From (3.2) and the definition of the $\mathfrak{B}_{(\alpha)}$'s

$$\text{(3.4)}\qquad F_n = \max_{\alpha=1,2,\cdots,n-1} \max\{|\hat{P}_H(B) - \alpha/n|;\ B \in \mathfrak{B}_{(\alpha)}\}.$$

Note that the right side of (3.4) may be expressed in terms of the D_{nj}'s: For $\alpha = 1$, since $D_{nn} = 1 - D_{n1} = \cdots = -D_{nn-1}$,

$$\begin{aligned}
\max\{|\hat{P}_H(B) - 1/n|;\ B \in \mathfrak{B}_{(1)}\} &= \max\{1/n - D_{n1}, D_{nn} - 1/n\} \\
&= \max\{1/n - D_{n1}, (n-1)/n \\
&\qquad - (D_{nl} + D_{n2} + \cdots + D_{nn-1})\};
\end{aligned}$$

and, in general,

$$\begin{aligned}
\text{(3.5)}\quad &\max\{|\hat{P}_H(B) - \alpha/n|;\ B \in \mathfrak{B}_{(\alpha)}\} \\
&= \max\{\alpha/n - (D_{n1} + \cdots + D_{n\alpha}), (n-\alpha)/n - (D_{n1} + \cdots + D_{nn-\alpha})\}.
\end{aligned}$$

The theorem is proved upon combining (3.4) and (3.5) to conclude (3.1).

A further simplification in the computation of F_n is obtained by selecting α^* to satisfy

$$1/n - D_{n1} > 1/n - D_{n2} > \cdots > 1/n - D_{n\alpha^*} > 0 > 1/n - D_{n(\alpha^*+1)} > \cdots > 1/n - D_{nn}.$$

It follows that we may compute

$$F_n = \alpha^*/n - (D_{n1} + D_{n2} + \cdots + D_{n\alpha^*}). \tag{3.6}$$

4. The null distribution of F_n. Let $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n-1)}$ be the order statistics from a random sample of size $n-1$ from the uniform distribution on (0, 1). The corresponding spacings are defined to be $Y_{(1)}, Y_{(2)} - Y_{(1)}, \cdots, Y_{(n-1)} - Y_{(n-2)}, 1 - Y_{(n-1)}$. Denote the ordered spacings by $U_{n1}, U_{n2}, \cdots, U_{nn}$. The null distribution of F_n is characterized in terms of these ordered spacings in

THEOREM 4.1. *Assume the conditions of Theorem* 2.1. *When $H: \hat{P} = \hat{P}_H$ is true, F_n is distributed as a function of ordered uniform spacings*:

$$F_n = \max\{\alpha/n - (U_{n1} + U_{n2} + \cdots + U_{n\alpha}); \alpha = 1, 2, \cdots, n-1\}.$$

In particular, the null distribution of F_n is both independent of the hypothesized $\hat{P}_H$ and the dimension p.

PROOF. From (3.1)

$$F_n = \max\{\alpha/n - (D_{n1} + D_{n2} + \cdots + D_{n\alpha}); \alpha = 1, 2, \cdots, n-1\}, \tag{4.1}$$

where $D_{n1}, D_{n2}, \cdots, D_{nn}$ are the ordered values of $\hat{P}_H(B_1), \hat{P}_H(B_2), \cdots, \hat{P}_H(B_n)$. That $\hat{P}_H(B_1), \hat{P}_H(B_2), \cdots, \hat{P}_H(B_n)$ are distributed as uniform spacings and that $D_{n1}, D_{n2}, \cdots, D_{nn}$ are, thus, ordered uniform spacings follows from Theorem 2.1. The proof is completed upon identifying D_{nj} with U_{nj} in (4.1) for $j = 1, 2, \cdots, n$.

To find the exact null distribution of F_n, note from Mauldon (1951) that the joint distribution of $U_{n1}, U_{n2}, \cdots, U_{nn-1}$ has density function $f_n(\mu_1, \mu_2, \cdots, \mu_{n-1}) = n!(n-1)!$ for $\mu_1, \mu_2, \cdots, \mu_{n-1}$ satisfying $0 < \mu_1 < \mu_2 < \cdots < \mu_{n-1}$ and $\mu_1 + \mu_2 + \cdots + \mu_{n-2} + 2\mu_{n-1} < 1$. The transformation of variables $\Delta_1 = 1/n - U_{n1}, \Delta_2 = 2/n - U_{n1} - U_{n2}, \cdots, \Delta_{n-1} = (n-1)/n - U_{n1} - \cdots - U_{n(n-1)}$ shows the joint density function of $\Delta_1, \Delta_2, \cdots, \Delta_{n-1}$ to be $g_n(\delta_1, \delta_2, \cdots, \delta_{n-1}) = n!(n-1)!$ for $1/n > \delta_1 > \delta_2 - \delta_1 > \delta_3 - \delta_2 > \cdots > \delta_{n-1} - \delta_{n-2} > -\delta_{n-1}$. Finally, $P(F_n \leqslant x)$ may be found from

$$P(F_n \leqslant x) = \textstyle\int_{-\infty}^{x} \cdots \int_{-\infty}^{x} g_n(\delta_1, \cdots, \delta_{n-1})\, d\delta_1 \cdots d\delta_{n-1}. \tag{4.2}$$

In the simplest cases the integration gives

$$\begin{aligned} P(F_3 \leqslant x) &= 1, \quad \text{for } x > \tfrac{2}{3} \\ &= 12x \min(\tfrac{1}{3}, x) - 12x \min(\tfrac{1}{3}, x/2) + 3x^2 - 3\min(\tfrac{1}{9}, x^2), \\ &\qquad\qquad \text{for } 0 < x < \tfrac{2}{3}, \\ &= 0, \quad \text{for } x < 0 \end{aligned}$$

and

$$\begin{aligned} P(F_4 \leqslant x) &= 1, \quad \text{for } x > \tfrac{3}{4} \\ &= 96\min^3\left(\tfrac{1}{4}, x/3\right) + 72x^2\min\left(\tfrac{1}{4}, x\right) + 4\min^2\left(\tfrac{1}{4}, x\right) \\ &\quad + 36x^2\min\left(\tfrac{1}{4}, x/3\right) + 12\min^2\left(\tfrac{1}{4}, x/3\right) - 108x^2\min\left(\tfrac{1}{4}, x/2\right) \\ &\quad + 144x\min^2\left(\tfrac{1}{4}, x/2\right) - 108x\min^2\left(\tfrac{1}{4}, x/3\right) - 36x\min^2\left(\tfrac{1}{4}, x\right) \\ &\quad - 48\min^3\left(\tfrac{1}{4}, x/2\right), \quad \text{for } 0 \leqslant x \leqslant \tfrac{3}{4} \\ &= 0, \quad \text{for } x < 0. \end{aligned}$$

The tedium in evaluating (4.2) increases rapidly with n. However, for large n the null distribution of F_n may be approximated using the normal approximation of

THEOREM 4.2. *Under the assumption of Theorem* 2.1, *when* $H: \hat{P} = \hat{P}_H$ *is true:*

$$\lim_{n\to\infty} \sup_{-\infty<x<\infty} \left| P\left[n^{\frac{1}{2}}(F_n - e^{-1})/(2e^{-1} - 5e^{-2})^{\frac{1}{2}} \leqslant x \right] - \int_{-\infty}^{x} (2\Pi)^{-\frac{1}{2}} e^{-\alpha^2/2} d\alpha \right| = 0.$$

REMARK 4.1. For the case $p = 1$, the Glivenko-Cantelli Theorem tells us that $\hat{P}$ may be estimated by $\hat{P}_n$ with uniform accuracy over all events of the form $B_x = (-\infty, x]$, $x \in R$: It concludes

$$\sup_{-\infty<x<\infty} |\hat{P}(B_x) - \hat{P}_n(B_x)| \to 0$$

almost surely as $n \to \infty$. Blum (1955) proves a related result for the general p-dimensional case. It is of interest to recognize that $\hat{P}$ may not be estimated by $\hat{P}_n$ uniformly over all of $\mathfrak{B}^p$, for Theorem 4.2 implies that

$$\sup_{B \in \mathfrak{B}^p} |\hat{P}(B) - \hat{P}_n(B)| \to e^{-1},$$

where the nonzero limit is in probability as $n \to \infty$.

PROOF OF THEOREM 4.2. Let $U_{n1}, U_{n2}, \cdots, U_{nn}$ be the ordered spacings from a random sample of $n - 1$ random variables distributed uniformly over (0, 1). Fix b to satisfy $1 - e^{-1} < b < 1$. With $[nb]$ representing the largest integer smaller than or equal to nb, the ordered spacings process is defined on $(1/n, [nb]/n)$ to be

$$U_n(t) = n^{\frac{1}{2}}[nU_{ni} + \log(1 - t)]$$

for $(i - 1)/n < t \leqslant i/n$, $2 \leqslant i \leqslant [nb]/n$.

Write U for the normal process on (0, 1) having continuous sample paths, mean value function 0, and covariance function $K(s, t) = s/(1 - s) - \log(1 - s)\log(1 - t)$, for $0 < s \leqslant t < 1$. Let $D = D(0, 1)$ be the space of right continuous functions on (0, 1) that have left hand limits, and extend U_n to the element U_n^* in D defined by

$$\begin{aligned} U_n^*(t) &= U_n(t), \qquad \text{if } 1/n \leqslant t < [nb]/n \\ &= U(t), \qquad \text{otherwise.} \end{aligned}$$

Let $Q(\nearrow)$ denote the class of positive, strictly increasing continuous functions q on [0, 1] for which $\int_0^1 q^{-2}(t)\,dt < \infty$. Let Q denote the class of all q such that $q(t) = q(1-t) = \bar{q}(t)$ for $0 \leqslant t \leqslant \frac{1}{2}$ and some $\bar{q} \in Q(\nearrow)$. Define the metric ρ_q on D by

$$\rho_q(f_1, f_2) = \sup_{0<t<1}|(f_1(t) - f_2(t))/q(t)|.$$

Finally, use the symbol $\rightarrow_P$ to denote convergence in probability.

Section 7 of Shorack (1972) shows that there is a probability space with special processes $\{D_n\}_{n\geqslant 1}$ and D defined thereon satisfying

$$\rho_{q/(1-I)}(D_n^*, D) \rightarrow_P 0 \tag{4.3}$$

as $n \rightarrow \infty$. The processes D_n^* all lie in $D(0, 1)$, and D lies in $C(0, 1)$. Furthermore, for each n, D_n and U_n generate the same measure on $D(0, 1)$, and D and U generate the same measure on $C(0, 1)$.

Let ρ be the Skorohod metric on D (cf. Billingsley 1968, page 111). To argue that (4.3) implies

$$\rho(D_n^*, D) \rightarrow_P 0, \tag{4.4}$$

take $q = [I(1 - I)]^{\frac{1}{4}} \in Q$, and observe that

$$\rho_{q/(1-I)}(D_n^*, D) = \sup_{0<t<1}\left|\frac{(1-t)(D_n^*(t) - D(t))}{[t(1-t)]^{\frac{1}{4}}}\right| \tag{4.5}$$

$$\geqslant \frac{1-b}{[b(1-b)]^{\frac{1}{4}}}\sup_{0<t<1}|D_n^*(t) - D(t)|.$$

From (4.3) and (4.5) we see that

$$\sup_{0<t<1}|D_n^*(t) - D(t)| \rightarrow_P 0 \tag{4.6}$$

as $n \rightarrow \infty$.

Since the sup metric dominates the Skorohod metric, (4.6) implies the result (4.4). Because the space D is separable in the metric ρ(cf. Billingsley 1968, 123), we may conclude from (4.4) that D_n^* converges weakly to the process D as $n \rightarrow \infty$ (cf. Billingsley, 1968, Theorem 4.1). Therefore, denoting this weak convergence in D by $\rightarrow_w$, we also conclude that

$$U_n^* \rightarrow_w U. \tag{4.7}$$

From Theorem 4.1 the test statistic F_n has representation

$$F_n = \max\{\alpha/n - (U_{n1} + U_{n2} + \cdots + U_{n\alpha});\ \alpha = 1, 2, \cdots, n-1\}.$$

Define F_{nb} to be a truncated form of F_n as

$$F_{nb} = \max\{\alpha/n - (U_{n1} + U_{n2} + \cdots + U_{n\alpha});\ \alpha = 1, 2, \cdots, [nb]/n\}.$$

We will show that F_n and F_{nb} have the same limiting distribution whenever b is greater than $1 - e^{-1}$ by showing

$$\lim_{n\rightarrow\infty} P(F_n \neq F_{nb}) = 0. \tag{4.8}$$

To prove (4.8), note from (3.6) that $F_n = \alpha^*/n - (U_{n1} + U_{n2} + \cdots + U_{n\alpha^*})$ for α^* satisfying $1/n - U_{n\alpha^*} > 0 > 1/n - U_{n(\alpha^*+1)}$. We thus see the equality of events $[F_n \neq F_{nb}]$, $[\alpha^* > [nb]/n]$, and $[1/n - U_{n([nb]/n+1)} > 0]$. We may therefore write

$$\begin{aligned} P(F_n \neq F_{nb}) &= P(1/n - U_{n([nb]/n+1)} > 0) \\ &= P\Big[n^{\frac{1}{2}}(nU_{n([nb]/n+1)} + \log(1 - b)) < n^{\frac{1}{2}}(1 + \log(1 - b))\Big], \end{aligned}$$

and (4.8) follows upon noting that $1 + \log(1 - b)$ is negative and upon noting from (4.7) that $n^{\frac{1}{2}}(nU_{n([nb]/n+1)} + \log(1 - b))$ is asymptotically normal with mean 0 and variance $b/(1 - b) - \log^2(1 - b)$.

To introduce more notation, let $[nt]$ be the largest integer smaller than or equal to nt. Define

$$\begin{aligned} \mu_t^n &= -\textstyle\int_{1/n}^{[nt]/n} \log(1 - s)\, ds, \\ \mu_t &= \lim_{n\to\infty} \mu_t^n = (1 - t)\log(1 - t) + t, \\ \sigma_t^2 &= 2t(1 - t) + 2(1 - t)^2 \log(1 - t) - (1 - t)^2 \log^2(1 - t), \\ l &= 1 - e^{-1} \\ \sigma_l^2 &= 2e^{-1} - 5e^{-2}. \end{aligned}$$

Also, for a functional $h : D \to D$ and for $x \in D$, denote the image function by $h(x)$. Further denote the value of the image function at t, $0 < t < 1$, by $h(x)(t)$.

The remainder of the proof involves representing F_{nb} as a functional of U_n^*:

$$n^{\frac{1}{2}}(F_{nb} - [nl]/n + \mu_l^n) = \sup\{ g_n(h_n(U_n^*))(t);\ 0 < t < b\} \tag{4.9}$$

where h_n and g_n are measurable mappings from D into D having measurable limits h and g. It will then follow that the left side of (4.9) is asymptotically distributed as $\sup_{0<t<b} g(h(U))(t)$. Finally, the distribution of $\sup_{0<t<b} g(h(U))(t)$ will be found to be normal with mean 0 and variance $2e^{-1} - 5e^{-2}$.

To proceed with the proof, define $h_n : D \to D$ by

$$h_n(x)(t) = \textstyle\int_{1/n}^{[nt]/n} x(s)\, ds.$$

Note that h_n is a continuous, thus measurable, mapping on the metric space D with metric ρ. The ρ-limit of h_n is the measurable functional h defined for each $x \in D$ by

$$h(x)(t) = \textstyle\int_0^t x(s)\, ds.$$

Next, define the measurable functional $g_n : D \to D$ by

$$g_n(x)(t) = \max\Big\{n^{\frac{1}{2}}\big([nt]/n - x(t)/n^{\frac{1}{2}} - \mu_t^n - [nl]/n + \mu_l^n\big),\ -x(l)\Big\}.$$

To evaluate $\lim g_n(x)$ for $x \in D$, note that $l = 1 - e^{-1}$ is chosen so that $t - \mu_t$ is less than $l - \mu_l$ for all $t \neq l$. Now, for $t \neq l$,

$$n^{\frac{1}{2}}([nt]/n - \mu_t^n - [nl]/n + \mu_l^n) \to -\infty$$

as $n \to \infty$, and we may argue that

$$\lim_{n\to\infty} g_n(x)(t) = -x(l)$$

for all t. Define $g : D \to D$ by

$$g(x)(t) = -x(l)$$

for all $0 < t < 1$.

Let E be the set of functions $x \in D$ such that

$$\rho(g_n(h_n(x_n)), g(h(x))) \to 0$$

fails to hold for some sequence $\{x_n\}$ approaching x in the ρ metric. A tedious but direct argument shows that E may contain only discontinuous functions $x \in D$. Since $U_n^* \to_w U$ and since U has continuous sample paths, it follows from Billingsley (1968, Theorem 5.5) that

$$g_n(h_n(U_n^*)) \to_w g(h(U)). \tag{4.10}$$

Since $\sup x(t)$ is a ρ-continuous function on D, it follows from (4.10) and Billingsley (1968, Theorem 5.1) that

$$\sup_{0<t<b} g_n(h_n(U_n^*))(t) \to_w \sup_{0<t<b} g(h(U))(t) = -h(U)(l). \tag{4.11}$$

To evaluate the left side of (4.11), write for $0 < t < b$

$$\begin{aligned} h_n(U_n^*)(t) &= \int_{1/n}^{[nt]/n} U_n(s)\, ds \\ &= n^{\frac{1}{2}}\sum_{i=1}^{[nt]} U_{ni} - n^{\frac{1}{2}}\int_{1/n}^{[nt]/n} \log(1/(1-s))\, ds \\ &= n^{\frac{1}{2}}\sum_{i=1}^{[nt]} U_{ni} - n^{\frac{1}{2}}\mu_t^n, \end{aligned}$$

and proceed to evaluate

(4.12)

$$\begin{aligned} \sup_{0<t<b} g_n(h_n(U_n^*))(t) &= \sup_{0<t<b} \max\Big\{n^{\frac{1}{2}}\big([nt]/n - \textstyle\sum_{i=1}^{[nt]} U_{ni} - [nl]/n + \mu_l^n\big), \\ &\qquad n^{\frac{1}{2}}\big([nl]/n - \textstyle\sum_{i=1}^{[nl]} U_{ni} - [nl]/n + \mu_l^n\big)\Big\} \\ &= n^{\frac{1}{2}} \max\big\{\alpha/n - \textstyle\sum_{i=1}^{\alpha} U_{ni} - [nl]/n + \mu_l^n; \\ &\qquad\qquad \alpha = 1, 2, \cdots, [nb]/n\big\} \\ &= n^{\frac{1}{2}}(F_{nb} - [nl]/n + \mu_l^n). \end{aligned}$$

To determine the distribution of $-h(U)(l)$ on the right side of (4.11), observe that $h(U)(l)$ is obtained by integrating the normal process:

$$h(U)(l) = \int_0^l U(s)\, ds.$$

It follows that $-h(U)(l)$ is itself normally distributed:

$$-h(U)(l) \sim N\Big(0, \int_0^l\int_0^l k(s, t)\, ds\, dt\Big), \tag{4.13}$$

where

$$\int_0^l\int_0^l k(s, t)\, ds\, dt = \sigma_l^2 = 2e^{-1} - 5e^{-2}$$

(cf. Parzen 1962, Theorem 3A, page 79 and Theorem 4B, page 91).

Returning to (4.11), we see from (4.12) and (4.13) that

$$n^{\frac{1}{2}}(F_{nb} - [nl]/n + \mu_l^n) \to_w - h(U)(l) \sim N(0, 2e^{-1} - 5e^{-2}).$$

Thus, also,

$$n^{\frac{1}{2}}(F_n - l + \mu_l) \to_w - h(U)(l) \sim N(0, 2e^{-1} - 5e^{-2}). \tag{4.14}$$

Theorem 4.2 is now a consequence of (4.14), the equality $l - \mu_l = e^{-1}$, and Polya's Theorem (cf. Rao, 1973, page 120).

5. Empirical results. For the univariate case ($p = 1$) the test statistic F_n may be computed by constructing statistically equivalent blocks $B_1, B_2, \cdots, B_n$ from cutting functions $h_\alpha(x) = x, \alpha = 1, 2, \cdots, n - 1$. With $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n-1)}$ the ordered sample, the blocks are then simply the intervals $B_1 = (-\infty, Y_{(1)}], B_2 = (Y_{(1)}, Y_{(2)}], \cdots, B_n = (Y_{(n-1)}, \infty)$. For testing the null hypothesis of a standard normal distribution, $\hat{P}_H$ is the corresponding normal probability measure; $D_{n1} < D_{n2} < \cdots < D_{nn}$ are the ordered values of $\hat{P}_H(B_1), \hat{P}_H(B_2), \cdots, \hat{P}_H(B_n)$; and F_n is constructed from the D_{nj}'s according to (3.6).

This proposed test for the standard normal distribution is compared with the Kolmogorov-Smirnov test and the chi-squared test as follows: Each of the three tests is performed on ten simulated samples of $n - 1 = 50$ observations from a "mixed uniform" distribution U^* that is coded as

$$\begin{aligned} U^* = {} & .1U(-1.6449 \pm .12) + .1U(-1.0364 \pm .12) + .1U(-.6745 \pm .12) \\ & + .1U(-.3858 \pm .12) + .1U(-.1257 \pm .12) + .1U(.1257 \pm .12) \\ & + .1U(.3853 \pm .12) + .1U(.6745 \pm .12) + .1U(1.0364 \pm .12) \\ & + .1U(1.6449 \pm .12). \end{aligned}$$

The notation $U(\Theta \pm .12)$ is for the uniform distribution from $\Theta - .12$ to $\Theta + .12$, and U^* above is a mixture with probabilities .1 each of the ten uniform distributions centered at standard normal percentials of 5%, 15%, 25%, $\cdots$, 85%, and 95%.

For constructing the chi-squared test, ten groups with an expected number of 5 in each group are taken. Intentionally, this chi-squared test is not expected to perform well against the alternative U^* as its power at U^* equals its significance level for testing the null hypothesis of a standard normal distribution. The Kolmogorov-Smirnov test is also found to perform poorly for the samples considered. The interest in the example is in providing (empirical) evidence that the proposed F_{51} test may perform better than chi-squared and Kolmogorov-Smirnov tests in some situations.

The results are included in Table 1. Entries in the table are values of the appropriate test statistic. Significance at the 5% and 1% levels are denoted by single and double asterisks respectively. The critical values for the three test statistics are obtained from the exact null distribution for the Kolmogorov-Smirnov statistic (K) and from the limiting null distributions for the chi-squared statistic (χ^2) and for the proposed statistic (F_{51}). The table also shows results for ten simulated samples of $n - 1 = 50$ observations from the hypothesized standard normal distribution.

TABLE 1

Empirical results for three goodness-of-fit tests on simulated samples

Population	Sample	X^2	K	F_{51}
		Test		
Mixed Uniform	1	6.4	.102	.428*
	2	5.2	.098	.470**
	3	3.6	.078	.460**
	4	7.6	.155	.456**
	5	4.4	.105	.456**
	6	9.6	.084	.454**
	7	14.0	.191*	.445*
	8	8.8	.119	.440*
	9	9.6	.100	.470**
	10	6.4	.126	.429*

Population	Sample	X^2	K	F_{51}
		Test		
Standard Normal	1	6.4	.092	.331
	2	16.4	.156	.410
	3	4.0	.108	.342
	4	8.8	.121	.404
	5	5.2	.126	.366
	6	17.6*	.199*	.351
	7	10.8	.135	.334
	8	10.4	.185	.380
	9	13.6	.230**	.330
	10	14.8	.175	.359

The critical values for the three statistics are:

	X^2	K	F_{51}		X^2	K	F_{51}
5%	16.92	.188	.424	1%	21.67	.226	.447

6. Concluding remarks. The literature provides a variety of tests for goodness-of-fit. Although no attempt is made here to review this literature, it should be pointed out that Durbin (1961) proposes goodness-of-fit statistics for the univariate case that are also functions of the variables $D_{n1}, D_{n2}, \cdots, D_{nn}$ that are defined in Theorem 3.1; these statistics thereby share a relationship with the test statistic F_n in (1.1) and in (3.1). Durbin (1961) also provides a randomization procedure whereby F_n may be used to test composite goodness-of-fit hypotheses including the hypothesis of a normal distribution with unknown mean and unknown variance.

Acknowledgment. The author would like to thank the referee for correcting an error in the proof of Theorem 4.2 and also for the apparent generous amount of time given to evaluating the paper.

REFERENCES

[1] ANDERSON, T. W. (1966). Some nonparametric procedures based on statistically equivalent blocks. *Proc. Internat. Symp. Multivariate Anal.* (P. R. Krishnaiah, ed.). 5-27. Academic, New York.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
[3] BLUM, J. R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* **26** 527–529.
[4] DURBIN, J. (1961). Some methods of constructing exact tests. *Biometrika.* **48** 41–55.
[5] MAULDON, J. G. (1951). Random division of an interval. *Proc. Camb. Phil. Soc.* **47** 331–336.
[6] PARZEN, E. (1962). *Stochastic Processes*. Holden-Day, San Francisco.
[7] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2d ed. Wiley, New York.
[8] SHORACK, GALEN, R. (1972). Convergence of quantile and spacings processes with applications. *Ann. Math. Statist.* **43** 1400–1411.

DEPARTMENT OF STATISTICS
VIRGINIA POLYTECHNIC INSTITUTE
& STATE UNIVERSITY
BLACKSBURG, VA 24061