A COMPARATIVE STUDY OF SEVERAL ONE-SIDED GOODNESS-OF-FIT TESTS[1]

By Douglas G. Chapman

University of Washington

0. Summary. Criteria for evaluating goodness-of-fit tests are reviewed and two additional criteria proposed. The several goodness-of-fit tests which have been proposed are studied in the light of these criteria. It is shown that it is relatively easy to evaluate the maximum and minimum power of those tests which are "partially ordered" against alternatives at a fixed "distance" from the hypothesis. A comparison is made of five tests on the basis of such minimum and maximum power functions.

1. Introduction. Let X be a real random variable with d.f. $\mathbf{F} \, \varepsilon \, \Omega_2$ the class of continuous distribution functions (d.f.) on R. The aim of this paper is a comparative study of some of the distribution-free tests of the hypothesis

$$H_0{:}F = F_0$$

(where F_0 is completely specified), against the alternative

$$F < F_0 .$$

The class of distributions belonging to Ω_2 that are less than F_0 will be denoted by $\tilde{\omega}$. (A distribution F is less than F_0 if $F(x) \leqq F_0(x)$ everywhere with the strict inequality holding on a set of positive F_0-measure.) Birnbaum and Scheuer [7] have called this problem that of testing goodness-of-fit against stochastically comparable alternatives. A list of a number of tests for this situation and for the case where the set of alternatives is $F \, \varepsilon \, \Omega_2$, $F \neq F_0$, as well as some of the considerations involved in designing such tests, have been given by Birnbaum [4].

If the goodness-of-fit test is merely a preliminary test to justify assumptions made for the purpose of further tests, its usefulness at the present time is debatable. As yet not enough is known of the effects of different types of deviations from assumptions on the behavior of statistical tests and estimates, nor of the effects of preliminary tests. Box and Andersen [9], however, have given examples which seem to indicate that the use of a preliminary test may leave the statistician in a less satisfactory position than if no preliminary test were made.

On the other hand the goodness-of-fit test is quite reasonable in validating a theoretical model. Moreover F, or functions of F, may enter into further developments of the whole problem so that it is desirable to have an explicit representation for it.

Received July 1, 1957; revised April 11, 1958.

[1] This research was initiated while the author held a Guggenheim Fellowship at the Readership in the Design and Analysis of Scientific Experiment at Oxford University. It was completed with the support of the O.N.R.

In many statistical applications tests are made for changes in a mean; equally well changes in the whole distribution may be of interest. In this case one-sided as well as two-sided alternatives to H_0 could be of interest to the statistician.

In [4] Birnbaum noted that it is desirable to introduce a metric into the space of distributions and he suggested a number of possibilities. The choice of the metric is to a large extent a metastatistical consideration. However, the metric

$$\rho(F, G) = \sup_{-\infty < x < \infty} | F(x) - G(x) |$$

or in the one-sided case

$$\rho^{-}(F, G) = \sup_{-\infty < x < \infty} (F(x) - G(x))$$

has been used extensively in probability and statistics. Furthermore these metrics seem appropriate in several of the situations discussed above where a test of H_0 is reasonable. We will consider only these distance functions and more especially the second which is appropriate to stochastically comparable alternatives. This study will be limited to those tests which have been proposed for this problem and for which the distribution theory of the test under the null hypothesis is known at least for the asymptotic case. For those tests that satisfy certain weak criteria, the maximum and minimum large sample power for alternatives whose distance from the hypothesis is equal to Δ, is determined. This approach of finding sharp upper and lower bounds for the power of a test for such alternatives was introduced by Birnbaum in [5].

The almost standard notation

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt$$

and Z_α for the root of the equation

$$\Phi(x) = \alpha$$

will be used.

$E_0 [f(X)]$ and $E_G [f(X)]$ will denote the expectation of the random function $f(X)$ when X has the distributions F_0 and G respectively.

2. Criteria for tests of H_0. A test of H_0 of size α is a measurable function $\varphi(X_1 \cdots X_n)$ or φ_n for short, from R_n to the interval (0, 1) such that

$$E_0(\varphi_n) \leqq \alpha.$$

Consider the alternative $G \varepsilon \bar{\omega}$. The power function $E_G(\varphi_n)$ will be denoted $\beta_{\varphi_n}(G)$.

The properties of admissibility, consistency and unbiasedness for a test are well known. We refer to Birnbaum and Rubin [6] for the concepts of tests of structure (d), distribution-free and strongly distribution-free tests, and recall that they showed that for all strictly monotone distributions in Ω_2 tests of structure (d) are strongly distribution free and conversely.

Since all tests we will consider are of structure (d) we may consider the problem in its canonical form, i.e., where F_0 is the uniform distribution on the interval (0, 1) and all distributions of $\tilde{\omega}$ are restricted to the unit interval.

To emphasize this, it will be convenient to let u be a sure number in (0, 1) and U an r.v. uniformly distributed in (0, 1). It will also be convenient to denote by $U_1, U_2, \cdots, U_n$ the *ordered* sample from this distribution. In some instances it will be convenient to introduce U_0 and U_{n+1}. These are set equal to 0 and 1 respectively.

We also introduce two more concepts, monotonicity and partial ordering, as applied to tests of the hypothesis H_0.

DEFINITION 1. φ is a monotone test of H_0 if

$$(1)\quad U_i \geqq V_i \quad (i = 1, 2, \cdots n) \Rightarrow (U_1, U_2, \cdots U_n) \geqq (V_1, V_2, \cdots V_n).$$

DEFINITION 2. φ is a partially ordered [p.o.] test of H_0 if

$$(2)\qquad G_1(u) \leqq G_2(u) \qquad \text{for all} \qquad u \,\varepsilon\, (0, 1) \Rightarrow \beta_\varphi(G_1) \geqq B_\varphi(G_2).$$

From the continuity theorem for Lebesgue-Stieltjes integrals we have the following obvious

REMARK. If φ is continuous except for a finite number of jumps and φ is p.o. then φ is unbiased.

The relationship between monotonicity and partial ordering will be useful later.

THEOREM. *Tests of structure* (d) *that are monotone are* p.o.

PROOF. Let $G_1(u) \leqq G_2(u) \leqq u$ and recall

$$(3)\qquad \beta_\varphi(G_i) = \int_0^1 \int_0^1 \cdots \int_0^1 \varphi(u_1, u_2, \cdots, u_n) \prod_{j=1}^n dG_i(u_j) \qquad (i = 1, 2).$$

Make the change of variables

$$y_j = G_i(u_j) \qquad (j = 1, 2, \cdots n)(i = 1, 2)$$

in the two integrals. The inverse is defined in the usual fashion, i.e.,

$$u_j = G_i^{-1}(y_j) = \inf_{0 \leqq x \leqq 1} [x : G_i(x) = y_j].$$

The two integrals become

$$(4)\qquad \int_0^1 \int_0^1 \cdots \int_0^1 \varphi[G_i^{-1}(y_1), \cdots, G_i^{-1}(y_n)] \prod_{j=1}^n dy_j \qquad (i = 1, 2).$$

Since $G_1 \leqq G_2$, $G_1^{-1} \geqq G_2^{-1}$; this, together with the monotonicity property of φ, implies the required inequality for $\beta_\varphi(G_1)$, $\beta_\varphi(G_2)$.

It may also be noted that any monotone test is admissible. This follows from a result of A. Birnbaum [2] (appendix) who considered this problem where the set of alternatives is restricted to d.f. with monotone densities. In this paper we determine which of the several tests of H_0 that have been suggested satisfy these criteria and then determine

$$\underline{\beta}_\varphi(\Delta) = \inf_{G \varepsilon \mathcal{G}(\Delta)} \beta_\varphi(G); \qquad \bar{\beta}_\varphi(\Delta) = \sup_{G \varepsilon \mathcal{G}(\Delta)} \beta_\varphi(G),$$

where

$$\mathcal{G}(\Delta) = [G : G \,\varepsilon\, \bar{\omega}, \qquad \rho^{-}(F_0, G) = \Delta], \qquad 0 < \Delta < 1,$$

for these several test functions φ.

To obtain sharp upper and lower bounds of the power of any p.o. test against all alternatives G such that

$$\rho^{-}(F_0, G) = \Delta, \tag{5}$$

we consider the alternatives

$$G_{mu_0}(u) = \begin{cases} 0, & u < 0, \\ u, & 0 \leqq u < u_0, \\ u_0, & u_0 \leqq u < u_0 + \Delta, \\ u, & u_0 + \Delta \leqq u < 1, \\ 1, & 1 \leqq u, \end{cases} \tag{6}$$

and

$$G_M(u) = \begin{cases} 0, & u < \Delta, \\ u - \Delta, & \Delta \leqq u < 1, \\ 1, & 1 \leqq u. \end{cases} \tag{7}$$

These distributions are not members of the family of alternatives $\bar{\omega}$, but it is possible to find distributions in $\bar{\omega}$ arbitrarily close to G_{mu_0} or G_M. Hence it follows from the continuity of the power functions, that if the test is p.o.

$$\underline{\beta}_\varphi(\Delta) = \inf_{0 \leqq u_0 \leqq 1-\Delta} \beta_\varphi(G_{mu_0}); \qquad \bar{\beta}_\varphi(\Delta) = \beta_\varphi(G_M).$$

Such bounds are given below for several of the tests of H that meet the criteria of admissibility, consistency, unbiasedness, monotonicity and partial orderedness.

3. Fisher and Pearson tests. The statistics

$$\pi = -2 \sum_{i=1}^{n} \ln U_i, \tag{8}$$

$$\pi' = -2 \sum_{i=1}^{n} \ln (1 - U_i) \tag{9}$$

were introduced in the problem of combining tests but are also suitable for testing H_0. If H_0 is true π and π' both are distributed as χ^2 with $2n$ d.f. Furthermore, the u.m.p. test of H_0 against the family of alternatives

$$G_k = u^k \qquad k > 1 \tag{10}$$

is obviously of the form: Reject H_0 if $\pi < c$. A similar statement may be made about π'.

Furthermore, such tests are obviously monotone and hence p.o.

It will be convenient to refer to the tests, reject H if $\pi < c$ or $\pi' > c$, simply as the tests π, π'. These are two of the class of likelihood ratio tests of the form: Reject H if $\sum_{i=1}^{n} \ln g_1(U_i) > c$, where g_1 is the derivative of a specified absolutely continuous alternative G_1.

If $E_0[\ln g_1(U)]^2 < \infty$, this test statistic is asymptotically normal and furthermore if $E_G[\ln g_1(U)]^2 < \infty$ and $E_0[\ln g_1(U)] < E_G[\ln g_1(U)]$ the usual argument shows that the test based on $\sum_{i=1}^{n} \ln g_1(U_i)$ is consistent for testing H_0 against the alternative G.

In particular for the tests π, π', we have

THEOREM. *The tests π, π' are consistent for the set of alternatives $\tilde{\omega}$.*

PROOF. In view of the remark above it is necessary to show that $E_0[\ln U]^2$, $E_G[\ln U]$ are finite and $E_0[\ln U] < E_G[\ln U]$. Now

$$E_G[\ln U]^2 = \int_0^1 (\ln u)^2 \, dG(u). \tag{11}$$

Let $1 > \epsilon > 0$; for every ϵ

$$\int_\epsilon^1 (\ln u)^2 \, dG(u) = (\ln u)^2 G(u) |_\epsilon^1 + 2 \int_\epsilon^1 G(u) \left| \frac{\ln u}{u} \right| du. \tag{12}$$

Since $G(u) \leqq u$ the first term on the right-hand side of (12) can be made arbitrarily small by appropriate choice of ϵ while for all ϵ the second integral is bounded by

$$\int_0^1 |\ln u| \, du = 1.$$

This shows that both $E_0[\ln U]^2$, $E_G[\ln U]^2$ exist and also validates the integration by parts in the next step.

For

$$\begin{aligned} E_G(\ln U) &= \int_0^1 (\ln u) \, dG(u) \\ &= \ln u G(u) |_0^1 - \int_0^1 \frac{G(u)}{u} \, du. \end{aligned} \tag{13}$$

The first term on the right-hand side of (13) is zero. Since $G(u) \leqq u$ with inequality holding on a set of positive measure

$$-\int_0^1 \frac{G(u)}{u} \, du > -\int_0^1 du = -1 = E_0[\ln U]$$

as required for the consistency of the π test.

The proof of the consistency of π' requires consideration of two cases.

CASE 1. $\qquad E_G[\ln (1 - U)] > -\infty.$

Since G is continuous and $G(u) < u$ on a set of positive measure, $\exists\ \epsilon'$ such that

$$\int_0^{1-\epsilon'} \frac{[u - G(u)]}{1 - u}\, du = 2\delta > 0.$$

Now in view of the finiteness of $E_G[\ln (1 - U)]$, $\exists\ \epsilon$ which may be chosen less than $\delta/3$ and ϵ' such that

$$|\ln \epsilon|\, [1 - G(1 - \epsilon)] < \delta/3$$

and also

$$\left| \int_0^1 \ln (1 - u)\, dG(u) - \int_0^{1-\epsilon} \ln (1 - u)\, dG(u) \right| < \delta/3.$$

Now

$$\begin{aligned}
\int_0^1 \ln (1 - u)\, dG(u) &< \int_0^{1-\epsilon} \ln (1 - u)\, dG(u) + \delta/3 \\
&= G(1 - \epsilon) \ln \epsilon + \int_0^{1-\epsilon} \frac{G(u)}{1 - u}\, du + \delta/3 \\
&= G(1 - \epsilon) \ln \epsilon + \int_0^{1-\epsilon} \frac{u}{1 - u}\, du - \int_0^{1-\epsilon} \frac{u - G(u)}{1 - u}\, du + \delta/3 \\
&= -1 + \epsilon + \ln \epsilon[G(1 - \epsilon) - 1] - 2\delta + \delta/3 < -1 - \delta.
\end{aligned}$$

Since the critical region of the π' test converges to: Reject H_0 if

$$-\frac{\pi'}{2n} < -1 - \frac{Z_\alpha}{\sqrt{n}},$$

while by Khintchine's theorem $-(\pi'/2n)$ converges almost surely to $E_G[\ln (1 - U)] \leqq -1 - \delta$ under the alternative G, the consistency follows.

CASE 2. $$E_G[\ln (1 - U)] = -\infty.$$

By well-known results in this case infinitely many of the sequence of the independent r.v. $\sum_{i=1}^n \ln (1 - U_i)$ $n = 1, 2, 3 \cdots$, are with probability 1 less than nA for any arbitrary A. Hence from the remark on the critical region the consistency is immediate.

As a consequence of this theorem it may be noted that π is asymptotically normal both under H_0 and all alternatives in $\tilde{\omega}$; it is trivial to give examples that this is not true for π'. This behavior is reversed for alternatives $G(u) \geqq u$.

The asymptotic normality of π permits an elementary derivation of $\beta_\pi(\Delta)$ and $\bar{\beta}_\pi(\Delta)$ for large samples. In particular

(14) $$E_M |\ln U| = 1 - \Delta(1 - \ln \Delta),$$

(15) $$\sigma_M^2 |\ln U| = 1 + 2\Delta^2 \ln \Delta - (\ln^2 \Delta)(\Delta + \Delta^2),$$

and

$$E_{mu_0}|\ln U| = 1 - \Delta + u_0 \ln \left(1 + \frac{\Delta}{u_0}\right) \tag{16}$$

$$\text{with} \quad \max_{u_0} E_{mu_0}|\ln U| = (1 - \Delta)[1 - \ln(1 - \Delta)]. \tag{17}$$

This maximum is attained when $u_0 = 1 - \Delta$. Also

$$\begin{aligned} E_{Mu_0}(\ln U)^2 = 2(1 - \Delta) - u_0[\ln^2(u_0 + \Delta) - \ln^2 u_0] \\ - 2(u_0 - \Delta)\ln(u_0 + \Delta) - 2u_0 \ln u_0 . \end{aligned} \tag{18}$$

A numerical study of the variance of $\ln U$ as a function of u_0 shows that the variance is maximized when $u_0 = 1 - \Delta$ though the changes with respect to u_0 are very slight.

For $u_0 = 1 - \Delta$

$$\begin{aligned} \sigma_m^2(\ln U) = (1 - \Delta)[2 - 2\ln(1 - \Delta) + \ln^2(1 - \Delta)] \\ - (1 - \Delta)^2[1 - \ln(1 - \Delta)]^2. \end{aligned} \tag{19}$$

Hence approximately for large n

$$\beta_\pi(\Delta) = \Phi\left(\frac{[Z_\alpha + \sqrt{n}\,[(1 - \Delta)\ln(1 - \Delta) + \Delta]}{\sigma_m(\ln u)}\right), \tag{20}$$

$$\bar{\beta}_\pi(\Delta) = \Phi\left(\frac{Z_\alpha + \sqrt{n}\,[\Delta(1 - \ln \Delta)]}{[1 + 2\Delta^2 \ln \Delta - \ln^2 \Delta(\Delta + \Delta^2)]^{1/2}}\right). \tag{21}$$

The minimum power of the π' test is attained against the alternative G_{m0}, i.e., the jump of height Δ is located at $u = \Delta$. Furthermore this minimum power is the same as the minimum power of the π test.

On the other hand π' will not be asymptotically normally distributed for G_M; in fact with probability $1 - (1 - \Delta)^n$, $\pi' = +\infty$ in which case rejection is immediate. However, under the condition that all the U_i are less than 1, π' is asymptotically normal so that

$$\bar{\beta}_{\pi'}(\Delta) \doteq 1 - (1 - \Delta)^n[1 - \Phi(x)], \tag{22}$$

where

$$x = \frac{Z_\alpha - \sqrt{n}\,\dfrac{\Delta \ln \Delta}{1 - \Delta}}{\left[1 - \dfrac{\Delta \ln^2 \Delta}{(1 - \Delta)^2}\right]^{1/2}}. \tag{23}$$

Tables giving numerical values of these minimum and maximum power functions are displayed in section 8 below where the several tests are compared.

3. D_n^- test. The empirical d.f. $F_n(u)$ is basic in many distribution-free tests of H_0. The use of the statistic

$$D_n^- = \sup_{0 \leq u \leq 1} [u - F_n(u)] \tag{24}$$

as a large sample test for H_0 became possible after Smirnov [20] obtained its limiting distribution. Subsequently Birnbaum and Tingey [3] gave a closed expression for the distribution of D_n^- for finite n.

These results are

$$(25) \qquad \lim_{n\to\infty} \Pr\,[\sqrt{n}\, D_n^- \leqq Z] = 1 - e^{-2Z^2}$$

and

$$(26) \qquad \Pr\,[D_n^- \leqq \epsilon] = 1 - \epsilon\left(\sum_{j=0}^{[n(1-\epsilon)]} \binom{n}{j}\left(1 - \epsilon - \frac{j}{n}\right)^{n-j}\left(\epsilon + \frac{j}{n}\right)^{j-1}\right),$$

where as usual $[x]$ is the greatest integer contained in x.

It is immediate from the definition that the test is monotone and hence p.o. and admissible, as well as being consistent.

Birnbaum in [5] gave upper and lower bounds for the power of the D_n^- test for alternatives of fixed distance Δ within the class of all continuous distribution functions. The upper bound is attained for the alternative labeled here G_M and we quote his result

$$(27) \qquad \bar{\beta}_{D_n^-}(\Delta) \begin{cases} = (\epsilon_n - \Delta) \displaystyle\sum_{i=0}^{[n(1-\epsilon_n+\Delta)]} \binom{n}{i}\left(1 - \epsilon_n + \Delta - \frac{i}{n}\right)^{n-1}\left(\epsilon_n + \Delta + \frac{i}{n}\right)^{i-1} \\ = 1 \begin{cases} \text{for } \epsilon_n \geqq \Delta, \\ \text{for } \epsilon_n < \Delta, \end{cases} \end{cases}$$

where ϵ_n is chosen so that

$$\Pr\,[D_n^- > \epsilon_n \mid H_0] = \alpha.$$

In view of Smirnov's result for large n

$$\bar{\beta}_{D_n^-} - (\Delta) \doteq e^{-2n(\epsilon_n - \Delta)^2} \qquad \text{for } \epsilon_n \geqq \Delta.$$

The lower bound of the power of the D_n^- test within the class of stochastically comparable alternatives was studied by Birnbaum and Scheuer [7]. Their result is given as a number of double and triple sums of terms of the same type as those in (26), and is not in a form useful for comparison or evaluation purposes.

The following approach does not yield a simple closed expression for the exact power, but an adequate approximation is obtained. We write

$$\begin{aligned} \beta(G_{mu_0}) &= \Pr\,[u_0 + \Delta - F_n(u_0 + \Delta - 0) \geqq \epsilon_n \mid G_{mu_0}) \\ (28) \qquad &+ \Pr\,[\sup_{0 \leqq u < u_0} \{u - F_n(u)\} \geqq \epsilon_n \mid u_0 + \Delta - F_n(u_0 + \Delta - 0) < \epsilon_n\,,\, G_{mu_0}] \\ &+ \Pr\,[\sup_{u_0+\Delta \leqq u < 1} \{u - F_n(u)\} \geqq \epsilon_n \mid \sup_{0 \leqq u_0 < u_0+\Delta} \{u - F_n(u)\} < \epsilon_n\,,\, G_{mu_0}]. \end{aligned}$$

It will be convenient to symbolize the three terms on the right hand of (28) by P_1, P_2, P_3 respectively. It is immediate that

$$(29) \qquad P_1 = \sum_{k=0}^{[n(u_0+\Delta-\epsilon_n)]} B(k; n, u_0),$$

where the right-hand summands denote binomial probabilities in the usual notation.

An examination of the integral representation of

$$P_n(\epsilon) = \Pr [\sup_{0 \leqq u \leqq 1} \{u - F_n(u)\} \leqq \epsilon]$$

given by Birnbaum and Tingey in [3]

$$P_n(\epsilon) = n! \int_0^{\epsilon} \int_{x_1}^{(1/n)+\epsilon} \int_{x_2}^{(2/n)+\epsilon} \cdots \int_{x_K}^{(K/n)+\epsilon} \int_{x_{K+1}}^{1} \cdots \int_{x_{n-1}}^{1} dx_n \cdots dx_{K+2}\, dx_{K+1} \cdots dx_3\, dx_2\, dx_1,$$

where $K = [n(1 - \epsilon)]$, shows immediately that P_2 and P_3 are bounded by α. Hence the dominant term in $\beta(G_{mu_0})$ is P_1 which is minimized when $U_0 = \frac{1}{2}$. This value has been used in making minimum power calculations for the D_n^- test.

However the actual values of P_2 can be determined in the large sample case. Consider

$$\Pr\left[\sup_{0 \leqq u \leqq u_0} \{u - F_n(u)\} \leqq \epsilon_n \mid F_n(u_0) = \frac{k}{n}, G_{mu_0}\right]$$

$$= k! \int_0^{\epsilon_n} \int_{u_1}^{1/u_0[(1/n)+\epsilon_n]} \cdots \int_{u_{K'}}^{1/u_0[(K'/n)+\epsilon_n]} \int_{u_{K'+1}}^{1} \cdots \int_{u_{k-1}}^{1} du_k \cdots du_{K'+2}\, du_{K'+1} \cdots du_2\, du_1, \tag{30}$$

where $K' = [n(u_0 - \epsilon_n)]$.

The integral form can be written down in a similar manner to that of $P_n(\epsilon)$ and the result given is obtained by a trivial change of variable. By a slight extension of the arguments used by Birnbaum and Tingey this can be expressed as a closed sum, viz

$$1 - \left(\frac{n\epsilon_n}{k}\right)\left(\frac{k}{nu_0}\right)^k \sum_{j=0}^{K'} \binom{k}{j} \left(\frac{nu_0}{k} - \frac{n\epsilon_n}{k} - \frac{j}{k}\right)^{k-j} \left(\frac{n\epsilon_n}{k} + \frac{j}{k}\right)^{j-1} \tag{31}$$

It is convenient to denote the function on the right-hand side

$$V(nu_0/k,\ n\epsilon_n/k,\ k).$$

The power of the D_n^- test against alternatives of the form

$$G_a(u) = \begin{cases} au, & 0 \leqq u < 1, 0 < a < \infty, \\ 1, & u \geqq 1, \end{cases}$$

can be expressed in terms of the function V. This can be seen by writing down the integral using the general power formula given by Birnbaum ([5], p. 486) or by a simple direct argument. In fact

$$\beta_{D_n^-}(G_a) = 1 - V\left(\frac{1}{a}, \epsilon_n, n\right). \tag{32}$$

While the sums in (31) can be evaluated by a straight-forward process, the process is tedious for large k, and we obtain instead an asymptotic result that yields a method of approximating V in this situation.

Let $G_{na}(u)$ denote a sequence of d.f. of the form

$$G_{na}(u) = \begin{cases} \left(1 + \frac{a}{\sqrt{n}}\right)^{-1} u, & 0 \leqq u < b, -\sqrt{n} < a < \sqrt{n}, \\ 1, & u \geqq b, \end{cases} \tag{33}$$

where

$$b = \min\left(1 + \frac{a}{\sqrt{n}}, 1\right),$$

Then

$$\beta_{D_n^-}(G_{na}) = \Pr\left[\sup_{0 \leqq u < b} \left(1 + \frac{a}{\sqrt{n}}\right) u - F_n(u) > \epsilon_n\right]. \tag{34}$$

Now we use Donsker's theorem [11] justifying Doob's heuristic approach to the Kolmogorov-Smirnov theorems [12] to validate the following steps:

$$\begin{aligned} \lim_{n \to \infty} \Pr [\sup_{0 \leqq u < b} \{\sqrt{n}(u - F_n(u) + au)\} > Z] \\ = \Pr [\sup_{0 \leqq u < 1} (X(u) + au) > Z], \end{aligned} \tag{35}$$

where $X(u)$ is a Gaussian process with the properties noted by Doob ([12], p. 397). Further, the transformation he made and his evaluation of

$$\Pr \{\sup [\zeta(t) - (at + b)] \geqq 0\}$$

may be used to evaluate this last probability. We have in fact

$$\begin{aligned} \Pr [\sup_{0 \leqq u < 1} [X(u) + au] > Z] &= \Pr\left[\sup_{0 \leqq u < \infty} \frac{\zeta(u) + au}{u + 1} > Z\right] \\ &= e^{-2Z(Z-a)}. \end{aligned} \tag{36}$$

In other words, putting $\epsilon_n = Z/\sqrt{n}$

$$\lim_{n \to \infty} V\left(1 + \frac{a}{\sqrt{n}}, \frac{Z}{\sqrt{n}}, n\right) = 1 - e^{-2Z(Z-a)}.$$

Hence if $n, k \to \infty$ with $\sqrt{k}[(n/2k) - 1] = (n - 2k)/2\sqrt{k}$ and n/k remaining finite and if u_0 is set equal to $\frac{1}{2}$

$$\begin{aligned} \lim_{n,k \to \infty} V\left(\frac{n}{2k}, \frac{n\epsilon_n}{k}, k\right) &= \lim_{n,k \to \infty} V\left(\frac{n}{2k}, \frac{Z}{\sqrt{k}}\sqrt{\frac{n}{k}}, k\right) \\ &= 1 - \exp\left[-2\frac{n}{k}Z^2 - \sqrt{\frac{n}{k}}\left(\frac{n - 2k}{\sqrt{k}}\right)Z\right] \end{aligned} \tag{37}$$

so that an approximate evaluation of P_2 is given by

$$(38) \qquad P_2 = \sum_{k=[n(\frac{1}{2}+\Delta-\epsilon_n)]}^{n} B(k; n, \tfrac{1}{2}) \left[\exp\left(-\frac{2_n^2\epsilon_n^2}{k} + 2n\epsilon_n - \frac{n^2\epsilon_n^2}{k}\right)\right].$$

This formula was used to evaluate P_2 for a number of values of n and Δ. These are shown in Table 1. The striking feature of the table is the negligible size of P_2.

Further, by making the change of variable $W = 1 - U$ and noting that

$$(39) \qquad P_3 \leqq \Pr[\sup_{u_0+\Delta<u<1} [u - F_n(u)] \geqq \epsilon \mid u_0 - F_n(u_0) < \epsilon_n, G_{mu_0}]$$

it is obvious that for $u_0 = \frac{1}{2}$, $P_3 \leqq P_2$.

Hence $\min_{u_0} \beta_{D_n^-}(G_{mu_0})$ is bounded between $P_1 + P_2$ and $P_1 + 2P_2$ for large samples.

5. Tests related to D_n^- test. Anderson and Darling [1] considered a class of tests based on the more general distance function

$$\sup_{-\infty<x<\infty} \sqrt{n} \mid F_n(x) - F(x) \mid \psi[F(x)],$$

where ψ is a non-negative weight function. The choice of $\psi = 1$ yields the Kolmogorov statistic. Anderson and Darling also studied

$$\psi(t) = \begin{cases} \dfrac{1}{t(1-t)}, & 0 < a \leqq t \leqq b < 1, \\ 0, & \text{otherwise,} \end{cases}$$

but the distribution function is not in usable form. The distribution of $\sup_{-\infty<x<\infty}\sqrt{n}\,[F(x) - F_n(x)]\psi[F(x)]$, when H_0 is true, has apparently only been obtained for the case $\psi = 1$. More recently, Pyke [17] has studied a class of tests based on a generalized one-sided distance, but again the distributions have not been given.

TABLE 1

$P_2 = \Pr[\sup_{0\leqq u<u_0} u - F_n(u) \geqq \epsilon_n \mid u_0 + \Delta - F_n(u_0 + \Delta - 0) < \epsilon_n]$*

Δ	n			
	50	100	200	400
0.05	.0081	.0062	.0035	.0011
0.10	.0027	.0001	—	—
0.20	.0002	—	—	—
0.30	—	—	—	—
0.40	—	—	—	—
0.50	—	—	—	—

* Calculations made using formula (38). Entries marked with — are less than .0001.

One asymptotic result of this type is known that could form the basis of a large sample test of H_0. This is the result due to Renyi [18], viz., if H_0 is true

$$\lim_{n\to\infty} P\left\{\sqrt{n} \sup_{a \leqq u} \frac{[F_n(u) - u]}{u} < Z\right\} = \sqrt{\frac{2}{\pi}} \int_0^{Z[a/(1-a)]^{\frac{1}{2}}} e^{-t^2/2}\, dt, \qquad Z > 0 \tag{40}$$

$$= 0, \qquad Z \leqq 0,$$

for arbitrary a, $0 < a < 1$.

The restriction $a \leqq u$ is unpleasant since it imposes an additional decision on the statistician, viz., the choice of a. Furthermore, it is apparent that the test based on this result cannot be consistent against alternatives which do not differ from $F_0(x)$ for the set $E[x:F_0(x) < a]$. On the other hand the test is consistent against all other alternatives in $\bar{\omega}$.

One feature of this test may be noted. The minimum power of the test may be studied in a manner parallel to that used for the D_n^- test. In particular the probability of rejection is the probability that the empirical d.f. $F_n(u)$ falls at some point below the line $u(1 + \epsilon_n) - \epsilon_n$ where ϵ_n is chosen to satisfy the size condition; i.e., approximately for large samples

$$\epsilon_n = Z_\alpha \sqrt{\frac{1-a}{an}}. \tag{41}$$

The primary term of the power function $\beta(G_{mu_0})$ is thus seen to be approximately

$$\Phi\left[\frac{\sqrt{n}\Delta - Z_\alpha \sqrt{\dfrac{1-a}{a}}\,(1 - u_0 - \Delta)}{[u_0(1 - u_0)]^{1/2}}\right], \tag{42}$$

which for sufficiently large n is minimized when

$$u_0 = 1 - a - \Delta.$$

Further this minimum power will be an increasing function of a; i.e., increasing a will increase the minimum power of the test within the class of d.f.'s for which the test is consistent but at the same time this class will be decreased.

6. Tests based on the integral criterion. To Cramér and Von Mises is due the idea of testing H_0 by a statistic based on the integral of the square of the difference between hypothetical and empirical distribution functions. Smirnov modified this by integrating with respect to the probability measure generated by $F(u)$. A more general form was given by Anderson and Darling [1] (this paper also gives references to the original authors which have been omitted here). This is

$$W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi[F(x)]\, dF.$$

The limiting distribution of this statistic with the weight function $\psi = 1$ was given first by Smirnov, then by Von Mises and later by Anderson and Darling.

They also gave a tabulation of the limiting distribution (cf. [1], p. 203). The latter authors also give the d.f. of W_n^2 for the weight function

$$\psi(t) = t(1 - t)$$

but the function is complex and no tabulation has been given.

Before discussing the classical form of W_n^2, it is of interest to note that since we are here considering one-sided alternatives, it is not unreasonable to introduce as a test statistic

$$W_n' = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]\, dF(x) = n \int_0^1 [F_n(u) - u]\, du. \tag{43}$$

It is seen at once that

$$W_n' = \sum_{i=1}^{n} U_i - \frac{n}{2} \tag{44}$$

so that the test is equivalent to one based on

$$\bar{U} = \frac{1}{n} \sum_{i=1}^{n} U_i. \tag{45}$$

Such a test has also been proposed by L. Moses.

For n large, under H_0, $\bar{U}$ is $N(\frac{1}{2}, 1/12n)$, while under th alteernative $G(u) < u$ it is normally distributed with mean $\int_0^1 u\, dG(u) > \frac{1}{2}$. The variance of $\bar{U}$ under the alternative is finite so that the test is consistent for all alternatives in $\bar{\omega}$. The test is also obviously monotone and hence p.o. For any alternative the large sample power is easily computed. In particular

$$\bar{\beta}_{\bar{U}}(\Delta) = \beta_{\bar{U}}(G_M) = 1 - \Phi\left(\frac{Z_\alpha - \sqrt{3n}\Delta(2 - \Delta)}{[1 - \Delta^2(6 - 8\Delta + 3\Delta)]^{1/2}}\right) \tag{46}$$

and

$$\beta_{\bar{U}}(G_{mu_0}) = 1 - \Phi\left(\frac{Z_\alpha - \sqrt{3n}\Delta^2}{[1 - \Delta^2(6 - 8\Delta + 3\Delta^2) + 12u_0\Delta^2]^{1/2}}\right). \tag{47}$$

The minimum of (47) is attained when $u_0 = 0$.

Consider now the classical integral criterion, i.e.,

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2\, dF(x) = \int_0^1 [F_n(u) - u]^2\, du. \tag{48}$$

It is well known that

$$\begin{aligned} n\omega^2 &= \frac{1}{12n} + \sum_{i=1}^{n} \left[U_i - \frac{2i - 1}{2n}\right]^2 \\ &= \sum_{i=1}^{n} U_i^2 - \frac{1}{n} \sum_{i=1}^{n} U_i(2i - 1) + \frac{n}{3} \end{aligned} \tag{49}$$

and that if H_0 is true

$$E(\omega^2) = \frac{1}{6n}, \qquad \sigma^2(\omega^2) = \frac{1}{n^2}\left(\frac{4n-3}{180n}\right). \tag{50}$$

It is known that if H_0 is true $n\omega^2$ has a limiting distribution which is not normal. However, if H_0 is false, the limiting distribution of ω^2, appropriately normalized, is normal.

For if the U_i have a d.f. $G(u)$

$$\begin{aligned} \omega^2 &= \int_0^1 [u - G_n(u)]^2\, du = \int_0^1 [u - G(u) + G(u) - G_n(u)]^2\, du \\ &= \int_0^1 \delta^2(u)\, du + 2\int_0^1 \delta(u)G(u) \\ &\qquad - 2\int_0^1 \delta(u)G_n(u) + \int_0^1 [G(u) - G_n(u)]^2\, du, \end{aligned} \tag{51}$$

where $G_n(u)$ has been written to emphasize that the sample has been drawn from the population with distribution G and where we have written $u - G(u) = \delta(u)$.

The notation

$$\int_0^u \delta(t)\, dt = D(u)$$

and

$$\int_0^1 \delta^2(u)\, du + 2\int_0^1 \delta(u)G(u) - 2D(1) + 2E[D(U)] = C(G)$$

will also be used.

From Kolmogorov's theorem that

$$\lim_{n\to\infty} [\Pr [\sqrt{n} \sup_{0<u<1} |G(u) - G_n(u)| \geqq Z]] = 2\sum_{v=1}^{\infty} (-1)^{v-1} e^{-2v^2Z^2} \tag{52}$$

it is easily seen that

$$\sqrt{n}\int_0^1 [G(u) - G_n(u)]^2\, du$$

tends to zero in probability. Also

$$\begin{aligned} \int_0^1 \delta(u)G_n(u)\, du &= \frac{1}{n}\sum_{i=1}^{n-1} i \int_{U_i}^{U_{i+1}} \delta(u)\, du \\ &= D(1) - \frac{1}{n}\sum_{i=1}^{n} D(U_i). \end{aligned} \tag{53}$$

Since $D(u) \leqq \frac{1}{2}$ for $0 \leqq u \leqq 1$, $E_G[D(U)]^2 < \infty$ and hence

$$\sqrt{n}\, \{1/n \textstyle\sum_{i=1}^n [D(U_i) - E[D(U)]]\}$$

is asymptotically normal with mean zero and variance given by the usual formula.

Finally then $\sqrt{n}\,(\omega^2 - C(G))$ is the sum of an asymptotically normal r.v. and one tending in probability to zero. It is therefore itself asymptotically normal with expectation zero.

Define ω_α by the equation $\Pr\,[n(\omega)^2 > \omega_\alpha \mid H_0] = \alpha$.

The ω^2 test (i.e., reject H_0 when $n\omega^2 > \omega_\alpha$) is consistent but not monotone. Its failure to be monotone arises from the fact that the test is two-sided and we are here considering one-sided alternatives. On the other hand, at least for n sufficiently large that the term

$$\int_{-\infty}^{\infty} [G(u) - G_n(u)]^2\, du$$

is negligible with respect to the other terms of ω^2, the test is p.o. This follows from the decomposition (51), since the other terms in this expression increase as G decreases.

The calculation of $E[D(U)]$, $\sigma^2[D(U)]$ is particularly simple for the alternatives G_{mu_0} and G_M. In fact, it is also possible in these cases to calculate straightforwardly $E(\omega^2)$.

Thus

$$E_{G_{mu_0}}(\omega^2) = \frac{\Delta^3}{3}\left(1 + \frac{1}{n}\right) + \frac{1}{n}\left(\tfrac{1}{6} + \Delta^2[u_0 - \tfrac{1}{2}]\right) \tag{54}$$

while

$$\sigma^2_{G_{mu_0}}(\omega^2) = \frac{u_0(1 - u_0)}{n}\,\Delta^4 + 0\left(\frac{1}{n^2}\right). \tag{55}$$

The value of u_0 which minimizes the function $\beta(G_{mu_0})$ is a rather complicated expression involving Δ, n and ω_α; however, it is easily seen that as $n \to \infty$ this minimizing value tends to $\frac{1}{2}$. For simplicity we have evaluated only the approximate large sample power function $\beta(G_{m,\frac{1}{2}})$:

$$\beta_{\omega^2}(G_{m,\frac{1}{2}}) = 1 - \Phi\left[\frac{2}{\Delta^2}\left(\frac{\omega_\alpha^2}{\sqrt{n}}\right) - \frac{2\Delta}{3}\left(1 + \frac{1}{n}\right)\sqrt{n} - \frac{1}{3\Delta^2\sqrt{n}}\right]. \tag{56}$$

Similarly evaluating $E(\omega^2)$ and $\sigma^2(\omega^2)$ to terms of order $1/n^2$

$$\bar{\beta}_{\omega^2}(\Delta) = \beta_{\omega^2}(G_M) = 1 - \Phi(x), \tag{57}$$

where

$$x = \frac{\dfrac{\omega_\alpha - \left(\dfrac{1}{6} - \dfrac{\Delta^2}{2} + \dfrac{\Delta^3}{3}\right)}{\sqrt{n}} - (\Delta^2 - \tfrac{2}{3}\Delta^2)\sqrt{n}}{2\Delta\left[\dfrac{1}{12} - \dfrac{\Delta^2}{2} + \dfrac{2}{3}\Delta^3 - \dfrac{\Delta^4}{4}\right]^{1/2}} \tag{58}$$

7. Other tests. A procedure that has been suggested for the problem of combining tests and which consequently could be adapted to the equivalent problem of testing H_0, is based on the minimum or maximum of the transformed observations, i.e., in our notation, U_1 or U_n. Even restricting the problem by choosing a simple univariate statistic such as U_1, does not yield a unique u. m. p. test. Moreover the "intuitive" test of H_0 against one-sided alternatives—i.e., reject H_0 when $U_1 > c$ for appropriately chosen c—is obviously not consistent. In fact, it is only consistent for those alternatives $G(u) < u$ such that $\lim_{u \to \infty} G(u)/u = 0$. Furthermore, the test—reject H when $U_n > c$—would be consistent for no alternatives of $\tilde{\omega}$.

Of more interest are a group of tests based on another class of statistics, the so-called spacing of the observations. It is convenient to define

$$S_i = U_i - U_{i-1}, \qquad i = 1, 2, \cdots n + 1. \tag{59}$$

Various tests based on the statistics S_i have been proposed by Sherman [19] and others. These tests are not p.o. and hence are excluded from the present study. The proof of this fact as well as some other properties of these tests will be given in a later paper.

TABLE 2A

Minimum power of several tests for alternatives whose distance from F_0 is Δ

Test	Δ										
	0.05	0.1	0.125	0.15	0.175	0.20	0.25	0.3	0.35	0.45	0.50
						$n = 50$					
π, π'	.052	.059	—	—	—	.072	—	.108	—	.179	.306
$\bar{U}$	.052	.059	.065	.073	.085	.102	.153	.250	.577	.676	.994
ω^2						.131	.448	.697	.842	.922	.981
D_n^-	.057	.156	.248	.372	.511	.648	.862	.964	—	1.000	
						$n = 100$					
π, π'	.053	.063	—	—	—	.080	—	.137	—	.254	.460
$\bar{U}$	.053	.065	—	.092	—	.148	.257	.457	.738	.949	1.000
ω^2				.054	.228	.449	.730	.914	.970	.990	.999
D_n^-	.086	.327	.521	.710	—	.940	—	1.000			
						$n = 200$					
π, π'	.054	.068	—	—	—	.094	—	.185	—	.382	.623
$\bar{U}$	.055	.075	—	.123	—	.232	.447	.756	.965	1.000	
ω^2			.069	.335	.617	.803	.956	.991	—	1.000	
D_n^-	.158	.649	.862	.964	—	.999	1.000				
						$n = 400$					
π, π'	.056	.076	—	—	—	.117	—	.270	—	.583	.902
$\bar{U}$	.058	.091	.125	.179	.264	.388	.727	.966	—	1.000	
ω^2	—	.054	.415	.757	.916	.974	.998				
D_n^-	.329	.940	—	1.000							

8. Comparison of the minimum and maximum powers of consistent, partially ordered tests. In the preceding sections it has been shown that the tests associated with the statistics $\pi(8)$, $\pi'(9)$, D_n^- (24), $\bar{U}$ (45) and ω^2 (48) are consistent, monotone and p.o. Furthermore, useful large sample approximations were found for $\underline{\beta}(\Delta)$ and $\bar{\beta}(\Delta)$ for each test. In view of the fact that most of these large sample power functions are expressed in terms of normal probabilities it would not be difficult to obtain inequalities between the power functions for the different tests. However, in not all cases does the same relationship between the power functions persist for all Δ or all n. Furthermore, such inequalities do not indicate the magnitude of the power differences.

As a more informative approach calculations have been made of $\underline{\beta}(\Delta)$ and $\bar{\beta}(\Delta)$ for each test for a range of values of n and Δ. These have been calculated for two test sizes, viz., $\alpha = 0.05$ and $\alpha = 0.01$. The minimum power was calculated for $\Delta = 0.05, 0.1, 0.2, 0.3, 0.4$ and 0.5 and where desirable, some intermediate values while the maximum power was calculated for $\Delta = 0.01$, (0.01) 0.10, 0.15, 0.20, 0.30, 0.40, and 0.50. A fixed sequence of sample sizes n was used, viz., $n = 50, 100, 200, 400, 600, 800, 1000, 2000, 4000, 6000, 8000, 10{,}000 \ldots$ with the stopping rule, stop whenever the absolute values of the normal deviate exceeded 3.

TABLE 2B

Maximum power of several tests for alternatives whose distance from F_0 is Δ

Δ	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.125	0.15	0.2
							$n = 50$						
π'	.959	.965	.972	.977	.983	.987	.990	.993	.995	.997	1		
π	.080	.125	.189	.272	.373	.484	.598	.706	.799	.871	.970	.996	
$\bar{U}$	.080	.123	.178	.246	.325	.412	.504	.595	.681	.758	.899	.968	.999
ω^2	—	—	—	.092	.193	.308	.412	.509	.589	.662	—	—	.973
D_n^-	.070	.096	.129	.170	.220	.278	.346	.420	.501	.586	.793	.948	
							$n = 100$						
π	.967	.980	.989	.949	.998	1							
π'	.111	.211	.354	.523	.689	.824	.916	.966	.988	.997	1		
$\bar{U}$	.096	.168	.267	.387	.518	.646	.759	.849	.914	.955	.994	1	
ω^2	—	—	.118	.277	.423	.562	.668	.754	.818	.869	—	—	.999
D_n^-	.080	123	.181	.257	.350	.459	.578	.698	.811	.905	1		
							$n = 200$						
Δ	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	1.0			
π	.169	.370	.633	.840	.950	.989	.999	1					
$\bar{U}$	.124	.250	.421	.609	.773	.889	.954	.984	.996	.999			
ω^2	—	.093	.317	.529	.675	.797	.873	.925	.955	.975			
D_n^-	.096	.171	.279	.421	.587	.755	.897	.983	1				

Some of the powers so calculated for $\alpha = 0.05$ are exhibited in Table 2A and 2B. It might be hoped that some of the tests could be eliminated by such a comparison—this would be the case if $\bar{\beta}$ for one test fell below $\underline{\beta}$ for some other test. However, this is not the case.

In general the tables indicate that the relationship of the tests is reversed from the minimum to the maximum power. Thus we have

$$\underline{\beta}_{\pi} = \underline{\beta}_{\pi'} < \underline{\beta}(\omega^2), \qquad \underline{\beta}_{\bar{U}} < \underline{\beta}_{D_n^-} < \bar{\beta}_{D_n^-} < \bar{\beta}(\omega^2), \qquad \bar{\beta}_{\bar{U}} < \bar{\beta}_{\pi} < \bar{\beta}_{\pi'} .$$

The relationship between the ω^2 and $\bar{U}$ tests varies with Δ and n.

It is evident that the π' test has the best maximum power of the tests considered, but its minimum power (and that of the π test) is extremely low. On the other hand the D_n^- test which has the lowest maximum power (of the tests considered) has the greatest minimum power. This raises the question whether there exists a non-trivial test which is p.o. and for which $\underline{\beta}(\Delta) = \bar{\beta}(\Delta)$.

An alternative comparison between the tests is given in Table 3, which shows the sample sizes necessary to achieve a pre-assigned power level β, for given Δ and for $\alpha = 0.05$. The values corresponding to $\beta = 0.95$ only are listed though corresponding values of n have been calculated for $\beta = 0.90$ and $\beta = 0.99$. The latter calculation emphasizes the poorness of the π, π' tests against alternatives G_{mu_0}—over 2,443,900 observations are required to insure $\underline{\beta}(0.05) = 0.99$. It should be noted that these values of n were calculated from the primary term

TABLE 3

Sample sizes necessary that $\underline{\beta}(\Delta)$ *and* $\bar{\beta}(\Delta) = 0.95$ *for several* p.o. *tests and for* $\alpha = 0.05$

Minimum Alternative

Test	Δ					
	0.05	0.1	0.2	0.3	0.4	0.5
D_n^-	1675	419	105	47	27	17
ω^2	14,038	2290	406	153	78	45
$\bar{U}$	569,067	34,233	1867	304	77	25
π, π'	1,677,025	102,081	23,903	4463	1325	511

Maximum Alternative

Test	Δ									
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
D_n^-	29,679	7420	3298	6855	1188	825	606	464	367	297
ω^2	4761	1057	540	302	204	160	104	80	65	53
$\bar{U}$	9108	2296	1027	583	375	261	193	148	117	95
π	3067	936	471	291	200	148	115	92	77	65

P_1[cf(29)] and consequently the required sample size with the D_n^- test is slightly over-estimated.

It is also to be noted that the smaller sample sizes indicated in Table 3 must not be construed too literally since they have been computed from asymptotic formulae.

Of these tests considered it appears that if no information is available on the possible alternatives to H_0 then from some minimax point of view, the D_n^- test is the most favorable.

Acknowledgment. The calculations on which Tables 1, 2, and 3 are based were made by Mrs. P. Freeman; some were also made by Mr. P. Stevens. Funds for these calculations were provided by the Agnes Anderson Fund of the University of Washington, for which the author is extremely grateful.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] Allan Birnbaum, "Combining independent tests of significance," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 559–574.

[3] Z. W. Birnbaum and F. Tingey, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[4] Z. W. Birnbaum, "Distribution free tests of fit for continuous distribution functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 1–8.

[5] Z. W. Birnbaum, "On the power of a one-sided test of fit for continuous distribution functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 284–489.

[6] Z. W. Birnbaum and H. Rubin, "On distribution free statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 593–598.

[7] Z. W. Birnbaum and Ernest M. Scheuer, *On the Power of a One-sided Test of Fit Against Stochastically Comparable Alternatives*, Technical Report No. 15, O.N.R. Contract, Laboratory of Statistical Research, University of Washington.

[8] Z. W. Birnbaum and Ronald Pyke, *On Some Distributions Related to the Statistic* D_n^+, Technical Report No. 23, O.N.R. Contract, Laboratory of Statistical Research, University of Washington.

[9] G. E. P. Box and S. L. Andersen, "Permutation theory in the derivation of robust criteria and the study of departures from assumption," *J. Roy. Stat. Soc.* (Series B), Vol. 17 (1955), pp. 1–26.

[10] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.

[11] Monroe D. Donsker, "Justification and extension of Dorb's heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277–281.

[12] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[13] W. Feller, *An Introduction to Probability Theory and Its Application*, John Wiley & Sons, New York, 1950.

[14] J. L. Hodges and E. L. Lehmann, "The efficiency of some nonparametric competitors of the *t*-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.

[15] E. L. Lehmann, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1956), pp. 23–43.

[16] P. MORAN, "The random division of an interval," *J. Roy. Stat. Soc. Suppl.*, Vol. 9 (1947), pp. 92–98.

[17] R. PYKE, "On one-sided distribution free statistics," Unpublished thesis, Dept. of Mathematics, University of Washington, 1956.

[18] A. RENYI, "On the theory of order statistics" *Acta. Math. Acad. Sci. Hung.*, Vol. 4 (1953), pp. 191–231.

[19] B. SHERMAN, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

[20] N. V. SMIRNOV, "On the deviation of the empirical distribution functions," *Rec. Math.* (*Mat. Sbornik*) (*NS*), Vol. 6 (1939), pp. 3–26.