EXACT BAHADUR EFFICIENCIES FOR THE KOLMOGOROV-SMIRNOV AND KUIPER ONE- AND TWO-SAMPLE STATISTICS[1]

By Innis G. Abrahamson

Columbia University

1. Introduction. In 1960, Bahadur [3] proposed two measures of the asymptotic performance of tests: one approximate measure, based on the limiting distribution of the test statistic; and one exact, based on the limiting form of the probability of a large deviation of the statistic from its asymptotic mean. Although knowledge of the exact measure is more desirable, it is often difficult to compute, whereas the approximate measure is usually trivially available. It is to be hoped that the approximate measure will nearly always be a good approximation to the exact measure, and there is considerable evidence in support of this conjecture, but counterexamples do exist (see e.g. [1], p. 20).

This paper explores the question for the Kolmogorov-Smirnov (K-S) and Kuiper statistics—two closely related measures of goodness-of-fit based on deviations of the sample distribution functions from the null case—in one- and two-sample situations. In the case of the weighted one-sample K-S statistic and the two-sample Kuiper statistic, it is possible to obtain exact measures although the corresponding approximate measures are not available. The relative efficiency (in a sense defined in Section 2) of the weighted one-sample K-S statistic to the unweighted K-S statistic is computed in a number of cases, and the relative efficiency of the Kuiper statistic to the unweighted K-S statistic is also examined, from which it appears that the Kuiper statistic is always at least as good (and often much better) than the K-S statistic.

In the cause of brevity, a great deal of the theory and all the computational detail (generally very tedious) has been omitted. Much of it is available in [1]. However, a short résumé of the theory of the Bahadur measures of performance and their properties is given in Section 2 for the sake of completeness, although the details here are also omitted.

2. Summary of the theory of Bahadur efficiency. Let X be a random quantity with sample space χ. We denote the sample space of all sequences of observations on X by $\boldsymbol{\chi} = (\chi \times \chi \times \cdots)$ with $\mathbf{x} = (x_1, x_2 \cdots)$ as a typical element of $\boldsymbol{\chi}$, where $x_i \varepsilon \chi$. Let $\{P_\theta : \theta \varepsilon \Omega\}$ be a collection of probability measures on $\boldsymbol{\chi}$, where Ω is an abstract parameter space, and let Ω_0 be a proper subset of Ω.

Received 1 August 1966; revised 11 April 1967.

[1] Some of the results presented in this paper formed part of a thesis submitted to the University of Chicago in fulfillment of the requirements for the Ph.D. The remaining research was done at Columbia University and some of the preparation was done at Bell Telephone Laboratories at Holmdel, New Jersey. This research was supported in part by Research Grant No. NSF-GP609 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation and in part by Research Grant No. NSF-GP-3694 at Columbia University.

Consider the hypothesis $H: \theta \varepsilon \Omega_0$. Let $\{T_n(\mathbf{x})\}$ be a sequence of non-negative real-valued functions on χ—i.e. statistics for use in testing H. Typically, T_n will be a function of the first n components of $\mathbf{x}$, i.e. of $(x_1, \cdots, x_n)$. $\{T_n(\mathbf{x})\}$ is called a *standard sequence* for testing H if the following three conditions hold:

I. Under H, $\{T_n\}$ converges in law to a non-degenerate random variable whose distribution is continuous and independent of $\theta \varepsilon \Omega_0$; i.e. there exists a non-degenerate, continuous distribution function, F, such that for all $\theta \varepsilon \Omega_0$ and real r,

$$\lim_{n\to\infty} P_\theta\{T_n < r\} = F(r).$$

II. There exists a constant $a \varepsilon (0, \infty)$ such that as $r \to \infty$,

$$\log [1 - F(r)] = -(ar^2/2)(1 + o(1)).$$

(This is true, for example, if the asymptotic distribution of $\{T_n\}$ is nornal).

III. There exists a non-negative, real-valued function, b, on Ω such that
(i) $b(\theta) > 0$ when $\theta \varepsilon \Omega - \Omega_0$, and
(ii) $T_n/n^{\frac{1}{2}} \to b(\theta)$ a.s. P_θ

$$(P_\theta\{T_n/n^{\frac{1}{2}} \to b(\theta)\} = 1).$$

From I and III, we see that $b(\theta) = 0$ when $\theta \varepsilon \Omega_0$ and $T_n(\mathbf{X}) \to \infty$ a.s. P_θ when $\theta \varepsilon \Omega - \Omega_0$. Since $\{T_n\}$ behaves when $\theta \varepsilon \Omega_0$ and blows up when $\theta \varepsilon \Omega - \Omega_0$, a statistician testing H by means of T_n would judge large values of T_n to be significant evidence against H. If he did not know (as is most often the case) the exact distribution of T_n under H, he could estimate the level attained by $T_n(\mathbf{X})$, which is the null probability of getting a more extreme value of T_n than he has, by the random variable

$$L_n = L(T_n(\mathbf{X})) = 1 - F(T_n(\mathbf{X})), \tag{2.1}$$

(ordinarily the statistician will know sufficient components of $\mathbf{X}$ to calculate $T_n(\mathbf{X})$).

If H does not hold, $L_n \to 0$ a.s. The behavior of L_n is better explored by examining the transformation

$$K_n = K(T_n) = -2 \log L(T_n).$$

It can then be shown that, for all $\theta \varepsilon \Omega$,

$$K_n/n \to a[b(\theta)]^2 \quad \text{a.s.} \quad P_\theta.$$

Thus K_n is asymptotically linear in n, which prompts the definition of the *approximate slope* (departing from the terminology of both [1] and [3]) of $\{T_n\}$ (or the test based on T_n) as the function

$$c(\theta) = a[b(\theta)]^2.$$

The approximation arises in (2.1) where the asymptotic null distribution of $\{T_n\}$ is used to compute L_n, which approximates the true level attained by T_n.

If $\theta \varepsilon \Omega - \Omega_0$ and ϵ is regarded as being a convincing level of significance, how large does N have to be for $L_N \leqq \epsilon < L_{N-1}$ (i.e., for the level ϵ to be attained by T_N)? N clearly varies with $\mathbf{X}$ and ϵ; but, in fact, when ϵ is small, N is essentially a large constant, for

$$\lim_{\epsilon\to 0} N/2 \log (1/\epsilon) = 1/c(\theta) \quad \text{a.s.} \quad P_\theta$$

or, more loosely, as $\epsilon \to 0$

$$N \sim 2 \log (1/\epsilon)/c(\theta) \quad \text{a.s.} \quad P_\theta, \theta \varepsilon \Omega - \Omega_0.$$

$c(\theta)$ can thus be interpreted as a measure of performance of $\{T_n\}$: the larger $c(\theta\}$, the faster $\{T_n\}$ rejects H by the approximate test based on the limiting F, for arbitrarily small ϵ.

The performances, as tests of H, of two standard sequences $\{T_n(\mathbf{X})\}$ and $\{U_n(\mathbf{X})\}$ with approximate slopes $c_T(\theta)$ and $c_U(\theta)$ respectively, may be compared by considering the ratio of the sample sizes needed by each to attain the same level using the same data. As $\epsilon \to 0$, we have for all $\theta \varepsilon \Omega - \Omega_0$,

$$N_T/N_U \to c_U(\theta)/c_T(\theta) \quad \text{a.s.} \quad P_\theta,$$

which suggests $\mathrm{E}_{U,T}(\theta) = c_U(\theta)/c_T(\theta)$ as an approximate measure of the asymptotic efficiency of $\{U_n\}$ relative to $\{T_n\}$. If Ω is a metric space and $\Omega - \Omega_0$ is dense in Ω, then for $\theta_0 \varepsilon \Omega_0$ and $\{\theta_\nu\}$ a sequence in $\Omega - \Omega_0$ with limit θ_0, we can define the approximate asymptotic limiting efficiency of $\{U\}$ relative to $\{T\}$ at θ_0 by

$$\mathcal{L}_{U,T}(\theta_0) = \lim_{\nu\to\infty} E_{U,T}(\theta_\nu),$$

assuming that the limit exists. $\mathcal{L}_{U,T}(\theta_0)$ is really a function of the path of approach, $\{\theta_\nu\}$, as well as θ_0. It is useful in that it is precisely those values of θ near θ_0 for which large sample sizes will usually be necessary for discrimination.

To get around the inexactness inherent in the foregoing analysis, more elaborate conditions are needed. $\{T_n\}$ is a *strictly standard sequence* if I and III hold and

II.′ There exists a non-negative function l on $[0, \infty]$ such that

(i) $l(z) > 0$ for $z \varepsilon (0, \infty)$, and

(ii) whenever $\{u_n\}$ is a sequence of real numbers for which $u_n^2/n \to z \varepsilon (0, \infty)$, we have uniformly for all $\theta \varepsilon \Omega_0$.

$$-\lim_{n\to\infty} (2/n) \log P_\theta\{T_n \geqq u_n\} = l(z).$$

Then the exact level attained by T_n is $L_n^*(\mathbf{X}) = \sup_{\theta\varepsilon\Omega_0} P_\theta\{\mathbf{x}: T_n(\mathbf{x}) \geqq T_n(\mathbf{X})\}$ and, with $K_n^*(\mathbf{X}) = -2 \log L_n^*(\mathbf{X})$, we have for all $\theta \varepsilon \Omega - \Omega_0$,

$$K_n^*/n \to l(b^2(\theta)) \quad \text{a.s.} \quad P_\theta.$$

The function $c^*(\theta) = l(b^2(\theta))$ is called the exact slope of $\{T_n\}$, and if, as before,

N^* is that sample size n at which $\{T_n\}$ "attains the level ϵ", it turns out that

$$N^*/2 \log (1/\epsilon) \to 1/c^*(\theta) \quad \text{a.s.} \quad P_\theta, \qquad \text{for all } \theta \varepsilon \Omega - \Omega_0 .$$

$E_{U,T}^*(\theta)$ and $\mathcal{L}_{U,T}^*$ are then defined as exact asymptotic and exact asymptotic limiting efficiencies, to correspond with E and $\mathcal{L}$ previously defined.

In general, it is not true that $c(\theta) = c^*(\theta)$. However, in most common cases where both $c(\theta)$ and $c^*(\theta)$ are available, we do have

$$c^*(\theta)/c(\theta) \to 1 \quad \text{as} \quad \theta \to \theta_0 \varepsilon \Omega_0 . \tag{2.2}$$

If this can be shown to be the case, then there is considerable justification for using $c(\theta)$ instead of $c^*(\theta)$ near Ω_0 (for the discussion of relative efficiencies, for example) when the former is more tractable, and if $\{U_n\}$ and $\{T_n\}$ are both standard and strictly standard and (2.2) holds for both, then $\mathcal{L}_{U,T} = \mathcal{L}_{U,T}^*$.

3. The one-sample problem.

3.1. The Kolmogorov-Smirnov Statistic. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables with common continuous distribution function $G(x)$, and let the sample distribution function determined by $X_1, \cdots, X_n$ be $F_n(x)$. For the purpose of testing the hypothesis $H: G = F$ (say), the (generalized) Kolmogorov-Smirnov statistic with weight function Ψ is defined by

$$K_{\Psi,n} = \text{Sup}_x\, n^{\frac{1}{2}}|F_n(x) - F(x)|\Psi(F(x)).$$

Anderson and Darling [2] have investigated the general asymptotic theory of $K_{\Psi,n}$; but they obtained explicit results only in the well-known case where $\Psi \equiv 1$, which yields the approximate slope (when the true distribution is $G \neq F$) for the standard sequence $\{K_{1,n}\}$ as $c_1(G, F) = 4d^2(G, F)$ where

$$d(G, F) = \text{Sup}_x |G(x) - F(x)|. \tag{3.1}$$

$c_\Psi(G, F)$ appears very difficult to obtain for $\Psi \not\equiv 1$, assuming that it exists, because the limiting distribution of $K_{\Psi,n}$ under the null hypothesis is not available.

The exact slope, $c_1^*(G, F)$, is implicit in the work of Sethuraman [11] and Hoadley [9] and has been explicitly obtained by the author ([1], p. 24) as

$$c_1^*(G, F) = c_1(G, F) + \tfrac{8}{9}[d^4(G, F)](1 + o(1))$$

when $d(G, F) \to 0$, which also demonstrates the local (near F) equivalence of c_1^* and c_1 of Equation (2.2).

We shall now find (under certain conditions) $c_\Psi^*(G, F)$, but we shall be unable to demonstrate (2.2) because of the lack of a suitable closed expression for the asymptotic distribution of $K_{\Psi,n}$.

The derivation depends on the following basic lemma:

LEMMA 1. *Let $Z_1, Z_2 \cdots$ be a sequence of independent observations on the random variable Z. Let $\bar{Z}_n$ be the mean of the first n observations, and $\varphi(t) = E(e^{tZ})$ be the moment generating function of Z. For a given $\epsilon > E(Z)$, suppose $E(e^{t(Z-\epsilon)}) =$*

$e^{-t\epsilon}\varphi(t)$ attains a minimum, ρ, at $t = \tau$ in the interior of $T = \{t:\varphi(t) < \infty\}$; i.e. *$\tau$ and ρ satisfy (where $d/dt\varphi(t) = \varphi'(t)$)*

$$\varphi'(\tau)/\varphi(\tau) = \epsilon \quad \text{and} \quad \rho = e^{-\tau\epsilon}\varphi(\tau).$$

Then

$$P\{\bar{Z}_n \geqq \epsilon\} \leqq \rho^n, \quad \text{and} \tag{3.2}$$

$$\lim_{n\to\infty} n^{-1} \log P\{\bar{Z}_n \geqq \epsilon\} = \log \rho. \tag{3.3}$$

For a proof of this Lemma, see e.g. [4].

It may be shown that no τ can be found if $\epsilon \leqq E(Z)$ and, indeed, if this were the case, $P\{\bar{Z}_n \geqq \epsilon\} = P\{\bar{Z}_n - E(Z) \geqq \epsilon - E(Z)\} \to 1$. The implicit function theorem guarantees that τ and hence ρ are differentiable functions of ϵ. Furthermore $\partial\rho/\partial\epsilon = -\tau e^{-\tau\epsilon}\varphi(\tau) < 0$, so that ρ is a decreasing function of ϵ which the result (3.3) would demand.

Returning to the one-sample problem, no generality is lost if we suppose $F(x) = x$ for $x \varepsilon (0, 1)$. With this understanding, consider a fixed x in $(0, 1)$ and define the independent binomial random variables $U_1, U_2 \cdots$ by

$$\begin{aligned} U_i &= 1 \quad \text{if } X_i \leqq x \\ &= 0 \quad \text{if } X_i > x. \end{aligned} \tag{3.4}$$

Then

$$F_n(x) = \bar{U}_n \tag{3.5}$$

and under H

$$\varphi_x(t) = E(e^{tU}) = xe^t + (1 - x). \tag{3.6}$$

It follows from Lemma 1 and (3.5) that for $\epsilon \varepsilon (0, 1)$,

$$\lim_{n\to\infty} n^{-1} \log P\{F_n(x) - x \geqq \epsilon\} = \log \rho_1(x, \epsilon), \quad \text{and} \tag{3.7a}$$

$$\lim_{n\to\infty} n^{-1} \log P\{F_n(x) - x \leqq -\epsilon\} = \log \rho_2(x, \epsilon), \quad \text{where} \tag{3.7b}$$

$$\begin{aligned} \rho_1(x, \epsilon) &= (x/(x + \epsilon))^{x+\epsilon}((1 - x)/(1 - x - \epsilon))^{1-x-\epsilon}, & 0 < x < 1 - \epsilon \\ &= 0, & 1 - \epsilon \leqq x < 1; \end{aligned} \tag{3.8a}$$

$$\begin{aligned} \rho_2(x, \epsilon)_1 &= 0, & 0 < x \leqq \epsilon \\ &= ((1 - x)/(1 - x + \epsilon))^{1-x+\epsilon}(x/(x - \epsilon))^{x-\epsilon}, & \epsilon < x < 1. \end{aligned} \tag{3.8b}$$

Notice that ρ_1 is a reflection of ρ_2 about $x = \frac{1}{2}$:

$$\rho_1(x, \epsilon) = \rho_2(1 - x, \epsilon). \tag{3.9}$$

Let

$$\begin{aligned} \rho(x, \epsilon) &= \max (\rho_i(x, \epsilon); i = 1, 2), & \epsilon \varepsilon (0, 1) \\ &= 0, & \epsilon \geqq 1. \end{aligned} \tag{3.10}$$

For $\epsilon > 0$, $\rho(x, \epsilon)$ is a decreasing and continuous function of ϵ.

It is then easy to prove:

LEMMA 2. *Under* H,

$$\lim_{n\to\infty} n^{-1} \log P\{|F_n(x) - x| \geqq \epsilon\} = \log \rho(x, \epsilon).$$

THEOREM 1. *If* $\Psi(x)$ *is a finite, positive-valued and continuous function of* x *in* $(0, 1)$; *and* $(1 - x)\Psi(x) \to 0$ *as* $x \to 1-$ and $x \to \Psi(x) \to 0$ *as* $x \to 0+$, *then under* H,

$$\lim_{n\to\infty} n^{-1} \log P\{K_{\Psi,n}/n^{\frac{1}{2}} \geqq \epsilon\} = \text{Sup}_{0<x<1} \log \rho(x, \epsilon/\Psi(x))$$

(*where* $\log 0$ *is defined as* $-\infty$).

PROOF. We certainly have

$$P\{K_{\Psi,n} \geqq n^{\frac{1}{2}}\epsilon\} \geqq P\{|F_n(x) - x|\Psi(x) \geqq \epsilon\}$$
$$= p_n(x) \text{ (say)}, \quad \text{for any } x.$$

Thus $\log P\{K_{\Psi,n} \geqq \epsilon n^{\frac{1}{2}}\} \geqq \log p_n(x)$ for any x, and hence for Lemma 2,

$$\liminf_{n\to\infty} n^{-1} \log P\{K_{\Psi,n} \geqq \epsilon n^{\frac{1}{2}}\} \geqq \lim_{n\to\infty} n^{-1} \log p_n(x) = \log \rho(x, \epsilon/\Psi(x))$$

for any x, and therefore

$$(3.11) \quad \liminf_{n\to\infty} n^{-1} \log \{K_{\Psi,n} \geqq \epsilon n^{\frac{1}{2}}\} \geqq \text{Sup}_x \log \rho(x, \epsilon/\Psi(x)) = \log \rho_\Psi^*(\epsilon).$$

Notice that no use has yet been made of the restrictions on Ψ.

$\rho_\Psi^*(\epsilon) = \text{Sup}_x\, \rho(x, \epsilon/\Psi(x))$ is attained for some $x \,\varepsilon\, (0, 1)$, for $\rho(x, \epsilon)$ is non-negative and jointly continuous in (x, ϵ) and therefore, by the continuity of $\Psi(x)$, $\rho(x, \epsilon/\Psi(x))$ is jointly continuous in (x, ϵ) for each fixed ϵ and hence attains its supremum in $[0, 1]$. In fact, this supremum must be attained in $(0,1)$ because $\rho \not\equiv 0$ and $\lim_{x\to 0} \rho(x, \epsilon/\Psi(x)) = \lim_{x\to 1} \rho(x, \epsilon/\Psi(x)) = 0$ by the conditions on Ψ.

Also, it is not hard to show that $\text{Sup}_x\, \rho(x, \epsilon/\Psi(x))$ is a continuous function of ϵ in $(0, 1)$. For if $\rho(x', \epsilon'/\Psi(x')) = \rho_\Psi^*(\epsilon')$ and $\rho(x'', \epsilon''/\Psi(x'')) = \rho_\Psi^*(\epsilon'')$, and $0 < \epsilon'' < \epsilon' < 1$, then

$$\rho(x', \epsilon''/\Psi(x')) - \rho(x', \epsilon'/\Psi(x')) \leqq \rho_\Psi^*(\epsilon'') - \rho_\Psi^*(\epsilon')$$
$$\leqq \rho(x'', \epsilon''/\Psi(x'')) - \rho(x'', \epsilon'/\Psi(x'')),$$

and since $\rho(x, \epsilon/\Psi(x))$ is continuous, $|\rho_\Psi^*(\epsilon'') - \rho_\Psi^*(\epsilon')| \to 0$ as $|\epsilon'' - \epsilon'| \to 0$. Now

$$(3.12) \quad \{K_{\Psi,n}/n^{\frac{1}{2}} \geqq \epsilon\}$$
$$\subset \bigcup_{0<x<1} \{|F_n(x) - x|\Psi(x) \geqq \epsilon\}$$
$$= \bigcup_{0<x<1} [\{F_n(x) - x \geqq \epsilon/\Psi(x)\} \cup \{F_n(x) - x \leqq -\epsilon/\Psi(x)\}].$$

Thus, for any positive integer N, since $F_n(x) - x \leqq F_n(i/N) - [(i - 1)/N]$ for $x \,\varepsilon\, [(i - 1)/N, i/N]$,

$$\begin{aligned}(3.13)\quad & \bigcup_{0<x<1}\{F_n(x) - x \geqq \epsilon/\Psi(x)\} \\ & \subset \bigcup_{i=1}^{N} \bigcup_{(i-1)/N<x\leqq i/N}\{F_n(x) - x \geqq \epsilon/\Psi(x)\} \\ & \subset [\bigcup_{i=2}^{N-1}\{F_n(i/N) - i/N \geqq \epsilon/\Psi_N^0(i/N) - 1/N\}] \\ & \cup [\bigcup_{(N-1)/N\leqq x<1}\{(1-x)\Psi(x) \geqq \epsilon\} \bigcup_{0<x\leqq 1/N}\{F_n(x) - x \geqq \epsilon/\Psi(x)\}]\end{aligned}$$

where $\Psi_N^0(i/N) = \text{Sup}\,(\Psi(x):(i-1)/N \leqq x < i/N)$, $i = 2, 3 \cdots N-1$.

Let M_n be defined by $1 - (1 - 1/M_n)^n < e^{-n^2}$. Then $M_n > N$ when n is sufficiently large. Define $\Psi_N^0(1/N) = \text{Sup}\,(\Psi(x):1/M_n \leqq x \leqq 1/N)$. Then

$$\bigcup_{0<x\leqq 1/N}\{F_n(x) - x \geqq \epsilon/\Psi(x)\} \subset [\bigcup_{0<x\leqq 1/M_n}\{F_n(x) \geqq \epsilon/\Psi(x)\}] \cup \{F_n(1/N) - 1/N \geqq \epsilon/\Psi_N^0(1/N) - 1/N\}.$$

But $A_{n,M} = \bigcup_{0<x<1/M_n}\{F_n(1/M_n) \geqq \epsilon/\Psi(x)\}$ occurs only if there is at least one observation X_i which is no larger than $1/M_n$, or else $F_n(1/M_n) = 0$, so that

$$(3.14)\qquad P(A_{n,M}) \leqq 1 - (1 - 1/M_n)^n < e^{-n^2} \qquad \text{by definition of } M_n.$$

By assumption, we have $(1-x)\Psi(x) \to 0$ as $x \to 1-$, so that for sufficiently large N it is impossible that $(1-x)\Psi(x) \geqq \epsilon$ for $x \in [(N-1)/N, 1]$.

Thus from (3.2), (3.8), (3.10), (3.13) and (3.14) we see that for N sufficiently large and n satisfying $M_n \geqq N$,

$$\begin{aligned} & P[\bigcup_{0<x<1}\{F_n(x) - x \geqq \epsilon/\Psi(x)\}] \\ & \leqq \textstyle\sum_{i=1}^{N-1} P\{F_n(i/N) - 1/N \geqq \epsilon/\Psi_N^0(i/N) - 1/N\} + [1 - (1 - 1/M_n)^n] \\ & \leqq \textstyle\sum_{i=1}^{N-1} \rho^n(1/N, \epsilon/\Psi_N^0(i/N) - 1/N) + e^{-n^2} \\ & \leqq N\, \text{Sup}_{1/M_n \leqq x \leqq 1-1/N}\, \rho^n(x, \epsilon/\Psi_N^*(x) - 1/N) + e^{-n^2}\end{aligned}$$

where $\Psi_N^*(x) = \{\text{Sup}\,\Psi(y): \max\,(1/M_n, x - 1/N) \leqq y \leqq x.\}$

A similar inequality may be proved for $P[\bigcup_x\{F_n(x) - x < -\epsilon/\Psi(x)\}]$ so that we have from (3.12)

$$P\{K_{\Psi,n}/n^{\frac{1}{2}} \geqq \epsilon\} \leqq 2N\ \text{Sup}_{1/M_n \leqq x \leqq 1-1/N}\, \rho^n(x, \epsilon/\Psi_N^*(x) - 1/N) + 2e^{-n^2}.$$

Then,

$$\begin{aligned} & n^{-1} \log P\{K_{\Psi,n}/n^{\frac{1}{2}} \geqq \epsilon\} \leqq (\log 2N)/n \\ & + n^{-1} \log \text{Sup}_{1/M_n \leqq x \leqq 1-1/N}\, \rho^n(x, \epsilon/\Psi_N^*(x) - 1/N) + n^{-1} \log (1 + o(1))\end{aligned}$$

so that

$$\limsup_{n\to\infty} n^{-1} \log P\{K_{\Psi,n}/n^{\frac{1}{2}} \geqq \epsilon\} \leqq \log \text{Sup}_{1/M_n \leqq x \leqq 1-1/N}\, \rho(x, \epsilon/\Psi_N^*(x) - 1/N).$$

But N is arbitrary, so that we must have

$$(3.15)\qquad \limsup_{n\to\infty} n^{-1} \log P\{K_{\Psi,n}/n^{\frac{1}{2}} \geqq \epsilon\} \leqq \log \rho_\Psi^*(\epsilon).$$

The theorem follows from (3.11) and (3.15).

Generally $\log \rho_\Psi^*(\epsilon)$ has no nice closed form, and one has to settle for the

TABLE I

	$F(x)$	$G(x)$	$\mathcal{L}_{1,\Psi_a}$
Shift alternatives	$\Phi(x)$	$\Phi(x-\theta)$	1
$\theta \to 0$	$C(x)$	$C(x-\theta)$	1
Scale alternatives	$\Phi(x)$	$\Phi(\theta x)$	.68
$\theta \to 1$	$C(x)$	$C(\theta x)$	.71
	$D_2(x)$	$D_2(\theta x)$	.74
	$D_4(x)$	$D_4(\theta x)$	.83
Mixture, $\theta \to 1-$	$\Phi(x)$	$\theta\Phi(x) + (1-\theta)\Phi(2x)$	.64

first one or two terms of the power series in ϵ, when ϵ is small. For example, when $\Psi = 1$,

$$(3.16) \qquad \log \rho_1^*(\epsilon) = -(2\epsilon^2 + \tfrac{4}{9}\epsilon^4 + c(\epsilon^4));$$

and when $\Psi(x) = [x(1-x)]^{-\frac{1}{2}} = \Psi_a(x)$ (say),

$$\log \rho_\Psi^*(\epsilon) = -(\tfrac{1}{2}\epsilon^2 + \tfrac{1}{12}\epsilon^4 + o(\epsilon^4)).$$

Thus the limiting efficiency of the statistic $K_{1,n}$ (the unweighted K-S) relative to

$$K_{\Psi_a,n} = n^{\frac{1}{2}} \operatorname{Sup}_x |F_n(x) - F(x)|/[F(x)\{1 - F(x)\}]^{\frac{1}{2}},$$

is

$$\mathcal{L}_{1,\Psi_a} = 4 \lim_{d(G,F)\to 0} \operatorname{Sup}_x |G(x) - F(x)|^2/\operatorname{Sup}_x [|G(x) - F(x)|^2/F(x)\{1 - F(x)\}].$$

Some values of $\mathcal{L}_{1,\Psi_a}$ for various F against alternatives G as $G \to F$, are given in Table I. ($\Phi(x)$ is the standard normal distribution function, $C(x) = 1/\pi \int_{-\infty}^{x} (1 + x^2)^{-1} dx$, $D_k(x) = P\{\chi_k^2 \leq x\}$.) Thus it appears that $K_{\Psi_a,n}$ is usually at least as good as the unweighted $K_{1,n}$.

3.2. The Kuiper Statistic: Kuiper [10] proposed the goodness-of-fit statistic for the one sample problem:

$$V_n = n^{\frac{1}{2}}[\operatorname{Sup}_x \{F_n(x) - F(x)\} - \operatorname{Inf}_x \{F_n(x) - F(x)\}]$$

in the notation of Figure 3.1, and found its limiting distribution. $\{V_n\}$ is a standard sequence with approximate slope

$$(3.17) \quad c_2(G, F) = 4[\operatorname{Sup}_x \{G(x) - F(x)\} - \operatorname{Inf}_x \{G(x) - F(x)\}]^2.$$

Using the methods of Section 3.1, it is possible to obtain the exact slope of $\{V_n\}$.

THEOREM 2. *With $\rho_1^*(\epsilon)$ defined by (3.11) when $\Psi \equiv 1$,*

$$\lim_{n\to\infty} n^{-1} \log P\{V_n/n^{\frac{1}{2}} \geqq \epsilon\} = \log \rho_1^*(\epsilon).$$

PROOF. The argument is basically the same as in Theorem 1, and will only be sketched below. We may assume $F(x) = x$ for $x \varepsilon (0, 1)$.

$$V_n/n^{\frac{1}{2}} = \text{Sup}_{x,y}\,[\{F_n(x) - x\} - \{F_n(y) - y\}]$$
$$\geqq |\{F_n(x) - F_n(y)\} - \{x - y\}| \qquad \text{for all } x > y.$$

Thus $\log P\{V_n/n^{\frac{1}{2}} \geqq \epsilon\} \geqq \text{Sup}_{x>y} \log P\{|\{F_n(x) - F_n(y)\} - \{x - y\}| \geqq \epsilon\}$. $F_n(x) - F_n(y)$ has the same distribution as $F_n(x - y)$, hence by Lemma 2,

$$(3.18) \qquad \liminf_{n\to\infty} n^{-1} \log P\{V_n/n^{\frac{1}{2}} \geqq \epsilon\} \geqq \text{Sup}_{x>y} \log \rho(x - y, \epsilon) = \log \rho_1^*(\epsilon).$$

Let $N > 2/\epsilon$. To each x, y in $(0, 1)$ correspond i, j such that $i/N < x \leqq (i + 1)/N$ and $j/N < y \leqq (j + 1)/N$, so that by the monotonicity of $F_n(x)$ and x,

$$P\{\{F_n(x) - F_n(y)\} - (x - y) \geqq \epsilon\}$$
$$\leqq P\{F_n(i/N) - F_n(j - 1/N) - (i - (j - 1))/N \geqq \epsilon - 2/N\},$$

$$P\{V_n/(n)^{\frac{1}{2}} \geqq \epsilon\}$$
$$\leqq P\{\textstyle\bigcup_{i,j}\{F_n(i/N) - F_n((j - 1)/N) - (i - (j - 1))/N \geqq \epsilon - 2/N\}\}$$
$$\leqq \textstyle\sum_{ij} P\{F_n(i/N) - F_n((j - 1)/N) - (i - (j - 1))/N \geqq \epsilon - 2/N\}$$
$$\leqq N^2[\rho_1^*(\epsilon - 2/N)]^n.$$

$$\limsup_{n\to\infty} n^{-1} \log P\{V_n/n^{\frac{1}{2}} \leqq \epsilon\} \leqq \lim_{n\to\infty} [n^{-1} \log N^2 + \log \rho_1^*(\epsilon - 2/N)];$$
$$= \log \rho_1^*(\epsilon - 2/N);$$

but N may be arbitrarily large, which implies, by the continuity of $\rho_1^*(\epsilon)$,

$$(3.19) \qquad \limsup_{n\to\infty} n^{-1} \log P\{V_n/n^{\frac{1}{2}} \geqq \epsilon\} \leqq \rho_1^*(\epsilon).$$

(3.18) and (3.19) prove Theorem 2.

Thus, from (3.16) the exact slope of $\{V_n\}$ is seen to be

$$c_2^*(G, F) = c_2(G, F) + \tfrac{1}{18}c_2^2(G, F) + o(d^4(G, F)) \quad \text{as} \quad d(G, F) \to 0.$$

The exact asymptotic efficiency of $\{V_n\}$ relative to $\{K_{1,n}\}$ is

$$E_{V,K}^*(F, G) = \log \rho_1^*(\text{Sup}\,(F - G) - \text{Inf}\,(F - G))/\log \rho_1^*\,(\text{Sup}\,|F - G|).$$

Since $\rho_1^*(\epsilon)$ is a decreasing function of ϵ and $\log \rho_1^*(\epsilon) < 0$, $E_{V,K}^*(F, G) \geqq 1$; i.e. V_n is always at least as efficient at $K_{1,n}$ against any alternative whatever. If, for definiteness, we suppose $\text{Sup}\,(G - F) = d(G, F)$, the limiting efficiency of $\{V_n\}$ relative to $\{K_{1,n}\}$ is

$$(3.20) \qquad \mathcal{L}_{V,K} = 1 + \lim_{|G-F|\to 0}[2\,\text{Sup}\,(F - G)/\text{Sup}\,(G - F) + \{\text{Sup}(F - G)/\text{Sup}\,(G - F)\}^2] \geqq 1.$$

Some values of $\mathcal{L}_{V,K}$ for various F against alternatives G are given in Table II.

$\mathcal{L}_{V,K}$ is graphed in Figures 1 and 2 as indicated. Figure 1 shows that, against scale alternatives, for Normal and Cauchy distributions $\mathcal{L}_{V,K}$ drops off quite

TABLE II

	$F(x)$	$G(x)$	$\mathfrak{L}_{V,K}$
Shift alternatives $\theta \to 0$	$F(x)$	$F(x + \theta)$	1
Scale alternatives $\theta \to 1$	$F(x) = 1 - F(-x)$	$F(x\theta)$	4
	$\Phi(x - \mu)(\mu > 0)$	$\Phi(\theta x - \mu)$	$[1 + \{(2r(\mu) - \mu)/(2r(\mu) + \mu)\} \cdot \exp(-\mu r(\mu))]^2$ $r(\mu) = [1 + \mu^2/4]^{\frac{1}{2}}$ (See Figure 1(a))
	$C(x - \mu)(\mu > 0)$	$C(\theta x - \mu)$	$4[\mu^2 + 1 - \mu(\mu^2 + 1)^{\frac{1}{2}}]^2$ (See Figure 1(b))
Mixtures	$F(x) = 1 - F(-x)$	$\theta F(x) + (1 - \theta)F(\alpha x)$ $(\alpha > 0)$	4
$\theta \to 1$	$\Phi(x)$	$\theta\Phi(x) + (1 - \theta)\Phi[\alpha x - \mu]$ $(\alpha > 1, \mu > 0)$	$[1 + \{\Phi(s - \alpha t) - \Phi(\alpha s - t)\}/\{\Phi(\alpha s + t) - \Phi(s + \alpha t)\}]^2$ $s = \alpha\mu/(\alpha^2 - 1)$ $t = \{s^2 + 2(\log\alpha)/(\alpha^2 - 1)\}^{\frac{1}{2}}$ (See Figure 2)

sharply from 4 until at about $\mu = 2$, there is almost nothing to choose between the two statistics. In Figure 2 for the case of Normal mixtures, we find that V_n is always at least four times as efficient as K_n.

These results may appear quite startling in that they depend on the fact that under H

$$n^{-1} \log P\{V_n \geqq n^{\frac{1}{2}}\epsilon\} \sim n^{-1} \log P\{K_{1,n} \geqq n^{\frac{1}{2}}\epsilon\} \tag{3.21}$$

which says that the distributions of V_n and $K_{1,n}$ are of the same exponential order in the tails. It is obvious that with high probability, $V_n > K_{1,n}$ and therefore the limiting distributions must be quite dissimilar in their main parts. Thus one might doubt the reliability of the conclusions in view of the fact that the tails and the main parts of a sequence of distributions do not necessarily have the same limiting properties. However, the Bahadur theory concerns itself with how well H "explains" the sequences $\{V_n\}$ or $\{K_{1,n}\}$ when, in fact, H is false and V_n and $K_{1,n}$ are growing roughly in proportion to $n^{\frac{1}{2}}$. The very fact that $V_n \geqq K_{1,n}$ assures that, according to (3.21) V_n "attains a smaller level of significance" than does K_n and therefore rejects H more emphatically.

4. The two-sample problem. It rarely happens that the null distribution $F(x)$, needed for $K_{\Psi,n}$ and V_n, is known. Most often, we will be confronted by the two-sample problem, in which it has to be decided whether two samples could reasonably be supposed to have been drawn from the same population.

4.1. The Kolmogorov-Smirnov Statistic: Let $X_1, X_2, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be independent samples of independent observations on X and Y respectively, with continuous distribution functions F and G respectively. Let F_m and G_n

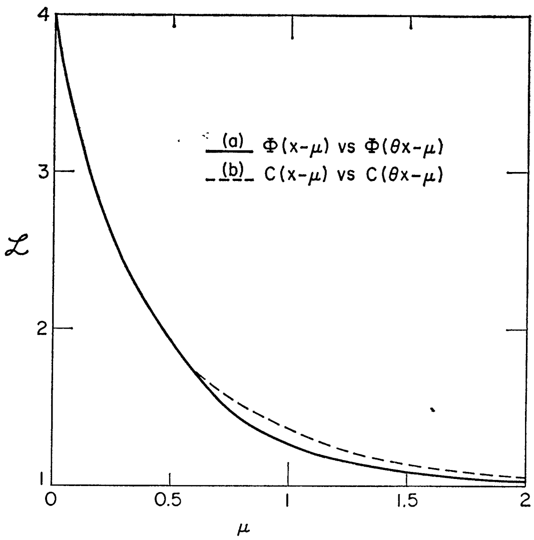

FIG. 1

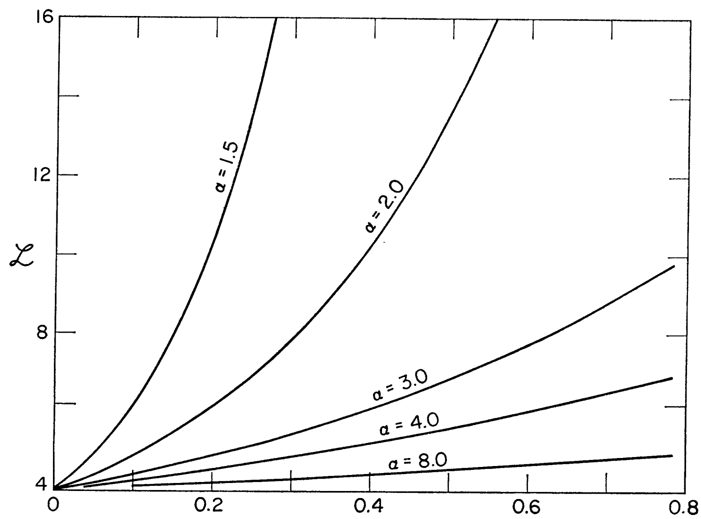

FIG. 2

be the corresponding sample distribution functions. For the purpose of testing the hypothesis H that $F \equiv G$, the usual K-S two-sample statistic is defined by

$$K_{m,n} = [mn/(m + n)]^{\frac{1}{2}} \operatorname{Sup}_x |F_m(x) - G_n(x)| = [mn/(m + n)]^{\frac{1}{2}} d(F_m, G_n).$$

It is possible to weight this in a similar way to the one-sample case, but the weight function would be of the form $\Psi(F_m, G_n)$—i.e., random—and no exact results are as yet available for this case.

We shall suppose that m and n increase in such a way that $m/(m + n) \to \nu \varepsilon (0, 1)$. In fact we shall require that ν be rational. $\nu = r/k$ where r and k are integers and $m = rM$, $n = (k - r)M$.
Even without this last restriction, it may be shown that when $F = G$,

$$K_{m,n}/(m + n)^{\frac{1}{2}} \to \{\nu(1 - \nu)\}^{\frac{1}{2}} d(F, G) \quad \text{a.s.}$$

Since the limiting distribution of $K_{m,n}$ is precisely that of $K_{1,n}$, the approximate slope of the standard sequence $\{K_{m,n}\}$ may be computed, and it is found to be

$$\begin{aligned} c_1(G, F, \nu) &= 4\nu(1 - \nu)\, d^2(F, G) \\ &= \nu(1 - \nu)c_1(G, F). \end{aligned} \tag{4.1}$$

In order to find the exact slope of $\{K_{m,n}\}$, it is necessary to develop a two-sample analogue of Lemma 1.

LEMMA 3. *Let $Z_1^{(i)}, Z_2^{(i)} \cdots$ be independent sequences of independent random variables $Z^{(i)}$, $i = 1, 2$. Let the mean of the first l observations in the* ith *sample be $\bar{Z}_l^{(i)}$. Let $\varphi_i(t) = E(e^{Z^{(i)}t})$.*

Suppose there exists a positive number τ in the interior of

$$T_\nu = \{t : \varphi_1(t/\nu)\varphi_2((-t)/(1 - \nu)) < \infty\}$$

such that, for a given number $\epsilon > E[Z^{(1)} - Z^{(2)}]$,

$$\varphi_1'(\tau/\nu)/\varphi_1(\tau/\nu) - \varphi_2'(-\tau/(1 - \nu))/\varphi_2(-\tau/(1 - \nu)) = \epsilon, \tag{4.2}$$

and let

$$\rho = \epsilon^{-\tau\epsilon}[\varphi_1(\tau/\nu)]^\nu[\varphi_2((-\tau)/(1 - \nu))]^{1-\nu}. \tag{4.3}$$

Then
$$P\{\bar{Z}_m^{(1)} - \bar{Z}_n^{(2)} \geqq \epsilon\} \leqq \rho^{n+m}$$
and
$$\lim_{m,n\to\infty} (m + n)^{-1} \log P\{\bar{Z}_m^{(1)} - \bar{Z}_n^{(2)} \geqq \epsilon\} = \log \rho$$
where $m, n \to \infty$ in such a way that $m/(m + n) = r/k = \nu$.

PROOF. Let $W^{(i)}$ be a random variable defined as the mean of $r(i = 1)$ or $k - r(i = 2)$ independent observations on $Z^{(i)}$. From the sequences $\{Z_l^{(i)}\}$, we can form independent sequences of independent observations on $W^{(i)}$, where the $(j + 1)$th observation is

$$W_{j+1}^{(1)} = [Z_{rj+1}^{(1)} + Z_{rj+2}^{(1)} + \cdots + Z_{r(j+1)}^{(1)}]/r$$
$$W_{j+1}^{(2)} = [Z_{(k-r)j+1}^{(2)} + Z_{(k-r)j+2}^{(2)} + \cdots + Z_{(k-r)(j+1)}^{(2)}]/(k - r)$$

so that

$$\begin{aligned} \textstyle\sum_{j=1}^{M} W_j^{(i)}/M = \bar{W}_M^{(i)} &= \bar{Z}_m^{(1)}, \quad \text{if } i = 1 \\ &= \bar{Z}_n^{(2)}, \quad \text{if } i = 2, \end{aligned}$$

and therefore

$$\bar{Z}_m^{(1)} - \bar{Z}_n^{(2)} = \bar{W}_M^{(1)} - \bar{W}_M^{(2)} = \overline{(W^{(1)} - W^{(2)})_M}. \tag{4.4}$$

The moment generating function of $k(W^{(1)} - W^{(2)})$ is

$$\varphi^*(t) = [\varphi_1(t/\nu)]^{\nu k}[\varphi_2(-t/(1-\nu))]^{(1-\nu)k}.$$

$\varphi^*(t) < \infty$ if and only if $t \varepsilon T_\nu$. Hence by Lemma 1,

$$P\{k(\bar{W}_M^{(1)} - \bar{W}_M^{(2)}) \geqq k\epsilon\} \leqq \rho^{kM}$$

where ρ is defined by (4.2) and (4.3), and

$$\lim_{M\to\infty} M^{-1} \log P\{k(\bar{W}_M^{(1)} - \bar{W}_M^{(2)}) \geqq k\epsilon\} = k \log \rho.$$

The desired result follows from (4.4) and the fact that $kM = m + n$.

Admittedly, Lemma 3 is restrictive in that it only handles rational values of ν and demands that the sample sizes increase in a special way. However, these conditions are not entirely unrealistic, and many of the existing results concerning the two-sample Kolmogorov-Smirnov statistic require $m = n$ or $r = 1$ (e.g. [5], [7], [8]).

Again, we may assume that under H, $F(x) = G(x) = x$ for $x \varepsilon (0, 1)$. If each $Z^{(i)}$ has the same distribution as U_i in (3.4) and we identify $F_m(x) = \bar{Z}_m^{(1)}$ and $G_n(x) = \bar{Z}_n^{(2)}$, Lemma 3 establishes that, in the two-sample case, H implies

$$\text{(4.5)} \quad \lim_{m,n\to\infty} (m+n)^{-1} \log P\{|F_m(x) - G_n(x)| \geqq \epsilon\} = \log \rho_0(x, \epsilon)$$

where $\rho_0(x, \epsilon) = \max(\rho_{0,1}(x, \epsilon), \rho_{0,2}(x, \epsilon))$ is defined by

$$\rho_{0,1}(x, \epsilon) = e^{-\epsilon\tau_1}[\varphi_x(\tau_1/\nu)]^\nu[\varphi_x(-\tau_1/(1-\nu))]^{1-\nu},$$

$$\rho_{0,2}(x, \epsilon) = e^{-\epsilon\tau_2}[\varphi_x(-\tau_2/\nu)]^\nu[\varphi_x(\tau_2/(1-\nu))]^{1-\nu},$$

$$\varphi_x(t) = xe^t + (1-x),$$

$$x\{e^{\tau_1/\nu}/\varphi_x(\tau_1/\nu) - e^{-\tau_1/(1-\nu)}/\varphi_x(-\tau_1/(1-\nu))\} = \epsilon,$$

$$x\{e^{-\tau_2/\nu}/\varphi_x(-\tau_2/\nu) - e^{\tau_2/(1-\nu)}/\varphi_x(\tau_2/(1-\nu))\} = -\epsilon,$$

THEOREM 3.

$$\lim_{m,n\to\infty} (m+n)^{-1} \log P\{K_{m,n}/(m+n)^{\frac{1}{2}} \geqq \epsilon\}$$
$$= \text{Sup}_x \log \rho_0(x, \epsilon/(\nu(1-\nu))^{\frac{1}{2}})$$
$$= \log \rho_0^*(\epsilon/(\nu(1-\nu))^{\frac{1}{2}}),$$

where $\rho_0^*(\epsilon/(\nu(1-\nu))^{\frac{1}{2}}) = \text{Sup}_x\, \rho_0(x, \epsilon/(\nu(1-\nu))^{\frac{1}{2}})$. The proof of Theorem 3 is similar to the proofs of Theorems 1 and 2 and slightly simpler than the proof of Theorem 4, and therefore we will not give it here.

It may be shown that

$$\text{(4.6)} \quad \log \rho_0^*(\epsilon/(\nu(1-\nu))^{\frac{1}{2}}) = -2\epsilon^2 + 0(\epsilon^4)$$

(for details, see [1], Section 3.2: the conjecture stated there can be proved to be correct). Thus

$$c_1^*(G, F, \nu) = 4\nu(1-\nu)d^2(F, G)(1 + o(1))$$

$$= c_1(G, F, \nu)(1 + o(1)) \tag{4.7}$$
$$= \nu(1 - \nu)c_1(G, F)(1 + o(1))$$

as $d(F, G) \to 0$, so that the approximate and exact slopes of $\{K_{m,n}\}$ are locally equivalent.

4.2. The two-sample Kuiper statistic:

The two-sample Kuiper statistic is defined as

$$V_{m,n} = [mn/(m + n)]^{\frac{1}{2}}[\mathrm{Sup}_x \{F_n(x) - G_m(x)\} - \mathrm{Inf}_x \{F_n(x) - G_m(x)\}].$$

Kuiper [10] derived the limiting distribution of $V_{m,n}$ for the case where $m = n$ and this distribution is identical with the limiting distribution in the one-sample case (Kuiper's normalization is different to that used here, which is more common, with the result that the two distributions he gives appear different). On the basis of our knowledge of one- and two-sample K-S and Cramer-Smirnov-von Mises statistics, and Theorem 4, we may conjecture that (under the usual restrictions on m and n) the asymptotic distribution of $V_{m,n}$ is in general identical with that of V_n. The non-availability of the general asymptotic distribution of $V_{m,n}$ means that we cannot compute the approximate slope of $\{V_{m,n}\}$, except in the case where $\nu = \frac{1}{2}$, in which case it turns out to be

$$c_2(G, F, \nu) = \nu(1 - \nu)c_2(G, F), \tag{4.8}$$

with $c_2(G, F)$ defined as in (3.17), which is consistent with the relationship (4.1). We will show the corresponding analogue of (4.7), viz.:

$$c_2^*(G, F, \nu) = \nu(1 - \nu)c_2(G, F)(1 + o(1))$$

as $d(G, F) \to 0$.

THEOREM 4.

$$\lim_{m,n\to\infty} (m + n)^{-1} \log P\{V_{m,n}/(m + n)^{\frac{1}{2}} \geqq \epsilon\} = \log \rho_0^*(\epsilon/(\nu(1 - \nu))^{\frac{1}{2}}).$$

PROOF.

$$\{V_{m,n}/(m + n)^{\frac{1}{2}} \geqq \epsilon\}$$
$$= \bigcup_{x>y} \{|[F_n(x) - G_m(x)] - [F_n(y) - G_m(y)]| \geqq \epsilon/(\nu(1 - \nu))^{\frac{1}{2}}\}$$
$$\supset \{|F_n(x) - F_n(y)] - [G_m(x) - G_m(y)]| \geqq \epsilon/(\nu(1 - \nu))^{\frac{1}{2}}\}$$

for any $x > y$. Thus from (4.5) we have

$$\liminf_{m,n\to\infty} (m + n)^{-1} \log P\{V_{m,n}/(m + n)^{\frac{1}{2}} \geqq \epsilon\}$$
$$\geqq \lim_{m,n\to\infty} (m + n)^{-1} \log P\{|F_n(x) - F_n(y) - G_m(x) + G_m(y)| \geqq \epsilon/(\nu(1 - \nu))^{\frac{1}{2}}\}$$
$$= \log \rho_0(x - y, \epsilon/(\nu(1 - \nu))^{\frac{1}{2}})$$

for any $x > y$, which implies

$$\liminf_{m,n\to\infty} (m+n)^{-1} \log P\{V_{m,n}/(m+n)^{\frac{1}{2}} \geqq \epsilon\}$$

$$\geqq \operatorname{Sup}_x \log \rho_0(x, \epsilon/(\nu(1-\nu))^{\frac{1}{2}}) = \log \rho_0^*(\epsilon/(\nu(1-\nu))^{\frac{1}{2}}). \tag{4.9}$$

Let N be a number so large that $N > 4(\nu(1-\nu))^{\frac{1}{2}}/\epsilon$ and let $x_1 < x_2 < \cdots < x_N$ be chosen so that $(i-1)/N \leqq G_m(x_i) < i/N$, $i = 1 \cdots N$ for $m > N$ (the x_i's will vary with m). Then we have

$$0 < G_m(x_i) - G_m(x_{i-1}) < 2/N. \tag{4.10}$$

When $x_{i-1} \leqq x < x_i$ and $x_{j-1} \leqq y < x_j$, monotonicity and (4.10) imply

$$[F_n(x) - F_n(y)] - [G_m(x) - G_m(y)]$$
$$\leqq [F_n(x_i) - F_n(x_{j-1})] - [G_m(x_{i-1}) - G_m(x_j)]$$
$$\leqq [F_n(x_i) - F_n(x_{j-1})] - [G_m(x_i) - G_m(x_{j-1}) - 4/N],$$

$$\{V_{m,n}/(m+n)^{\frac{1}{2}} \geqq \epsilon\} \subset \bigcup_{i=1}^{N} \bigcup_{j=2}^{N+1} \{[F_n(x_i) - F_n(x_{j-1})] - [G_m(x_i) - G_m(x_{j-1})] \geqq \epsilon/(\nu(1-\nu))^{\frac{1}{2}} - 4/N\}$$

$$P\{V_{m,n}/(m+n)^{\frac{1}{2}} \geqq \epsilon\} \leqq \sum_{i=1}^{N} \sum_{j=2}^{N+1} P\{[F_n(x_i) - F_n(x_{j-1})] - [G_m(x_i) - G_m(x_{j-1})] \geqq \epsilon/(\nu(1-\nu))^{\frac{1}{2}} - 4/N\}$$
$$\leqq N^2[\rho_0^*(\epsilon/(\nu(1-\nu))^{\frac{1}{2}} - 4/N)]^{m+n}.$$

Thus

$$\limsup_{m,n\to\infty} (m+n)^{-1} \log P\{V_{m,n}/(m+n)^{\frac{1}{2}} \geqq \epsilon\} \leqq \log \rho_0^*(\epsilon/(\nu(1-\nu))^{\frac{1}{2}} - 4/N).$$

This result is true for any sufficiently large N, and $\rho_0^*(\epsilon)$ is continuous in ϵ (by an argument similar to that for $\rho^*(\epsilon)$ in Theorem 1), so that we must have

$$\limsup_{m,n\to\infty} (m+n)^{-1} \log P\{V_{m,n}/(m+n)^{\frac{1}{2}} \geqq \epsilon\} \leqq \log \rho_0^*(\epsilon/(\nu(1-\nu))^{\frac{1}{2}}).$$

This, together with (4.9), proves Theorem 4.

Theorems 3 and 4 could also be proved using Theorem 1 of [9] and the calculus of variations.

From (4.6) it is evident that

$$c_2^*(G, F, \nu) = 4\nu(1-\nu)[\operatorname{Sup}_x \{F(x) - G(x)\} + \operatorname{Sup}_x \{G(x) - F(x)\}]^2 + o(d^2(F, G)).$$

Thus the expression for the limiting efficiency of $V_{m,n}$ relative to $K_{m,n}$ will be the same as the limiting efficiency of V_n relative to K_n in the one-sample case.

5. Acknowledgments. The author is deeply indebted to: Professor R. R. Bahadur for his advice and encouragement which made this paper possible;

Dr. Bruce Hoadley for his valuable suggestions; the referee for his very careful reading and constructive comments; and many others who have devoted some of their time to the preparation of this paper.

REFERENCES

[1] ABRAHAMSON, I. G. (1965). On the stochastic comparison of tests of hypotheses. University of Chicago. Unpublished dissertation.

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

[3] BAHADUR, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31** 276–295.

[4] BAHADUR, R. R. AND RANGA RAO, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[5] BLACKMAN, J. (1958). Correction to "an extension of the Kolmogorov distribution". *Ann. Math. Statist.* **29** 318–324.

[6] DARLING, D. A. (1957). Kolmogorov-Smirnov, Cramér-von Mises tests. *Ann. Math. Statist.* **28** 823–834.

[7] GNEDENKO, B. V. AND RVACEVA, E. R. (1952). On a problem of the comparison of two distributions. *Math. Statist. Prob.* **1** (1961) 69–72.

[8] GNEDENKO, B. V. (1952). Some results on the maximum discrepancy between two empirical distributions. *Select. Transl. Math. Statist. Prob.* **1** (1961) 73–75.

[9] HOADLEY, A. B. (1967). On the probability of large deviations of functions of several empirical cdf's. *Ann. Math. Statist.* **38** 360–381.

[10] KUIPER, N. H. (1960). Tests concerning random points on a circle. *Proc. Konink. Ned. Akad. van Wettenschaffen, A.* **63** 38–47.

[11] SETHURAMAN, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316.