ON THE PROBABILITY OF LARGE DEVIATIONS OF FUNCTIONS OF SEVERAL EMPIRICAL CDF'S[1]

By A. Bruce Hoadley[2]

University of California, Berkeley

1. Summary. In [14], Sanov proved that if F_N is the empirical cumulative distribution function (cdf) of a sample drawn from a population whose true cdf is F_0 and Ω is a set of cdf's which satisfies certain regularity conditions and does not contain F_0, then $P\{F_N \varepsilon \Omega\}$ is roughly $\exp\{-N \inf_{F\varepsilon\Omega} \int \ln (dF/dF_0)\, dF\}$. This theory is extended to the c-sample case and to the case where the set of cdf's in question depends on N. These extensions are used to estimate the probability of a large deviation of those statistics which are, or can be approximated by, uniformly continuous functions of the empirical cdf's. As an example, the main result is applied to the Wilcoxon statistic, and the resulting formula is used to compute the exact Bahadur efficiency of the Wilcoxon test relative to the t-test.

2. Introduction. If $\{T_N\}_{N=1}^{\infty}$ is a sequence of statistics, defined on some probability space $(\mathfrak{X}, S, P)$, for which $T_N \rightarrow_P \theta$ as $N \rightarrow \infty$, then we shall say that the event $\{T_N \geqq \theta + \epsilon_N\}$ is a large deviation of T_N if $\lim_{N\rightarrow\infty} N^{-1} \ln P\{T_N \geqq \theta + \epsilon_N\} = -I$ with $0 < I < \infty$. To facilitate discussion, I will be called the index of the large deviation.

In recent years, several definitions of asymptotic relative efficiency between two tests of some hypothesis (see [2], [4], and [10]) have been proposed which require the computation of indices of large deviations for their application. Until recently, the computation of indices had been studied extensively only in the case of sums of independent random variables. The theory of large deviations has now been extended by Sanov [14], Sethuraman [15], [16], Abrahamson [1], and others, to include statistics which are not sums of independent random variables.

The approach taken by Sethuraman and Abrahamson was to extend the result for the sample mean to statistics that can be written as $\sup_{f\varepsilon\mathfrak{F}} N^{-1} \sum_{j=1}^{N} f(X_j)$, where $X_1, \cdots, X_N$ are independent identically distributed random variables in a metric space $\mathfrak{X}$, and $\mathfrak{F}$ is some well-behaved class of real valued functions on $\mathfrak{X}$. In [1], Abrahamson used the above approach to compute the exact Bahadur efficiency for a large number of classical statistics including the F-statistic, studentized range, between-sample sum of squares, and 1 and 2 sample Kol-

Received 24 March 1966; revised 11 October 1966.

[1] This paper is part of a Doctoral Dissertation submitted by the author to the Graduate Division of the University of California, Berkeley; and was prepared with the partial support of the National Science Foundation, Grant GP-2593 (at the Department of Statistics, University of California, Berkeley). Also part of the work was done while the author was an NSF Cooperative Graduate Fellow. The revisions and the computations were done at Bell Telephone Laboratories, Inc.

[2] Now at Bell Telephone Laboratories, Inc., Holmdel, New Jersey.

mogorov-Smirnov statistics. Also, Klotz [12] utilized a theorem of Feller [7] on the large deviation of sums of independent, non-identically distributed random variables to compute the exact Bahadur efficiency of the one-sample Wilcoxon and normal scores tests.

In [14], Sanov shows that if F_N is the empirical cumulative distribution function (cdf) of a sample drawn from a population with cdf F_0, and Ω is a certain well-behaved set of cdf's, not containing F_0, then $\lim_{N\to\infty} N^{-1} \ln P\{F_N \varepsilon \Omega\} = -\inf_{F\varepsilon\Omega} \int \ln (dF/dF_0)\, dF$. In this paper Sanov's approach to large deviation theory is amended and extended to yield theorems which give the indices for large deviations of statistics which are functions of several empirical cdf's.

Fundamental to Sanov's approach to large deviation theory is the concept of information, which is discussed in [13]. Let D denote the set of cdf's on $(-\infty, +\infty)$. For F and G in D, let F and G be absolutely continuous with respect to μ (e.g., $\mu = (F + G)/2$). Let $f = dF/d\mu$ and $g = dG/d\mu$, and let

$$\begin{aligned} f \ln (f/g) &= 0 && \text{if } f = 0 \\ &= +\infty && \text{if } f > 0 \text{ and } g = 0 \\ &= \text{itself} && \text{if } f > 0 \text{ and } g > 0. \end{aligned}$$

Then

$$I(F, G) = \int f \ln (f/g)\, d\mu \tag{2.1}$$

is well defined and has the following well-known properties (see [13]): (i) $0 \leqq I(F, G) \leqq \infty$. (ii) *if* $I(F, G) < \infty$, *then* $F \ll G$ *and* $I(F, G) = \int \ln (dF/dG)\, dF$. (iii) $I(F, G) = 0$ *if and only if* $F = G$. In [13], $I(F, G)$ is called the mean information for discrimination between F and G per observation from F.

Let D^c denote the cartesian product of D with itself c-times; then for any $Q = (F_1, \cdots, F_c)$ and $R = (G_1, \cdots, G_c)$ in D^c, and $\rho = (\rho_1, \cdots, \rho_c)$ we define

$$I_\rho^c(Q, R) = \textstyle\sum_{j=1}^{c} \rho_j I(F_j, G_j). \tag{2.2}$$

If $\Sigma \subset D$ and $\Omega \subset D^c$, we introduce the special notation

$$\begin{aligned} I(\Sigma, F_0) &= \inf_{F\varepsilon\Sigma} I(F, F_0), \\ I_\rho^c(\Omega, Q_0) &= \inf_{Q\varepsilon\Omega} I_\rho^c(Q, Q_0). \end{aligned} \tag{2.3}$$

The ρ and c will be dropped from I_ρ^c whenever it is clear from the context what is meant.

We will consider the c-sample set-up, which can be described as follows: Let $X_{j,1}, \cdots, X_{j,n_j}$, $j = 1, \cdots, c$, denote c independent samples, where $X_{j,1}, \cdots, X_{j,n_j}$ is a sample drawn from a population with continuous cdf $F_{j,0}$. Let $N = \sum_{j=1}^{c} n_j$, and, for the asymptotic theory, let $N \to \infty$ in such a way that $|n_j/N - \rho_j| = O(\ln N/N)$ as $N \to \infty$, and $\rho_j > 0$, for $j = 1, \cdots, c$. Let $Q_0 =$

$(F_{1,0}, \cdots, F_{c,0})$ and $S_N = (F_{1,n_1}, \cdots, F_{c,n_c})$, where F_{j,n_j} is the empirical cdf of the jth sample; i.e. $F_{j,n_j}(x) = n_j^{-1}$[no. of X_{ji}'s $\leqq x$].

Since we shall be dealing with functions of several empirical cdf's, it is useful to introduce a metric space on which these functions can be defined. Let E denote the normed linear space of functions of bounded variation on $(-\infty, +\infty)$, with $\|h\| = \sup_x |h(x)|$ for $h \varepsilon E$. Let E^c be the cartesian product of E with itself c-times, and for $H = (h_1, \cdots, h_c) \varepsilon E^c$, define $\|H\|_c = \max_{1 \leqq j \leqq c} \{\|hj\|\}$. $\|\cdot\|_c$ is a norm on E^c; hence we can consider E^c as a metric space with the metric induced by this norm.

Suppose T is a real valued function with domain D^c, then let

$$\Omega_r = \{Q \varepsilon D^c \colon T(Q) \geqq r\}, \qquad I(r) = I(\Omega_r, Q_0); \tag{2.4}$$

and let $\{\xi_N\}_{N=1}^{\infty}$ be a sequence of numbers for which $\lim_{N\to\infty} \xi_N = 0$.

THEOREM 1. *If T is uniformly continuous, then for every $r > T(Q_0)$ at which $I(r)$ is continuous,*

$$\lim_{N\to\infty} N^{-1} \ln P\{T(S_N) \geqq r + \xi_N\} = -I(\Omega_r, Q_0).$$

Theorem 1 is applicable to estimating the probability of large deviations of many statistics which arise in the theory of testing. For example, in the one-sample case, it handles the Kolmogorov-Smirnov one- and two-sided tests of fit, and the Cramér-Von Mises test of fit, because they are uniformly continuous functions of the empirical cdf. In the two sample case, it applies to the Chernoff-Savage statistics, (see [5]), when the kernal function J is bounded; for example in the Wilcoxon case. However, the sample mean, which can be written $T(F_N) = \int x\, dF_N$, where F_N is the empirical cdf of the sample, is not a uniformly continuous function of the empirical cdf, so Theorem 1 is not applicable. But $T^{(B)}(F_N) = \int_{|x| \leqq B} x\, dF_N$ is uniformly continuous and approximates $\int x\, dF_N$. This idea is explored to yield a more general theorem which gives the index of a large deviation of any statistic T_N which can be suitably approximated by a sequence of uniformly continuous functions of the empirical cdf's.

For each positive integer B, let $T^{(B)}$ be a uniformly continuous real valued function with domain D^c for which $r_0 = \lim_{B\to\infty} T^{(B)}(Q_0)$ exists. Let

$$\Omega_y^{(B)} = \{Q \varepsilon D^c \colon T^{(B)}(Q) \geqq y\}, \qquad I^{(B)}(y) = I(\Omega_y^{(B)}, Q_0). \tag{2.5}$$

If $\mathbf{X}_N$ denotes the c-samples of observations, we consider statistics of the form $T_N = T_N(\mathbf{X}_N)$, where T_N is Borel measurable. Define

$$d_N^{(B)} = T^{(B)}(S_N) - T_N. \tag{2.6}$$

THEOREM 2. *Suppose r and r_1 are numbers with $r_0 < r < r_1$ for which conditions* (I), (II), and (III) *hold* (*see Section* 4.); *then*

$$\lim_{N\to\infty} N^{-1} \ln P\{T_N \geqq r + \xi_N\} = -\lim_{B\to\infty} I^{(B)}(r).$$

Conditions (I) and (II) are conditions involving the continuity of the two functions $I^{(B)}(y)$ and $I(y) = \lim_{B\to\infty} I^{(B)}(y)$. Condition (III) states: for each sequence $B_N \to \infty$, there exists a sequence $\delta_N \downarrow 0$ such that if P_N denotes the conditional distribution of $d_N^{(B_N)}$ given $\{T_N \geqq y\}$ and Q_N denotes the conditional distribution of $d_N^{(B_N)}$ given $\{T^{(B_N)}(S_N) \geqq y + \delta_N\}$, then $\int_{-\delta_N}^{\delta_N} dP_N$ and $\int_{-\delta_N}^{\delta_N} dQ_N$ do not tend to zero exponentially or faster with N.

This theorem gives an alternative approach to finding the index of a large deviation of the sample mean, a problem which has been solved by Cramér [6], Chernoff [4], Bahadur and Rao [3], and others. In the two sample case, it can be applied to the difference of the sample means (for details see [9]); and the author conjectures that the Chernoff-Savage statistics with unbounded J can be handled if suitable restrictions are imposed.

3. Proof of Theorem 1. The proof of Theorem 1 requires a considerable amount of construction and preliminary lemmas. Some of the ideas and definitions are due to Sanov [14]. These will be pointed out as the discussion progresses.

We start out by introducing a certain kind of elementary set of cdf's which will be used to approximate more complicated sets. For F_0 a fixed continuous cdf, let $(a)_M = \{-\infty = a_0 < a_1 < \cdots < a_{M-1} < a_M = +\infty\}$ be a partition of $(-\infty, +\infty)$ which is defined by

$$a_k = \inf\{x: F_0(x) = k/M\}, \quad \text{for } k = 1, \cdots, M-1. \tag{3.1}$$

Let

$$\begin{aligned} I_k &= ((k-1)/M, k/M], & k &= 2, \cdots, M, \\ &= [0, 1/M], & k &= 1, \end{aligned} \tag{3.2}$$

be M sub-intervals of [0, 1]. The partition, $(a)_M$, of $(-\infty, +\infty)$ and the partition, $\{0 < 1/M < \cdots < (M-1)/M < 1\}$, of [0, 1] form a grid on $(-\infty, +\infty) \times [0, 1]$, which shall be called an *F_0-grid of size M* (see Fig. 3.1). We use this grid to associate with each cdf, H, a set of cdf's which is defined as follows: For $k = 1, \cdots, M-1$, suppose $H(a_k) \varepsilon I_{j(k)}$, then $\{F: F(a_k) \varepsilon I_{j(k)}$ for $k = 1, \cdots, M-1\}$ shall be called a *F_0-strip of size M* (see Fig. 3.1). This concept is closely related to that of ϵ-neighborhood introduced in [14]. The functions

$$\begin{aligned} \underline{H}(x) &= 0 & &\text{for } x \varepsilon (-\infty, a_1) & & \\ &= (j(k)-1)/M & &\text{for } x \varepsilon [a_k, a_{k+1}), & k &= 1, \cdots, M-1, \\ \bar{H}(x) &= j(k)/M & &\text{for } x \varepsilon (a_{k-1}, a_k], & k &= 1, \cdots, M-1, \\ &= 1 & &\text{for } x \varepsilon (a_{M-1}, \infty) & & \end{aligned}$$

shall be called the *lower and upper border functions* of the F_0 strip; and $\bar{H}(a_k) - \underline{H}(a_k) = 1/M$ shall be called the width of the *F_0-strip* at a_k (see Fig. 3.1).

As in [14], the set of cdf's F, for which $F(a_k) = v_k$, where $0 = v_0 \leqq v_1 \leqq \cdots \leqq v_{M-1} \leqq v_M = 1$, shall be called a *cylinder of size M*; v_k shall be called the vertex

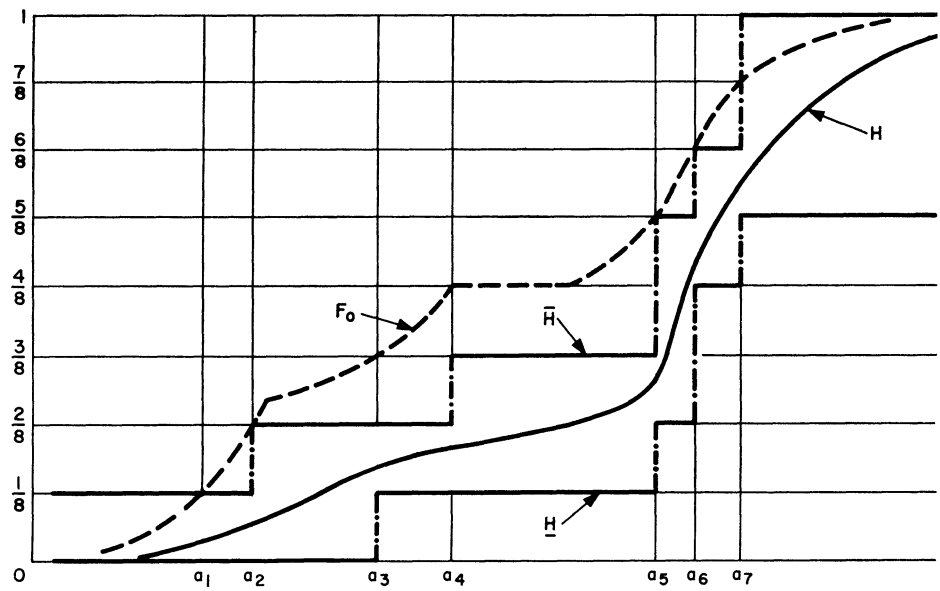

FIG. 3.1. F_0-grid of size 8 and F_0-strip of size 8.

of the cylinder at a_k ; and $\Delta_k v = v_k - v_{k-1}$ shall be called the *increments of the vertices* (see Fig. 3.2). If $v_k = l/N$ with l an integer and $0 \leqq l \leqq N$, then the cylinder is called an *N-cylinder*. Note that empirical cdf's fall in N-cylinders.

For any cdf H, let $\Delta_k H$ denote $H(a_k) - H(a_{k-1})$. A cdf, F, is said to be *F_0-piecewise-linear* on $(a)_M$, if for $k = 1, \cdots, M$,

$$(3.3) \quad F(x) = F(a_{k-1}) + (\Delta_k F/\Delta_k F_0)[F_0(x) - F_0(a_{k-1})] \quad \text{for } x \in (a_{k-1}, a_k)\cdot$$

We shall call $m_k = (\Delta_k F/\Delta_k F_0)$ the *F_0-slope of F on (a_{k-1}, a_k)*. (see Fig. 3.2).

LEMMA 3.1. *If F is F_0-piecewise-linear on $(a)_M$, with F_0-slope m_k on (a_{k-1}, a_k), then* (i) $m_k = M(\Delta_k F)$, (ii) $\sum_{k=1}^M m_k = M$, (iii) $I(F, F_0) = M^{-1} \sum_{k=1}^M m_k \ln m_k \leqq \ln M$.

PROOF. (i) By construction $\Delta_k F_0 = M^{-1}$, hence $m_k = (\Delta_k F/\Delta_k F_0) = M(\Delta_k F)$.

(ii) $\sum_{k=1}^M m_k = \sum_{k=1}^M M(\Delta_k F) = M$.

(iii) $I(F, F_0) = \sum_{k=1}^M (\Delta_k F) \ln (\Delta_k F/\Delta_k F_0) = M^{-1} \sum_{k=1}^M m_k \ln m_k$.

Also,

$$\begin{aligned} I(F, F_0) &= \textstyle\sum_{k=1}^M (\Delta_k F) \ln \Delta_k F - \sum_{k=1}^M (\Delta_k F) \ln \Delta_k F_0 \\ &= \textstyle\sum_{k=1}^M (\Delta_k F) \ln \Delta_k F + \ln M \\ &\leqq \ln M. \end{aligned}$$

Suppose Δ is a cylinder with vertices $v_1, \cdots, v_{M-1}$. Then by definition, Δ is just the class of all cdf's which pass through the points $(a_1, v_1), \cdots, (a_{M-1}, v_{M-1})$. The unique F_0-piecewise-linear cdf, which passes through these points,

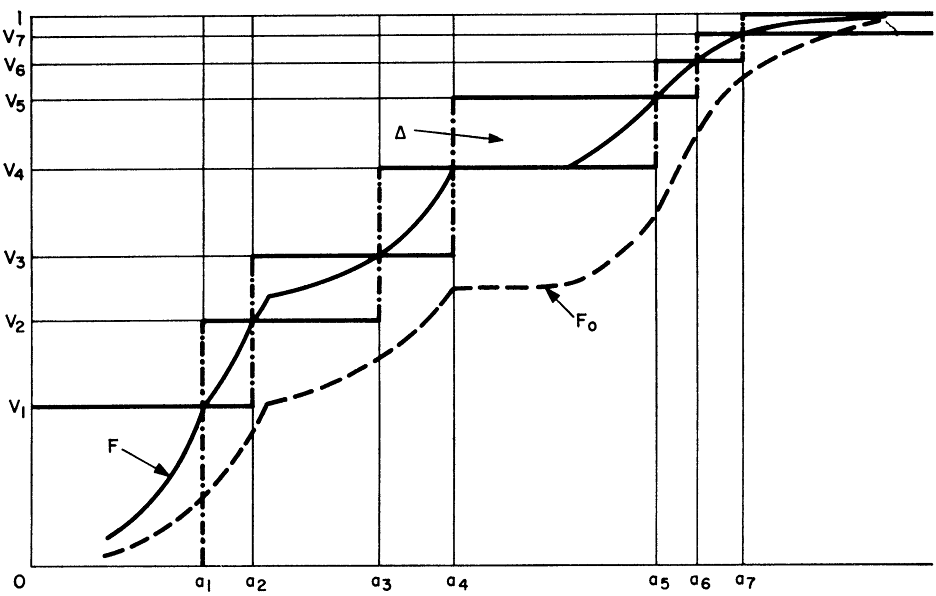

FIG. 3.2. F is F_0-piecewise-linear and is the F_0-extremal cdf of the cylinder Δ.

is called the *F_0-extremal* cdf *of* Δ(see Fig. 3.2), and we have

LEMMA 3.2 (*Theorem* 6 *of Sanov* [14]). *If* Δ *is the cylinder with vertices* $v_1, \cdots, v_{M-1}$ *with respect to* $(a)_M$, *then*

$$I(\Delta, F_0) = I(\Phi_0, F_0) = \sum_{k=1}^{M} \Delta_k v \ln (\Delta_k v/\Delta_k F_0),$$

where Φ_0 *is the* F_0*-extermal* cdf *of* Δ.

The next step is to derive an asymptotic expression for $P\{F_N \varepsilon V_N\}$, where V_N is a sequence of F_0-strips of size M_N and $M_N \to \infty$ at a certain rate. Such an expression is given in [14], p. 239, line 5, for a certain sequence of ϵ-neighborhoods; but the proof does not seem to be valid for the generality considered there. However, the basic ideas in Sanov's proof apply to the proof of the lemma below

LEMMA 3.3. *Let* $X_1, \cdots, X_N$ *be a sample drawn from a population with continuous* cdf F_0, *and let* F_N *be the empirical* cdf. *If* $\{V_N\}$ *is a sequence of* F_0*-strips of size* $M_N = o(N/\ln N)$, *then*

$$P\{F_N \varepsilon V_N\} = \exp \{-N[I(V_N, F_0) + O(M_N \ln N/N)]\}$$

uniformly in V_N *as* $N \to \infty$

PROOF. For notational convenience, let $M = M_N$. The width of V_N at a_k is M^{-1}; hence, since $M = o(N/\ln N)$, it is clear that for sufficiently large N, there is an N-cylinder contained in V_N. Now let Δ_N be any N-cylinder contained in V_N with vertices $v_1, \cdots, v_{M-1}$. It follows from the multinomial distribution, Sanov [14], pp. 214–215, formulae (3) and (4), and Lemma 3.2 that

$$\begin{aligned} P\{F_N \,\varepsilon\, \Delta_N\} &= \exp\{-N[\textstyle\sum_{k=1}^{M} (\Delta_k v) \ln (\Delta_k v/\Delta_k F_0) + O(M \ln N/N)]\} \\ &= \exp\{-N[I(\Delta_N, F_0) + O(M \ln N/N)]\} \qquad (3.4)\\ &\leqq \exp\{-N[I(V_N, F_0) + O(M \ln N/N)]\} \end{aligned}$$

uniformly in V_N as $N \to \infty$.

Let $\{\Delta_{N,k}\}_{k=1}^{s}$ be the set of all distinct N-cylinders contained in V_N. The number of possible vertices of N-cylinders at a_k, $k = 1, \cdots, M-1$, is $N+1$; so $s < (N+1)^M$, and hence

$$\begin{aligned} P\{F_N \,\varepsilon\, V_N\} &= \textstyle\sum_{k=1}^{s} P\{F_N \,\varepsilon\, \Delta_{N,k}\} \\ &\leqq (N+1)^M \exp\{-N[I(V_N, F_0) + O(M \ln N/N)]\} \qquad (3.5)\\ &= \exp\{-N[I(V_N, F_0) + O(M \ln N/N)]\} \end{aligned}$$

uniformly in V_N as $N \to \infty$.

Let $\tau_N = O(M \ln N/N)$, and let Φ_N^* be a cdf in V_N for which

$$I(V_N, F_0) \leqq I(\Phi_N^*, F_0) < I(V_N, F_0) + \tau_N.$$

Also let Δ_N^* be a cylinder of size M defined by the points

$$(a_1, \Phi_N^*(a_1)), \cdots, (a_{M-1}, \Phi_N^*(a_{M-1})),$$

and let Φ_N be the F_0-extremal cdf of Δ_N^*. Then $\Phi_N \,\varepsilon\, V_N$ and

$$I(V_N, F_0) \leqq I(\Phi_N, F_0) \leqq I(\Phi_N^*, F_0) < I(V_N, F_0) + \tau_N. \qquad (3.6)$$

Since $M = o(N/\ln N)$, it is clear that for sufficiently large N, there is an N-cylinder, Δ_N', contained in V_N, with vertices $v_1', \cdots, v_{M-1}'$, for which

$$|v_k' - \Phi_N(a_k)| \leqq N^{-1}, \qquad k = 1, \cdots, M-1.$$

It is generally true that for $x, y \,\varepsilon\, [0, 1]$,

$$|x - y| < 2/N \Rightarrow |x \ln x - y \ln y| < \epsilon_N = O(\ln N/N).$$

This along with (3.6) yields:

$$\begin{aligned} 0 &\leqq I(\Delta_N', F_0) - I(V_N, F_0) \\ &\leqq I(\Delta_N', F_0) - I(\Phi_N, F_0) + \tau_N \\ &= \textstyle\sum_{k=1}^{M} [\Delta_k v' \ln (\Delta_k v'/\Delta_k F_0) - \Delta_k \Phi_N \ln (\Delta_k \Phi_N/\Delta_k F_0)] + \tau_N \\ &\leqq \textstyle\sum_{k=1}^{M} |\Delta_k v' \ln \Delta_k v' - \Delta_k \Phi_N \ln \Delta_k \Phi_N| \\ &\quad + \textstyle\sum_{k=1}^{M} |\Delta_k \Phi_N - \Delta_k v'| \, |\ln \Delta_k F_0| + \tau_N \\ &\leqq M\epsilon_N + (2/N)M \ln M + \tau_N \\ &= O(M \ln N/N), \end{aligned}$$

uniformly in V_N as $N \to \infty$. Combining this with (3.4) gives

$$\begin{aligned} P\{F_N \varepsilon V_N\} &\geqq P\{F_N \varepsilon \Delta_N'\} \\ &= \exp\{-N[I(\Delta_N', F_0) + O(M \ln N/N)]\} \\ &= \exp\{-N[I(V_N, F_0) + O(M \ln N/N)]\} \end{aligned} \tag{3.7}$$

uniformly in V_N as $N \to \infty$. The lemma follows from (3.5) and (3.7).

We now generalize Lemma 3.3 to the c-sample set-up. If, for $j = 1, \cdots, c$, V_j is an $F_{j,0}$-strip of size M, then $U = V_1 \times \cdots \times V_c \subset D^c$ will be called a *Q_0-product strip of size M.*

LEMMA 3.4. *In the c-sample set-up, if $\{U_N\}$ is a sequence of Q_0-product strips of size $M_N = o(N/\ln N)$, then*

$$P\{S_N \varepsilon U_N\} = \exp\{-N[I(U_N, Q_0) + O(M_N \ln N/N)]\},$$

uniformly in U_N as $N \to \infty$.

PROOF. U_N is a Q_0-product strip; hence it can be written $U_N = V_{1,N} \times \cdots \times V_{c,N}$, where $V_{j,N}$ is an $F_{j,0}$-strip of size M_N. By independence and Lemma 3.3, we get

$$\begin{aligned} P\{S_N \varepsilon U_N\} &= P\{(F_{1,n_1}, \cdots, F_{c,n_c}) \varepsilon V_{1,N} \times \cdots \times V_{c,N}\} \\ &= \textstyle\prod_{j=1}^c P\{F_{j,n_j} \varepsilon V_{j,N}\} \\ &= \textstyle\prod_{j=1}^c \exp\{-N[(n_j/N)I(V_{j,N}, F_{j,0}) + O(M_N \ln N/N)]\} \\ &= \exp\{-N[I(U_N, Q_0) + \textstyle\sum_{j=1}^c (n_j/N - \rho_j)I(V_{j,N}, F_{j,0}) \\ &\qquad + O(M_N \ln N/N)]\} \end{aligned} \tag{3.8}$$

uniformly in U_N as $N \to \infty$. But $V_{j,N}$ contains a cylinder and thus contains the the $F_{j,0}$-extremal cdf (call it G_j) of that cylinder; hence by Lemma 3.1

$$I(V_{j,N}, F_{j,o}) \leqq I(G_j, F_{j,o}) \leqq \ln M_N.$$

Therefore

$$|\textstyle\sum_{j=1}^c (n_j/N - \rho_j)I(V_{j,N}, F_{j\cdot 0})| \leqq \ln M_N \textstyle\sum_{j=1}^c |n_j/N - \rho_j| = O(M_N \ln N/N),$$

uniformly in U_N as $N \to \infty$. Applying this to (3.8) yields the lemma.

In order to get an asymptotic expression for $P\{S_N \varepsilon \Omega_N\}$, where Ω_N is a sequence of sets in D^c, we must put regularity conditions on Ω_N. These conditions amount to saying that Ω_N can be approximated by Q_0-product strips.

DEFINITION 3.1. A sequence, $\{\Omega_N\}_{N=1}^{\infty}$, of subsets of D^c is said to be a Q_0-regular sequence if for each $\rho = (\rho_1, \cdots, \rho_c)$, with $\rho_j > 0$, the following conditions are satisfied:

(A) $I(\Omega_N, Q_0) < \infty$, and $\lim_{N\to\infty} I(\Omega_N, Q_0) = I < \infty$;

(B) for each $\eta > 0$, there exists $K + 1 = K(\eta, N) + 1$ Q_0-product strips of size $M_N = o(N/\ln N)$, $U_N, U_{1,N}, \cdots, U_{K,N}$, such that

(i) $U_N \subset \Omega_N$,

(ii) $I(U_N, Q_0) < I(\Omega_N, Q_0) + \eta$,
(iii) $\Omega_N \subset \bigcup_{i=1}^{K} U_{i,N}$,
(iv) $I(U_{i,N}, Q_0) > I(\Omega_N, Q_0) - \eta$.

In the one sample case, this is closely related to the concept of F_0-distinguishability, which was introduced by Sanov [14], p. 233. The difference is that here we use F_0-strips of size $M_N = o(N/\ln N)$ instead of ϵ-neighborhoods, and we drop the condition that $K = o(N)$.

Lemma (3.5). *For the c-sample set-up, if* $\{\Omega_N\}$ *is a* Q_0*-regular sequence, then*

$$\lim_{N\to\infty} N^{-1} \ln P\{S_N \varepsilon \Omega_N\} = -I.$$

Proof. Let $U_N, U_{1,N}, \cdots, U_{K,N}$ be the Q_0-product strips given in Definition 3.1. By Lemma 3.4 and condition (B)-(i), (ii) of Definition 3.1, we have

$$\begin{aligned} P\{S_N \varepsilon \Omega_N\} &\geqq P\{S_N \varepsilon U_N\} \\ &= \exp\{-N[I(U_N, Q_0) + O(M_N \ln N/N)]\} \\ &\geqq \exp\{-N[I(\Omega_N, Q_0) + \eta + O(M_N \ln N/N)]\}. \end{aligned} \tag{3.9}$$

From each $F_{j,0}$-grid of size M, one can construct no more than $(M+2)^{M+1} F_{j,0}$-strips of size M; hence

$$K \leqq (M_N + 2)^{c[M_N+1]}. \tag{3.10}$$

So by condition (B)-(iii), (iv) of Definition 3.1, (3.10), and Lemma 3.4 we have

$$\begin{aligned} P\{S_N \varepsilon \Omega_N\} &\leqq \textstyle\sum_{i=1}^{K} P\{S_N \varepsilon U_{i,N}\} \\ &= \textstyle\sum_{i=1}^{K} \exp\{-N[I(U_{i,N}, Q_0) + O(M_N \ln N/N)]\} \\ &\leqq K \max_{1\leqq i\leqq K} \exp\{-N[I(U_{i,N}, Q_0) + O(M_N \ln N/N)]\} \\ &\leqq (M_N + 2)^{c[M_N+1]} \exp\{-N[\min_{1\leqq i\leqq K} I(U_{i,N}, Q_0) \\ &\qquad + O(M_N \ln N/N)]\} \\ &\leqq \exp\{-N[I(\Omega_N, Q_0) - \eta + O(M_N \ln N/N)]\}. \end{aligned} \tag{3.11}$$

Putting (3.9) and (3.11) together yields

$$\begin{aligned} O(M_N \ln N/N) - \eta - I(\Omega_N, Q_0) & \\ &\leqq N^{-1} \ln P\{S_N \varepsilon \Omega_N\} \\ &\leqq -I(\Omega_N, Q_0) + O(M_N \ln N/N) + \eta \end{aligned} \tag{3.12}$$

as $N \to \infty$. Letting $N \to \infty$, noting that $\eta > 0$ is arbitrary, and using condition (A) of Definition 3.1 proves Lemma 3.5.

Before getting into the proof of Theorem 1, we require one more lemma.

Lemma 3.6. *Let* F *be an* F_0*-piecewise linear* cdf *with* F_0*-slope* m_k *on* (a_{k-1}, a_k). *Let* $\hat{m} = \max_{1\leqq k\leqq M} m_k$. *If* $\hat{m} > M\alpha_M$, *where* $\alpha_M \downarrow 0$ *in such a way that* $\alpha_M^{-1} = o(\ln M)$ *as* $M \to \infty$, *then* $I(F, F_0) \geqq b_M$, *where* $\lim_{M\to\infty} b_M = +\infty$.

PROOF. Let $m_l = \hat{m}$. Then from Lemma 3.1, we have that

$$I(F, F_0) = M^{-1}[\sum_{k \neq l} m_k \ln m_k + \hat{m} \ln \hat{m}].$$

It is generally true that if $u_1, \cdots, u_\nu, v_1, \cdots, v_\nu \geqq 0$, then

$$\sum_{i=1}^{\nu} v_i \ln (v_i/u_i) \geqq (\sum_i v_i) \ln [(\sum_i v_i)/(\sum_i u_i)].$$

(See [8], p. 97, inequality no. 17.) Using this inequality, and the fact that $\sum_{k \neq l} m_k = M - \hat{m}$, the above becomes

$$I(F, F_0) \geqq M^{-1}\{(M - \hat{m}) \ln [(M - \hat{m})/(M - 1)] + \hat{m} \ln \hat{m}\}. \tag{3.13}$$

Applying the hypothesis that $\hat{m} > M\alpha_M$ to (3.13) yields, for $M\alpha_M \geqq 1$,

$$\begin{aligned} I(F, F_0) &\geqq (1 - \hat{m}/M)\{\ln (1 - \hat{m}/M) + \ln[M/(M - 1)]\} \\ &\qquad + M^{-1}(M\alpha_M) \ln M\alpha_M \\ &\geqq -[\exp \{1\}]^{-1} + \alpha_M \ln M + \alpha_M \ln \alpha_M \\ &= b_M . \end{aligned} \tag{3.14}$$

But $\alpha_M \to 0 \Rightarrow \alpha_M \ln \alpha_M \to 0$; and $\alpha_M^{-1} = o(\ln M) \Rightarrow \alpha_M \ln M \to +\infty$; hence $\lim_{M \to \infty} b_M = +\infty$.

We are now in a position to prove Theorem 1. The first step is to show that the constant sequence $\Omega_r = \{Q \varepsilon D^c: T(Q) \geqq r\}$ is a Q_0-regular sequence for every $r > T(Q_0)$ at which $I(r)$ is continuous.

CONDITION (A). $I(\Omega_r, Q_0) = I(r) < \infty$, because I is continuous at r.

CONDITION (B). Since Ω_r does not depend on N, we need not consider N in the verification of condition (B). Fix $\eta > 0$. By the continuity of I at r, we may choose $\delta > 0$ such that $I(r + \delta) < I(r) + \eta/2$. Now choose $Q^* = (F_1^*, \cdots, F_c^*) \varepsilon \Omega_{r+\delta}$ such that

$$\begin{aligned} I(Q^*, Q_0) < I(\Omega_{r+\delta}, Q_0) + \eta/2 = I(r + \delta) + \eta/2 < I(r) + \eta \\ = I(\Omega_r, Q_0) + \eta. \end{aligned} \tag{3.15}$$

Since T is continuous at Q^*, we can choose $\epsilon > 0$ such that

$$\|Q - Q^*\|_c < \epsilon \Rightarrow T(Q) > T(Q^*) - \delta \geqq r. \tag{3.16}$$

For $j = 1, \cdots, c$, $I(F_j^*, F_{j,0}) < \infty$ implies that F_j^* is absolutely continuous with respect to $F_{j,0}$, hence there exists $\theta > 0$ such that $\Delta_k F_{j,0} < \theta \Rightarrow \Delta_k F_j^* < \epsilon/2$. Since $F_{j,0}$ is continuous, we can choose M so large that $\Delta_k F_{j,0} = 1/M < \theta$, for $j = 1, \cdots, c$ and $k = 1, \cdots, M$, and hence $\Delta_k F_j^* < \epsilon/2$. Let $V_{j,M}^*$ be the $F_{j,0}$-strip of size M which contains F_j^* and let $\bar{F}_j^*(\underline{F}_j^*)$ be the upper (lower) border function of $V_{j,M}^*$. By construction

$$\|\bar{F}_j^* - \underline{F}_j^*\| \leqq \max_{1 \leqq k \leqq M} \Delta_k F_j^* + 2/M;$$

so if we also choose M so large that $2/M < \epsilon/2$, then

$$\|\bar{F}_j^* - \underline{F}_j^*\| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Thus for any F_j and G_j in $V^*_{j,M}$, $\|F_j - G_j\| < \epsilon$; and hence for any Q in $U_M{}^* = V^*_{1,M} \times \cdots \times V^*_{c,M}$, $\|Q - Q^*\|_c < \epsilon$, which implies by (3.16) that $T(Q) > r$; so $U_M{}^* \subset \Omega_r$. Applying (3.15) yields

$$(3.17) \qquad I(U_M{}^*, Q_0) \leqq I(Q^*, Q_0) < I(\Omega_r, Q_0) + \eta;$$

hence $U_M{}^*$ is that product-strip required in condition (B)-(i), (ii).

For parts (iii) and (iv) fix $\eta > 0$; then by the continuity of I at r, we can choose $\delta > 0$ such that

$$(3.18) \qquad I(r - \delta) > I(r) - \eta/2.$$

Since T is uniformly continuous, we can choose $\epsilon > 0$ such that

$$(3.19) \qquad \|Q - R\|_c < \epsilon \Rightarrow T(R) > T(Q) - \delta/2.$$

Consider the sequences $\{\alpha_M\}$ and $\{b_M\}$ defined in Lemma 3.6, and choose M so large that for $m \geqq M$,

$$(3.20) \qquad \alpha_m + 2/m < \epsilon \quad \text{and} \quad b_m > (\min_{1 \leqq j \leqq c} \rho_j)^{-1} I(r).$$

Consider now the finite class of all Q_0-product strips of size M, and let $\mathcal{C}$ be the sub-class of product strips U for which either $\inf_{Q \varepsilon U} T(Q) \geqq r$ or $\inf_{Q \varepsilon U} T(Q) < r \leqq \sup_{Q \varepsilon U} T(Q)$. Label the members of $\mathcal{C}$, $U_1, \cdots, U_K$. We proceed now to show that $U_1, \cdots, U_K$ are the product strips required by condition (B)-(iii), (iv).

Every $F_{j,0}$-strip V_j of size M contains a $F_{j,0}$-piecewise linear cdf; hence by Lemma 3.1, $I(V_j, F_{j,0}) \leqq \ln M$. For any Q_0-product strip, $U = V_1 \times \cdots \times V_c$, it is true that, $I(U, Q_0) = \sum_{j=1}^{c} \rho_j I(V_j, F_{j,0}) \leqq \ln M$; thus $I(U_i, Q_0) < \infty$ for $i = 1, \cdots, K$. From the definition of $\mathcal{C}$, it follows that $\Omega_r = \{Q: T(Q) \geqq r\} \subset \bigcup_{i=1}^{K} U_i$; hence it remains only to show that

$$(3.21) \qquad I(\Omega_r, Q_0) - \eta < I(U_i, Q_0), \qquad \text{for } i = 1, \cdots, K.$$

For those $U \varepsilon \mathcal{C}$ for which $\inf_{Q \varepsilon U} T(Q) \geqq r$, we have $U \subset \Omega_r$; hence (3.21) holds. The remaining members of $\mathcal{C}$ need special attention.

Let

$$(3.22) \qquad \mathcal{S} = \{U \varepsilon \mathcal{C}: \inf_{Q \varepsilon U} T(Q) < r \leqq \sup_{Q \varepsilon U} T(Q)\}.$$

Each $U \varepsilon \mathcal{S}$ can be written $U = V_1 \times \cdots \times V_c$, where V_j is an $F_{j,0}$-strip of size M. It is clear that there exists $F_j{}^* \varepsilon V_j$ such that

$$(3.23) \qquad I(F_j{}^*, F_{j,0}) < I(V_j, F_{j,0}) + \eta/2,$$

and $F_j{}^*$ is $F_{j,0}$-piecewise linear with an $F_{j,0}$-slope of $m_{j,k}$ on $(a_{j,k-1}, a_{j,k})$. Let $\hat{m}_j = \max_{1 \leqq k \leqq M} m_{j,k}$. Note that $Q^* = (F_1{}^*, \cdots, F_c{}^*) \varepsilon U$, and by (3.23)

$$(3.24) \quad I(Q^*, Q_0) = \sum_{j=1}^{c} \rho_j I(F_j{}^*, F_{j,0}) < \sum_{j=1}^{c} \rho_j [I(V_j, F_{j,0}) + \eta/2] = I(U, Q_0) + \eta/2.$$

Now V_j is an $F_{j,0}$-strip with border step functions $\bar{H}_j$ and $\underline{H}_j$. It is clear from the construction of $\bar{H}_j$ and $\underline{H}_j$ that

$$\Delta_k F_j^* \geqq \bar{H}_j(a_{j,k}) - \underline{H}_j(a_{j,k-1}) - 2/M. \tag{3.25}$$

Since $\Delta_k F_j^* = M^{-1} m_{j,k}$, (3.25) becomes

$$\bar{H}_j(a_{j,k}) - \underline{H}_j(a_{j,k-1}) \leqq M^{-1}(m_{j,k} + 2); \tag{3.26}$$

hence

$$\begin{aligned} \sup_x [\bar{H}_j(x) - \underline{H}_j(x)] \\ = \max_{1 \leqq k \leqq M} [\bar{H}_j(a_{j,k}) - \underline{H}_j(a_{j,k-1})] \leqq M^{-1}(\hat{m}_j + 2). \end{aligned} \tag{3.27}$$

We now decompose $\mathcal{S}$ into the two disjoint sets:

$$\begin{aligned} \mathcal{S}_1 &= \{U \varepsilon \mathcal{S} : \hat{m}_j \leqq M\alpha_M, \text{ for } j = 1, \cdots, c\}; \\ \mathcal{S}_2 &= \mathcal{S} - \mathcal{S}_1. \end{aligned} \tag{3.28}$$

For $U \varepsilon \mathcal{S}_1$, by (3.27) and (3.20), we have, for $j = 1, \cdots, c$,

$$\begin{aligned} \sup_x [\bar{H}_j(x) - \underline{H}_j(x)] \leqq M^{-1}(\hat{m}_j + 2) \leqq M^{-1}(M\alpha_M + 2) \\ = \alpha_M + 2/M < \epsilon; \end{aligned} \tag{3.29}$$

hence for Q and R in U, $\|Q - R\|_c < \epsilon$, which implies by (3.19) that

$$T(R) > T(Q) - \delta/2. \tag{3.30}$$

From (3.22), we see that one may choose $\tilde{Q} \varepsilon U$ such that

$$T(\tilde{Q}) > r - \delta/2. \tag{3.31}$$

Since Q^* is also in U, (3.30) and (3.31) yield $T(Q^*) > T(\tilde{Q}) - \delta/2 > r - \delta$; i.e.

$$Q^* \varepsilon \Omega_{r-\delta}. \tag{3.32}$$

Combining (3.24), (3.32), and (3.18), yields

$$\begin{aligned} I(U, Q_0) > I(Q^*, Q_0) - \eta/2 \geqq I(\Omega_{r-\delta}, Q_0) - \eta/2 \\ = I(r - \delta) - \eta/2 > I(r) - \eta = I(\Omega_r, Q_0) - \eta; \end{aligned}$$

hence (3.21) is satisfied for those $U \varepsilon \mathcal{S}_1$. For $U \varepsilon \mathcal{S}_2$, we still have from (3.24) that

$$I(U, Q_0) > I(Q^*, Q_0) - \eta/2 = \textstyle\sum_{j=1}^{c} \rho_j I(F_j^*, F_{j,0}) - \eta/2. \tag{3.33}$$

Since $U \varepsilon \mathcal{S}_2$, there is an l, $1 \leqq l \leqq c$, such that $\hat{m}_l > M\alpha_m$; hence by Lemma 3.6,

$$I(F_l^*, F_{l,0}) \geqq b_M. \tag{3.34}$$

Combining (3.33), (3.34), and (3.20) yields

$$\begin{aligned} I(U, Q_0) > \rho_l I(F_l^*, F_{l,0}) - \eta/2 \geqq \rho_l b_M - \eta/2 \\ > (\min_{i \leqq j \leqq c} \rho_j)^{-1} \rho_l I(r) - \eta/2 > I(r) - \eta = I(\Omega_r, Q_0) - \eta; \end{aligned}$$

hence (3.21) is satisfied for $U \varepsilon \mathcal{S}_2$; and therefore for $U \varepsilon \mathcal{C}$.

We have now proved that Ω_r is a constant Q_0-regular sequence; hence by Lemma 3.5,

$$\lim_{N\to\infty} N^{-1} \ln P\{T(S_N) \geqq r\} = -I(r). \tag{3.35}$$

The above result is now used to prove Theorem 1. I is non-decreasing for $r > T(Q_0)$; hence it has at most countably many discontinuities. Therefore we may choose a sequence, $\{\delta_k > 0\}_{k=1}^{\infty}$, such that $r \pm \delta_k$ are continuity points of I, and $\lim_{k\to\infty} \delta_k = 0$. For each k, choose N so large that $|\xi_N| < \delta_k$; hence we have

$$\begin{aligned} N^{-1} \ln P\{T(S_N) \geqq r + \delta_k\} &\leqq N^{-1} \ln P\{T(S_N) \geqq r + \xi_N\} \\ &\leqq N^{-1} \ln P\{T(S_N) \geqq r - \delta_k\}. \end{aligned} \tag{3.36}$$

For k suitably large, we can apply (3.35) to (3.36) and get

$$\begin{aligned} -I(r + \delta_k) &\leqq \liminf_{N\to\infty} N^{-1} P\{T(S_N) \geqq r + \xi_N\} \\ &\leqq \limsup_{N\to\infty} N^{-1} P\{T(S_N) \geqq r + \xi_N\} \leqq -I(r - \delta_k). \end{aligned}$$

Letting $k \to \infty$, the continuity of I at r gives Theorem 1.

4. Conditions and proof for Theorem 2. The conditions for Theorem 2 are

(I) *$I^{(B)}(y) < \infty$, for $y \in [r_0, r_1)$, and $I^{(B)}$ is continuous at r, for suitably large B.*

(II) $I(y) = \lim_{B\to\infty} I^{(B)}(y) < \infty$, *for $y \in [r_0, r_1)$, and I is continuous at r.*

(III) *For each sequence $B_N \to \infty$, there exists a sequence $\delta_N \downarrow 0$ such that for each $y \in [r_0, r_1)$*

(i) $P\{T^{(B_N)}(S_N) \geqq y + \delta_N\} > 0$ *for suitably large N,*

(ii) $P\{T_N \geqq y\} > 0$ for suitably large N,

(iii) $\lim_{N\to\infty} N^{-1} \ln [1 - P\{|d_N^{(B_N)}| > \delta_N \mid T^{(B_N)}(S_N) \geqq y + \delta_N\}] = 0$,

(iv) $\lim_{N\to\infty} N^{-1} \ln [1 - P\{|d_N^{(B_N)}| > \delta_N \mid T_N \geqq y\}] = 0$.

As with Theorem 1, we first prove Theorem 2 for the case $\xi_N \equiv 0$. Choose $\delta > 0$ and $B^* > 0$ such that for all $B > B^*$ we have

(4.1)
(i) $T^{(B)}(Q_0) < r - \delta$.
(ii) $I^{(B)}(y)$ is non-decreasing for $y \in [r - \delta, r + \delta]$, and $I^{(B)}$ is cont
tinous at r.

Let

$$\Lambda = \{y \in [r - \delta, r + \delta]: I^{(B)} \text{ is continuous at } y \text{ for all } B > B^*, \text{ and } I \text{ is continuous at } y\}. \tag{4.2}$$

By (4.1), it is clear that $[r - \delta, r + \delta] - \Lambda$ is countable; hence Λ is dense in $[r - \delta, r + \delta]$. Now consider a fixed $y \in \Lambda$ and a fixed $\eta > 0$. For $B > B^*$, $T^{(B)}$ is uniformly continuous, and y is a continuity point of $I^{(B)}$; hence Theorem 1 as well as its proof are applicable. From the proof of Theorem 1, we get Q_0-product strips of size $M = M(B)$, $U^{(B)}, U_1^{(B)}, \cdots, U_K^{(B)}$, which have the following properties:

$$(4.3)\qquad \begin{aligned} &\text{(i)} && U^{(B)} \subset \Omega_y^{(B)}, \\ &\text{(ii)} && I(U^{(B)}, Q_0) < I(\Omega_y^{(B)}, Q_0) + \eta, \\ &\text{(iii)} && \Omega_y^{(B)} \subset \textstyle\bigcup_{i=1}^{K} U_i^{(B)}, \\ &\text{(iv)} && I(U_i^{(B)}, Q_0) > I(\Omega_y^{(B)}, Q_0) - \eta. \end{aligned}$$

It is clear that we can choose a sequence $\{B_N\}_{N=1}^{\infty}$, such that

$$(4.4)\qquad \lim_{N\to\infty} B_N = \infty \quad\text{and}\quad M_N = M(B_N) = o(N/\ln N)$$

as $N \to \infty$. From assumptions (I) and (II), (4.3), and (4.4) we see that $\Omega_y^{(B_N)}$ is a Q_0-regular sequence; hence by Lemma 3.5, we have, letting $B = B_N$,

$$(4.5)\qquad \begin{aligned} \lim_{N\to\infty} N^{-1} \ln P\{S_N \,\varepsilon\, \Omega_y^{(B)}\} &= \lim_{N\to\infty} N^{-1} \ln P\{T^{(B)}(S_N) \geqq y\} \\ &= -\lim_{N\to\infty} I(\Omega_y^{(B)}, Q_0) = -I(y), \end{aligned}$$

for each $y \,\varepsilon\, \Lambda$.

To complete the proof of Theorem 2 in the case $\xi_N \equiv 0$, we need to show that

$$(4.6)\qquad \lim_{N\to\infty} N^{-1} \ln P\{T_N \geqq r\} = \lim_{N\to\infty} N^{-1} \ln P\{T^{(B)}(S_N) \geqq r\}.$$

To do this, assumption (III) will play the dominant role. Let $\delta_N \downarrow 0$ be the sequence guaranteed by assumption (III). Then for suitably large N, we have,

$$(4.7)\qquad \begin{aligned} &P\{T^{(B)}(S_N) \geqq r + \delta_N\} \\ &\quad= P\{T^{(B)}(S_N) \geqq r + \delta_N,\, |d_N^{(B)}| \leqq \delta_N\} \\ &\qquad + P\{T^{(B)}(S_N) \geqq r + \delta_N,\, |d_N^{(B)}| > \delta_N\} \\ &\quad\leqq P\{T^{(B)}(S_N) \geqq r + d_N^{(B)}\} \\ &\qquad + P\{T^{(B)}(S_N) \geqq r + \delta_N\} P\{|d_N^{(B)}| > \delta_N \mid T^{(B)}(S_N) \geqq r + \delta_N\}; \end{aligned}$$

hence, by noting (2.6), we see that

$$(4.8)\qquad \begin{aligned} P\{T_N \geqq r\} \geqq{}& P\{T^{(B)}(S_N) \geqq r + \delta_N\} \\ &\cdot[1 - P\{|d_N^{(B)}| > \delta_N \mid T^{(B)}(S_N) \geqq r + \delta_N\}]. \end{aligned}$$

On the other hand, for suitably large N, we have

$$(4.9)\qquad \begin{aligned} P\{T_N \geqq r\} &= P\{T_N \geqq r,\, |d_N^{(B)}| > \delta_N\} \\ &\quad + P\{T^{(B)}(S_N) \geqq r + d_N^{(B)},\, |d_N^{(B)}| \leqq \delta_N\} \\ &\leqq P\{T_N \geqq r\} P\{|d_N^{(B)}| > \delta_N \mid T_N \geqq r\} \\ &\quad + P\{T^{(B)}(S_N) \geqq r - \delta_N\}; \end{aligned}$$

therefore

$$(4.10)\qquad P\{T_N \geqq r\} \leqq P\{T^{(B)}(S_N) \geqq r - \delta_N\}[1 - P\{|d_N^{(B)}| > \delta_N \mid T_N \geqq r\}]^{-1}.$$

An application of condition (III) to (4.8) and (4.10) yields

$$\begin{aligned}(4.11)\qquad &\liminf_{N\to\infty} N^{-1}\ln P\{T^{(B)}(S_N) \geqq r + \delta_N\}\\ &\leqq \liminf_{N\to\infty} N^{-1}\ln P\{T_N \geqq r\}\\ &\leqq \limsup_{N\to\infty} N^{-1}\ln P\{T_N \geqq r\}\\ &\leqq \limsup_{N\to\infty} N^{-1}\ln P\{T^{(B)}(S_N) \geqq r - \delta_N\}.\end{aligned}$$

Since the complement of Λ is countable, we can choose a sequence $\lambda_k \downarrow 0$ such that $r \pm \lambda_k \varepsilon \Lambda$. Then for k fixed, it is clear from (4.5) that

$$\begin{aligned}(4.12)\qquad -I(r+\lambda_k) &\leqq \liminf_{N\to\infty} N^{-1}\ln P\{T^{(B)}(S_N) \geqq r + \delta_N\}\\ &\leqq \limsup_{N\to\infty} N^{-1}\ln P\{T^{(B)}(S_N) \geqq r - \delta_N\} \leqq -I(r-\lambda_k).\end{aligned}$$

Combining (4.12) with (4.11), letting $k \to \infty$, and recalling that I is continuous at r, we get

$$\lim_{N\to\infty} N^{-1}\ln P\{T_N \geqq r\} = -I(r) = -\lim_{B\to\infty} I(\Omega_r^{(B)}, Q_0);$$

which completes the proof of Theorem 2 in the case $\xi_N \equiv 0$.

To remove the restriction $\xi_N \equiv 0$, one uses the same argument that was used for Theorem 1.

5. Application of Theorem 1 to the Wilcoxon statistic. Consider the 2-sample set-up, where we adopt the following special notation: $m = n_1$, $n = n_2$, $F_0 = F_{1,0}$, $G_0 = F_{2,0}$, $F_m = F_{n_1}$, and $G_n = F_{n_2}$. Let T be defined by $T(F, G) = \int F\,dG$; then $T(S_N) = \int F_m\,dG_n$ is just the Mann-Whitney form of the Wilcoxon statistic. The purposes of this section are to first show that Theorem 1 is applicable to this statistic, and then to derive a useful formula for the probability of a large deviation.

To show that Theorem 1 is applicable, we need only show that T is uniformly continuous. Integration by parts yields

$$\begin{aligned}\left|\int F_1\,dG_1 - \int F_2\,dG_2\right| &= \left|\int (F_1 - F_2)\,dG_1 + \int F_2\,dG_1 - \int F_2\,dG_2\right|\\ &\leqq \int |F_1 - F_2|\,dG_1 + \left|\int F_2\,d(G_1 - G_2)\right|\\ &= \int |F_1 - F_2|\,dG_1 + \left|\int (G_1(x-0) - G_2(x-0))\,dF_2\right|\\ &\leqq \int |F_1 - F_2|\,dG_1 + \int |G_1(x-0) - G_2(x-0)|\,dF_2,\end{aligned}$$

which proves the uniform continuity of T.

We now compute the index of a large deviation of the Wilcoxon statistic, which can be done explicitly if $F_0 = G_0$. Since the Wilcoxon statistic is distribution free, we can, without loss of generality, assume $F_0 = G_0 = U_0$, where U_0 is the cdf of the uniform distribution on [0, 1]. In this case, $T(Q_0) = \int U_0\,dU_0 = \frac{1}{2}$, and $\sup \int F\,dG = 1$; hence we consider $\frac{1}{2} < r < 1$. The problem is to minimize $\rho_1 I(F, U_0) + \rho_2 I(G, U_0)$ subject to the condition that $\int F\,dG \geqq r$. A necessary condition for $I(F, U_0)$ to be finite (see Section 2) is that F be absolutely continuous with respect to U_0; hence we may restrict attention to those F and G which are absolutely continuous cdf's on [0, 1]. If we let f and g denote the densities of F and G, then we seek the minimum of $\rho_1 \int_0^1 f \ln f\, dx + \rho_2 \int_0^1 g \ln g\, dx$ subject to $\int_0^1 Fg\, dx \geqq r$. For notational convenience, let $H(f, g) = \rho_1 \int_0^1 f \ln f\, dx + \rho_2 \int_0^1 g \ln g\, dx$. Now for $\lambda > 0$, $\lambda[\int_0^1 Fg\, dx - r] \geqq 0$; hence

$$(5.1) \quad -H(f, g) \leqq \int_0^1 \rho_1 f \ln (1/f)\, dx + \int_0^1 \rho_2 g \ln (1/g)\, dx + \lambda[\int_0^1 Fg\, dx - r],$$

where $y \ln y = 0 = y \ln (1/y)$ when $y = 0$. Using integration by parts and Jensen's inequality, we get from (5.1) that

$$(5.2) \quad \begin{aligned} -H(f, g) \leqq\ & \rho_1[\int_0^1 f \ln (1/f)\, dx - \int_0^1 (\lambda/\rho_1) Gf\, dx] \\ & + \lambda(1 - r) + \rho_2 \int_0^1 g \ln (1/g)\, dx \\ =\ & \rho_1 \int_0^1 \ln [(1/f) \exp \{-(\lambda/\rho_1) G\}] f\, dx + \lambda(1 - r) \\ & + \rho_2 \int_0^1 g \ln (1/g)\, dx \\ \leqq\ & \rho_1 \ln \int_0^1 \exp \{-(\lambda/\rho_1) G\}\, dx + \lambda(1 - r) + \rho_2 \int_0^1 g \ln (1/g)\, dx, \end{aligned}$$

with equality if and only if

$$(5.3) \quad \begin{aligned} &\text{(i)} \quad \int_0^1 Fg\, dx = r, \\ &\text{(ii)} \quad (1/f) \exp \{-(\lambda/\rho_1) G\} \equiv a \quad \text{constant}. \end{aligned}$$

Notice that the last term in (5.2) does not depend on f; hence for a fixed g, $H(f, g)$ is minimized by any solution, f, of (5.3).

Interchanging the role of F and G when manipulating (5.1) we obtain:

$$(5.4) \quad \begin{aligned} -H(f, g) \leqq\ & \rho_2[\int_0^1 g \ln (1/g)\, dx + \int_0^1 (\lambda/\rho_2) Fg\, dx] - \lambda r \\ & + \int_0^1 \rho_1 f \ln (1/f)\, dx \\ =\ & \rho_2[\int_0^1 \ln [(1/g) \exp \{(\lambda/\rho_2) F\}] g\, dx - \lambda r + \int_0^1 \rho_1 f \ln (1/f)\, dx \\ \leqq\ & \rho_2 \ln \int_0^1 \exp (\lambda/\rho_2) F\, dx - \lambda r + \int_0^1 \rho_1 f \ln (1/f)\, dx, \end{aligned}$$

with equality if and only if

$$(5.5) \quad \begin{aligned} &\text{(i)} \quad \int_0^1 Fg\, dx = r, \\ &\text{(ii)} \quad (1/g) \exp \{(\lambda/\rho_2) F\} \equiv a \quad \text{constant}. \end{aligned}$$

Notice again that the last term in (5.4) does not depend on g; hence for a fixed f, $H(f, g)$ is minimized by any solution g of (5.5).

We proceed now to show that (5.3) combined with (5.5) has a unique solution; and hence the minimum of $H(f, g)$ is attained by this unique solution. The details of the following are routine but lengthy, and will therefore be left out. They can be found in [9]. Letting $r = \frac{1}{2} + \epsilon$, we seek a solution of

$$\text{(i)} \quad \int_0^1 F(x) G'(x)\, dx - \tfrac{1}{2} = \epsilon,$$

$$(5.6)\qquad \text{(ii)}\quad G'(x) = K_1 \exp\{(\lambda/\rho_2)F(x)\},$$
$$\text{(iii)}\quad F'(x) = K_2 \exp\{-(\lambda/\rho_1)G(x)\}.$$

(5.6)-(ii) implies that $G(x) = \int_a^x K_1 \exp\{(\lambda/\rho_2)F(t)\}\,dt$; and since $F(t)$ is continuous, we get from the fundamental theorem of calculus that G is differentiable on $(0, 1)$. Likewise F is differentiable; hence G is twice differentiable, which implies that F is twice differentiable. In fact, both F and G have derivatives of all orders on $(0, 1)$.

Applying elementary techniques to (5.6), one finds that the solution is

$$(5.7)\quad F_\lambda(x) = -(\rho_2/\lambda)\ln[\rho_1\mu\exp\{-(\lambda/\rho_1\rho_2)x\} + (1 - \rho_2\mu)],$$
$$G_\lambda(x) = (1/\rho_2)(x + (\rho_1\rho_2/\lambda)\ln[\rho_1\mu\exp\{-(\lambda/\rho_1\rho_2)x\} + (1 - \rho_1\mu)]),$$

where $\lambda > 0$ is a solution to $\int_0^1 F_\lambda G_\lambda'\,dx - \frac{1}{2} = \epsilon$, and

$$(5.8)\qquad \mu = (1/\rho_1)[(\exp\{-(\lambda/\rho_2)\} - 1)(\exp\{-(\lambda/\rho_1\rho_2)\} - 1)^{-1}].$$

By routine methods, one can show that $W(\lambda) = \int_0^1 F_\lambda G_\lambda'\,dx - \frac{1}{2}$ is a strictly increasing function on $(0, \infty)$, with $\lim_{\lambda\to 0} W(\lambda) = 0$ and $\lim_{\lambda\to\infty} W(\lambda) = \frac{1}{2}$; hence $W(\lambda) = \epsilon$ has a unique solution, $\lambda(\epsilon) > 0$, for $0 < \epsilon < \frac{1}{2}$, and (5.6) has the unique solution $(F_{\lambda(\epsilon)}, G_{\lambda(\epsilon)})$, which minimizes $\rho_1\int_0^1 F'\ln F'\,dx + \rho_2\int_0^1 G'\ln G'\,dx$ subject to $\int_0^1 FG'\,dx - \frac{1}{2} \geqq \epsilon$.

The computation of the minimum is also routine, and the result is stated below.

THEOREM 5.1. *If U_0 is the $R(0, 1)$ cdf, then for $0 < \epsilon < \frac{1}{2}$,*

$$\inf_{\int FdG - \frac{1}{2} \geqq \epsilon}[\rho_1 I(F, U_0) + \rho_2 I(G, U_0)] = \rho_1\ln(1/\rho_1) + \rho_2\ln(1/\rho_2)$$
$$+ \ln[(1 - \exp\{-\lambda/\rho_2\})^{\rho_1}(1 - \exp\{-\lambda/\rho_1\})^{\rho_2}(1 - \exp\{-\lambda/\rho_1\rho_2\})^{-1}]$$
$$+ \lambda[2\epsilon - 1],$$

where $\lambda = \lambda(\epsilon) > 0$ is the unique solution of

$$W(\lambda) = -(1/\rho_1\rho_2)\int_0^1 [x\exp\{-(\lambda/\rho_1\rho_2)x\}$$
$$\cdot(\exp\{-(\lambda/\rho_1\rho_2)x\} + [(1/\rho_1\mu) - 1])^{-1}]\,dx + (1/2\rho_2) = \epsilon,$$

and

$$\mu(\lambda) = (1/\rho_1)[(1 - \exp\{-\lambda/\rho_2\})(1 - \exp\{-\lambda/\rho_1\rho_2\})^{-1}].$$

Furthermore, the minimum is attained by

$$F_\lambda(x) = -(\rho_2/\lambda)\ln[\rho_1\mu\exp\{-(\lambda/\rho_1\rho_2)x\} + (1 - \rho_1\mu)],$$
$$G_\lambda(x) = (1/\rho_2)[x + (\rho_1\rho_2/\lambda)\ln(\rho_1\mu\exp\{-(\lambda/\rho_1\rho_2)x\} + (1 - \rho_1\mu))].$$

6. Interpretation of results. At this point, it is useful to give an interpretation of the result in the Wilcoxon case, which one can extrapolate to the general case. To do this, we appeal to [14], Theorem 13, page 242. This theorem says that if F_N is the empirical cdf of a sample drawn from a population with cdf F_0, and F is a cdf which is absolutely continuous with respect to F_0, then

$$\lim_{\delta\to 0} \lim_{N\to\infty} N^{-1} \ln P\{\|F_N - F\| \leqq \delta\} = -I(F, F_0). \tag{6.1}$$

Hence $I(F, F_0)$ is an asymptotic measure of the unlikelihood of the sample having a cdf near F when the true cdf is F_0. If $I(\hat{F}, F_0) = I(\Omega, F_0)$, where Ω is some F_0-regular class of cdf's, then, in the above sense, $\hat{F}$ is the most likely cdf for the empirical cdf to be near, given that $F_N \varepsilon \Omega$. Furthermore, according to Lemma 3.5 and (6.1), we have $\lim_{\delta\to 0} \lim_{N\to\infty} N^{-1} \ln P\{\|F_N - \hat{F}\| \leqq \delta\} = \lim_{N\to\infty} N^{-1} \ln P\{F_N \varepsilon \Omega\}$, which says intuitively that given $F_N \varepsilon \Omega$, it is very likely to be near $\hat{F}$.

Applying these ideas to the Wilcoxon example, we see that given $\int F_m\, dG_n - \frac{1}{2} \geqq \epsilon$, the empircal cfd's are very likely to be near $F_{\lambda(\epsilon)}$ and $G_{\lambda(\epsilon)}$ respectively, in the sense that $\lim_{\delta\to 0} \lim_{N\to\infty} N^{-1} \ln P\{\|F_m - F_{\lambda(\epsilon)}\| \leqq \delta \text{ and } \|G_n - G_{\lambda(\epsilon)}\| \leqq \delta\} = \lim_{N\to\infty} N^{-1} \ln P\{\int F_m\, dG_n - \frac{1}{2} \geqq \epsilon\}$. It is interesting to note that by applying l'Hospital's rule to (5.7), we get that for $0 \leqq x \leqq 1$,

$$\lim_{\epsilon\to 0} F_{\lambda(\epsilon)}(x) = \lim_{\epsilon\to 0} G_{\lambda(\epsilon)}(x) = x; \tag{6.2}$$

and a routine computation shows that

$$\begin{aligned} \lim_{\epsilon\to\frac{1}{2}} F_{\lambda(\epsilon)}(x) &= x/\rho_1, & 0 \leqq x \leqq \rho_1, \\ &= 1, & \rho_1 \leqq x \leqq 1; \\ \lim_{\epsilon\to\frac{1}{2}} G_{\lambda(\epsilon)}(x) &= 0, & 0 \leqq x \leqq \rho_1, \\ &= (1/\rho_2)(x - \rho_1), & \rho_1 \leqq x \leqq 1. \end{aligned} \tag{6.3}$$

So given $\int F_m\, dG_n - \frac{1}{2} \geqq 0$, the empirical cdf's are very likely to be near U_0, which is what one would expect, since $\{\int F_m\, dG_n - \frac{1}{2} \geqq 0\}$ is not even a large deviation. However, given the largest deviation possible, namely $\int F_m\, dG_n - \frac{1}{2} \geqq \frac{1}{2}$, the empirical cdf's are very likely to be near uniform on $(0, \rho_1)$ for F_m, and uniform on $(\rho_1, 1)$ for G_n, which again is what one would expect.

7. Some other applications of Theorems 1 and 2. For the one sample case, let T be defined by $T(F) = \sup_x |F(x) - F_0(x)|$. Then $T(F_N)$ is the Kolmogorov-Smirnov two-sided test of fit statistic. Since T is uniformly continuous, Theorem 1 applies. In [9], $I(r) = \inf \{I(F, F_0) : \sup_x |F(x) - F_0(x)| \geqq r\}$ is computed for $0 < r < 1$ and shown to be continuous; hence an expression for

$$\lim_{N\to\infty} N^{-1} P\{\sup_x |F_N(x) - F_0(x)| \geqq r\}$$

is obtained, which agrees with the one derived by Sethuraman [15] using a different method.

If T is defined by $T(F) = \sup_x [F_0(x) - F(x)]$, then $T(F_N)$ is the Kolmogorov-Smirnov one-sided test of fit; and in [9], an expression is obtained for $\lim_{N\to\infty} N^{-1} P\{\sup_x [F_0(x) - F_N(x)] \geqq r\}$. In order to compute the Chernoff efficiency (see [4]) of $D_N^- = \sup_x [F_0(x) - F_N(x)]$, it is necessary to consider the probability of a large deviation of D_N^- when the sample is drawn from a population with a continuous cdf $G \neq F_0$. If for example,

$$G(x) \leqq F_0(x) \quad \text{and} \quad \sup_x [F_0(x) - G(x)] = d > 0,$$

then the large deviation of interest is $\{D_N^- \leqq r\}$ with $0 < r < d$. Its index is computed in [9].

In the 2-sample case, Theorem 2 can be applied to the difference of the sample means if some fairly restrictive conditions are imposed. Using the special notation of Section 5, we assume

(7.1)
(i) F_0 and G_0 are absolutely continuous with densities f_0 and g_0;
(ii) $\int \exp\{tx\}\, dF_0(x) < \infty$;
(iii) $\int \exp\{tx\}\, dG_0(x) < \infty$ for all t,
$\int x\, dF_0(x) = \int x\, dG_0(x) = 0$.

For each positive integer B, define

$$\gamma_B(x) = x \quad \text{for } |x| \leqq B$$
$$(7.2) \qquad = B \quad \text{for } x > B$$
$$= -B \quad \text{for } x < -B$$

and

$$(7.3) \qquad T^{(B)}(F, G) = \int \gamma_B\, d(G - F).$$

In [9], it is shown that Theorem 2 is applicable to $T_N = \int x\, d(G_n - F_m) = \bar{Y} - \bar{X}$, and

$$\lim_{B\to\infty} I^{(B)}(r) = \lim_{B\to\infty} [\inf \{\rho_1 I(F, F_0) + \rho_2 I(G, G_0): \int \gamma_B\, d(G - F) \geqq r\}]$$

is computed for $r > r_0 = \lim_{B\to\infty} T^{(B)}(F_0, G_0) = 0$. Using a completely different method, Abrahamson [1] derived the result for $\bar{Y} - \bar{X}$ under different conditions.

8. Exact Bahadur efficiency. The purpose of this section is to show how the results of the paper can be used to compute exact Bahadur efficiency. As a first step, it is shown how the index of a large deviation of a statistic enters into the formula for the exact slope of that statistic. (For a description of Bahadur efficiency and a definition of exact slope, see [2].) Let $(\mathfrak{X}, \alpha, P_\theta)$ be a probability space, where $\theta \varepsilon \Theta$, the parameter set. Let Θ_0 be a subset of Θ, and consider the problem of testing the hypothesis $\theta \varepsilon \Theta_0$, based on N observations on $\mathfrak{X}$. Let $\{T_N\}_{N=1}^{\infty}$ be a sequence of statistics defined on $\mathfrak{X}$ for which the following conditions are satisfied:

I. There exists a continuous cdf, F, such that for every $\theta \varepsilon \Theta_0$,

$$\lim_{N\to\infty} P_\theta\{T_N < x\} = F(x), \qquad -\infty < x < \infty.$$

II. There exists a function b on $\Theta - \Theta_0$, with $0 < b(\theta) < b_1$, (b_1 could be $+\infty$), such that for every $\theta \varepsilon \Theta - \Theta_0$, $T_N/N^{\frac{1}{2}} \to_{P_\theta} b(\theta)$.

III. For every sequence, $\epsilon_N \to \epsilon$, with $0 < \epsilon < b_1$, and every $\theta \varepsilon \Theta_0$, $P_\theta\{T_N/N^{\frac{1}{2}} \geqq \epsilon_N\} = \exp\{-N[I(\epsilon) + o(1)]\}$ as $N \to \infty$; e.g. $I(\epsilon)$ might be the "$\lim_{B\to\infty} I(\Omega_\epsilon^{(B)}, Q_0)$" of Theorem 2).

When these conditions are satisfied, $\{T_N\}$ is a standard sequence in the strict sense, (see [2], p. 282), and the exact slope of $\{T_N\}$ is given by

$$(8.1) \qquad \begin{aligned} c(\theta) &= 0 && \text{for } \theta \varepsilon \Theta_0 \\ &= 2I(b(\theta)) && \text{for } \theta \varepsilon \Theta - \Theta_0 . \end{aligned}$$

This is easy to see, because the equation at the bottom of p. 281 [2] becomes

$$(8.2) \qquad 2N^{-1} \ln P_\theta\{T_N/N^{\frac{1}{2}} \geqq u_N/N^{\frac{1}{2}}\} = -2I(z^{\frac{1}{2}})[1 + o(1)]$$

as $N \to \infty$; hence $f(z) = 2I(z^{\frac{1}{2}})$, and (8.1) follows from [2], p. 282, (6*).

As an example, we compute the exact Bahadur efficiency of the Wilcoxon test relative to the t-test in the 2-sample location problem. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be independent samples drawn from normal populations with common unknown variance τ^2, $EX = \mu$, and $EY = \mu + \theta$. We consider the problem of testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. Let

$$\begin{aligned} T_N^{(1)} &= N^{\frac{1}{2}}[\textstyle\int F_m \, dG_n - \frac{1}{2}], \\ (8.3) \quad T_N^{(2)} &= (nm/N)^{\frac{1}{2}}(\bar{Y} - \bar{X}) \\ &\quad \cdot [(\textstyle\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2)/(N-2)]^{-\frac{1}{2}} \end{aligned}$$

denote the statistics to be compared. $\{T_N^{(1)}\}$ and $\{T_N^{(2)}\}$ are both standard sequences in the strict sense with

$$(8.4) \qquad \begin{aligned} b_1(\theta, \tau) &= \textstyle\int \Phi(x/\tau) \, d\Phi((x - \theta)/\tau) - \frac{1}{2}, \\ b_2(\theta, \tau) &= (\rho_1\rho_2)^{\frac{1}{2}}(\theta/\tau), \end{aligned}$$

where Φ is the standard normal cdf.

To compute the exact slopes of $\{T_N^{(1)}\}$ and $\{T_N^{(2)}\}$, one must first compute the $I_1(\epsilon)$ and $I_2(\epsilon)$ of condition III. By Theorem 1 and Section 5, we have, under H_0,

$$\begin{aligned} P\{T_N^{(1)}/N^{\frac{1}{2}} \geqq \epsilon_N\} &= P\{\textstyle\int F_m \, dG_n - \frac{1}{2} \geqq \epsilon_N\} \\ &= \exp\{-N(\inf_{\int F dG - \frac{1}{2} \geqq \epsilon} [\rho_1 I(F, U_0) + \rho_2 I(G, U_0)] + o(1))\}, \end{aligned}$$

as $N \to \infty$; hence

$$(8.5) \qquad I_1(\epsilon) = \inf_{\int F dG - \frac{1}{2} \geqq \epsilon} [\rho_1 I(F, U_0) + \rho_2 I(G, U_0)],$$

which is given in Theorem 5.1. Also under H_0, we have

$$P\{T_N^{(2)}/N^{\frac{1}{2}} \geqq \epsilon_N\} = P\{F_{1,N-2}/N \geqq \epsilon_N^2\},$$

where $F_{1,N-2}$ is a statistic which has a F-distribution with 1 and $N - 2$ degrees of freedom. To compute this, we appeal to Abrahamson [1], p. 120, (4.8.11), and obtain

$$P\{T_N^{(2)}/N^{\frac{1}{2}} \geqq \epsilon_N\} = \exp\{-N([\ln(1 + \epsilon^2)]/2 + o(1))\};$$

hence

$$(8.6) \qquad I_2(\epsilon) = [\ln(1 + \epsilon^2)]/2.$$

By use of (8.1), we can now compute the exact slopes, and hence the exact

Bahadur efficiency, which is just the ratio of the slopes. The result is

$$(8.7) \qquad E_{1,2}(\theta, \tau) = 2I_1(b_1(\theta, \tau))/2I_2(b_2(\theta, \tau)) = 2I_1(b_1(\theta, \tau))/\ln [1 + (\rho_1\rho_2\theta^2/\tau^2)],$$

where $b_1(\theta, \tau)$ is given in (8.4) and I_1 given in (8.5).

In [9], p. 95, it is deduced that for fixed τ,

$$(8.8) \qquad E_{1,2}(\theta, \tau) = 3/\pi + o(\theta) \quad \text{as} \quad \theta \to 0.$$

One recognizes $3/\pi$ as the Pitman efficiency. Formula (8.8) actually gives more than the Pitman efficiency, because it contains the slope of the efficiency curve at $\theta = 0$, namely zero; hence $3/\pi$ is a good local approximation to the efficiency curve.

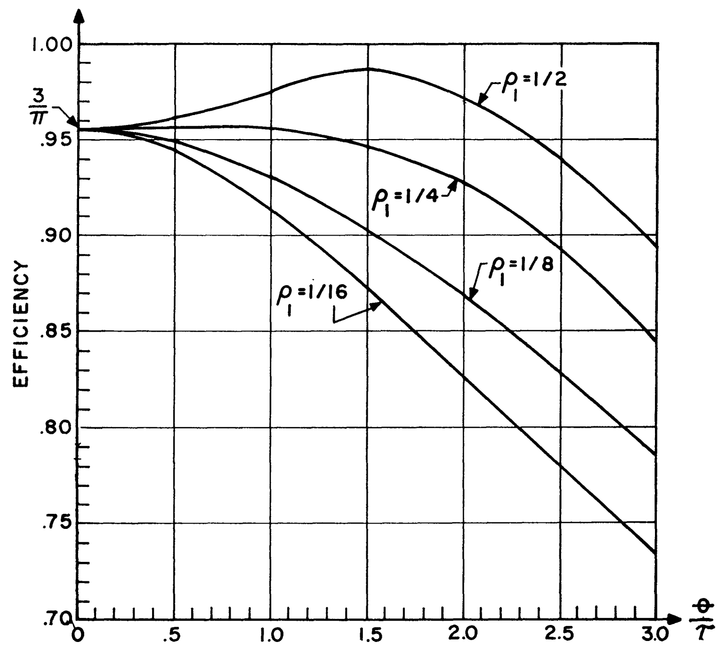

FIG. 8.1. Exact Bahadur efficiency.

TABLE I

Exact Bahadur efficiency of the Wilcoxon test relative to the t-test for the 2-sample location problem with normal samples

ρ_1	θ/τ							
	0	.1	.5	1.0	1.5	2.0	2.5	3.0
$\frac{1}{16}$	.9549	.9549	.9435	.9132	.8704	.8252	.7812	.7336
$\frac{1}{8}$	.9549	.9545	.9485	.9300	.9018	.8671	.8277	.7799
$\frac{1}{4}$	.9549	.9549	.9557	.9555	.9469	.9264	.8921	.8454
$\frac{1}{2}$	.9549	.9552	.9615	.9746	.9808	.9694	.9385	.8929

Note that $b_1(\theta, \tau)$ and $b_2(\theta, \tau)$, and hence $E_{1,2}(\theta, \tau)$ depend on θ and τ through θ/τ. Also, if ρ_1 and ρ_2 are interchanged, $E_{1,2}(\theta, \tau)$ remains the same.

A series of FORTRAN IV programs have been written by the author to carry out the computation of $E_{1,2}(\theta, \tau)$. The computations were performed at Bell Telephone Laboratories, Incorporated, Holmdel, New Jersey; and the results follow in Table I and Figure 8.1.

The referee pointed out to me that by comparing the above graph with Figure 1 of [12], one can see that the efficiency at θ/τ in the two-sample case with $\rho_1 = \frac{1}{2}$ is the same as the efficiency at μ in the one-sample case when $\mu = \theta/2\tau$. It would be interesting to have an analytical verification of this numerical agreement. The reader should note that the efficiencies computed by Klotz in [12] are just exact Bahadur efficiencies.

The main conclusion that can be reached on the basis of Figure 8.1 is that the performance of the Wilcoxon test becomes worse as the sample sizes become more unequal.

9. Acknowledgment. I wish to express my appreciation for the guidance and advice of Professors E. L. Lehmann and L. LeCam.

REFERENCES

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

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

[3] Bahadur, R. R. and Ranga Rao, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[4] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[5] Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain non-parametric test statistics. *Ann. Math. Statist.* **29** 972–994.

[6] Cramér, H. (1938). Sur un Noveau theoreme limite de la theorie des probabilites. *Actualités Sci, Ind.* 736.

[7] Feller, W. (1943). Generalization of a probability limit theorem of Cramér. *Trans. Amer. Math. Soc.* **54** 361–372.

[8] Hardy, G. H., Littlewood, J. E., and Polya, G. (1964). *Inequalities.* University Press, Cambridge.

[9] Hoadley, A. B. (1965). The theory of large deviations with statistical applications. University of California, Berkeley. Unpublished dissertation.

[10] Hodges, J. L., Jr. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t test. *Ann. Math. Statist.* **27** 324–335.

[11] Khinchin, A. I. (1957). *Mathematical Foundations of Information Theory.* Dover, New York.

[12] Klotz, J. (1965). Alternative efficiencies for signed rank tests. *Ann. Math. Statist.* **36** 1759–1766.

[13] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22** 79–86.

[14] Sanov, I. N. (1957). On the probability of large deviations of random variables (Russian) *Mat. Sbornik* N.S. **42(84)**, 11–44.

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

[16] Sethuraman, J. (1965). On the probability of large deviations of the mean for random variables in $D[0, 1]$. *Ann. Math. Statist.* **36** 280–285.