ASYMPTOTIC DISTRIBUTION OF THE LOG LIKELIHOOD RATIO BASED ON RANKS IN THE TWO SAMPLE PROBLEM[1]

I. R. SAVAGE[2]
and
J. SETHURAMAN
FLORIDA STATE UNIVERSITY

1. Introduction

Let $X_1, X_2, \cdots, Y_1, Y_2, \cdots$, be independent random variables where the X (Y) have common *strictly increasing* and *continuous* distribution function F_1^* (F_2^*). Let $N = 2n$ and $W_{N,1} \leqq W_{N,2} \leqq \cdots \leqq W_{N,N}$ be a rearrangement of $X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_n$ in increasing order of magnitude, $n = 1, 2, \cdots$. Define

$$(1.1) \qquad Z_{N,i} = \begin{cases} 0 & \text{if } W_{N,i} \text{ is an } X, \\ 1 & \text{if } W_{N,i} \text{ is a } Y, \end{cases}$$

$i = 1, \cdots, N$, and let $\mathbf{Z}_N = (Z_{N,1}, \cdots, Z_{N,N})$.

Let F_1 and F_2 be two arbitrary *strictly increasing continuous* distribution functions. Let

$$(1.2) \qquad L_N = L_N(\mathbf{z}_N) = \frac{P(\mathbf{Z}_N = \mathbf{z}_N | F_1^* = F_1, F_2^* = F_2)}{P(\mathbf{Z}_N = \mathbf{z}_N | F_1^* = F_2^* = F_1)}.$$

Note that the denominator in the above does not depend on F_1 and is equal to $1/\binom{N}{n}$ and that the numerator is unchanged if F_1 and F_2 are replaced by F_1K^{-1} and F_2K^{-1}, where K is a strictly increasing continuous distribution function. Let

$$(1.3) \qquad \ell_N = \ell_N(\mathbf{z}_N) = \log L_N(\mathbf{z}_N).$$

From now on $P(E)$ will stand for the probability of the event E when the common distribution of X and Y are F_1^* and F_2^*, respectively. Our main aim is to prove the asymptotic normality of $\ell_N(\mathbf{Z}_N)$ under suitable conditions (see Theorems 5.1 and 5.2). The conditions imposed are $A1$, $A2$, $A3$, and B or $\bar{A}1$, $\bar{A}2$, and $\bar{B}$ (see

[1] Research supported by the Army, Navy and Air Force under the Office of Naval Research Contract No. NONR 988(08), Task Order NR 042-004. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.
[2] Now visiting at the Center for Advanced Studies in the Behavioral Sciences, Stanford.

Sections 2 and 4). Conditions B and $\bar{B}$ are similar and in our opinion they form the crucial condition. In Section 6 we exhibit an example in which Condition $\bar{B}$ is not satisfied and the asymptotic distribution of ℓ_N is not normal.

In many statistical contexts it is of great importance to study the distribution of a likelihood ratio like L_N. It seems however that in nonparametric settings like ours other statistics like linear rank statistics have been studied more than the likelihood ratio. Laws of large numbers and bounds to the large deviation probabilities relating to linear rank statistics have been obtained earlier than those relating to ℓ_N. Problems of asymptotic normality of linear rank statistics have been well studied by now. We now describe some of the known results for ℓ_N.

Savage and Sethuraman [8] have shown, when $F_2 = F_1^A$ where A is a constant, that there is a number $I(F_1^*, F_2^*; F_1, F_1^A)$ such that for each $\varepsilon > 0$ there is a $\rho < 1$ with

$$P\left\{\left|\frac{1}{N}\ell_N(\mathbf{Z}_N) - I(F_1^*, F_2^*; F_1, F_1^A)\right| > \varepsilon\right\} < \rho^N, \tag{1.4}$$

for all large N. More generally, under some mild conditions, Berk and Savage [1] showed that there is a constant $I(F_1^*, F_2^*, F_1, F_2)$ such that for each $\varepsilon > 0$ there is a $\rho < 1$ with

$$P\left\{\left|\frac{1}{N}\ell_N(\mathbf{Z}_N) - I(F_1^*, F_2^*; F_1, F_2)\right| > \varepsilon\right\} < \rho^N, \tag{1.5}$$

for all large N. These results were used by these authors in their study of sequential tests based on ℓ_N. Some exact large deviation results for ℓ_N have recently been obtained by Hájek [5] when $F_1^* = F_2^*$ and F_1 and F_2 satisfy some mild conditions. The only results known about the asymptotic normality of ℓ_N were obtained simultaneously by Sethuraman [9] and Govindarajulu [4], for the special case when $F_2 = F_1^A$ where A is a constant. In this case $\sqrt{N}\,[\ell_N/N - I(F_1^*, F_2^*; F_1, F_1^A)]$ has a limiting normal distribution. This result was arrived at rather directly since in this special case $\ell_N(\mathbf{Z}_N)$ turned out to be a slightly modified Chernoff-Savage statistic.

In this paper we consider the asymptotic distribution of ℓ_N in the general case. We basically have two different proofs of the asymptotic distribution which are applicable to different situations. These two situations do not exhaust all possible cases and by the time we have imposed all of our conditions they do not cover the case treated in Sethuraman [9]. In the first situation it is assumed among other things that the density functions of F_1 and F_2 are bounded away from 0. In the second situation the density functions of F_1 and F_2 are histograms with a finite number of feet. There are some other conditions imposed on F_1^* and F_2^*. The proof of the asymptotic normality of ℓ_N uses some hints from game theory in the first situation and relies heavily on the results of Berk and Savage [1] in the second situation. A heuristic proof which applies to the first situation is sketched in Section 3 and we hope that these heuristics can be justified in situations more general than what we have been able to do.

Section 2 deals with the notation and various preliminaries. Section 4 contains some intermediate lemmas and theorems and the main results are in Section 5 in the form of Theorems 5.1 and 5.2.

It may be noted that we are trying to establish the asymptotic normality of ℓ_N in the two sample problem when the ratio of the first sample size to the combined sample size, λ_N, is identically equal to $\frac{1}{2}$. It is believed that our conclusions are valid at least when $\lambda_N = \lambda + o(1/\sqrt{N})$ where $0 < \lambda < 1$, but we do not treat these extensions here.

2. Definitions and notations

Let Δ be the class of all probability measures on $[0, 1]$. An element in Δ may also be viewed as a distribution function (left continuous or right continuous). Thus when we say that P belongs to Δ we may mean either that P is a probability measure on $[0, 1]$ or that $P = P(x)$ is a distribution function on $[0, 1]$. This ambiguity does not cause confusion and helps in reducing the amount of notation. We say that a sequence $\{P_n, n = 0, 1, \cdots\}$ in Δ converges to P_0 when P_n converges to P_0 weakly. Endowed with this topology, Δ becomes a compact metric space (for example, see Gnedenko and Kolmogorov [3], Chapter 2). If $\{\mu_n, n = 0, 1, \cdots\}$ is a sequence of probability measures on Δ (on the Borel σ-field generated by the open sets in Δ) we say that $\mu_n \to \mu_0$ if $\int_\Delta g\, d\mu_n \to \int_\Delta g\, d\mu_0$ for every bounded continuous function g on Δ. With this notation of convergence (which is the usual weak convergence for measures) the space of probability measures on Δ becomes another compact metric space (for instance, see Billingsley [2], Section 6).

When the probability density function (p.d.f.) of a distribution function $P(x)$ exists we denote it by $p(x)$, the corresponding lower case letter. Let Δ^* be the subset of Δ consisting of distribution functions $P(x)$, with continuous probability density functions $p(x)$ and with $p(x) \geqq \delta$ on $[0, 1]$ for some $\delta > 0$. Let Δ_N be the subset of Δ which consists of distributions that give masses in multiples of $1/N$ to at most N distinct points of $[0, 1]$, $N = 1, 2, \cdots$.

Now let $0 = a_0 < a_1 < \cdots < a_R = 1$ be fixed and generate a partition of $[0, 1]$. Relative to this partition we define $\bar{\Delta}$ to be the subset of Δ consisting of distribution functions $P(x)$ with p.d.f. $p(x)$ satisfying

$$p(x) = p_r \geqq 0, \qquad x \in [a_{r-1}, a_r), \qquad r = 1, \cdots, R. \tag{2.1}$$

Let $\bar{\Delta}^*$ be the subset of $\bar{\Delta}$ consisting of distribution functions $P(x)$ with p.d.f. $p(x)$ satisfying

$$p(x) = p_r > 0, \qquad x \in [a_{r-1}, a_r), \qquad r = 1, \cdots, R. \tag{2.2}$$

Finally let $\bar{\Delta}_N$ be the subset of $\bar{\Delta}$ consisting of distribution functions $P(x)$ with p.d.f. $p(x)$ satisfying

$$p(x) = p_r, \qquad p_r(a_r - a_{r-1}) = \text{a multiple of } \frac{1}{N}, \qquad x \in [a_{r-1}, a_r), \qquad r = 1, \cdots, R. \tag{2.3}$$

Notice that when we talk of $\bar{\Delta}$ or $\bar{\Delta}^*$ or $\bar{\Delta}_N$, we always have the partition generated by $\{a_0, a_1, \cdots, a_{R_j}\}$ in mind and that this partition is arbitrary but fixed throughout this paper.

We have remarked earlier that ℓ_N and L_N are unchanged if we replace F_1 and F_2 by F_1K^{-1} and F_2K^{-1}, where K is strictly increasing continuous distribution function on $[0, 1]$. For the same reason the distributions of ℓ_N and L_N remain unchanged if we replace F_1^* and F_2^* by $F_1^*K^{-1}$ and $F_2^*K^{-1}$. We will therefore normalize F_1, F_2, F_1^*, and F_2^* by appropriately choosing K as follows: put

$$\begin{aligned} H(x) &= \tfrac{1}{2}[F_1(x) + F_2(x)], \\ U_j(t) &= F_j(H^{-1}(t)), \\ H^*(x) &= \tfrac{1}{2}[F_1^*(x) + F_2^*(x)], \end{aligned} \tag{2.4}$$

and

$$U_j^*(t) = F_j^*(H^{*-1}(t)), \qquad 0 \leqq t \leqq 1, \quad j = 1, 2. \tag{2.5}$$

Note that

$$U_1(t) + U_2(t) = U_1^*(t) + U_2^*(t) = 2t, \qquad 0 \leqq t \leqq 1, \tag{2.6}$$

and that U_1, U_2, U_1^*, and U_2^* have probability density functions satisfying

$$u_1(t) + u_2(t) = u_1^*(t) + u_2^*(t) = 2, \qquad 0 \leqq t \leqq 1. \tag{2.7}$$

When $u_1(t)$ and $u_2(t)$ do not vanish on $[0, 1]$, define

$$M(t) = \log \frac{u_1(t)}{u_2(t)}. \tag{2.8}$$

We now state conditions concerning F_1, F_2, F_1^*, and F_2^*, in terms of U_1, U_2, U_1^*, and U_2^*.

CONDITION A.

A1. *$U_1^*(t)$ is strictly convex on* $[0, 1]$ *and $u_1^*(t)$ is continuous.*

A2. *U_1, U_2 belong to Δ^*.*

A3. *$M(t)$ is strictly increasing on* $[0, 1]$.

Condition A3 is equivalent to $u_1(t)$ being strictly increasing on $[0, 1]$ which is the same as $U_1(t)$ being strictly convex. Conditions A2 and A3 together imply that M is of bounded variation on $[0, 1]$. A general example where A2 and A3 are satisfied is as follows: $F_2 = \phi(F_1)$ where ϕ is a distribution function on $[0, 1]$ with ϕ' strictly decreasing and continuous with $\phi' \geqq \delta$ for some $\delta > 0$. As a further special case we can put $\phi(t) = (1 - \lambda)t + \lambda t^A$ for some $A < 1$ and $0 < \lambda < 1$.

Condition A1 is restrictive and is not satisfied if $U_1^*(t) = t$, $0 \leqq t \leqq 1$, which corresponds to the null hypothesis in the two sample problem. Condition A2 is similarly restrictive. It is not satisfied in the case $F_2 = F_1^A$ though we know from Sethuraman [9] that ℓ_N is asymptotic normal in this case.

Note that the roles X and Y can be reversed without changing the asymptotic distribution of ℓ_N. Thus if U_1, U_2, U_1^*, and U_2^* do not satisfy Condition A one could reverse the roles of X and Y and verify whether U_2, U_1, U_2^*, and U_1^* satisfy Condition A.

CONDITION $\bar{\mathrm{A}}$.

$\bar{\mathrm{A}}$1. *The p.d.f.* u_1^* *is continuous on* $[0, 1]$.

$\bar{\mathrm{A}}$2. *There is a partition* $0 = a_0 < a_1 < \cdots < a_R = 1$ *of* $[0, 1]$ *and* U_1, U_2 *belong to* $\bar{\Delta}^*$, *that is*

$$u_j(t) = u_{j,r} > 0, \qquad t \in [a_{r-1}, a_r), \quad r = 1, \cdots, R, \quad j = 1, 2. \tag{2.9}$$

Condition $\bar{\mathrm{A}}$2 is used by Berk and Savage [1] when obtaining bounds on large deviations probabilities of ℓ_N. They were successful in removing these restrictions later on in their paper. We have not been able to do this. However Condition $\bar{\mathrm{A}}$2 is a useful point to start an investigation of the properties of ℓ_N.

Let $F_{1,N}(x)$, $F_{2,N}(x)$, and $H_N(x)$ be the right continuous empirical distribution functions of $(X_1, \cdots, X_n)$, $(Y_1, \cdots, Y_n)$ and $(X_1, \cdots, X_n, Y_1, \cdots, Y_n)$, respectively. Clearly $H_N(x) = \frac{1}{2}[F_{1,N}(x) + F_{2,N}(x)]$. For $t \in [0, 1]$ define

$$H_N^{-1}(t) = \inf\{x : H_N(x) \geqq t\}. \tag{2.10}$$

Then H_N^{-1} is left continuous. Define

$$U_{j,N}(t) = F_{j,N}(H_N^{-1}(t)), \qquad 0 \leqq t \leqq 1, \quad j = 1, 2. \tag{2.11}$$

Note that $U_{1,N}$ and $U_{2,N}$ are left continuous and

$$U_{1,N}(t) = \frac{1}{n} \sum_{i < Nt+1} (1 - Z_{N,i}), \tag{2.12}$$

$$U_{2,N}(t) = \frac{1}{n} \sum_{i < Nt+1} Z_{N,i}, \tag{2.13}$$

and

$$U_{1,N}(t) + U_{2,N}(t) = 2t, \qquad t = 0, \frac{1}{N}, \cdots, \frac{N}{N}. \tag{2.14}$$

With probability one, $U_{1,N}$ and $U_{2,N}$ assume $\binom{N}{n}$ different values, each, in Δ.

Let $P \in \Delta$ and $Q \in \Delta^*$. Equation (3.1) motivates the following definitions. Let

$$i(P, Q; U_{1,N}) = \frac{1}{2} \sum_{j=1}^{2} \int_0^1 \log \frac{u_j(t)}{q(t)} \, dU_{j,N}(P(t)) \tag{2.15}$$

and

$$i(P, Q; U_1^*) = \frac{1}{2} \sum_{j=1}^{2} \int_0^1 \log \frac{u_j(t)}{q(t)} \, dU_j^*(P(t)). \tag{2.16}$$

Notice that these integrals always exist under Condition A or $\bar{\mathrm{A}}$ since $\log u_j/q$ is a bounded function, $j = 1, 2$. For any $P \in \Delta$, the element P_N of Δ_N is defined as

$$P_N(t) = \begin{cases} \dfrac{i}{N}, & P^{-1}\left(\dfrac{i}{N}\right) \leqq t < P^{-1}\left(\dfrac{i+1}{N}\right), i = 0, 1, \cdots, N-1 \\ 1, & P^{-1}(1) \leqq t \leqq 1. \end{cases} \tag{2.17}$$

Then

$$P_N^{-1}\left(\frac{i}{N}\right) = P^{-1}\left(\frac{i}{N}\right), \qquad i = 0, 1, \cdots, N, \tag{2.18}$$

and

$$|P_N(t) - P(t)| \leqq \frac{1}{N}, \qquad 0 \leqq t \leqq 1. \tag{2.19}$$

Thus

$$\begin{aligned} i(P, Q; U_{1,N}) = \\ \frac{1}{2n}\left\{\sum_{i=1}^{N} \log \frac{u_1\left(P^{-1}\left(\frac{1}{N}\right)\right)}{q\left(P^{-1}\left(\frac{i}{N}\right)\right)}(1 - Z_{N,i}) + \sum_{i=1}^{N} \log \frac{u_2\left(P^{-1}\left(\frac{i}{N}\right)\right)}{q\left(P^{-1}\left(\frac{i}{N}\right)\right)} Z_{N,i}\right\} \\ = i(P_N, Q; U_{1,N}). \end{aligned} \tag{2.20}$$

Let

$$W_N(t) = \sqrt{N}\,[U_{1,N}(t) - U_1^*(t)], \qquad 0 \leqq t \leqq 1 \tag{2.21}$$

and

$$W_N^1(t) = \begin{cases} W_N(t) & \text{for } \dfrac{1}{N} \leqq t \leqq 1, \\ 0 & \text{for } 0 \leqq t < \dfrac{1}{N}. \end{cases} \tag{2.22}$$

Let D^- be the space of left continuous function on $[0, 1]$ which are right continuous at 0. This space becomes a complete metric space under the Skorohod topology. If $u_1^*(t)$ is continuous (this is Condition $\bar{\text{A}}1$ and it is a part of Condition A1) it follows from Pyke and Shorack [7] (Theorem 4.1(a)) that the distributions of $\{W_N^1(t), 0 \leqq t \leqq 1\}$ in D^- converge weakly to the distribution of a Gaussian process $\{W(t), 0 \leqq t \leqq 1\}$ with mean function 0 and variance-covariance function

$$K(t, s) = \tfrac{1}{2}U_1^*(t)[1 - U_1^*(s)]u_2^*(t)u_2^*(s) + \tfrac{1}{2}U_2^*(t)[1 - U_2^*(s)]u_1^*(t)u_1^*(s). \tag{2.23}$$

Since u_1^* is continuous, it is easily seen from the form of $K(t, s)$ in (2.23) that $\{W(t), 0 \leqq t \leqq 1\}$ has continuous path functions with probability 1. We can therefore state Lemma 2.1 below without proof in which for any function h on

$[0, 1]$ and any $\delta > 0$,

$$\omega(h, \delta) = \sup_{|t-s| \leqq \delta, 0 \leqq t, s \leqq 1} |h(t) - h(s)|. \tag{2.24}$$

LEMMA 2.1. *Let u_1^* be continuous. Then*

(i) *the distributions of $\{W_N^1(t), 0 \leqq t \leqq 1\}$ converge weakly to the distribution of $\{W(t), 0 \leqq t \leqq 1\}$ and*

(ii) *for each $\varepsilon > 0$,*

$$\lim_N \sup P\{\omega(W_N^1, \delta) \geqq \varepsilon\} \to 0 \tag{2.25}$$

as $\delta \to 0$.

We would next like to investigate the asymptotic distribution of $i(P, Q; U_{1,N})$. Let $\log q$ be of bounded variation and let either $\bar{\text{A}}2$ or A2 and A3 hold. Then, from (2.20),

(2.26)

$$\begin{aligned}
&\sqrt{N}[i(P, Q; U_{1,N}) - i(P, Q; U_1^*)] \\
&= \sqrt{N}[i(P_N, Q; U_{1,N}) - i(P_N, Q; U_1^*)] + \sqrt{N}[i(P_N, Q; U_1^*) - i(P, Q; U_1^*)] \\
&= \frac{1}{2}\sqrt{N} \sum_{j=1}^{2} \int \log \frac{u_j}{q} d(U_{j,N}P_N - U_j^*P_N) + \frac{1}{2}\sqrt{N} \sum_{j=1}^{2} \int \log \frac{u_j}{q} d(U_j^*P_N - U_j^*P) \\
&= \frac{1}{2}\sqrt{N} \int \log \frac{u_1}{u_2} d(U_{1,N}P_N - U_1^*P_N) - \frac{1}{2}\sqrt{N} \sum_{j=1}^{2} \int (U_j^*P_N - U_j^*P)\, d \log \frac{u_j}{q}.
\end{aligned}$$

Using (2.19), this becomes

$$-\frac{1}{2}\int W_N(P_N)\, dM + O\left(\frac{1}{\sqrt{N}}\right) = -\frac{1}{2}\int W_N^1(P_N)\, dM + O_p\left(\frac{1}{\sqrt{N}}\right). \tag{2.27}$$

LEMMA 2.2. *Let u_1^* be continuous and* A2 *and* A3 *hold. Let $P^{(N)}$ be a random element in Δ, $N = 1, 2, \cdots$, such that the distribution of $P^{(N)}$ converges to the degenerate distribution at P^*, where $P^* \in \Delta^*$. Let $Q \in \Delta^*$ and $\log q$ be of bounded variation. Then*

$$\sqrt{N}[i(P^{(N)}, Q; U_{1,N}) - i(P^{(N)}, Q; U_1^*)] \tag{2.28}$$

has a limiting normal distribution with mean 0 *and variance $\sigma_{P^*}^2$ given by*

$$\sigma_{P^*}^2 = \frac{1}{4}\iint K(t, s)\, dMP^{*-1}(t)\, dMP^{*-1}(s). \tag{2.29}$$

(Note that the P^* above and in Lemma 2.3 can be any element of Δ^* and not necessarily the P^* defined in Condition B which follows later.)

PROOF. Using the Skorohod representation for sequences of random elements in $\Delta \times D^-$ (as Pyke and Shorack [7] do in D^-), we can assume that $(P^{(N)}, W_N) \to (P^*, W)$ in $\Delta \times D^-$ with probability 1. We can then imitate the proof of Theorem 4.1(b) of Pyke and Shorack [7] for our case as follows. From (2.26)

$$\sqrt{N}\,[i(P^{(N)}, Q; U_{1,N}) - i(P^{(N)}, Q; U_1^*)] + \frac{1}{2}\int W(t)\, dMP^{*-1}(t) \tag{2.30}$$

$$= -\frac{1}{2}\int [W_N^1(P_N^{(N)}) - W(P^*)]\, dM + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Using Lemma 2.1, this becomes

$$-\frac{1}{2}\int \big(W_N^1(P_N^{(N)}) - W(P_N^{(N)}) + W(P_N^{(N)}) - W(P^*)\big)\, dM + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{2.31}$$

$$= o_p(1).$$

This completes the proof of Lemma 2.2.

COROLLARY. *Under the conditions of Lemma* 2.2 *the limiting distribution of*

$$\sqrt{N}[i(P^*, Q; U_{1,N}) - i(P^*, Q; U_1^*)] \tag{2.32}$$

is normal with mean 0 *and variance* $\sigma^2_{P^*}$.

When $\bar{A}$ holds we can simplify the expressions for $i(P, Q; U_{1,N})$, and so on whenever $Q \in \bar{\Delta}^*$. Let $Q \in \bar{\Delta}^*$. Then

$$i(P, Q; U_{1,N}) = \frac{1}{2}\sum_{j=1}^{2}\int \log\frac{u_j(t)}{q(t)}\, dU_{j,N}(P(t)) \tag{2.33}$$

$$= \frac{1}{2}\sum_{j=1}^{2}\sum_{r=1}^{R} \log\frac{u_{j,r}}{q_r}\,[U_{j,N}(P(a_r)) - U_{j,N}(P(a_{r-1}))],$$

which depends on P only through its values $P(a_r)$, $r = 0, 1, \cdots, R$. Thus a general P in Δ may be replaced by the unique $\bar{P}$ in $\bar{\Delta}$ with $P(a_r) = \bar{P}(a_r)$, $r = 0, 1, \cdots, R$, in $i(P, Q; U_{1,N})$ without changing its value. Therefore it is natural to restrict P to belong to $\bar{\Delta}$ and Q to $\bar{\Delta}^*$ in dealing with $i(P, Q; U_{1,N})$, and so forth, when $(\bar{A})$ holds. Also note that with the convention $0 \log 0 = 0$, $i(P, P; U_{1,N})$ can be defined for all $P \in \bar{\Delta}$. The following lemma and its corollary are analogous to Lemma 2.2 and its corollary.

LEMMA 2.3. *Let* $\bar{A}1$ *and* $\bar{A}2$ *hold. Let* $P^{(N)}$ *be a random element of* $\bar{\Delta}$, $N = 1, 2, \cdots$, *such that the distributions of* $P^{(N)}$ *in* Δ *converge weakly to the degenerate distribution at* P^* *where* $P^* \in \bar{\Delta}^*$. *Then*

$$\sqrt{N}\,[i(P^{(N)}, P^{(N)}; U_{1,N}) - i(P^{(N)}, P^{(N)}; U_1^*)] \tag{2.34}$$

has a limiting normal distribution with mean 0 *and variance* $\sigma^2_{P^*}$ *given in* (2.29).

PROOF. Using the Skorohod representation as in Lemma 2.2 we may assume that $(P^{(N)}, W_N^1) \to (P^*, W)$ with probability 1. Since $P^{(N)} \in \bar{\Delta}$ and $P^* \in \bar{\Delta}^*$, it follows that the total variation of $\log p^{(N)}$ is bounded in probability for all large N and converges to the total variation of $\log p^*$. Thus using a slight extension of (2.27) and Lemma 2.1,

(2.35)

$$\sqrt{N}\,[i(P^{(N)}, P^{(N)}; U_{1,N}) - i(P^{(N)}, P^{(N)}; U_1^*)] + \frac{1}{2}\int W(t)\,dMP^{*-1}(t)$$

$$= -\frac{1}{2}\int [W_N^1(P_N^{(N)}) - W(P^*)]\,dM + O_p\left(\frac{1}{\sqrt{N}}\right)$$

$$= -\frac{1}{2}\int [W_N^1(P_N^{(N)}) - W(P_N^{(N)}) + W(P_N^{(N)}) - W(P^*)]\,dM + O_p\left(\frac{1}{\sqrt{N}}\right)$$

$$= o_p(1).$$

This completes the proof of Lemma 2.3.

COROLLARY. *Let the conditions of Lemma* 2.3 *hold. The limiting distribution of*

$$\sqrt{N}\,[i(P^*, P^*; U_{1,N}) - i(P^*, P^*; U_1^*)] \tag{2.36}$$

is normal with mean 0 *and variance* $\sigma^2_{P^*}$.

3. A heuristic proof of the asymptotic normality of ℓ_N

An expression for L_N may be written as below using a formula from Hoeffding [6]. Fix $Q \in \Delta^*$ and write

$$\begin{aligned}
(3.1)\quad L_N(\mathbf{Z}_N) &= N! \int_{0 \leqq w_1 \leqq \cdots \leqq w_N \leqq 1} \cdots \int \prod_{i=1}^N \{[u_1(w_i)]^{1-Z_{N,i}}[u_2(w_i)]^{Z_{N,i}}\,dw_1 \cdots dw_N \\
&= N! \int_{0 \leqq w_1 \leqq \cdots \leqq w_N \leqq 1} \cdots \int \exp\left\{\sum_{i=1}^N \left[(1 - Z_{N,i}) \log \frac{u_1(w_i)}{q(w_i)}\right.\right. \\
&\qquad\qquad \left.\left. + Z_{N,i} \log \frac{u_2(w_i)}{q(w_i)}\right]\right\} dQ(w_1) \cdots dQ(w_N) \\
&= N! \int_{0 \leqq w_1 \leqq \cdots \leqq w_N \leqq 1} \cdots \int \exp\{Ni(P_{\mathbf{w}}, Q; U_{1,N})\}\,dQ(w_1) \cdots dQ(w_N),
\end{aligned}$$

where $P_{\mathbf{w}}$ is the empirical distribution function of $w_1, \cdots, w_N$ and $i(P_{\mathbf{w}}, Q; U_{1,N})$ is as defined in (2.15). Note that $P_{\mathbf{w}}$ is in Δ_N. Expression (3.1) is the first place we have written down L_N explicitly and is fundamental to this investigation. A simple consequence of (3.1) is

$$L_N(\mathbf{Z}_N) \leqq \exp\{N \sup_{P \in \Delta_N} i(P, Q; U_{1,N})\} \tag{3.2}$$

since

$$N! \int_{0 \leqq w_1 \leqq \cdots \leqq w_N \leqq 1} \cdots \int dQ(w_1) \cdots dQ(w_N) = 1. \tag{3.3}$$

Since Q is arbitrary, (3.2) implies that

$$L_N(\mathbf{Z}_N) \leqq \exp\{N \inf_{Q \in \Delta^*} \sup_{P \in \Delta_N} i(P, Q; U_{1,N})\} \tag{3.4}$$

$$= \exp\{N \inf_{Q \in \Delta^*} \sup_{P \in \Delta^*} i(P, Q; U_{1,N})\}$$

in view of (2.20). Again, since the integrand in (3.1) is an exponential raised to the Nth power, we may for large N act as if

$$\ell_N = \log L_N \sim N \inf_{P \in \Delta^*} \sup_{Q \in \Delta^*} i(P, Q; U_{1,N}), \tag{3.5}$$

where $\sim$ means that the expressions on both sides of it are equal in some asymptotic sense. Let us look at the expression on the right side with $U_{1,N}$ replaced by its limit which is U_1^*. We get

$$N \inf_{Q \in \Delta^*} \sup_{P \in \Delta^*} i(P, Q; U_1^*) \tag{3.6}$$

which may be replaced by

$$N \sup_{P \in \Delta^*} \inf_{Q \in \Delta^*} i(P, Q; U_1^*), \tag{3.7}$$

if the appropriate result concerning the equality of a min max to a max min holds. This last expression can be shown (see Lemma 4.1) to be equal to

$$N \sup_{P \in \Delta^*} i(P, P; U_1^*). \tag{3.8}$$

Let us assume that there is a unique P^* in that Δ^* such that the above is equal to $Ni(P^*, P^*; U_1^*)$. This assumption will be made precise in Section 4 as Condition B. The simplification which occurs when we use the limit U_1^* leads us to write

$$\ell_N \sim Ni(P^*, P^*; U_{1,N}). \tag{3.9}$$

If the above were true, then from the corollary to Lemma 2.2 it follows that

$$\frac{1}{\sqrt{N}} [\ell_N - Ni(P^*, P^*; U_1^*)] \tag{3.10}$$

has a limiting normal distribution with mean 0 and variance $\sigma_{P^*}^2$.

4. Some intermediate results for i (P, Q; ·)

This section deals with some properties of $i(P, Q; U_1^*)$ and $i(P, Q: U_{1,N})$. These results are used in the proof of the main theorems of Section 5.

LEMMA 4.1. *Let Condition* A2 *or* $\bar{\text{A}}$2 *hold. Let* $P \in \Delta^*$. *Then*

$$i(P, Q; U_1^*) > i(P, P; U_1^*) \tag{4.1}$$

for all $Q \in \Delta^*$ *with* $Q \neq P$.

PROOF. Condition A2 or $\bar{\text{A}}2$ implies that $i(P, Q; U_1^*)$ is finite for $P \in \Delta^*$ and $Q \in \Delta^*$. For $Q \in \Delta^*$ with $Q \neq P$,

$$(4.2) \qquad i(P, Q; U_1^*) - i(P, P; U_1^*) = \frac{1}{2} \sum_{j=1}^{2} \int \left(\log \frac{u_j}{q} - \log \frac{u_j}{p}\right) dU_j^* P$$
$$= \int \log \frac{p}{q}\, dP > 0.$$

This proves (4.1).

LEMMA 4.2. *Let $Q \in \Delta^*$. Let A2 hold. Then $i(P, Q; U_1^*)$ is a continuous function of P with P varying in Δ.*

PROOF. Let $P^m \in \Delta$, $m = 1, 2, \cdots$, and $P^m \to P$. Then

$$(4.3) \qquad i(P^m, Q; U_1^*) - i(P, Q; U_1^*) = \frac{1}{2} \sum_{j=1}^{2} \int \log \frac{u_j}{q}\, d(U_j^* P^m - U_j^* P) \to 0,$$

since $\log u_j/q$ is bounded and continuous and $U_j^* P^m \to U_j^* P$ in Δ, $j = 1, 2$.

COROLLARY. *Under the conditions of Lemma* 4.1

$$(4.4) \qquad \sup_{P \in \Delta^*} i(P, Q; U_1^*) = \sup_{P \in \Delta} i(P, Q; U_1^*).$$

REMARK. Let (V_1, V_2) be a pair of distribution functions on $[0, 1]$. Define

$$(4.5) \qquad i(P, Q; V_1, V_2) = \frac{1}{2} \sum_{j=1}^{2} \int \log \frac{u_j}{q}\, dV_j(P),$$

whenever $P \in \Delta$ and $Q \in \Delta^*$. One can see in a manner similar to the proof of Lemma 4.2 that for fixed $Q \in \Delta^*$, $i(P, Q; V_1, V_2)$ is a continuous function of $(V_1(P), V_2(P))$.

LEMMA 4.3. *Let $Q \in \Delta^*$. Let* A1, A2, *and* A3 *hold. Then $i(P, Q; U_1^*)$ is a strictly concave function of P with P varying in Δ.*

PROOF. Let $P_1, P_2, \in \Delta$, $P_1 \neq P_2$, $0 < \alpha < 1$. Then

$$(4.6) \qquad i(\alpha P_1 + (1 - \alpha)P_2, Q; U_1^*) - \alpha i(P_1, Q; U_1^*) - (1 - \alpha) i(P_2, Q; U_1^*)$$
$$= \frac{1}{2} \sum_{j=1}^{2} \int \log \frac{u_j}{q}\, d[U_j^*(\alpha P_1 + (1 - \alpha)P_2) - \alpha U_j^*(P_1) - (1 - \alpha) U_j^*(P_2)]$$
$$= \frac{1}{2} \int \log \frac{u_1}{u_2}\, d[U_1^*(\alpha P_1 + (1 - \alpha)P_2) - U_1^*(P_1) - (1 - \alpha) U_1^*(P_2)]$$
$$= -\frac{1}{2} \int [U_1^*(\alpha P_1 + (1 - \alpha)P_2) - \alpha U_1^*(P_1) - (1 - \alpha) U_1^*(P_2)]\, dM$$
$$> 0,$$

since U_1^* is strictly convex and M is strictly increasing.

COROLLARY. *If $Q \in \Delta^*$, U_1^* is convex, $\log u_1/u_2$ is nondecreasing and bounded then $i(P, Q; U_1^*)$ is concave in P.*

We now state one form of the critical and hard to verify condition of this paper.

CONDITION B. *There exists a unique P^* in Δ^* such that*

$$\max_{P \in \Delta^*} i(P, P; U_1^*) = i(P^*, P^*; U_1^*) \tag{4.7}$$

and $\log p^*$ *is of bounded variation on* $[0, 1]$.

Notice that $i(P, P; U_1^*)$ may be written as $\int_0^1 \xi(t, P, p)\, dt$ with

$$\xi(t, P, p) = \frac{1}{2} \sum_{j=1}^{2} [\log u_j - \log p] u_j^*(P) p. \tag{4.8}$$

Thus one might use the Euler formula of the calculus of variations to verify B. The following lemma also can be useful in verifying Condition B.

LEMMA 4.4. *Let U_1^* be convex and* $\log u_1/u_2$ *be nondecreasing. Let* A2 *hold. Then $i(P, P; U_1^*)$ is a strictly concave function of P with P varying in Δ^*. Thus if there is a P_1 with*

$$i(P_1, P_1; U_1^*) = \max_{P \in \Delta^*} i(P, P; U_1^*), \tag{4.9}$$

then P_1 is unique.

PROOF. Let $P_1, P_2, \in \Delta^*, P_1 \neq P_2$ and $0 < \alpha < 1$. Then $\alpha P_1 + (1 - \alpha) P_2 \in \Delta^*$ and

$$\begin{aligned} &i(\alpha P_1 + (1 - \alpha) P_2, \alpha P_1 + (1 - \alpha) P_2; U_1^*) \\ &\quad - \alpha i(P_1, P_1; U_1^*) - (1 - \alpha) i(P_2, P_2; U_1^*) \\ &\quad = -\frac{1}{2} \int [U_1^*(\alpha P_1 + (1 - \alpha) P_2) - \alpha U_1^*(P_1) - (1 - \alpha) U^*(P_2)]\, dM \\ &\quad - \int [(\alpha p_1 + (1 - \alpha) p_2) \log (\alpha p_1 + (1 - \alpha) p_2) - \alpha p_1 \log p_1 \\ &\qquad\qquad - (1 - \alpha) p_2 \log p_2]\, dt > 0, \end{aligned} \tag{4.10}$$

The first term is $\geqq 0$ since U_1^* is convex and M is nondecreasing and the second term is > 0 since $x \log x$ is strictly convex.

LEMMA 4.5. *Let Condition* B *hold. Let U_1^* be convex, M be nondecreasing and Condition* A2 *hold. Then*

$$\max_{P \in \Delta \text{ or } \Delta^*} \min_{Q \in \Delta^*} i(P, Q; U_1^*) = \min_{Q \in \Delta^*} \max_{P \in \Delta \text{ or } \Delta^*} i(P, Q; U_1^*) \tag{4.11}$$

and

$$\max_{P \in \Delta \text{ or } \Delta^*} i(P, P^*; U_1^*) = i(P^*, P^*; U_1^*), \tag{4.12}$$

where the P^* is as specified in Condition B.

PROOF. We first establish (4.12). Only this conclusion of the lemma will be used later. The equality in (4.11) is aesthetically pleasing and is implied by (4.12) in the presence of (4.1) and (4.4) as we shall see later.

Let $P \in \Delta^*$. From Lemma 4.1

$$i(P, P; U_1^*) = \min_{Q \in \Delta^*} i(P, Q; U_1^*). \tag{4.13}$$

From Condition B

$$i(P^*, P^*; U_1^*) = \max_{P \in \Delta^*} i(P, P; U_1^*) = \max_{P \in \Delta^*} \min_{Q \in \Delta^*} i(P, Q; U_1^*). \tag{4.14}$$

Now, let $0 < \alpha < 1$, $P \in \Delta^*$, $P \neq P^*$. Then for $Q \in \Delta^*$, $i(P', Q; U_1^*)$ is a concave function of P' from the corollary to Lemma 4.3. Thus

$$\begin{aligned} i\big((1-\alpha)P^* + \alpha P, Q; U_1^*\big) &\geqq (1-\alpha)i(P^*, Q; U_1^*) + \alpha i(P; Q; U_1^*) \\ &\geqq (1-\alpha)i(P^*, P^*; U_1^*) + \alpha i(P, Q; U_1^*) \end{aligned} \tag{4.15}$$

from (4.1). Put $Q = Q_1$ where $Q_1 = (1-\alpha)P^* + \alpha P$. The above reduces to

$$i(Q_1, Q_1; U_1^*) \geqq (1-\alpha)i(P^*, P^*; U_1^*) + \alpha i(P, Q_1; U_1^*). \tag{4.16}$$

In view of Condition B we must have

$$i(P, Q_1; U_1^*) \leqq i(P^*, P^*; U_1^*) \tag{4.17}$$

for all $P \in \Delta^*$. Now let $\alpha \to 0$. We show later in this proof that

$$i(P, Q_1; U_1^*) \to i(P, P^*; U_1^*). \tag{4.18}$$

Thus

$$i(P, P^*; U_1^*) \leqq i(P^*, P^*; U_1^*) \tag{4.19}$$

for all $P \in \Delta^*$. The conclusion (4.12) of Lemma 4.5 follows from (4.19) and (4.4).

Again, from (4.12) and (4.14)

$$\begin{aligned} \min_{Q \in \Delta^*} \max_{P \in \Delta \text{ or } \Delta^*} i(P, Q; U_1^*) &\leqq \max_{P \in \Delta \text{ or } \Delta^*} i(P, P^*; U_1^*) \\ &= i(P^*, P^*; U_1^*) \\ &\leqq \max_{P \in \Delta \text{ or } \Delta^*} \min_{Q \in \Delta^*} i(P, Q; U_1^*). \end{aligned} \tag{4.20}$$

But a min max is always greater than or equal to a max min. Thus there is equality throughout (4.20) which now establishes (4.11).

To make this proof complete it therefore remains to establish (4.18) which can be done as follows:

$$\begin{aligned} i(P, Q_1; U_1^*) - i(P, P^*; U_1^*) &= \frac{1}{2} \sum_{j=1}^{2} \int \left(\log \frac{u_j}{q_1} - \log \frac{u_j}{p^*}\right) dU_j^* P \\ &\doteq \int \log \frac{p^*}{(1-\alpha)p^* + \alpha p}\, p\, dt, \end{aligned} \tag{4.21}$$

which tends to 0 as α tends to 0 by the dominated convergence theorem since the integrand in the last expression tends to 0 and is bounded in modulus for P and P^* belonging to Δ^*.

COROLLARY. *Let Condition* B *hold. Let* A1, A2, *and* A3 *hold. Then*

$$i(P, P^*; U_1^*) < i(P^*, P^*; U_1^*) \tag{4.22}$$

for all $P \neq P^*$, $P \in \Delta$. *Further if* P^{**} *is any random element in* Δ *with*

$$i(P^{**}, P^*; U_1^*) \geqq i(P^*, P^*; U_1^*) \tag{4.23}$$

with probability 1 *then* $P^{**} = P^*$ *with probability* 1.

This corollary follows readily from (4.12) in Lemma 4.5 and the strict concavity of $i(P, P^*; U_1^*)$ in P established in Lemma 4.3.

THEOREM 4.1. *Let Conditions* B, A1, A2, *and* A3 *hold. Then there exist random elements* $P^{(N)}$ *in* Δ_N, $N = 1, 2, \cdots$, *such that* (*i*)

$$i(P^{(N)}, P^*; U_{1,N}) \geqq \sup_{P \in \Delta_N} i(P, P^*; U_{1,N}) - \frac{1}{N} \tag{4.24}$$

and such that (*ii*) *the probability measures induced by* $P^{(N)}$ *converge to the degenerate probability measure at* P^*. *Here* P^* *is as given in Condition* B.

PROOF. With probability 1, $U_{1,N}$ is one of the $\binom{N}{n}$ distribution functions that give masses $1/n$ to n of the N points $1/N, 2/N, \cdots, N/N$. For each value of $U_{1,N}$ we can fix a $P^{(N)}$ in Δ_N such that

$$i(P^{(N)}, P^*; U_{1,N}) \geqq \sup_{P \in \Delta_N} i(P, P^*; U_{1,N}) - \frac{1}{N}. \tag{4.25}$$

Notice that this supremum is always finite since

$$\left|i(P, P^*; U_{1,N})\right| \leqq \frac{1}{2} \sum_{j=1}^{2} \sup_t \left|\log \frac{u_j(t)}{p^*(t)}\right|. \tag{4.26}$$

Thus we obtain random elements $P^{(N)}$ in Δ_N, $N = 1, 2, \cdots$, satisfying (4.24). Since Δ is compact and metric, the probability measures induced by $\{P^{(N)}\}$ have limit points. Let the probability measure induced by P^{**} be one such limit point. Since $\sup_t |U_{1,N}(t) - U_1^*(t)| \to 0$ with probability 1, the probability measure of $U_1^*(P^{**})$ is a limit point of the probability measures of $\{U_{1,N}(P^{(N)})\}$. (All these random elements are in Δ.) From the remark following Lemma 4.2, the distribution of $i(P^{**}, P^*; U_1^*)$ is a limit point of the distributions of $\{i(P^{(N)}, P^*; U_{1,N})\}$. But, from (2.20),

$$\begin{aligned} i(P^{(N)}, P^*; U_{1,N}) &\geqq \sup_{P \in \Delta N} i(P, P^*; U_{1,N}) - \frac{1}{N} \\ &= \sup_{P \in \Delta} i(P, P^*; U_{1,N}) - \frac{1}{N} \\ &\geqq i(P^*, P^*; U_{1,N}) - \frac{1}{N}. \end{aligned} \tag{4.27}$$

Again $i(P^*, P^*; U_{1,N})$ converges in probability to the constant $i(P^*, P^*; U_1^*)$. Hence

$$i(P^{**}, P^*; U_1^*) \geqq i(P^*, P^*; U_1^*) \tag{4.28}$$

with probability 1. From the corollary to Lemma 4.5, $P^{**} = P^*$ with probability 1. This establishes the second part of Theorem 4.1.

REMARK. Under the conditions of Theorem 4.1.

$$\sup_{P \in \Delta_N} i(P, P^*; U_{1,N}) \to i(P^*, P^*; U_1^*) \tag{4.29}$$

in probability.

COROLLARY. *Under the conditions of Theorem* 4.1 *and with the sequence* $\{P^{(N)}\}$ *as defined there*

$$\sqrt{N}\,[i(P^{(N)}, P^*; U_{1,N}) - i(P^{(N)}, P^*; U_1^*)] \tag{4.30}$$

has a limiting normal distribution with mean 0 *and variance* $\sigma_{P^*}^2$ *as given in* (2.29).

The above corollary follows from an application of Lemma 2.2.

We now state a condition analogous to the Condition B which will be used when Conditions $\bar{\text{A}}1$ and $\bar{\text{A}}2$ hold.

CONDITION $\bar{\text{B}}$. *There exists a unique* P^* *in* $\bar{\Delta}^*$ *such that*

$$i(P^*, P^*; U_1^*) = \max_{P \in \bar{\Delta}} i(P, P; U_1^*). \tag{4.31}$$

THEOREM 4.2. *Let Conditions* $\bar{\text{B}}$, $\bar{\text{A}}1$, *and* $\bar{\text{A}}2$ *hold. Then there exist random elements* $P^{(N)}$ *in* $\bar{\Delta}_N$, $N = 1, 2, \cdots$, *such that* (*i*)

$$i(P^{(N)}, P^{(N)}; U_{1,N}) = \max_{P \in \bar{\Delta}_N} i(P, P; U_{1,N}), \tag{4.32}$$

and such that (*ii*) *the probability measures induced by* $\{P^{(N)}\}$ *converge to the degenerate probability measure at* P^* *where* P^* *is as given in Condition* $\bar{\text{B}}$.

PROOF. Notice that $\bar{\Delta}_N$ contains only a finite number of elements. For each one of the $\binom{N}{n}$ possible values of $U_{1,N}$ we can fix a $P^{(N)}$ in $\bar{\Delta}_N$ to satisfy (4.32). These $P^{(N)}$ are random elements in $\bar{\Delta}_N$, $N = 1, 2, \cdots$. Since Δ is compact and metric, the probability measures of $\{P^{(N)}\}$ have limit points. Notice that these limit points can only be probability measures in $\bar{\Delta}$. Let the probability measure of P^{**} be one such limit point. In a fashion similar to the proof of Theorem 4.1, the distribution of $i(P^{**}, P^{**}; U_1^*)$ is a limit point of the distributions of $i(P^{(N)}, P^{(N)}; U_{1,N})$. From Berk and Savage [1] (page 1668, line 9)

$$\begin{aligned} \max_{P \in \bar{\Delta}_N} i(P, P; U_{1,N}) &\geqq \sup_{P \in \bar{\Delta}} i(P, P; U_{1,N}) + O\left(\frac{\log N}{N}\right) \\ &\geqq i(P^*, P^*; U_{1,N}) + O\left(\frac{\log N}{N}\right). \end{aligned} \tag{4.33}$$

Combining these facts,

$$i(P^{**}, P^{**}; U_1^*) \geqq i(P^*, P^*; U_1^*) \tag{4.34}$$

with probability 1. From Condition $\bar{\text{B}}$ it now follows that $P^{**} = P^*$ with probability 1. This establishes the second part of Theorem 4.2.

COROLLARY. *Under the conditions of Theorem* 4.2 *and with the sequence as defined there*

$$\sqrt{N}\,[i(P^{(N)}, P^{(N)}; U_{1,N}) - i(P^{(N)}, P^{(N)}; U_1^*)] \tag{4.35}$$

has a limiting normal distribution with mean 0 *and variance* $\sigma_{P^*}^2$ *as given in* (2.29).

This corollary follows immediately by an application of Lemma 2.3.

5. The proof of the asymptotic normality of ℓ_N

We present two theorems on the asymptotic normality of ℓ_N, one when B, A1, A2, and A3 hold and the other when $\bar{\text{B}}$, $\bar{\text{A}}1$, and $\bar{\text{A}}2$ hold. It may be noted once again that Sethuraman [9] has established the asymptotic normality of ℓ_N for a case not covered by the two theorems alluded to above. In Section 6 we have an example where $\bar{\text{A}}1$ and $\bar{\text{A}}2$ hold but $\bar{\text{B}}$ does not hold and the asymptotic distribution of ℓ_N is not normal.

The proof of the asymptotic normality when $\bar{\text{B}}$, $\bar{\text{A}}1$, and $\bar{\text{A}}2$ hold is short and is facilitated considerably by the results of Berk and Savage [1]. The proof when B, A1, A2, and A3 hold is more lengthy and uses the results of Section 4 heavily. We feel that it might be possible to generalize this method of proof to include the case when $\bar{\text{B}}$, $\bar{\text{A}}1$ and $\bar{\text{A}}2$ hold and also to include other cases of interest which are now excluded by the restrictive conditions B, A1, A2, and A3. We have not been able to do this so far.

THEOREM 5.1. *Let* B, A1, A2, *and* A3 *hold. Then*

$$\frac{1}{\sqrt{N}}\left[\ell_N - Ni(P^*, P^*; U_1^*)\right] \tag{5.1}$$

has a limiting normal distribution with mean 0 *and variance* $\sigma_{P^*}^2$ *as given in* (2.29). *Here* P^* *is as specified by Condition* B.

PROOF. Substituting $Q = P^*$ in (3.2) we obtain

$$\ell_N \leqq N \sup_{P \in \Delta_N} i(P, P^*; U_{1,N}). \tag{5.2}$$

Let $P^{(N)}$ be the random element in Δ_N given by Theorem 4.1. Using equalities (4.12) and then (4.24), we have

$$\begin{aligned}\frac{1}{\sqrt{N}}&\left[\ell_N - Ni(P^*, P^*; U_1^*)\right] \\ &\leqq \sqrt{N}\left[\sup_{P \in \Delta_N} i(P, P^*; U_{1,N}) - \sup_{P \in \Delta} i(P, P^*; U_1^*)\right] \\ &\leqq \sqrt{N}\left[i(P^{(N)}, P^*; U_{1,N}) - i(P^{(N)}, P^*; U_1^*)\right] + \frac{1}{\sqrt{N}}.\end{aligned} \tag{5.3}$$

The corollary to Theorem 4.1 states that the above expression has a limiting normal distribution with mean 0 and variance $\sigma_{P^*}^2$. Thus

$$\liminf_N P\{[\ell_N - Ni(P^*, P^*; U_1^*)]/\sqrt{N} \leqq x\} \geqq \Phi\left(\frac{x}{\sigma_{P^*}}\right), \tag{5.4}$$

where $\Phi(x)$ is the distribution function of the standard normal random variable.

For any function h on $[0, 1]$, define

$$\|h\| = \sup_{0 \leqq t \leqq 1} |h(t)|. \tag{5.5}$$

Substituting $Q = P^*$ in (3.1) and letting $S = \{0 \leqq w_1 \leqq \cdots \leqq w_N \leqq 1 : \|P_{\mathbf{w}} -$

$P^*\| \leqq \varepsilon/\sqrt{N}\}$ we have, for any $\varepsilon > 0$,

$$L_N \geqq N! \int \underset{S}{\cdots} \int \exp\{Ni(P_{\mathbf{w}}, P^*, U_{1,N})\}\, dP^*(w_1) \cdots dP^*(\omega_N) \tag{5.6}$$

$$\geqq K_N(\varepsilon) \exp\{N \inf_P [i(P, P^*; U_{1,N}) : P \in \Delta_N, \|P - P^*\| \leqq \varepsilon/\sqrt{N}]\},$$

where

$$K_N(\varepsilon) = \int \cdots \int dP^*(w_1) \cdots dP^*(w_N) \tag{5.7}$$

with the integral taken over the set $\{w_1, \cdots, w_N : \|P_{\mathbf{w}} - P^*\| \leqq \varepsilon/\sqrt{N}\}$. From the Kolmogorov-Smirnov theorem (Billingsley [2], Section 16)

$$K_N(\varepsilon) \to K(\varepsilon) > 0. \tag{5.8}$$

Consider

(5.9)

$$\sqrt{N}\Big|\inf_{P\in\Delta_N,\, \|P-P^*\| \leqq \varepsilon/\sqrt{N}} i(P, P^*; U_{1,N}) - i(P^*, P^*; U_{1,N})\Big|$$

$$\leqq \sqrt{N} \sup_{P\in\Delta,\, \|P-P^*\| \leqq \varepsilon/\sqrt{N}} |i(P, P^*; U_{1,N}) - i(P^*, P^*; U_{1,N})|$$

$$= \sqrt{N} \sup_{\|P-P^*\| \leqq \varepsilon/\sqrt{N}} |i(P, P^*; U_{1,N}) - i(P, P^*; U_1^*)$$

$$- i(P^*, P^*; U_{1,N}) + i(P^*, P^*; U_1^*) + i(P, P^*; U_1^*) - i(P^*, P^*; U_1^*)|$$

$$\leqq \sup_{\|P-P^*\| \leqq \varepsilon/\sqrt{N}} \tfrac{1}{2}\Big|\int [W_N^1(P_N) - W_N^1(P_N^*)]\, dM\Big|$$

$$+ \sup_{\|P-P^*\| \leqq \varepsilon/\sqrt{N}} \frac{\sqrt{N}}{2}\left|\int [U_1^*(P) - U_1^*(P^*)]\, dM\right| + O_p\left(\frac{1}{\sqrt{N}}\right)$$

Thus using (2.29) in which P_N and P_N^* are also defined, we have

$$\leqq \tfrac{1}{2}C\, \omega\left(W_N^1, \frac{\varepsilon}{\sqrt{N}} + \frac{2}{N}\right) + C\varepsilon + O_p\left(\frac{1}{\sqrt{N}}\right), \tag{5.10}$$

where $C = \text{Var}\,(M)$ and in which we have used the fact that $\|P_N - P_N^*\| \leqq \varepsilon/\sqrt{N} + 2/N$ and $\|u_1^*\| \leqq 2$. From Lemma 2.1 (ii), the first term in the above is $o_p(1)$. Thus from (5.6), (5.8), and (5.10), we have

(5.11)

$$\frac{1}{\sqrt{N}}[\ell_N - Ni(P^*, P^*; U_1^*)]$$

$$\geqq \sqrt{N}\,[i(P^*, P^*; U_{1,N}^*) - i(P^*, P^*; U_1^*)] + o_p(1) - C\varepsilon.$$

Using the corollary to Lemma 2.2 in the above and noting that $\varepsilon > 0$ is arbitrary, we have

$$\limsup_N P\{[\ell_N - Ni(P^*, P^*; U_1^*)]/\sqrt{N} \leqq x\} \leqq \Phi\left(\frac{x}{\sigma_{P^*}}\right). \tag{5.12}$$

Theorem 5.1 now follows from (5.4) and (5.12).

The next theorem applies to the case when u_1 and u_2 are histograms.

THEOREM 5.2. *Let* $\bar{\mathrm{B}}$, $\bar{\mathrm{A}}1$, and $\bar{\mathrm{A}}2$ *hold. Then* $[\ell_N - Ni(P^*, P^*; U_1^*)]/\sqrt{N}$ *has a limiting normal distribution with mean* 0 *and variance* $\sigma_{P^*}^2$ *as given in* (2.29). *Here* P^* *is as specified in Condition* $\bar{\mathrm{B}}$.

PROOF. The expression for L_N in (3.1) can be simplified in this case. Putting Q = the uniform distribution on $[0, 1]$ in (3.1),

$$\begin{aligned} L_N &= N! \int_{0 \leqq w_1 \leqq \cdots \leqq w_N \leqq 1} \cdots \int \prod_{i=1}^{N} \{[u_1(w_i)]^{1 - Z_{N,i}}[u_2(w_i)]^{Z_{N,i}}\} \, dw_1 \cdots dw_N \\ &= N! \sum_{P \in \Delta_N} \prod_{r=1}^{R} \left\{ \prod_{j=1}^{2} \bar{u}_{j,r}^{N[U_{j,N}(P(a_r)) - U_{j,N}(P(a_{r-1}))]} \right\} (N[P(a_r) - P(a_{r-1})])! \\ &= N! \sum_{P \in \bar{\Delta}_N} \exp\{Ni(P, P; U_{1,N})\} \prod_{r=1}^{R} \frac{[P(a_r) - P(a_{r-1})]^{N[P(a_r) - P(a_{r-1})]}}{(N[P(a_r) - P(a_{r-1})])!}, \end{aligned} \tag{5.13}$$

where $\bar{u}_{j,r} = u_{j,r}(a_r - a_{r-1})$, $j = 1, 2$, $r = 1, \cdots, R$. Using Stirling's formula and simplifying (5.13), Berk and Savage [1] (page 1667, line 17 and page 1668, line 9) have shown that

$$|\ell_N - N \sup_{P \in \Delta_N} i(P, P; U_{1,N})| = O(\log N) \tag{5.14}$$

and

$$|\sup_{P \in \bar{\Delta}_N} i(P, P; U_{1,N}) - \sup_{P \in \bar{\Delta}} i(P, P; U_{1,N})| = O\left(\frac{\log N}{N}\right). \tag{5.15}$$

Now, using (5.14) and (5.15),

$$\begin{aligned} &\sqrt{N}[i(P^*, P^*; U_{1,N}) - i(P^*, P^*; U_1^*)] \\ &\quad \leqq \sqrt{N}[\sup_{P \in \bar{\Delta}} i(P, P; U_{1,N}) - i(P^*, P^*; U_1^*)] \\ &\quad \leqq \frac{1}{\sqrt{N}}[\ell_N - Ni(P^*, P^*; U_1^*)] + O\left(\frac{\log N}{\sqrt{N}}\right) \\ &\quad = \sqrt{N}[\sup_{P \in \bar{\Delta}_N} i(P, P; U_{1,N}) - \sup_{P \in \bar{\Delta}} i(P, P; U_1^*)] + O\left(\frac{\log N}{\sqrt{N}}\right) \\ &\quad \leqq \sqrt{N}[i(P^{(N)}, P^{(N)}; U_{1,N}) - i(P^{(N)}, P^{(N)}; U_1^*)] + O\left(\frac{\log N}{\sqrt{N}}\right), \end{aligned} \tag{5.16}$$

where $P^{(N)}$ is a random element in $\bar{\Delta}_N$ chosen as in Theorem 4.2. The top expression in (5.16) has a limiting normal distribution with mean 0 and variance σ_{P*}^2 from the corollary to Lemma 2.3. From the corollary to Theorem 4.2 the last expression in (5.16) has the same limiting distribution. This establishes that $[\ell_N - Ni(P^*, P^*; U_1^*)]/\sqrt{N}$ has a limiting normal distribution with mean 0 and variance σ_{P*}^2.

6. An example where ℓ_N has a nonnormal asymptotic distribution

An example is given below where with the usual normalization ℓ_N has an asymptotic distribution which is not normal. In this example U_1 and U_2 have probability density functions which are histograms and U_1^* and U_2^* have continuous probability density functions. Thus Conditions $\bar{\text{A}}1$ and $\bar{\text{A}}2$ are automatically satisfied. We establish that $\bar{\text{B}}$ is not satisfied. Finally we evaluate the asymptotic distribution of ℓ_N.

Let $R = 2$, $a_0 = 0$, $a_1 = \frac{1}{2}$, $a_2 = 1$. Let $\alpha > 0$, $\beta > 0$ and $\alpha + \beta = 2$. Let $\lambda = \log(\alpha/\beta) > 0$, $\mu = \log(\frac{1}{4}\alpha\beta)^{1/2}$. Let

$$u_1(t) = \begin{cases} \alpha & \text{if } 0 \leqq t \leqq \frac{1}{2} \\ \beta & \text{if } \frac{1}{2} < t \leqq 1, \end{cases} \tag{6.1}$$

$$u_2(t) = 2 - u_1(t), \qquad 0 \leqq t \leqq 1. \tag{6.2}$$

For $P \in \bar{\Delta}$ set $P(\frac{1}{2}) = \pi$. Then

$$p(t) = \begin{cases} 2\pi & \text{if } 0 \leqq t \leqq \frac{1}{2} \\ 2(1 - \pi) & \text{if } \frac{1}{2} < t < 1. \end{cases} \tag{6.3}$$

Also, $0 \leqq \pi \leqq 1$ and π parametrizes the class $\bar{\Delta}$. Now, for any U_1^*, U_2^*,

$$\begin{aligned} i(\pi) &= i(P, P; U_1^*) \\ &= \frac{1}{2}\left\{U_1^*(\pi)\log\frac{\alpha}{2\pi} + [1 - U_1^*(\pi)]\log\frac{\beta}{2(1-\pi)}\right. \\ &\qquad \left. + U_2^*(\pi)\log\frac{\beta}{2\pi} + [1 - U_2^*(\pi)]\log\frac{\alpha}{2(1-\pi)}\right\} \\ &= [U_1^*(\pi) - \pi]\log\frac{\alpha}{\beta} \\ &\qquad + [-\pi\log\pi - (1-\pi)\log(1-\pi)] + \log\left(\frac{\alpha\beta}{4}\right)^{1/2} \\ &= \lambda[U_1^*(\pi) - \pi] + L(\pi) + \mu, \end{aligned} \tag{6.4}$$

where $L(\pi) = -\pi\log\pi - (1-\pi)\log(1-\pi)$. Notice that $\ell(\pi) = L'(\pi) = \log(1-\pi)/\pi$. Let $\lambda_0 = 1/(1 + e^{\lambda})$, $\lambda^0 = (1 - \lambda_0) = e^{\lambda}/(1 + e^{\lambda})$. Then $0 < \lambda_0 < \frac{1}{2} < \lambda^0 < 1$ and $\ell(\lambda_0) = -\ell(\lambda^0) = \lambda$. We choose U_1^* and U_2^* as follows.

$$u_1^*(t) = \begin{cases} 2 - 2t/\lambda_0 & \text{if } 0 \leqq t \leqq \lambda_0, \\ 1 - \ell(t)/\lambda & \text{if } \lambda_0 \leqq t \leqq \lambda^0, \\ 2 - 2(t - \lambda^0)/(1 - \lambda^0) & \text{if } \lambda^0 \leqq t \leqq 1, \end{cases} \tag{6.5}$$

and

$$u_2^*(t) = 2 - u_1^*(t), \qquad 0 \leqq t \leqq 1. \tag{6.6}$$

Then u_1^* is continuous on $[0, 1]$, $0 \leqq u_1^*(t) \leqq 2$, and $\int_0^1 u_1^*(t)\,dt = 1$. Thus u_1^* and u_2^* are continuous probability density functions. An explicit form for U_1^* is

$$U_1^*(t) = \begin{cases} 2t - t^2/\lambda_0 & \text{if } 0 \leqq t \leqq \lambda_0, \\ t + c - L(t)/\lambda & \text{if } \lambda_0 \leqq t \leqq \lambda^0, \\ 2t - \lambda^0 - (t - \lambda^0)^2/(1 - \lambda^0) & \text{if } \lambda^0 \leqq t \leqq 1, \end{cases} \tag{6.7}$$

where $c = L(\lambda_0)/\lambda = L(\lambda^0)/\lambda$. Substituting this U_1^* in (6.4) we obtain

$$i(\pi) - \mu = \begin{cases} \lambda(\pi - \pi^2/\lambda_0) + L(\pi) & \text{if } 0 \leqq \pi \leqq \lambda_0, \\ \lambda c & \text{if } \lambda_0 \leqq \lambda \leqq \lambda^0, \\ \lambda\big(\pi - \lambda^0 - (\pi - \lambda^0)^2/(1 - \lambda^0)\big) + L(\pi) & \text{if } \lambda^0 \leqq \pi \leqq 1. \end{cases} \tag{6.8}$$

It can be shown (for instance by the sign of the derivatives, and so on) that $i(\pi) - \mu < c\lambda$ for $\pi < \lambda_0$ and $\pi > \lambda^0$, and that for any $\theta > 0$ there exists a $\delta > 0$ such that

$$i(\pi) - \mu \leqq c\lambda - \delta \qquad \text{for all} \quad \pi < \lambda_0 - \theta, \pi > \lambda^0 + \theta. \tag{6.9}$$

Thus

$$\max_{0 \leqq \pi \leqq 1} i(\pi) = \lambda c + \mu = i(t) \qquad \text{for any} \quad t \in [\lambda_0, \lambda^0]. \tag{6.10}$$

This means that Condition $\bar{\text{B}}$ is not satisfied.

For $\theta > 0$ and x real let

$$G_\theta(x) = P\{\sup_{\pi \in [\lambda_0 - \theta, \lambda^0 + \theta]} W(\pi) \leqq x\}, \tag{6.11}$$

where $\{W(t), 0 \leqq t \leqq 1\}$ is the Gaussian process in D^- defined with mean function 0 and variance-covariance $K(t, s)$ as defined in (2.23).

Recalling relation (5.14) we have

$$|\ell_N - N \sup_{P \in \bar{\Delta}N} i(P, P; U_{1,N})| = O(\log N). \tag{6.12}$$

Further, by a simplification similar to (6.4)

$$i(P, P; U_{1,N}) = \lambda\big(U_{1,N}(\pi) = \pi\big) + L(\pi) + \mu, \tag{6.13}$$

whenever $\pi = P(\frac{1}{2})$ is a multiple of $1/N$. Thus, from (5.14) and (6.13),

$$\sqrt{N}\,[i(P, P; U_{1,N}) - i(P, P; U_1^*)] = \lambda W_N(\pi), \tag{6.14}$$

whenever $P \in \bar{\Delta}_N$, that is when $\pi = P(\frac{1}{2})$ is a multiple of $1/N$.

Now

(6.15)

$$\sqrt{N}\,[\sup_{P\in\bar\Delta_N} i(P, P; U_{1,N}) - (\lambda c + \mu)]$$
$$\geqq \sqrt{N} \sup_P \{[i(P, P; U_{1,N}) - i(P, P; U_1^*)]; \lambda_0 \leqq \pi \leqq \lambda^0, P \in \bar\Delta_N\}$$
$$= \lambda \sup_P \{W_N(\pi); \lambda_0 \leqq \pi \leqq \lambda^0, \pi \text{ a multiple of } 1/N\},$$

using (6.14). From Lemma 2.1 (i), comparing (5.14) and (6.15), we have that

$$\limsup_N P\{(\ell_N - N[\lambda c + \mu])/\sqrt{N} \leqq x\} \leqq G_0(x). \tag{6.16}$$

Again, for any $\theta > 0$, $\sqrt{N}\,[\sup_{P\in\bar\Delta_N} i(P, P; U_{1,N}) - (\lambda c + \mu)]$ is equal to the maximum of

$$\sqrt{N}\,[\sup_p \{i(P, P; U_{1,N}) - (\lambda c + \mu): P \in \bar\Delta_N, \pi \notin [\lambda_0 - \theta, \lambda^0 + \theta]\}] \tag{6.17}$$

and

$$\sqrt{N} \sup \{i(P, P; U_{1,N}) - (\lambda c + \mu): P \in \bar\Delta_N, \pi \in [\lambda_\theta - \theta, \lambda^\theta + \theta]\} \tag{6.18}$$

which is smaller than the maximum of

$$\sqrt{N} \sup \{[i(P, P; U_{1,N}) - i(P, P; U_1^*)]: P \in \bar\Delta_N, \pi \notin [\lambda_0 - \theta, \lambda^0 + \theta]\} - \delta\sqrt{N} \tag{6.19}$$

and

$$\sqrt{N} \sup_P \{[i(P, P; U_{1,N}) - i(P, P; U_1^*)]: P \in \bar\Delta_N, \pi \in [\lambda_0 - \theta, \lambda^0 + \theta]\}, \tag{6.20}$$

in view of (6.9) and (6.10). The first term in the above tends to $-\infty$ in probability. The limiting distribution function of the second term is $G_\theta(x)$ by an application of Lemma 2.1 (i). Thus

$$\liminf_N P\{[\ell_N - N(\lambda c + \mu)]/\sqrt{N} \leqq x\} \geqq G_\theta(x). \tag{6.21}$$

For the Gaussian process W with continuous path functions one can show that

$$\lim_{\theta\to 0} G_\theta(x) = G_0(x). \tag{6.22}$$

Since $\theta > 0$ is arbitrary in (6.21), using (6.22) and then comparing it with (6.16) we find that

$$\lim_N P\{\ell_N - N(\lambda c + \mu)]/\sqrt{N} \leqq x\} = G_0(x) \tag{6.23}$$

which is the probability that the supremum of the Gaussian process W on $[\lambda_0, \lambda^0]$ is less than or equal to x. The distribution G_0 is clearly not normal. This completes the example.

◇ ◇ ◇ ◇ ◇

We thank Professors R. R. Bahadur and J. Hájek for several helpful remarks and comments.

REFERENCES

[1] R. Berk and I. R. Savage, "The information in a rank-order and the stopping time of some associated SPRT's," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 1661–1674.

[2] P. Billingsley, *Convergence of Probability Measures*, New York, Wiley, 1968.

[3] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (translated by K. L. Chung), Cambridge, Addison-Wesley, 1954.

[4] Z. Govindarajulu, "Asymptotic normality and efficiency of two sample rank order sequential probability ratio tests based on a Lehmann alternative," unpublished, 1968.

[5] J. Hájek, oral communication.

[6] W. Hoeffding, "Optimum nonparametric tests," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 83–92.

[7] R. Pyke and G. Shorack, "Weak convergence of a two-sample empirical process and a new proof to the Chernoff-Savage theorem," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 755–771.

[8] I. R. Savage and J. Sethuraman, "Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives," *Ann. Math. Statist.*, Vol. 37 (1966), pp. 1154–1160. (Correction note, *Ann. Math. Statist.*, Vol. 38 (1967), p. 1309.)

[9] J. Sethuraman, "Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives—II," *Ann. Math. Statist.*, Vol. 41 (1970), pp. 1322–1333.