The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1661–1674

THE INFORMATION IN A RANK-ORDER AND THE STOPPING TIME OF SOME ASSOCIATED SPRT'S[1]

BY ROBERT H. BERK AND I. RICHARD SAVAGE

University of Michigan and Florida State University

1. Introduction. The following is a contribution to the theory of certain SPRT's based on ranks. Such procedures were suggested by Wilcoxon and further discussed by several writers [1], [2], [4]. We consider the two-sample problems of sequentially testing the null hypothesis $G = F$ against the alternative $\phi: F = \phi_1(K), G = \phi_2(K)$, where ϕ_1 and ϕ_2 are given CDF's on [0, 1] that specify ϕ and K ranges through all CDF's. (In cases of interest, as in the null hypothesis, ϕ is usually of the form $G = \phi(F)$, where ϕ is a CDF on [0, 1].) At stage n, one has n observations from each population, which provide the usual rank-order statistic. It is well-known that under either hypothesis, the distribution of the rank-order depends only on the corresponding ϕ; hence one may compute $\mathbf{L}_n$, the likelihood ratio for the rank-order based on the two hypotheses. One continues sampling until $\mathbf{L}_n$ leaves an interval (B, A), $0 < B < 1 < A < \infty$.

The theory presented by Hall, Wijsman and Ghosh (1965) shows that these procedures are indeed SPRT's: i.e., under either hypothesis, the nth rank-order is sufficient for the first n sampling stages. Theoretical knowledge about these procedures is limited, although some Monte Carlo studies are reported by Bradley, Merchant and Wilcoxon (1966). Sethuraman (1967), improving results obtained jointly with Savage (1966), established (with little restriction on the true distributions) that when ϕ is a Lehmann alternative, the stopping variable defined by the SPRT has a finite moment generating function (in a neighborhood of zero). As a further contribution along these lines, we consider more general ϕ and establish the almost-sure convergence of $n^{-1} \ln \mathbf{L}_n$. The convergence is rapid enough to insure that the stopping variable has a finite moment generating function when the limit (which depends on the true distributions) is not zero. The limit is identified as the difference of two information numbers and in the particular case of a Lehmann alternative, with that obtained by Savage and Sethuraman. A by-product of our development is a strengthening of the Glivenko-Cantelli theorem and a similar theorem for a statistic equivalent to the rank-order (Section 5).

2. The main theorem. Let $(\mathbf{X}_1, \cdots, \mathbf{X}_n; \mathbf{Y}_1, \cdots, \mathbf{Y}_n)$ denote the two samples at stage n and F^* and G^*, the sampled populations. (F^*, G^*) need not be in either of the hypotheses defining the sequential rank-test. Let $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_{2n})$ denote the rank-order of the combined sample: $\mathbf{Z}_k = 0$ or 1 according as the kth observation in the combined ordered sample is an X or a Y.

Received 26 July 1967.

[1] Work supported by the Office of Naval Research Contract No. Nonr 988(13) and National Science Foundation Grant GP-6008.

Let $L_n(z)$ denote the likelihood ratio for $\mathbf{Z}$ computed under the hypotheses that the samples are independent and that $F = \phi_1(K)$, $G = \phi_2(K)$ or $G = F$ (we suppose the term corresponding to ϕ is in the numerator). Consider $n^{-1} \ln L_n(\mathbf{Z})$; in the spirit of statistical information theory, this might be termed the information per (X, Y) observation contained in an observed rank-order. (We turn to information theoretic considerations in Section 7.) Our main effort is devoted to proving

2.1. THEOREM. *Suppose* $(\mathbf{X}_1, \mathbf{X}_2, \cdots; \mathbf{Y}_1, \mathbf{Y}_2, \cdots)$ *are independent and identically distributed with continuous distributions* F^* *and* G^*. *Then if* ϕ *belongs to a certain class* Φ, *there exists a real number* I^* *and for all* $\delta > 0$, *there is a* $\rho < 1$ *so that for n sufficiently large,* $P^*(|n^{-1} \ln L_n(\mathbf{Z}) - I^*| > \delta) < \rho^n$. P^* denotes probability under (F^*, G^*); the class Φ is described in Definition 6.3 and I^* is given by (6.2).

In the sequel, it will be convenient to define ϕ implicitly by choosing specific representatives F and G. In this connection, note the following: A monotonic increasing transformation of the data leaves $\mathbf{Z}$ unchanged. Hence we can suppose that all observations are on the unit interval. The following normalization is sometimes convenient. If $H^* = (F^* + G^*)/2$, let $U^* = F^*H^{-*}$ and $V^* = G^*H^{-*}$ (where $H^{-*} = H^{*-1}$ and for specificity, for any CDF H, we define $H^{-1}(t) = \inf \{x: H(x) > t\}$). This corresponds to applying the transformation H^* (which, with probability one, is increasing) to the data. As F^* (resp. G^*) $\ll H^*$, U^* (resp. V^*) $\ll \lambda =$ Lebesgue measure on [0, 1]. Moreover, $U^*(t) + V^*(t) = 2t$, hence the corresponding densities satisfy $u^* + v^* = 2$ and consequently are bounded. Similarly, beginning with any representatives F and G, we may obtain representatives $U = FH^{-1}$ and $V = GH^{-1}$ $(H = (F + G)/2)$ which have bounded densities. Thus when convenient and without loss of generality, we may assume F and G have this property. U and V are a possible choice for ϕ_1 and ϕ_2. Note also that the continuity assumption for F^* and G^* is not essential; they just may not have common discontinuities (so that rank-orders are, with probability one, well-defined). Some discussion of these points is contained in [3], Section 2. In the sequel, the assumption of mutual independence for $\mathbf{X}_1, \mathbf{X}_2, \cdots; \mathbf{Y}_1, \mathbf{Y}_2$ can be weakened to the assumption that $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2), \cdots$ are independently and identically distributed bivariate observations with $\Pr[\mathbf{X} = \mathbf{Y}] = 0$.

3. The histogram case. In Section 6 we prove Theorem 2.1 by first assuming there are representatives F and G with densities f and g that are histograms. That is, there are a finite number of intervals on which f and g are constant. (It is not essential in our calculations that the total area of the histograms be one.) We then extend the results to f and g that can be suitably approximated by histograms. The following analysis is basic to the discussion.

For any representatives f and g, it is known that

$$(3.1) \qquad L_n(z) = (2n)! \int_{0<x_1<\cdots<x_{2n}<1} \prod_1^{2n} f(x_i)^{1-z_i} g(x_i)^{z_i} \, dx_i .$$

Suppose f and g are histograms constant on the J intervals defined by

$0 = a_0 < a_1 < \cdots < a_J = 1$. Equation (3.1) may be written

$$L_n(z) = \sum_{(N_1,\cdots,N_J)} (2n)! \int_{(N_1,\cdots,N_J)} \prod_1^{2n} f(x_i)^{1-z_i} g(x_i)^{z_i}\, dx_i$$

where $\int_{(N_1,\cdots,N_J)}$ means the integration is carried out with the first $N_1 x$'s restricted to (a_0, a_1), the next N_2 to (a_1, a_2), etc. and the summation is over all J-partitions, $(N_1, \cdots, N_J)$ of $2n$. Since f and g are constant on each (a_j, a_{j+1}), the last expression becomes

$$L_n(z) = \sum_{(N_1,\cdots,N_J)} (2n)! \prod_1^J f_j^{m_j} g_j^{n_j}/N_j!,$$

where f_j (resp. g_j) is the area of f (resp. g) over (a_j, a_{j+1}) and $n_j = N_j - m_j$ is the sum of the N_j z's in the jth piece of z when partitioned into J pieces of lengths $N_1, \cdots, N_J$ respectively (from left to right). Since $\sum N_j = 2n$,

$$L_n(z) = \sum_{(N_1,\cdots,N_J)} \binom{2n}{N_1,\cdots,N_J} \exp\{\textstyle\sum_1^J m_j \ln f_j + n_j \ln g_j\}. \tag{3.2}$$

We wish to consider the logarithm of the nth root of L_n. Asymptotically, the nth root of a sum of (not too many) positive terms behaves like the nth root of the largest summand, in that their ratio approaches one. For this, it is sufficient that K_n, the number of summands at stage n, satisfy: $\ln K_n = o(n)$. As the number of summands in (3.2) is $O(n^{J+1})$, asymptotically,

$$\begin{aligned} &n^{-1} \ln L_n(z) \\ &\sim \max_{(N_1,\cdots,N_J)} \{n^{-1} \ln \binom{2n}{N_1,\cdots,N_J} + \textstyle\sum_1^J (m_j/n) \ln f_j + (n_j/n) \ln g_j\} \\ &= I_n'(z), \end{aligned} \tag{3.3}$$

in that their difference approaches zero.

We now wish to replace z by $\mathbf{Z}$ in (3.3) and consider the probabilistic behavior of the resulting random variable. To facilitate this, we let $\mathbf{V}_n(t) = n^{-1} \sum_1^{[2nt]} \mathbf{Z}_k$; $0 \leqq t \leqq 1$ and $[x]$ is the largest integer equal to or less than x. $\mathbf{V}_n(t)$ is the proportion of the Y-sample that falls below the t-fractile of the combined sample. We define $\mathbf{U}_n$ similarly for the X-sample. Thus $\mathbf{U}_n(t) + \mathbf{V}_n(t) = [2nt]/n$. Let $P_0 = 0$ and $P_j = (N_1 + \cdots + N_j)/2n$. Then if $\mathbf{n}_j$ is obtained from $\mathbf{Z}$ as is n_j from z, $\mathbf{n}_j/n = \mathbf{V}_n(P_j) - \mathbf{V}_n(P_{j-1}) = \Delta\mathbf{V}_n(P_j)$ and

$$\begin{aligned} n^{-1} \ln \mathbf{L}_n = n^{-1} \ln L_n(\mathbf{Z}) &\sim \max_{(N_1,\cdots,N_J)} \{n^{-1} \ln \binom{2n}{N_1,\cdots,N_J} \\ &\quad + \textstyle\sum_1^J [(\ln f_j)\Delta\mathbf{U}_n(P_j) + (\ln g_j)\Delta\mathbf{V}_n(P_j)]\} \\ &= \mathbf{I}_n'. \end{aligned} \tag{3.4}$$

4. A limit theorem for $\mathbf{U}_n$. Let $\mathbf{F}_n$ (resp. $\mathbf{G}_n$) denote the empirical CDF of the X (resp. Y) sample and let $\mathbf{H}_n = (\mathbf{F}_n + \mathbf{G}_n)/2$. Then the fractiles of the combined sample are given by $\mathbf{H}_n^{-1}$, and $\mathbf{U}_n = \mathbf{F}_n\mathbf{H}_n^{-1}$. For large n, the combined sample resembles a sample of size $2n$ from the mixed population with CDF $H^* = (F^* + G^*)/2$. The t-fractile of such a sample is approximately $H^{-*}(t)$ and one expects a proportion $U^*(t) = F^*H^{-*}(t)$ of the X-sample to fall below this value.

4.1. THEOREM. *For every* $\epsilon > 0$, *there is a* $\rho < 1$ *so that for* n *sufficiently large*, $P^*(\sup_t |\mathbf{U}_n(t) - U^*(t)| > \epsilon) < \rho^n$.

PROOF. We remark first that as indicated in Section 2, U^* and V^* are absolutely continuous and have densities satisfying $u^* + v^* = 2$. Then $|\mathbf{U}_n - U^*| \leqq |\mathbf{F}_n\mathbf{H}_n^{-1} - F^*\mathbf{H}_n^{-1}| + |F^*\mathbf{H}_n^{-1} - U^*|$. Considering the last term :

$$\sup |F^*\mathbf{H}_n^{-1} - U^*| = \sup |U^*H^*\mathbf{H}_n^{-1} - U^*| \leqq 2\sup_t |H^*\mathbf{H}_n^{-1}(t) - t|$$
$$= 2\sup_t |\mathbf{H}_n H^{-*}(t) - t| = 2\sup |\mathbf{H}_n - H^*|.$$

(Note that for any CDF H on [0, 1], $\sup_t |H(t) - t| = \sup_t |H^{-1}(t) - t|$.) Thus $\sup |\mathbf{U}_n - U^*| \leqq \sup |\mathbf{F}_n - F^*| + 2\sup |\mathbf{H}_n - H^*| \leqq 2\sup |\mathbf{F}_n - F^*| + \sup |\mathbf{G}_n - G^*|$ and the theorem follows from Sethuraman's large-deviation result for the empirical CDF [9]. □

REMARK: It would be of theoretical interest to obtain the limiting value for ρ above. We mention in passing that $n^{\frac{1}{2}}(\mathbf{U}_n - U^*)$ converges in law to a Gaussian process [7]. Further evidence that $\mathbf{U}_n$ behaves like an empirical CDF is given in the next section.

5. Distribution and concentration functions. We collect here some relevant results concerning the Kolmogorov metric for CDF's and the concentration function. The main result is a strengthening of the Glivenko-Cantelli theorem. All CDF's considered have domain [0, 1] and are right-continuous. (Thus for any CDF F, define $F^{-1}(t) = \inf\{x: F(x) > t\}$.) In this section, F, G and H do not have the connotation of the previous sections. We note first that for CDF's F and G,

5.1. LEMMA. $\forall x, |F(x) - G(x)| \leqq \epsilon \Leftrightarrow \forall t, F^{-1}(t - \epsilon) \leqq G^{-1}(t) \leqq F^{-1}(t + \epsilon)$.

PROOF. Since $G \leqq F + \epsilon$, $\{x: G(x) > t\} \subset \{x: F(x) > t - \epsilon\} \Rightarrow G^{-1}(t) \geqq F^{-1}(t - \epsilon)$. Also, $\{x: F(x) > t + \epsilon\} \subset \{x: G(x) > t\}$. □

Let

$$c(\epsilon, F) = \sup_x F(x + \epsilon) - F(x - \epsilon) \tag{5.1}$$

be the concentration function defined by F.

5.2. LEMMA. *F is continuous* $\Leftrightarrow c(\cdot, F)$ *is continuous at zero.*

PROOF. If F is continuous, the mapping $(x, \epsilon) \rightarrow F(x + \epsilon) - F(x - \epsilon)$ is jointly continuous on $[0, 1] \times [0, 1]$ (say) and hence is uniformly continuous there. This implies the continuity of $c(\cdot, F)$. Conversely, if F jumps, it is clear that $c(\epsilon, F) >$ length of the jump for all $\epsilon > 0$. □

5.3. COROLLARY. *If* $\lim_n \sup_x |F_n(x) - F(x)| = 0$ *and F is strictly increasing, then* $\lim_n \sup_t |F_n^{-1}(t) - F^{-1}(t)| = 0$.

PROOF. For by (5.1) and Lemma 5.1, $\sup |F_n - F| \leqq \epsilon \Rightarrow \sup |F_n^{-1} - F^{-1}| \leqq c(\epsilon, F^{-1}) \downarrow 0$ as $\epsilon \downarrow 0$ since F^{-1} is continuous. □

For use in the sequel, we define

$$\begin{aligned} c'(\epsilon, F) = \sup_x \{&\sup \{F(y + \epsilon) - F(y - \epsilon): |y - x| \leqq \epsilon\} \\ &- \sup \{F(y) - F(y-): |y - x| \leqq \epsilon\}\}. \end{aligned} \tag{5.2}$$

We note that $c'(\cdot, F) \leqq c(\cdot, F)$, with equality if F is continuous. In addition, we have always

5.4. LEMMA. *For any F, $c'(\cdot, F)$ is continuous at zero.*

PROOF. We may write $F = F_C + F_D$, where F_C is the continuous part of F and F_D, the discrete part. Clearly $c'(\cdot, F) \leqq c(\cdot, F_C) + c'(\cdot, F_D)$, hence we may suppose that F is discrete. Suppose F jumps at $x_1, x_2, \cdots$, by amounts $p_1 \geqq p_2 \geqq \cdots$. Choose $\delta > 0$ and m so that $\sum_{m+1}^{\infty} p_k < \delta$. Choose $\epsilon > 0$ so that $2\epsilon < \min\{|x_i - x_j|: 1 \leqq i, j \leqq m\}$. Choose $x \varepsilon [0, 1]$. Suppose first that for some $i \leqq m$, $|x - x_i| \leqq \epsilon$. Then $\sup\{F(y) - F(y-): |y - x| \leqq \epsilon\} = p_i$. Since for $1 \leqq j \neq i \leqq m$, $x_j \not\varepsilon [x - \epsilon, x + \epsilon]$, it is clear that $\sup\{F(y + \epsilon) - F(y - \epsilon): |y - x| \leqq \epsilon\} \leqq p_i + \delta$. If, on the other hand, $|x - x_i| > \epsilon$ for $i = 1, \cdots, m$, then $\sup\{F(y + \epsilon) - F(y - \epsilon): |y - x| \leqq \epsilon\} \leqq \delta$. Thus $c'(\epsilon, F) \leqq \delta$, showing that $c'(\cdot, F)$ is continuous at zero. □

Consider now $\mathbf{F}_n$, the empirical CDF based on n independent observations from F^*. We extend the Glivenko-Cantelli theorem by noting that, in addition to convergence in Kolmogorov-distance, there is, asymptotically, a uniform correspondence between the intervals over which $\mathbf{F}_n$ and F^* are constant. (In general, if F^* is constant over some interval, even though $\sup|F_n - F^*| \to 0$, F_n can be strictly increasing for every n. This does not happen with the empirical CDF.) In the sequel, an F-interval denotes one over which F is constant and which is not contained in any larger such interval. For any two CDF's F and G, we define their maximum discrepancy, $\nu(F, G)$, as follows: Let $R_1, R_2, \cdots$, be the (non-degenerate) F-intervals (take all $R_i = \emptyset$ if there are none) and let $S_1, S_2, \cdots$, be the G-intervals. Then, letting Δ denote symmetric difference and λ, Lebesgue measure on [0, 1],

$$\nu(F, G) = \sup_i \inf_j \lambda(R_i \Delta S_j) + \sup_j \inf_i \lambda(R_i \Delta S_j). \tag{5.3}$$

(Note that sup and inf are really max and min.) ν measures the maximum non-overlap between corresponding F and G-intervals. If $\nu < \epsilon$ and x_1 and x_2 are in some F-interval, then there exist y_1 and y_2 in a G-interval so that $|x_i - y_i| < \epsilon$, $i = 1, 2$ (and vice versa).

5.5. LEMMA. *ν is a pseudometric.*

PROOF. Symmetry is obvious; we verify the triangle inequality. Let R, S and T be three intervals. Then

$$\lambda(R \Delta S) \leqq \lambda(R \Delta T) + \lambda(S \Delta T). \tag{5.4}$$

This follows immediately from the fact that $R\Delta S \subset (R\Delta T) \cup (S\Delta T)$, which in turn follows from $R - S \subset (R - T) \cup (T - S)$.

Let F, G and H be three CDF's and $\{R_i\}$, $\{S_i\}$, $\{T_i\}$ be their intervals. Let $a_{ij} = \lambda(R_i \Delta S_j)$, $b_{ik} = \lambda(R_i \Delta T_k)$, $c_{kj} = \lambda(T_k \Delta S_j)$. Let $a_i = \inf_j a_{ij}$ and $a = \sup_i a_i$, etc. Then (5.4) implies that $a_{ij} \leqq b_{ik} + c_{kj} \Rightarrow a_i \leqq b_{ik} + c_k \Rightarrow a_i - c_k \leqq b_{ik} \Rightarrow a_i - c \leqq b_i \Rightarrow a_i \leqq b_i + c \Rightarrow a \leqq b + c$. Letting $a_j^* = \inf_i a_{ij}$ and $a^* = \sup_j a_j^*$, we find similarly that $a^* \leqq b^* + c^*$. Finally, $\nu(F, G) = a + a^*$. □

For $\nu(\mathbf{F}_n, F^*)$, (5.3) simplifies. For if $\{R_i\}$ denote the F^*-intervals, wp 1, there are no observations in R_i and $\mathbf{F}_n$ will be constant over R_i too. Thus every F^*-interval is contained in some $\mathbf{F}_n$ interval and $\nu(\mathbf{F}_n, F^*) = \sup_i \inf_j (\mathbf{S}_i - R_j)$, where $\{\mathbf{S}_j\}$ are the $\mathbf{F}_n$ intervals.

5.6. THEOREM. *For all $\delta > 0$ there is a $\rho < 1$ so that for n sufficiently large, $P^*(\nu(\mathbf{F}_n, F^*) > \delta) < \rho^n$.*

PROOF. Choose $\delta > 0$. We will show that there is an $\epsilon > 0$ so that $\sup |\mathbf{F}_n - F^*| \leqq \epsilon \Rightarrow \nu(\mathbf{F}_n, F^*) \leqq \delta$. Let t_i be the value of F^* on R_i and choose m so that $\sum_{m+1}^{\infty} \lambda R_i < \delta/2$. Choose $\epsilon > 0$ so that $c'(\epsilon, F^{-*}) < \delta/2$ and $2\epsilon < \min\{|t_i - t_j|: 1 \leqq i, j \leqq m\}$; by Lemma 5.4, this is always possible. Suppose that $\sup |\mathbf{F}_n - F^*| \leqq \epsilon$ and that $\mathbf{F}_n$ assumes the value t on one of its intervals; we distinguish three cases.

CASE 1: For some $i \leqq m$, $|t - t_i| \leqq \epsilon$ and the $\mathbf{F}_n$-interval contains R_i. Then, since for $j \neq i, j \leqq m$, $|t - t_j| > \epsilon$, the amount by which this $\mathbf{F}_n$-interval exceeds λR_i is bounded above by $c'(\epsilon, F^*) \leqq \delta$ (see Lemma 5.1 and (5.2)).

CASE 2: For some $i \leqq m$, $|t - t_i| \leqq \epsilon$ but the $\mathbf{F}_n$-interval does not contain R_i. Then the length of this $\mathbf{F}_n$-interval cannot exceed $\max\{F^{-*}(t_i + 2\epsilon) - F^{-*}(t_i), F^{-*}(t_i-) - F^{-*}(t_i - 2\epsilon)\} \leqq c(\epsilon, F_C^{-*}) + \delta/2 \leqq \delta$, where F_C^{-*} is the continuous part of F^{-*} and the $\delta/2$ reflects the maximum variation of F_D^{-*} above $F^{-*}(t_i)$ or below $F^{-*}(t_i-)$.

CASE 3: For all $i \leqq m$, $|t - t_i| > \epsilon$. Then $\mathbf{F}_n^{-1}(t) - \mathbf{F}_n^{-1}(t-) \leqq F^{-*}(t + \epsilon) - F^{-*}(t - \epsilon) \leqq c(\epsilon, F_C^{-*}) + \delta/2 \leqq \delta$. Thus, provided $\sup |\mathbf{F}_n - F^*| \leqq \epsilon$, the maximum discrepency $\nu(\mathbf{F}_n, F^*) \leqq \delta$. The proof is concluded by an appeal to Sethuraman's large-deviation result for the empirical CDF. □

The relation between $\mathbf{U}_n$ and U^* intervals is not quite as simple as for $\mathbf{F}_n$ and F^*; yet:

5.7. COROLLARY. *For all $\delta > 0$ there is a $\rho < 1$ so that for n sufficiently large, $P^*(\nu(\mathbf{U}_n, U^*) > \delta) < \rho^n$.*

PROOF. We may suppose that $H^*(x) = x$ (thus $F^* = U^*$). We then argue that

$$\nu(\mathbf{F}_n, F^*) \leqq \delta/3 \quad \text{and} \quad \sup_t |\mathbf{H}_n^{-1}(t) - t| \leqq \delta/3 \Rightarrow \nu(\mathbf{U}_n, U^*) \leqq \delta.$$

For then, if (a, b) is an $\mathbf{F}_n$-interval, it contains an F^*-interval of length $\geqq b - a - \nu(\mathbf{F}_n, F^*) \geqq b - a - \delta/3$. Moreover, $\mathbf{U}_n = \mathbf{F}_n\mathbf{H}_n^{-1}$ will have a corresponding interval containing $(a + \delta/3, b - \delta/3)$ and contained in $(a - \delta/3, b + \delta/3)$. Thus the discrepency is $\leqq \delta$. The proof is concluded by noting that the large-deviation result for $\mathbf{H}_n$ implies one for $\mathbf{H}_n^{-1}$ (either via Corollary 5.3 or because $\sup |\mathbf{H}_n - t| = \sup |\mathbf{H}_n^{-1} - t|$). □

We may thus extend the Glivenko-Cantelli theorem as follows. Let

$$d(F, G) = \sup |F - G| + \nu(F, G); \tag{5.5}$$

d defines a metric on CDF's. Then Theorem 5.6 and the result of Sethuraman referred to there imply:

5.8. COROLLARY. *For all $\delta > 0$ there is a $\rho < 1$ so that for n sufficiently large,*

$P^*(d(\mathbf{F}_n, F^*) > \delta) < \rho^n$. *Theorem* 4.1 *and Corollary* 5.7 *imply a similar result for* $d(\mathbf{U}_n, U^*)$.

6. Proof of Theorem 2.1. We return to the main development after establishing a preliminary lemma.

6.1. LEMMA. *Let* $\mathcal{U}$ *be a metric space metrized by* d. *Let* $(p, u) \to i(p, u)$ *be a mapping of* $\mathcal{P} \times \mathcal{U}$ *into* R, *where* $\mathcal{P}$ *is arbitrary. A sufficient condition that* $I(\cdot) = \sup\{i(p, \cdot) : p \in \mathcal{P}\}$ *be continuous is that for all* $\epsilon > 0, \exists \delta > 0$ *so that when* $I(u) - i(p, u) < \epsilon$ *and* $d(u, u') < \delta, \exists p' \in \mathcal{P}$ *so that* $|i(p, u) - i(p', u')| < \epsilon$.

REMARK. Under the given conditions, I turns out to be uniformly continuous.

PROOF. Given $\epsilon > 0$ and u, choose p so that $I(u) - i(p, u) < \epsilon/2$. Choose δ as in the hypothesis so that when $d(u, u') < \delta$, $\exists p'$ satisfying $|i(p, u) - i(p', u')| < \epsilon/2$. Then if $d(u, u') < \delta$, $I(u) - I(u') = [I(u) - i(p, u)] + [i(p, u) - i(p', u')] + [i(p', u') - I(u')] \leqq \epsilon$. Reversing the roles of u and u' (i.e., choosing first p' so that $I(u') - i(p', u') < \epsilon/2$) gives, similarly, $I(u') - I(u) \leqq \epsilon$. Thus $d(u, u') < \delta \Rightarrow |I(u) - I(u')| \leqq \epsilon$. □

To this point, we have not carried the main development beyond (3.3). In view of (3.2), we have that $|n^{-1} \ln \mathbf{L}_n - \mathbf{I}_n'| \leqq O(n^{-1} \ln n)$, where the bound is non-stochastic, being determined by the number of summands in (3.2). Thus the limiting behavior of $n^{-1} \ln \mathbf{L}_n$ and $\mathbf{I}_n'$ coincide. In view of (3.4) and Theorem 4.1, one might expect $\mathbf{I}_n'$ to have a non-stochastic limit. This is a consequence of Lemma 6.2 below. Let $P = (P_0, P_1, \cdots, P_J)$, $0 = P_0 \leqq P_1 \leqq \cdots \leqq P_J = 1$ and $p_j = P_j - P_{j-1}$. Let

$$i(P, U^*) = \sum [\ln f_j \Delta U^*(P_j) + \ln g_j \Delta V^*(P_j) - 2p_j \ln p_j]; \tag{6.1}$$

$$I(U^*) = I^* = \sup\{i(P, U^*) : 0 \leqq P_1 \leqq \cdots \leqq P_J = 1\}. \tag{6.2}$$

6.2. LEMMA. *Suppose* f *and* g *are histograms. Then for all* $\epsilon > 0$ *there is a* $\rho < 1$ *so that for* n *sufficiently large,* $P^*(|n^{-1} \ln \mathbf{L}_n - I^*| > \epsilon) < \rho^n$.

PROOF. Again letting $P_j = (N_1 + \cdots + N_j)/2n$, so that $p_j = N_j/2n$, we first show that the constant on the right hand side of (3.4) becomes uniformly close to $-2 \sum p_j \ln p_j$ as n increases. To this end, we use the following modified Stirling approximation (that is valid uniformly in $N \geqq 0$ as n increases):

$$\begin{aligned} n^{-1} \ln N! &= n^{-1}[\ln (N + 1)! - \ln (N + 1)] \\ &= n^{-1}[-N + (N + \tfrac{1}{2}) \ln (N + 1)] + O(n^{-1}). \end{aligned}$$

Then

$$n^{-1} \ln \binom{2n}{N_1, \cdots, N_J} = -\sum (p_j + \tfrac{1}{4}n) \ln (p_j + \tfrac{1}{2}n) + O(n^{-1}).$$

Since

$$|n^{-1} \sum \ln (p_j + \tfrac{1}{2}n)| \leqq n^{-1} J \ln 2n \quad \text{and}$$

$$\begin{aligned} \sum p_j(\ln (p_j + \tfrac{1}{2}n) - \ln p_j) &= \sum p_j \ln (1 + \tfrac{1}{2}np_j) \leqq \ln \sum p_j(1 + \tfrac{1}{2}np_j) \\ &= \ln (1 + J/2n), \end{aligned}$$

we see that $|n^{-1} \ln \binom{2n}{N_1, \cdots, N_J} + 2\sum p_j \ln p_j| = O(n^{-1} \ln n)$. Thus asymptotically, we may replace the multinomial term in (3.4) by $-2\sum p_j \ln p_j$. We now replace U^* and V^* in (6.1) by $\mathbf{U}_n$ and $\mathbf{V}_n$ and denote the resulting expression by $i(P, \mathbf{U}_n)$. Let

$$I(\mathbf{U}_n) = \mathbf{I}_n = \sup_P i(P, \mathbf{U}_n), \tag{6.3}$$

where, in accordance with (3.4), the supremum (actually, maximum) is over all P whose coordinates are multiples of $\frac{1}{2}n$. Since $\mathbf{U}_n$ and $\mathbf{V}_n$ increase only at multiples of $\frac{1}{2n}$, we may take the supremum over all P as in (6.1). In view of the foregoing, $|\mathbf{I}_n - \mathbf{I}_n'| = O(n^{-1} \ln n)$, so the limiting behavior of $n^{-1} \ln \mathbf{L}_n$ and $\mathbf{I}_n$ coincide.

We now argue that for large n, the respective maxima of $i(\cdot, \mathbf{U}_n)$ and $i(\cdot, U^*)$ are close, except with exponentially small probability. We apply Lemma 6.1 to i defined by (6.1), using the metric d given by (5.5). We show that the condition of Lemma 6.1 is satisfied for this metric. Corollary 5.8 will then establish a large-deviation result for $\mathbf{I}_n$. Choose P so that $i(P, U^*) > -\infty$. Then $f_j = 0 \Rightarrow \Delta U^* P_j = 0$ (resp., $g_j = 0 \Rightarrow \Delta V^* P_j = 0$). If $d(\mathbf{U}_n, U^*) \leqq \delta$ and $d(\mathbf{V}_n, V^*) \leqq \delta$, we may choose P' so that $|P_j - P_j'| < \delta, j = 1, \cdots, J$ and so that $\Delta U^* P_j = 0 \Rightarrow \Delta \mathbf{U}_n P_j' = 0$ (resp., $\Delta V^* P_j = 0 \Rightarrow \Delta \mathbf{V}_n P_j' = 0$). (This is because the intervals of constancy for $\mathbf{U}_n$ and U^* (resp. $\mathbf{V}_n$ and V^*) have a maximum symmetric difference of length at most δ.) Moreover,

$$\begin{aligned} |\Delta U^* P_j - \Delta \mathbf{U}_n P_j'| &\leqq |\Delta \mathbf{U}_n P_j' - \Delta U^* P_j'| + |\Delta U^* P_j - \Delta U^* P_j'| \\ &\leqq 2 \sup |\mathbf{U}_n - U^*| + 4 \max_j |P_j - P_j'| \\ &\leqq 6\delta. \end{aligned}$$

Furthermore, $|\sum (p_j \ln p_j - p_j' \ln p_j')| \leqq 2J\delta|\ln 2\delta|$; hence

$$|i(P, U^*) - i(P', \mathbf{U}_n)| \leqq 6\delta[\textstyle\sum_{f_j>0} |\ln f_j| + \sum_{g_j>0} |\ln g_j|] + 4J\delta|\ln 2\delta|.$$

Thus we may choose δ so that this last quantity is less than a given ϵ. From Lemma 6.1 we see that I is (uniformly) continuous in the d metric and Corollary 5.8 (for $\mathbf{U}_n$ and $\mathbf{V}_n$) completes the proof of Lemma 6.2. □

We complete the proof of Theorem 2.1 by extending Lemma 6.2 to densities that can be suitably approximated by histograms. For this purpose, we make explicit the dependence of i, $\mathbf{I}_n$, etc. on f (and g) by writing $i(\cdot, \cdot; f)$, $\mathbf{I}_n(f)$, etc. Consider (6.1). Since $U^*(t) + V^*(t) = 2t$,

$$\Delta U^*(P_j) + \Delta V^*(P_j) = 2p_j \text{ and}$$

$$i(P, U^*; f) = \sum [\ln (f_j/p_j) \Delta U^*(P_j) + \ln (g_j/p_j) \Delta V^*(P_j)].$$

Let p be the histogram having the same support as f and g and with areas p_j; let P be the corresponding CDF. We may write (6.1) as

$$\begin{aligned} i(P, U^*; f) &= \int [\ln (f/p)\, dU^*P + \ln (g/p)\, dV^*P] \\ &= \int [\ln f\, dU^*P + \ln g\, dV^*P - 2 \ln p\, dP] \\ &= \int [\ln (dF/dP)\, dU^*P + \ln (dG/dP)\, dV^*P], \end{aligned} \tag{6.4}$$

where $U^*P(t) = U^*(P(t))$. We henceforth identify a vector $P = (P_1, \cdots, P_J)$ with its corresponding integrated histogram $P(\cdot)$. Thus (6.2) becomes

$$I(U^*; f) = I^*(f) = \sup_P i(P, U^*; f), \tag{6.5}$$

where the supremum is over all integrated histograms P. We will show that for a large class of (f, g), $|\mathbf{I}_n(f) - I^*(f)|$ satisfies a large deviation result. To this end we say:

6.3. Definition. *The alternative ϕ is in Φ if representatives f and g (on* [0, 1]*) may both be chosen to satisfy:*

(i) *f is of bounded variation.*

(ii) *For δ sufficiently small, $[0 < f < \delta]$ is a (possibly empty) interval and $[f = 0]$ is a (possibly empty) finite union of intervals.*

(iii) $\exists \gamma_f$, $0 < \gamma_f < 1$ so that

$$\limsup_{\delta \to 0} \lambda[0 < f < \gamma_f \delta]/\lambda[0 < f < \delta] = 1 - 2\alpha_f < 1.$$

Above, $[\cdot]$ means $[x: \cdot]$; in (iii), we take the expression to be zero if the numerator vanishes. Condition (iii) requires that the representative densities not decrease too rapidly to zero. If the representatives F and G (on $(-\infty, \infty)$) are normalized to give representatives U and V, then $u = dU/dt = 2fH^{-1}/(fH^{-1} + gH^{-1})$; $H = (F + G)/2$. Hence if f and g satisfy (i) and (ii), so do their normalizations. A useful condition that implies (iii) is

(iii′) $\exists \beta > 0$ *so that in a neighborhood of zero, $\lambda[0 < f < \delta]/\delta^\beta$ is non-decreasing in δ.*

(Under (iii′), (iii) holds for every $0 < \gamma < 1$.) Condition (iii′) is essentially a condition on the likelihood ratio g/f and may be so stated for arbitrary representatives: From above, $u = 2/(1 + gH^{-1}/fH^{-1})$ and $[0 < u < \delta] = [gH^{-1}/fH^{-1} > k]$, where $k = (2 - \delta)/\delta$. Thus $\lambda[0 < u < \delta] = \lambda[gH^{-1}/fH^{-1} > k] = H[g/f > k]$, where the last set is contained in $(-\infty, \infty)$. That u satisfies (iii′) is equivalent to the existence of a $\beta > 0$ so that $H[g/f > k]k^\beta$ is eventually non-increasing in k. Since $F[g/f > k] \leqq G[g/f > k]/k$, this is equivalent to $G[g/f > k]$ satisfying this condition. That is, under G, the likelihood ratio g/f must have a distribution that decreases as fast as k^β, for some $\beta > 0$. (Or, $G[\ln(g/f) > x]$ must eventually decrease (at least) as fast as $e^{-\beta x}$.) Of course f/g must behave similarly under F. (From the above it follows too that if representatives f and g on [0, 1] satisfy (iii) (or (iii′)), so do their normalizations u and v.)

6.4. Lemma. *If ϕ is in Φ, then for all $\epsilon > 0$ there is a $\rho < 1$ so that for n sufficiently large, $P^*(|\mathbf{I}_n(f) - I^*(f)| > \epsilon) < \rho^n$.*

Proof. Choose $\delta > 0$ and representatives satisfying (i)–(iii) above. Since $\ln \max(f, \delta)$ is of bounded variation, there are (sign generalized) histograms r_δ and r^δ such that $r_\delta \leqq \ln \max(f, \delta) \leqq r^\delta$ and $r^\delta - r_\delta \leqq \delta$. By (ii), we can suppose that $r_\delta = r^\delta = \ln \delta$ on $[0 < f < \delta]$. Let $f^\delta = \chi_{[f>0]} \exp r^\delta$, $f_\delta = \chi_{[f \geqq \delta]} \exp r_\delta + (\delta\gamma_f)\chi_{[\delta\gamma_f \leqq f < \delta]}$, where χ_A is the characteristic function of a set A. Then $f_\delta \leqq f \leqq f^\delta$ and $\ln(f^\delta/f_\delta) \leqq \delta$ on $[f \geqq \delta]$. Similarly, we obtain histograms g_δ and g^δ. (6.3) and (6.4) show that

$$\mathbf{I}_n(f_\delta) \leqq \mathbf{I}_n(f) \leqq \mathbf{I}_n(f^\delta). \tag{6.6}$$

Moreover, since f^δ etc. are histograms, it follows from Lemma 6.2 that both $|\mathbf{I}_n(f_\delta) - I^*(f_\delta)|$ and $|\mathbf{I}_n(f^\delta) - I^*(f^\delta)|$ satisfy a large deviation result. From (6.4) and (6.5), we see that

$$I^*(f_\delta) \leqq I^*(f) \leqq I^*(f^\delta). \tag{6.7}$$

We now show that $I^*(f^\delta) - I^*(f_\delta) \to 0$ as $\delta \to 0$. Without loss of generality, we may assume f and g are bounded above by 2 and that $A = [0 < f < \delta]$ and $B = [0 < g < \delta]$ are disjoint. From 6.4 we see that for any measurable $C \subset [0, 1]$,

$$\begin{aligned} i(P, U^*; f) \leqq 2 \ln 2 - 2 \textstyle\int \ln p \, dP \leqq 2 \ln 2 &- 2PC \ln (PC/\lambda C) \\ &- 2(1 - PC) \ln [(1 - PC)/(1 - \lambda C)]. \end{aligned} \tag{6.8}$$

The inequality follows from the convexity of $-p \ln p$. Choose an integrated histogram P^δ so that $i(P^\delta, U^*; f^\delta) = I^*(f^\delta)$. Letting $C = A$ in (6.8), we see that $I^*(f^\delta) \leqq 2 \ln 2 - 2P^\delta A \ln (P^\delta A/\lambda A) - 2(1 - P^\delta A) \ln [(1 - P^\delta A)/(1 - \lambda A)]$. As the LHS of this inequality is uniformly (in δ) bounded away from $-\infty$ (by $I^*(f)$, e.g.) and as $\delta \to 0$, $\lambda A \to 0$, we see that $P^\delta A \to 0$ also. Similarly, $P^\delta B \to 0$. Modify P^δ to get P_δ by transferring all P^δ mass on $[0 < f < \delta\gamma_f]$ (resp. $[0 < g < \delta\gamma_g]$) to $A' = [\delta\gamma_f \leqq f < \delta]$ (resp. $B' = [\delta\gamma_g \leqq g < \delta]$), retaining the histogram structure. As this merely redistributes mass inside two intervals, P_δ and P^δ agree off $A \cup B$. Then,

$$\begin{aligned} I^*(f^\delta) - I^*(f_\delta) &\leqq i(P^\delta, U^*; f^\delta) - i(P_\delta, U^*; f_\delta) \\ &= \textstyle\int \ln f^\delta \, dU^*P^\delta + \int \ln g^\delta \, dV^*P^\delta - 2 \int \ln p^\delta \, dP^\delta - \int \ln f_\delta \, dU^*P_\delta \\ &\quad - \textstyle\int \ln g_\delta \, dV^*P_\delta + 2 \int \ln p_\delta \, dP_\delta \\ &\leqq 2\delta + [\textstyle\int_A \ln \delta \, dU^*P^\delta - \int_{A'} \ln \delta\gamma_f \, dU^*P_\delta] \\ &\quad + [\textstyle\int_B \ln \delta \, dV^*P^\delta - \int_{B'} \ln \delta\gamma_g \, dV^*P_\delta] \\ &\quad + 2[\textstyle\int_{A' \cup B'} \ln p_\delta \, dP_\delta - \int_{A \cup B} \ln p^\delta \, dP^\delta] \\ &\leqq 2\delta + 2P^\delta A \ln (1/\gamma_f) + 2P^\delta B \ln (1/\gamma_g) \\ &\quad + 2[P_\delta A' \ln (P_\delta A'/\lambda A') + P_\delta B' \ln (P_\delta B'/\lambda B') \\ &\quad - P^\delta A \ln (P^\delta A/\lambda A) - P^\delta B \ln (P^\delta B/\lambda B)] \\ &\leqq 2[\delta + P^\delta A \ln (1/\gamma_f) + P^\delta B \ln (1/\gamma_g) + P^\delta A \ln (\lambda A/\lambda A') \\ &\quad + P^\delta B \ln (\lambda B/\lambda B')] \\ &\leqq 2[\delta + P^\delta A\{\ln (1/\gamma_f) + \ln (1/\alpha_f)\} \\ &\quad + P^\delta B\{\ln (1/\gamma_g) + \ln (1/\alpha_g)\}] \\ &\to 0 \quad \text{as} \quad \delta \to 0. \end{aligned}$$

(The above reduction is facilitated by noting that off A (resp., B), $\ln (f^\delta/f_\delta) \leqq \delta$ (resp., $\ln (g^\delta/g_\delta) \leqq \delta$) and that p^δ and p_δ coincide off $A \cup B$, while p_δ vanishes on

$A \cup B - (A' \cup B')$. Also, on A (resp. B), $p^\delta = P^\delta A/\lambda A$ (resp. $P^\delta B/\lambda B$); a similar fact holds for p_δ and $P_\delta A' = P^\delta A$. Finally, for δ sufficiently small, condition (iii) guarantees that $\lambda A/\lambda A' \leqq 1/\alpha_f$ (resp. $\lambda B/\lambda B' \leqq 1/\alpha_g$.) Hence $I^*(f^\delta) - I^*(f_\delta) \to 0$ and the lemma follows from the large deviation results for $\mathbf{I}_n(f^\delta)$ and $\mathbf{I}_n(f_\delta)$ and (6.7). □

We note that this proof can be adapted to the case when $[0 < f < \delta]$ and $[0 < g < \delta]$ are finite unions of intervals, provided condition (iii) holds inside each one of them. This concludes the proof of Theorem 2.1.

7. The limit expressed by information numbers. As shown above, when ϕ is in Φ, $\lim n^{-1} \ln \mathbf{L}_n = I^*(f)$. Below, we interpret this limit as the difference of two (Kullback-Leibler) information numbers. The following lemmas show that in obtaining $I^*(f)$ by maximizing $i(\cdot, U^*, f)$, the restriction to integrated histograms is not necessary.

7.1. LEMMA. *If f and g are histograms, $I^*(f) = \sup_P i(P, U^*; f)$, where the supremum is over all continuous* CDF*'s that dominate both F and G* (i.e., $F + G$).

PROOF. Choose P satisfying the hypothesis and let P_J be the integrated histogram agreeing with P at the partition points $(a_1, \cdots, a_J)$. Then $F, G \ll P_J \ll P$ and writing $dF/dP = (dF/dP_J)(dP_J/dP)$, etc., from (6.4) we see that $i(P, U^*; f) = \int [\ln (dF/dP_J)\, dU^*P + \ln (dG/dP_J)\, dV^*P] + 2 \int \ln (dP_J/dP)\, dP$. Since dF/dP_J (resp. dG/dP_J) is constant on each (a_{j-1}, a_j), we may replace P by P_J in the first integral. Hence

$$i(P, U^*; f) = i(P_J, U^*; f) + 2 \int \ln (dP_J/dP)\, dP \leqq i(P_J, U^*; f).$$

The lemma follows. □

7.2. LEMMA. *Lemma* 7.1 *continues to hold for any f and g representing any ϕ in Φ.*

PROOF. Choosing f^δ etc. as in Lemma 6.4, by the previous lemma we have $I^*(f_\delta) \leqq I^*(f) \leqq \sup_P i(P, U^*; f) \leqq I^*(f^\delta)$, where the supremum is over all continuous P dominating F and G. As shown in Lemma 6.4, $I^*(f^\delta) - I^*(f_\delta) \to 0$ and the desired conclusion follows. Thus for the cases of interest to us, (6.5) is subsumed under the definition:

$$I^*(f) = \sup_P i(P, U^*; f) \tag{7.1}$$

where the supremum is over all continuous CDF's dominating F and G.

We now proceed to identify $I^*(f)$ as the difference of two *information* numbers. The authors are indebted to Professor Herman Chernoff for this interpretation. Since $F \ll P \Leftrightarrow FP^{-1} \ll \lambda$, letting $Q = P^{-1}$, we see from (7.1) and (6.5) that

$$I^*(f) = \sup_Q \int [\ln (dFQ/d\lambda)\, dU^* + \ln (dGQ/d\lambda)\, dV^*],$$

where the supremum is over all Q satisfying $FQ \ll \lambda$ and $GQ \ll \lambda$. Suppose ϕ is absolutely continuous and consider

$$I(\phi^*; \phi) = \inf_{\hat{F}} \int \ln [d(U^* \times V^*)/d(\phi_1(\hat{F}) \times \phi_2(\hat{F}))]\, d(U^* \times V^*),$$

where $\phi^* = (U^*, V^*)$, $U^* \times V^*$ is the product measure defined by U^* and V^*

and the infimum is over all $\hat{F} \ll \lambda$. Following Kullback (1959), $I(\phi^*; \phi)$ may be called the minimum information per (X, Y) observation for discriminating against ϕ.

7.3. LEMMA. *If ϕ is represented by f and g, $I^*(f) = I(\phi^*; \phi_0) - I(\phi^*; \phi)$ where ϕ_0 is the hypothesis $[G = F]$.*

PROOF. Note first that when $f = g$, $I^*(f) = \sup_P \int \ln (dF/dP)\, d(U^*P + V^*P) = 2 \sup \int \ln (dF/dP)\, dP = 0$ (choose $P = F$). Hence

$$\begin{aligned} I^*(f) &= \sup_Q \int \ln (d(FQ \times GQ)/d(\lambda \times \lambda))\, d(U^* \times V^*) \\ &\quad - \sup_Q \int \ln (d(FQ \times FQ)/d(\lambda \times \lambda))\, d(U^* \times V^*) \\ &= \inf_{\hat{F}} \int \ln (d(U^* \times V^*)/d(\hat{F} \times \hat{F}))\, d(U^* \times V^*) \\ &\quad - \inf_{\hat{F}} \int \ln (d(U^* \times V^*)/d(\phi_1(\hat{F}) \times \phi_2(\hat{F})))\, d(U^* \times V^*) \\ &= I(\phi^*; \phi_0) - I(\phi^*; \phi). \end{aligned}$$

(Note that in the penultimate integral, $\hat{F} \ll \lambda$ and in the final integral, $\phi_1\hat{F} \ll \lambda$ and $\phi_2\hat{F} \ll \lambda$.)

8. Application to SPRT'S. The results of the last section may be interpreted in terms of the asymptotic behavior of $\mathbf{L}_n$. Asymptotically, $\mathbf{L}_n$ moves toward whichever hypothesis is less distinguishable from (F^*, G^*), as measured by $I(\phi^*; \cdot)$. Thus if $I(\phi^*; \phi) > I(\phi^*; \phi_0)$, $[G = F]$ is the less distinguishable hypothesis and Theorem 2.1 shows that $\lim n^{-1} \ln \mathbf{L}_n < 0$, hence $\mathbf{L}_n \to 0$. This has immediate application to the SPRT based on $\mathbf{L}_n$. In fact, the large deviation result of Theorem 2.1 tells us more.

8.1. COROLLARY. *If ϕ is in Φ, then $\mathbf{N}$, the stopping time defined by the* SPRT *based on $\mathbf{L}_n$, has a moment generating function that is finite in a neighborhood of the origin under any (F^*, G^*) for which $I^*(f) \neq 0$.*

PROOF. By a standard argument (see, e.g., Theorem 2.1 of [8]), for some $\epsilon > 0$, $P^*(\mathbf{N} > n) \leqq P^*(|n^{-1} \ln \mathbf{L}_n - I^*(f)| > \epsilon) < \rho^n$ for n sufficiently large. Thus under (F^*, G^*), $\mathbf{N}$ has a non-degenerate moment generating function. □

If ϕ is a Lehmann alternative, it is easily checked that ϕ is in Φ. Hence we provide an alternative derivation of the result given by Savage and Sethuraman (1966). In Section 9, we identify their limit with our I^*. If f and g are normal densities differing by a shift, it is immediate that $G[\ln (g/f) > x]$ decreases exponentially. Hence this alternative belongs to Φ and our results apply to that particular case of interest as well.

9. Lehmann alternatives. We now establish that the limit for $n^{-1} \ln \mathbf{L}_n$ given here coincides with that given by Savage and Sethuraman (1966) when ϕ is a Lehmann alternative. We take $\phi(t) = t^m$, $m > 1$ and choose representatives $F(t) = t$, $G(t) = t^m$. We then wish to evaluate

$$I^*(m) = \sup_P \int [\ln (dF/dP)\, dU^*P + \ln (dG/dP)\, dV^*P]. \tag{9.1}$$

Since ϕ is in Φ, the supremum may be taken over all integrated histograms dominating F and G (i.e., that dominate Lebesgue measure) or some larger class. If $\lambda \ll P$, then $P^{-1} = Q \ll \lambda$ and we consider those P such that Q has an absolutely continuous density q. This class contains the histograms described above. Substituting $Q = P^{-1}$, (9.1) becomes

$$(9.2) \qquad I(m) = \sup_Q \{\ln m + \int [(m-1) \ln Q \, dV^* + 2 \ln q \, dt]\}.$$

Applying Euler's equation to the quantity in brackets yields $(m-1)v^*/Q = -2q'/q^2$, where $q' = dq/dt$. Letting $S = \ln Q$ and $s = S'$, $q = se^S$ and we obtain $(m-1)v^*s = -2(s'/s + s)$, or $(m-1)v^* + 2 = -2s'/s^2 = 2\,d(1/s)/dt$. Integration yields $2/s = (m-1)V^* + 2(t+a)$ (a is an arbitrary constant); thus $S(t) = b - \int_t^1 2\,dr/[(m-1)V^*(r) + 2(r+a)]$, where b is another arbitrary constant. Hence the stationary Q is given by

$$Q(t) = c \exp\{-\int_t^1 2\,dr/[(m-1)V^*(r) + 2(r+a)]\}; \qquad c = e^b.$$

Since $Q(1) = 1$, $c = 1$. In order that $Q(0) = 0$, the integral defining Q must diverge at zero. Since $V^*(r) \leqq 2r$, this can only happen if $a = 0$, since $1/(r+a) \geqq 2/[(m-1)V^*(r) + 2(r+a)] \geqq 1/(Ar+a)$. Thus

$$Q(t) = \exp\{-\int_t^1 2\,dr/[2r + (m-1)V^*(r)]\}$$

makes the expression inside the brackets in (9.2) stationary. Substituting this in (9.2) (note that $q(t) = 2Q/[2t + (m-1)V^*]$) yields

$$I^*(m) = \ln m + \int_0^1 \{(m-1)v^*(t) \ln Q + 2 \ln 2Q - 2 \ln (2t + (m-1)V^*)\}\, dt$$
$$= \ln 4m - \int_0^1 \int_t^1 \{2[(m-1)v^*(t) + 2]/[(m-1)V^*(r) + 2r]\}\, dr\, dt$$
$$- 2 \int_0^1 \ln [(m-1)V^*(t) + 2t]\, dt.$$

Upon changing the order of integration, the middle term is seen to be 2 and $I^*(m) = \ln 4m - 2 - 2\int_0^1 \ln (2t + (m-1)V^*)\, dt$. Now, substituting $t = H^*(x)$ and recalling that $V^*H^* = G^*$, we obtain $I^*(m) = \ln 4m - 2 - \int \ln (F^* + mG^*)\, d(F^* + G^*)$, as given in (12) of Savage and Sethuraman (1966), with our m corresponding to their A.

REFERENCES

[1] Bradley, Ralph A. (1967). Topics in rank-order statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 609–638. Univ. of California Press.

[2] Bradley, Ralph A., Merchant, Sarla D. and Wilcoxon, Frank. (1966). Sequential ranks tests II. Modified two-sample procedures. *Technometrics* **8** 615–623.

[3] Govindarajulu, Z., LeCam, L. and Raghavachari, M. (1967). Generalizations of theorems of Chernoff and Savage on the asymptotic normality of test statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 593–607. Univ. of California Press.

[4] Hall, W. J., Wijsman, R. A. and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36** 575–614.

[5] KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York.
[6] LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23–43.
[7] PYKE, R. and SHORACK, G. (1968). Weak convergence of a two-sample empirical process and a generalized Chernoff-Savage theorem. *Ann. Math. Statist.* **39**, 755–771.
[8] SAVAGE, I. R. and SETHURAMAN, J. (1966). Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives. *Ann. Math. Statist.* **37** 1154–1160.
[9] SETHURAMAN, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316.
[10] SETHURAMAN, J. (1967). Stopping time of a rank order sequential probability ratio test based on Lehmann alternatives II. Technical report No. 4/67. Indian Statist. Inst., Calcutta.