The Annals of Statistics
1977, Vol. 5, No. 2, 330–341

WEAK CONVERGENCE OF PROCESSES RELATED TO LIKELIHOOD RATIOS[1]

BY W. J. HALL[2] AND R. M. LOYNES

University of Rochester and Sheffield University

Material in Chapter VI of Hájek and Šidák's book is extended to a sequential analysis setting: conditions are given under which a sequence of *log-likelihood-ratio processes* (log-likelihood-ratios for sequential sampling, represented as jump processes in continuous time) converges weakly to a Wiener process with drift, the drift parameter depending on which hypothesis, in a suitable neighorhood of a null hypothesis, prevails. Conditions for convergence of other "test statistic" processes, related to likelihood ratios, are also given. Asymptotic sequential tests can thereby be constructed. Some "two-sample problem" examples are treated.

1. Sequential analysis motivation. In order to construct tests of hypotheses from "large samples," the asymptotic distribution of a suitable test statistic is needed when the null hypothesis is true. For a more complete description of the behavior of such tests, for example to approximate the power or to study the asymptotic efficiency of one test relative to another, the asymptotic distribution is needed under alternative hypotheses. To get beyond generalities the problem must be specified more closely, and this can be done in various ways; we choose the Pitman approach, in which the null hypothesis and error probabilities are fixed, but the alternative hypothesis tends towards the null, thereby ensuring large samples: then power can be investigated in the neighborhood of the null hypothesis.

An elegant and powerful way of structuring this approach was introduced by Le Cam (1960), and utilized in the book of Hájek and Šidák (1967) (hereafter H–S), using the notion of *contiguous hypotheses*. Roughly speaking, it goes like this:

Consider a sequence of specific alternative hypotheses, and a simple null hypothesis H_0, so chosen that the log-likelihood-ratio statistic $\log L_n$ is asymptotically normal under H_0 with mean equal to minus one half of the variance, as the sample size n tends to infinity. Let us say the hypotheses are then "close." (According to Le Cam, this implies the weaker concept of "contiguity." This happens, for example, in one-parameter exponential families if the parameter under the alternative tends to the parameter under H_0 at the rate $n^{-\frac{1}{2}}$.) This log-likelihood-ratio statistic is of course suitable, even optimal, for hypothesis testing.

Received September 1975; revised July 1976.

[1] Supported by the U.S. Army Research Office and the U.K. Scientific Research Council.

[2] Part of this material was presented in a Special Invited Address at the 37th Annual Meeting in Edmonton, August 1974.

AMS 1970 *subject classifications.* Primary 62L10; Secondary 60F05, 62E20.

Key words and phrases. Weak convergence, sequential analysis, likelihood ratios, Le Cam's lemmas, two-sample problem, sign test.

Now consider another test statistic S_n. Typically, if it is to be comparable to $\log L_n$, it too must be asymptotically normal under H_0. Suppose slightly more: namely, suppose S_n and $\log L_n$ are asymptotically bivariate normal, rather than just marginally asymptotically normal. (The Cramér–Wold device, as described in Billingsley (1968), is the usual tool, requiring only consideration of the one-dimensional distributions of linear combinations.) Le Cam then showed that S_n and $\log L_n$ must also be asymptotically bivariate normal under the sequence of alternatives; moreover, the asymptotic second-order central moments are the same, the asymptotic mean of S_n is shifted by the covariance, and the asymptotic mean of $\log L_n$ has its sign changed. Thus, asymptotic marginal distributions under the alternatives are inferred from those under H_0. If S_n is standardized to have the same H_0-limit as $\log L_n$, then it also has the same limit under the sequence of alternatives except its mean is possibly smaller—by the amount $(1 - \rho)\sigma^2$, where ρ and σ^2 are the correlation and variance in the bivariate normal limit distribution. Thus, if $\rho = 1$, S_n and $\log L_n$ have the same asymptotic distribution under both hypotheses. Otherwise, ρ^2 turns out to be a convenient index of asymptotic efficiency (*Pitman efficiency*), of a test based on S_n relative to one based on $\log L_n$.

Our goal is to extend all of this to a sequential analysis setting. A SPRT, as a typical example, is based on the sequence $\{\log L_n\}$, with fixed stopping barriers, n being the sample size. If we are to consider asymptotic behavior we shall have a sequence of alternative hypotheses, indexed by an integer ν. (In the fixed sample-size situation discussed above, ν and n could in effect be identified, but here this simplification is no longer possible.) Thus we have a sequence $\{L_{\nu n}\}$ for each ν. Our method of treating the problem is to set each sequence $\{L_{\nu n}\}$ into continuous time by writing $L_\nu(t) = L_{\nu n}$ at $n = [\nu t]$ (later in the text, for greater generality, we write $n = [n_\nu t]$ for some increasing n_ν). Then frequently $\{\log L_{\nu n}\}$ (or more precisely $\log L_\nu$) can be shown to behave asymptotically ($\nu \to \infty$) like a Wiener process with drift (under H_0), and if the alternative hypotheses are "close" to the null, the drift parameter may be expected to be minus one half the variance parameter; the change to a sequential setting has forced us to deal with random processes rather than random variables. If we now have another sequence $\{S_n\}$ (or more precisely $\{S_{\nu n}\}$) on which we may want to base a sequential test, it too may, when similarly embedded in continuous time, behave asymptotically like a Wiener process with drift (under H_0). By extending the Le Cam theory, we show that if $\{S_n, \log L_n\}$ behaves jointly like a "bivariate Wiener process" under H_0 then it will do likewise under the sequence of alternatives, with different drift parameters, paralleling the results reviewed above. Hence a test for the drift of a Wiener process can be used to provide a sequential test based on $\{S_n\}$, whose asymptotic properties are those derived from Wiener process theory. Such asymptotic sequential analysis is described more fully in Hall (1975).

In this reformulation in continuous time, it is enough (for proving limit theorems) to consider finite time intervals $[0, K]$ for a sequence of K-values tending to infinity; the processes are then in the function space $D[0, K]$ (or $D[0, \infty)$—see Appendix). It turns out, in developing the theory summarized above, that it is enough to consider $(S_{[\nu t]}, \log L_{[\nu K]})$, a "process" in $D[0, K] \times R_1$, rather than the bivariate process $(S_{[\nu t]}, \log L_{[\nu t]})$ in $D[0, K] \times D[0, K]$. The corresponding limit process, with one "marginal" being a Wiener process and the other a normal random variable (with appropriate "joint" structure), we call a *W-N process*. These processes, and their role as weak limits of processes related to likelihood ratios, are described in Section 2. This material constitutes a natural extension of *Le Cam's third lemma* (described in H–S, and in Hall and Loynes, 1977—hereafter H–L-I). *Le Cam's second lemma* is extended to this stochastic process setting in Section 3: it is useful for verifying the hypotheses (convergence under H_0) in Le Cam's third lemma.

Armed with this machinery, the weak convergence of the *log-likelihood-ratio process* is verified in Section 4, in the case of an H_0 hypothesis specifying i.i.d. observations and an alternative specifying various location shifts ("regression in location," H–S); weak convergence is also verified under other "nearby" regression alternatives.

Sections 5 and 6 present applications to the two-sample problem, both *parametric* and *sign-test* versions. Another paper by us (referred to as H-L-III) will present *linear rank statistic* applications. A *t-test* application appears in Hall (1973).

Some topological and weak convergence questions are summarized in the Appendix.

2. Wiener-normal processes and convergence to them. In this section we define certain stochastic processes with paths in spaces such as $D[0, \infty) \times R_1$, and this enables use of an extension of *Le Cam's third lemma* (H–S) to prove that various processes related to log-likelihood-ratios converge to Wiener processes with drift.

We first consider *bivariate Wiener processes* $\mathbf{Z}$ on $C[0, \infty)^2$ or $C[0, K]^2$. For each t, label the coordinates $Z_1(t)$ and $Z_2(t)$. The process $\mathbf{Z}$ has independent increments and multinormal finite-dimensional distributions. Also $EZ_1(t) = \mu t$, $EZ_2(t) = \lambda t$, and the covariance matrix of $\mathbf{Z}(t)$ is Σt where

$$\Sigma = \begin{pmatrix} \tau^2 & \gamma \\ \gamma & \sigma^2 \end{pmatrix} \quad \text{(written hereafter as } \Sigma = \Sigma(\tau, \gamma, \sigma)) ;$$

$\mathbf{Z}(0) = (0, 0)$.

Now consider $\mathbf{Z}$ on $C[0, K]^2$, relabel the first coordinate Z (dropping the subscript 1) and replace the second coordinate by $Z^K = Z_2(K)$. We refer to (Z, Z^K) as a *W-N* (*Wiener-normal*) *process* in $C[0, K] \times R_1$, with parameters $(\mu, \lambda; \Sigma)$; note that Z is a Wiener process W_{μ,τ^2} with drift μ and variance τ^2, Z^K is a $N(K\lambda, K\sigma^2)$ rv, and $\operatorname{Cov}(Z(t), Z^K) = \gamma t$. The process is characterized by the joint distribution of $Z(t_1), \cdots, Z(t_k), Z^K$ for every k and $0 \leqq t_1 < \cdots < t_k \leqq K$;

it is multinormal with means $\mu t_1, \cdots, \mu t_k, \lambda K$ and covariance matrix with (i, j)-element $\tau^2 \min(t_i, t_j)$ for i and $j \leq k$, and with $(k+1)$th row $\gamma t_1, \gamma t_2, \cdots, \gamma t_k, \sigma^2 K$.

We prove weak convergence of various processes (Z_ν, Z_ν^K) in $D[0, K] \times R_1$ to (Z, Z^K), for a sequence of K's $\uparrow \infty$. It is sufficient to prove weak convergence of Z_ν to Z in $D[0, K]$, weak convergence of Z_ν^K to Z^K, and convergence of the finite-dimensional distributions of (Z_ν, Z_ν^K) to those of (Z, Z^K) (because "joint tightness" follows from "marginal tightness"—see Billingsley (1968), page 41, Problem 6).

We also need the following:

LEMMA 1. *If (Z, Z^K) is W-$N(\mu, \lambda; \Sigma(\tau, \gamma, \sigma))$ on $C[0, K] \times R_1$ under P with $\lambda = -\frac{1}{2}\sigma^2$, and if $dQ = \exp(Z^K)\, dP$, then (Z, Z^K) is W-$N(\mu + \gamma, -\lambda; \Sigma)$ under Q.*

PROOF. First note that Q is a probability measure since $\int dQ = \exp(\lambda K + \frac{1}{2}\sigma^2 K) = 1$. It is now sufficient to evaluate the characteristic function of $Z(t_1), \cdots, Z(t_k), Z^K$ under Q and see that it is consistent with the conclusion of the lemma.

Now let $X_1, X_2, \cdots$ be rv's (real- or vector-valued) on a fixed measurable space $(\Omega, \mathscr{A})$, and write $\mathscr{A}_n$ for the subfield generated by the first n X's. Let $\{\mu_\nu\}$ be a sequence of measures thereon and write $\mu_{\nu n}$ for the restriction to $\mathscr{A}_n$. Let P_ν, Q_ν and R_ν be absolutely continuous probability measures specifying densities (w.r.t. $\mu_{\nu n}$) for $X_1, \cdots, X_n$ denoted by $p_{\nu n}$, $q_{\nu n}$, and $r_{\nu n}$, respectively, and write $L_{\nu n} = q_{\nu n}/p_{\nu n}$ and $M_{\nu n} = r_{\nu n}/p_{\nu n}$ (with $L_{\nu n} = 1$ if $q_{\nu n} = p_{\nu n} = 0$ or $n = 0$, $= \nu$ if $q_{\nu n} > p_{\nu n} = 0$ and $n > 0$, and $= 1/\nu$ if $p_{\nu n} > q_{\nu n}$ and $n > 0$, and similarly for $M_{\nu n}$). Let $S_{\nu n}$ be an $\mathscr{A}_n$-measurable (real) function. Finally, let $\{n_\nu\}$ be a monotone sequence of integers increasing to ∞ with ν. (Ω, $\mathscr{A}$ and X_i could also have suffices ν with minor changes in what follows.)

We set various (real) $\mathscr{A}_n$-measurable functions in continuous time as follows, as elements of $D[0, \infty)$: L_ν at t is $L_\nu(t) = L_{\nu n}$ at $n = [n_\nu t]$; write L_ν^K for the restriction of L_ν to $D[0, K]$; likewise for M_ν, S_ν, etc. The *log-likelihood-ratio process* (Q_ν to P_ν) is then defined as $\log L_\nu$ in $D[0, \infty)$, representing at time t the log-likelihood-ratio of the first $[n_\nu t]$ X's. Write P_ν^K for the restriction of P_ν to $\mathscr{A}_{[n_\nu K]}$, etc.

An extended form of what H–S call *Le Cam's third lemma* appeared as Theorem 2 in H–L-I; part of that theorem in the present setting (using Lemma 1) is:

THEOREM 1. *If $(S_\nu^K, \log L_\nu(K)) \Rightarrow W$-$N(\mu, \lambda; \Sigma)$ in $D[0, K] \times R_1$ under $\{P_\nu\}$ where $\Sigma = \Sigma(\tau, \gamma, \sigma)$, and if $\lambda = -\frac{1}{2}\sigma^2$, then $(S_\nu^K, \log L_\nu(K)) \Rightarrow W$-$N(\mu + \gamma, -\lambda; \Sigma)$ under $\{Q_\nu\}$.*

A version of this theorem with the process L_ν rather than the rv $L_\nu(K)$ is possible but unnecessary here. Indeed, W–N convergence under $\{Q_\nu\}$ almost requires the hypothesis of the theorem, and mutual contiguity of $\{P_\nu^K\}$ and $\{Q_\nu^K\}$ (but not of $\{P_\nu\}$ and $\{Q_\nu\}$) can also be inferred; see H–L-I.

Taking S_ν to be $\log L_\nu$, and eliminating the redundancy, we obtain a corollary (analogous to part of Corollary 1 in H–L-I and to Theorem 7.2 in Roussas (1972)):

COROLLARY 1. *If* $\log L_\nu \Rightarrow W_{\delta,\sigma^2}$ *in* $D[0,\infty)$ *under* $\{P_\nu\}$ *with drift parameter* $\delta = -\frac{1}{2}\sigma^2$ *for some* $\sigma \geqq 0$, *then it converges under* $\{Q_\nu\}$ *to* $W_{-\delta,\sigma^2}$.

As a second immediate corollary, we have

COROLLARY 2. *If* $(\log L_\nu^K, \log M_\nu(K)) \Rightarrow W\text{-}N(-\frac{1}{2}\tau^2, -\frac{1}{2}\sigma^2; \Sigma(\tau,\gamma,\sigma))$ *in* $D[0,K] \times R_1$ *under* $\{P_\nu\}$ *for every* K, *then* $\log L_\nu \Rightarrow W_{\gamma-\frac{1}{2}\tau^2,\tau^2}$ *in* $D[0,\infty)$ *under* $\{R_\nu\}$.

In sequential analysis applications, Corollary 1 enables construction of tests of P_ν vs. Q_ν, based on $\log L_\nu$, with asymptotic properties determinable under both hypotheses; Corollary 2 enables determination of asymptotic properties under another hypothesis R_ν. Of necessity (see H–L-I), these various hypotheses are (essentially mutually) contiguous.

An example of Corollaries 1 and 2, and Theorem 1, is provided by considering exponential family rv's with parameters, within order $n_\nu^{-\frac{1}{2}}$ of each other (as in Hall (1975)). Specifically, let the parameter be θ_0 under $P_\nu = P$, $\theta_0 + \delta/\nu^{\frac{1}{2}}$ under Q_ν, and $\theta_0 + c\delta/\nu^{\frac{1}{2}}$ under R_ν, with the origin so chosen that $\mathscr{E}_P X = 0$, and writing $\mathscr{E}_P X^2 = \sigma^2$ $(n_\nu = \nu)$. It may be shown that $(\log L_\nu^K, \log M_\nu(K)) \Rightarrow W\text{-}N((\lambda - \frac{1}{2})\sigma^2, -\frac{1}{2}c^2\sigma^2; \sigma^2\Sigma(1, c, |c|))$ under P (with $\lambda = 0$), by Donsker's theorem. This implies convergence of $\log L_\nu$ under Q_ν $(\lambda = 1)$ by Corollary 1 and under R_ν $(\lambda = c)$ by Corollary 2. Letting $S_{\nu n}$ be the *sign test statistic* defined in Hall (1975)—specifically, $S_{\nu n} = \nu^{-\frac{1}{2}} \sum_{i=1}^n \operatorname{sgn}(X_i - m)$ where m is a median (but not an atom) under P—Donsker's theorem and the Cramér–Wold device enable verification of the hypothesis of Theorem 1; we thus obtain the conclusion that $S_\nu \Rightarrow W_{\lambda\rho\sigma,1}$ under P, Q_ν, and R_ν (with $\lambda = 0$, 1 and c, respectively) where $\rho = \operatorname{corr}(\operatorname{sgn}(X - m), X)$ under P. Implications in sequential analysis are discussed there.

Further applications appear in Sections 5 and 6 below, in Hall (1973) and in H–L-III; all are in the spirit of applications of Le Cam's third lemma in H–S.

We mention one final fact here. By proving the joint weak convergence of $\log L_\nu$ and $\log M_\nu$ to a bivariate Wiener process under P_ν (as an extension of Theorem 1 or Corollary 2), the joint weak convergence under both Q_ν and R_ν can be concluded. By differencing the two coordinate processes, we thus obtain the weak convergence of the log-likelihood-ratio process of Q_ν to R_ν under both Q_ν and R_ν. Thus, a sequential test of Q_ν vs. R_ν could be constructed and evaluated, only carrying out convergence proofs under P_ν.

3. Le Cam's second lemma. The hypotheses of the theorems and corollaries of Section 2 may be verified in special cases using Donsker's theorem or a theorem of Loynes (1976). These latter theorems require second moments (or moments of order one, at least). Some of the variables to which we shall apply the previous section will, however, be log-likelihood-ratios, which will not in general have such moments. It is therefere convenient to extend *Le Cam's second lemma* (H–S, page 205) to cover weak convergence of processes.

Consider an infinite sequence $X_1, X_2, \cdots$ of rv's on a fixed measurable space $(\Omega, \mathscr{A})$, as before. According to P_ν, they are independent with densities (Lebesgue measure) $f_{\nu 1}, f_{\nu 2}, \cdots$, and according to Q_ν they are independent with densities $g_{\nu 1}, g_{\nu 2}, \cdots$. The whole development parallels H–S very closely, and we begin by writing

$$Z_{\nu n} = 2 \sum_{i=1}^{n} \{[g_{\nu i}(X_i)/f_{\nu i}(X_i)]^{\frac{1}{2}} - 1\} \tag{1}$$

with its continuous-time analog $Z_\nu(t) = Z_\nu(t)$ at $n = [n_\nu t]$. The *uniform asymptotic negligibility* of the summands needs slight strengthening to

$$\lim_{\nu \to \infty} \max_{1 \leqq i \leqq n_\nu K} P_\nu \left(\left| \frac{g_{\nu i}(X_i)}{f_{\nu i}(X_i)} - 1 \right| > \varepsilon \right) = 0 \qquad \text{for every} \quad K . \tag{2}$$

We then have

LEMMA 2. *If* (2) *holds and* $Z_\nu \Rightarrow W_{-\frac{1}{4}\sigma^2, \sigma^2}$ *in* $D[0, \infty)$ *under* $\{P_\nu\}$ *for some* $\sigma \geqq 0$, *then* $P_\nu(\sup_{0 \leqq t \leqq K} |\log L_\nu(t) - Z_\nu(t) + \frac{1}{4}\sigma^2 t| > \varepsilon) \to 0$ *for every* K *and* $\varepsilon > 0$, *and* $\log L_\nu \Rightarrow W_{-\frac{1}{2}\sigma^2, \sigma^2}$ *in* $D[0, \infty)$ *under* $\{P_\nu\}$.

The first part may be proved in parallel to H–S, taking care that convergence is uniform for $t \in [0, K]$, and the second part follows from the first.

4. Weak convergence of log-likelihood-ratio processes. We now apply the results of previous sections to obtain conditions for the weak convergence of log-likelihood-ratio processes, the null hypothesis specifying i.i.d. observations and the (contiguous) alternatives specifying location shifts (or *regression in location*, H–S). The treatment parallels that of Section VI.2.1 of H–S, with modest changes. Our rv's X are now real.

ASSUMPTION A. F is an absolutely continuous df with density f with finite Fisher information $I(f)$.

Write $\phi(u) = \phi(u, f) = -f'[F^{-1}(u)]/f[F^{-1}(u)]$ so that $\int \phi(u)\,du = 0$ (H–S, middle of page 17 and Lemma I.2.4a) and $I = \int \phi(u)^2\,du < \infty$.

For given f, consider a sequence of *location alternatives* $\{Q_\nu\}$; Q_ν is a probability measure on $(\Omega, \mathscr{A})$ according to which $X_1, \cdots, X_n$ have joint density

$$q_{\nu n} = \prod_{i=1}^{n} f(x_i - d_{\nu i}) \qquad n = 1, 2, \cdots .$$

The following assumption is made about the $d_{\nu i}$ (limits are as $\nu \to \infty$):

ASSUMPTION B. For some d_ν, some positive integers $n_\nu \to \infty$, some $\delta > 0$, and for each positive t,

(i) $\max_{1 \leqq i \leqq n_\nu t} (d_{\nu i} - d_\nu)^2 \to 0$;
(ii) $\bar{d}_{\nu t} \equiv n^{-1} \sum_{i=1}^{n} d_{\nu i}(\text{at } n = [n_\nu t]) = d_\nu + o(n_\nu^{-\frac{1}{2}})$;
(iii) $\sum_{i=1}^{n} (d_{\nu i} - d_\nu)^2(\text{at } n = [n_\nu t]) \to \delta^2 t$.

Write $d'_{\nu i} = d_{\nu i} - d_\nu$, and $\sigma = \delta I^{\frac{1}{2}}$. (H–S require d_ν to be our $\bar{d}_{\nu t}$, but they only need consider n_ν (their N_ν) X's whereas we have an infinite sequence. Our n_ν is simply an index which could be replaced by ν; it has no special interpretation. Also, by rescaling, δ could be taken to be unity.)

Let P_ν be a probability measure on $(\Omega, \mathscr{A})$ according to which $X_1, \cdots, X_n$ have joint density

$$p_{\nu n} = \prod_{i=1}^n f(x_i - d_\nu) \qquad n = 1, 2, \cdots .$$

Thus, P_ν is consistent with the null hypothesis of i.i.d. observations, and it may be shown that $\{p_{\nu n_\nu}\}$ and $\{q_{\nu n_\nu}\}$ are mutually contiguous. Obviously, P_ν depends on ν only through the location parameter d_ν.

Let $L_{\nu n} = q_{\nu n}/p_{\nu n}$ (with the same conventions as in Section 2 when $p_{\nu n}$ or $q_{\nu n}$ or $n = 0$). Let $\log L_\nu(t) = \log L_{\nu n}$ at $n = [n_\nu t]$; $\log L_\nu$ is the *log-likelihood-ratio process*, clearly in $D[0, \infty)$ for each ν. (Alternatively, and asymptotically equivalently, we could set $L_\nu(t)$ equal to $L_{\nu n}$ when $t = \sum_{i=1}^n d_{\nu i}'^2$, but equal spacing of observations seems more natural when motivated by sequential analysis.)

The main result of this section is the following theorem, giving sufficient conditions for the convergence of the log-likelihood-ratio process.

THEOREM 2. *Under Assumptions A and B, and with* $\sigma^2 = \delta^2 I(f)$, $\log L_\nu \Rightarrow W_{-\frac{1}{2}\sigma^2, \sigma^2}$ *under* $\{P_\nu\}$ *and* $\log L_\nu \Rightarrow W_{\frac{1}{2}\sigma^2, \sigma^2}$ *under* $\{Q_\nu\}$ *in* $D[0, \infty)$.

Note that the limit behavior depends only on $\delta^2 I(f)$: thus the $d_{\nu i}$ enter only through δ, and f only through $I(f)$. It may also be noted in passing that the theorem implies as a corollary that $\{p_{\nu n_\nu}\}$ and $\{q_{\nu n_\nu}\}$ are mutually contiguous (H-L-I).

PROOF. This is analogous to Theorem VI.2.1 of H-S, and our proof closely parallels theirs: we introduce a process Y_ν which is easily studied, and by this means show that the hypotheses of Lemma 2 are satisfied. Application of Corollary 1 of Section 2 then completes the proof.

Let

$$(3) \qquad Y_{\nu n} = \sum_{i=1}^n d_{\nu i}' \phi(U_i) , \qquad \text{where} \quad U_i = F(X_i - d_\nu) ,$$

and define the process Y_ν by $Y_\nu(t) = Y_{\nu n}$ at $n = [n_\nu t]$. Under P_ν the U_i are independent and uniformly distributed on [0, 1]. (Our $Z_{\nu n}$ and $Y_{\nu n}$ (defined in (1) and (3)) corresponds to W_d and T_d of H-S VI.2.1 (3) and (13).)

Condition (2) is a consequence of Assumption B, and it is therefore sufficient (by Lemma 2) to show that under P_ν, $Z_\nu \Rightarrow W_{-\frac{1}{4}\sigma^2, \sigma^2}$ in $D[0, K]$ for every K, and this follows easily from the corollary in Loynes (1976). His conditions (i) and (ii) are easy. Conditions (iii) and (iv) with Y_ν replacing Z_ν (W_{0, σ^2} in (iv)) are easily verified, the latter by using the Cramér-Wold device and Theorem V.1.2 of H-S; Lemma VI.2.1b of H-S and Lemma 3 below then allow one to conclude that conditions (iii) and (iv) are also satisfied for Z_ν, completing the proof.

To obtain asymptotic behavior under other (contiguous) alternatives, suppose

ASSUMPTION B(iv). e_ν, $e_{\nu i}$ and ε satisfy Assumptions B(i) (ii) (iii) (replacing d_ν, $d_{\nu i}$ and δ), for the same n_ν's; and

$$\text{B(v)}: \quad \text{for each} \quad t\,, \quad \textstyle\sum_{i=1}^{n} d'_{\nu i} e'_{\nu i} \quad (\text{at } n = [n_\nu t]) \to \eta t \quad (|\eta| \leqq \delta\varepsilon)\,.$$

Let R_ν be the probability measure corresponding to Q_ν but with the $d_{\nu i}$'s replaced by $e_{\nu i}$'s. Proceeding very much as in H–S (Section VI.2.4) and above, we obtain from Corollary 2 of Section 2

THEOREM 3. *Under Assumptions* A *and* B(i)—(v), *and with* $\sigma^2 = \delta^2 I(f)$ *and* $\gamma = \eta I(f)$,

$$\log L_\nu \Rightarrow W_{\gamma - \frac{1}{2}\sigma^2, \sigma^2} \quad \text{under} \quad \{R_\nu\} \quad \text{in} \quad D[0, \infty)\,.$$

Simple sufficient conditions for B(i)—(v) are:

ASSUMPTION B′. For some $\delta > 0$, $\varepsilon > 0$ and some η, and bounded sequences $\{d_i''\}$ and $\{e_i''\}$,

(a) for arbitrary d_ν, $d_{\nu i} = d_\nu + d_i''/\nu^{\frac{1}{2}}$, $n_\nu = \nu$,

$$n^{-1} \textstyle\sum_{i=1}^{n} d_i'' \to 0 \qquad \text{and} \qquad n^{-1} \sum_{i=1}^{n} d_i''^2 \to \delta^2 \quad \text{as} \quad n \to \infty\,;$$

(b) (a) holds with d's and δ replaced by e's and ε;

(c) $n^{-1} \sum_{i=1}^{n} d_i'' e_i'' \to \eta$ as $n \to \infty$.

According to the above results weak convergence of log-likelihood-ratios is guaranteed under Assumptions A and B. Joint finite-dimensional distributions with other statistics need to be evaluated from time to time, however, and it is then often convenient to apply the following lemma (essentially implied by H–S Theorem VI.2.1 and Lemma VI.2.1.b): it implies that for such purposes any one of $\log L_\nu(t)$, $Y_\nu(t) - \frac{1}{2}\sigma^2 t$, $Z_\nu(t) - \frac{1}{4}\sigma^2 t$ may be replaced by another.

LEMMA 3. *Under Assumptions* A *and* B(i)—(iii), *for each t the quantities* (a), (b), (c) *differ by amounts which tend to* 0 *in* P_ν*-probability*:

$$\text{(a)} \quad \log L_\nu(t)\,; \qquad \text{(b)} \quad Y_\nu(t) - \tfrac{1}{2}\sigma^2 t\,; \qquad \text{(c)} \quad Z_\nu(t) - \tfrac{1}{4}\sigma^2 t\,.$$

Stronger results are true (the difference between (b) and (c) tends to 0 in mean square, and all differences of the corresponding processes converge weakly to 0), but do not seem of great value.

5. The two-sample problem—parametric version. Observations, the X's, now come (independently) from one of two populations; we say they are Y's or Z's respectively. Let p_n be the proportion of Y's among the first n observations and $q_n = 1 - p_n$. We assume that p_n does not depend on the values of the oberva-tions and $p_n \to p$ for some $p \in (0, 1)$; $q = 1 - p$.

According to $P_\nu = P$, all observations have a specified density f, satisfying Assumption A of Section 4; according to Q_ν, the Y's have density $f(x + qd)$ and the Z's $f(x - pd)$, and we set $d = \delta(\nu pq)^{-\frac{1}{2}}$ (δ fixed). Also, consider R_ν, analog-ous to Q_ν but with d replaced by cd for some $c \neq 0$. Then Q_ν and R_ν are loca-tion shift alternatives, with population means differing by d and cd, respectively.

Assumption B′ is readily verified, with $\varepsilon^2 = c^2\delta^2$ and $\eta = c\delta^2$ (and $d_\nu = 0$, $d_i'' = -\delta(q/p)^{\frac{1}{2}}$ or $\delta(p/q)^{\frac{1}{2}}$, and $e_i'' = cd_i''$).

Under these assumptions, Theorems 2 and 3 imply that the log-likelihood-ratio-process for Q_ν to P converges weakly to a Wiener process with drift $(\lambda - \frac{1}{2})\sigma^2$ and variance $\sigma^2 = \delta^2 I(f)$, with $\lambda = 0$ under P, $=1$ under Q_ν, and $=c$ under R_ν.

This result is not of much practical interest in this form since the P-hypothesis is simple: f needs to be completely specified in order to calculate the log-likelihood-ratio statistic. In the remainder of this section, we assume f to be $N(\mu, \sigma^2)$ and show how the various nuisance parameters can be estimated. A second version of the problem, invariant under common shifts, is given in the next section. Variations, based on linear rank statistics, appear in H–L–III.

Assuming now that f is a $N(\mu, \sigma^2)$ density ($I = \sigma^{-2}$), we find, after n' Y's and n'' Z's ($n = n' + n''$),

$$\log L_{\nu n} = -n'qd\sigma^{-2}(\bar{Y}_{n'} - \mu + \tfrac{1}{2}qd) + n''pd\sigma^{-2}(\bar{Z}_{n''} - \mu - \tfrac{1}{2}pd)$$

where $\bar{Y}_{n'}$ and $\bar{Z}_{n''}$ are the respective sample averages. Setting this in continuous time ($n = [\nu t]$), we have a Wiener process limit with drift $(\lambda - \frac{1}{2})\delta^2/\sigma^2$ and variance δ^2/σ^2. If μ is unknown (but σ and p known), we now replace it by a pooled estimate $\mu_n = p\bar{Y}_{n'} + q\bar{Z}_{n''}$. Also, we delay "start-up" of the process (a device which enables collection of sufficient data to estimate parameters well (Hall, 1975)) and thus define, for some $C > 0$,

$$\begin{aligned} S_{\nu n} &= I(n \geqq [C\nu^{\frac{1}{2}}]) \cdot \log L_{\nu n}|_{\mu=\mu_n} \\ &= I(n \geqq [C\nu^{\frac{1}{2}}]) \cdot (n'q^2 + n''p^2)d\sigma^{-2}(\bar{Z}_{n''} - \bar{Y}_{n'} - \tfrac{1}{2}d)\,. \end{aligned}$$

Now let $D_\nu(t) = |\log L_{\nu n} - S_{\nu n}|$ at $n = [\nu t]$ and $J_\nu = \{t \,|\, C/\nu^{\frac{1}{2}} \leqq t \leqq K\}$ for fixed K. Then $D_\nu \equiv \sup_{J_\nu} D_\nu(t) = \sup |\mu_n - \mu| \cdot |n''p - n'q|\, d\sigma^{-2}$ where $d = \delta/(\nu pq)^{\frac{1}{2}}$, $n' = np_n$, $n'' = nq_n$ and $n = [\nu t]$. We find

$$D_\nu \leqq A \sup |p_n - p| \cdot \{p\nu^{\frac{1}{2}}| \sup \bar{Y}_{n'} - \mu| + q\nu^{\frac{1}{2}} \sup |\bar{Z}_{n''} - \mu|\}$$

where $A = \delta K\sigma^{-2}(pq)^{-\frac{1}{2}}$. Since the term in curly brackets can be shown to converge weakly to a rv, and since $\sup_{J_\nu} |p_n - p| = \max_{[C\nu^{\frac{1}{2}}] \leqq n \leqq \nu K} |p_n - p| \to 0$ as $\nu \to \infty$, we have that $D_\nu \to 0$ in probability, for every K. This is sufficient, by Lemma A1 (with $\mu = 0$ and $V_\nu = \theta = 1$), to assure that $\log L_\nu$ and S_ν have the same weak limit under P.

By considering the finite-dimensional distributions of $(S_\nu^K, \log L_\nu(K))$, we can now verify that the hypothesis of Theorem 1 holds with $\sigma^2 = \tau^2 = \gamma$. Hence, S_ν and $\log L_\nu$ have the same weak limit under Q_ν also. Replacing L_ν above by M_ν, the same holds under R_ν. Hence, $S_{\nu n}$ may be used for constructing sequential tests as if it were $\log L_{\nu n}$, without affecting the asymptotic properties under P, Q_ν or R_ν.

If p, q and σ^2 are also unknown, they may be estimated by p_n, q_n and S_n^2 (the usual pooled estimate) for n not too small. By similar methods, based on

Lemma A1, it can be shown that $S_{\nu n}$ with p, q and σ^2 replaced by these estimates has the same asymptotic behavior as $\log L_{\nu n}$. Alternatively, the methods of Hall (1975) may be applied directly to the sequence $S_{\nu n}$ (with estimates throughout).

6. The two-sample problem—sign test version. We now suppose that one observation is taken from each population at each stage, and let $X_n = (-1)^n(Z_n - Y_n)$. We confine attention to procedures based on the sequence of X's which are invariant under common location shifts.

Under $P_\nu = P$, Y_n and Z_n have identical densities $g(x - \theta_n)$ (θ_n unspecified) so that X_n has density $f(x) = \int g(x + u)g(u)\,du$. Under Q_ν, Y_n has density $g(x - \theta_n)$ and Z_n density $g(x - \theta_n - \Delta_\nu)$, so that X_n has density $f(x - d_{\nu n})$ with $d_{\nu n} = (-1)^n\Delta_\nu$; thus Δ_ν is the hypothesized *location shift*, and it will be convenient to write $\Delta_\nu = \delta/\nu^{\frac{1}{2}}$. Define R_ν similarly to Q_ν, with $d_{\nu n}$ replaced by $cd_{\nu n}$. We assume Assumption A for f; Assumption B′ is readily verified ($d_n'' = (-1)^n\delta$, $\varepsilon = c\delta$, $\eta = c\delta^2$).

Thus, by Theorems 2 and 3, the log-likelihood-ratio process converges to a Wiener process with drift $(\lambda - \frac{1}{2})\sigma^2$ and variance $\sigma^2 = \delta^2 I(f)$, under P, Q_ν and R_ν (with $\lambda = 0$, 1 and c).

We now proceed to give a *sign-statistic* analog of the above. Let $S_n = \operatorname{sgn}(Z_n - Y_n) = (-1)^n \operatorname{sgn} X_n$, $S_{\nu n} = \nu^{-\frac{1}{2}} \sum_{i=1}^n S_i$, and $S_\nu(t) = S_{\nu n}$ at $n = [\nu t]$. Now $S_\nu \Rightarrow W$ under P (by Donsker's theorem); the finite-dimensional distributions of $(S_\nu^K, \log L_\nu(K))$ converge to those of W–$N(0, -\frac{1}{2}\sigma^2; \Sigma(1, \gamma, \sigma))$ where $\gamma = \operatorname{cov}_F(\operatorname{sgn} X, -\delta f'(X)/f(X)) = -\delta \int \operatorname{sgn} x \cdot f'(x)\,dx = 2\delta f(0) = 2\delta \int g(x)^2\,dx$. Hence, Theorem 1 applies and $S_\nu \Rightarrow W_{\gamma,1}$ under Q_ν. Replacing L_ν by M_ν, $S_\nu \Rightarrow W_{c\gamma,1}$ under R_ν.

Equivalently, letting $T_{\nu n} = 2f(0)\Delta_\nu \sum_{i=1}^n S_i - 2nf(0)^2\Delta_\nu^2$, we have $T_\nu \Rightarrow W_{(\lambda-\frac{1}{2})\gamma^2,\gamma^2}$ in $D[0, \infty)$ under P, Q_ν and R_ν (with $\lambda = 0$, 1 and c). Hence, assuming $f(0)$ is known, a sequential test may be carried out based on $T_{\nu n}$, with Wald's constants as stopping barriers. The hypotheses tested are $H_0\colon \{Z_n\} =_{\mathscr{L}} \{Y_n\}$ vs. $H_1\colon \{Z_n\} =_{\mathscr{L}} \{Y_n + \Delta_\nu\}$, the various densities being translations of g with $f(0) = \int g^2\,dx$ assumed known. The test will have *asymptotic efficiency* $\rho^2 = \gamma^2/\sigma^2 = 4f(0)^2/I(f)$, relative to the optimal (invariant) SPRT based on $\log L_\nu$ (see Hall (1975)).

Now suppose $f(0) = \int g^2\,dx$ is not known but the "type" of g is assumed known—i.e., g is known except for a scale parameter, say $g(x) = h(x/\xi)/\xi$ for a given density h. Then, with $k = \int h(x)^2\,dx$ (assumed known), we have $f(0) = k/\xi$ and only ξ is unknown. We can replace ξ in the statistic $T_{\nu n}$ by a strongly consistent location-invariant scale estimate $\hat{\xi}_n$, if we introduce a "delayed start-up" factor $I_{\nu n} = I(n > C\nu^{\frac{1}{2}})$, and the asymptotic behavior of $T_{\nu n}$ will be unchanged. The location invariance assures that the distribution of $\hat{\xi}_n$ is constant over ν, P, Q_ν and R_ν, and the strong consistency (a.s. convergence) is then sufficient for the application of Lemma A1 (see remark following it). For example, we can take $\hat{\xi}_n = 2^{-\frac{1}{2}}\sigma_h^{-1}\hat{\sigma}_n$ where σ_h is the (known) standard deviation of the density h and $\hat{\sigma}_n$ is the sample standard deviation of $Z_1 - Y_1, \cdots, Z_n - Y_n$. Of course, the statistic $T_{\nu n}$ is no longer simply a *sign statistic*.

Now suppose even the type of g is unknown, and hence $f(0)$ is completely unknown. Then we can replace $f(0)$ in $T_{\nu n}$ by a strongly consistent location-invariant estimate $\hat{f}_n$, as in H–L-III, if we also introduce a delayed start-up factor $I_{\nu n}$, and the asymptotic properties again remain unchanged. A suitable estimate is $\hat{f}_n = (\hat{f}_{1n} + \hat{f}_{2n})/2$ where $\hat{f}_{1n} = n^{-2}[\Sigma[3 - ((Y_i - Y_j)/a_n)^2]a_n^{-1}\phi((Y_i - Y_j)/a_n)]^+$, $\hat{f}_{2n}$ is identical with Z's replacing the Y's, and $a_n = an^{-\frac{1}{4}}$, ϕ is the standard normal density, and the summation is over $\{i, j \mid 1 \leqq i < j \leqq n\}$; it is shown in H–L-III that it is strongly consistent whenever g has a uniformly continuous and absolutely integrable derivative.

Finally, in the special case when g is assumed to be $N(0, \xi^2)$, then $Z - Y$ is $N(\lambda\Delta_\nu, 2\xi^2)$ (with $\lambda = 0$, 1 or c), $\gamma = \delta/(\xi\pi^{\frac{1}{2}})$, $I = 1/(2\xi^2)$ and $\rho^2 = 2/\pi$. The corresponding sequential sign test is based on $T_{\nu n} = \pi^{-\frac{1}{2}}\xi^{-1}\Delta_\nu \sum_{i=1}^n S_i - n(2\pi\xi^2)^{-1}\Delta_\nu^2$, where ξ may be replaced by $\hat{\xi}_n = [n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2]^{\frac{1}{2}}$ (or possibly $1/2\pi^{\frac{1}{2}}\hat{f}_n$)), and "start-up" delayed until $n \geqq n_0 = O(\nu^{\frac{1}{2}})$. A sequential test based on this statistic would have asymptotic efficiency (relative to a normal-theory SPRT) of $\rho^2 = 2/\pi$, as in nonsequential theory.

APPENDIX

In this paper we deal with weak convergence of random (right-continuous) functions on $[0, \infty)$; it is convenient to carry out the analysis in the space $D[0, \infty)$ with the Skorokhod J_1-topology (see Billingsley (1968) and Stone (1964)). Since our limiting measures are always those for Wiener processes, the details of the topology are unimportant. Relying on Theorem 15.5 of Billingsley (1968) and the theorem of Stone (1964) (with (2′) replacing (2)), we see that weak convergence to a Wiener process in $D[0, \infty)$ occurs if it occurs in $D[0, K]$ for a sequence of K-values increasing to infinity.

The following lemma is useful in sequential analysis applications since it asserts that, if a process converges weakly, then modest perturbation of it (approximate additional drift, approximate rescaling, and truncation on the time axis) does not destroy its convergence. The proof is straightforward (similar to Hall (1975)) and omitted.

LEMMA A1. *Suppose U_ν, V_ν and X_ν are in $D[0, \infty)$, and $0 \leqq t_\nu' \downarrow 0$ and $\infty \geqq t_\nu'' \uparrow \infty$; denote $I_\nu(t) = I(t_\nu' \leqq t < t_\nu'')$. If*

(i) $X_\nu \Rightarrow W_{\lambda,1}$ *in* $D[0, \infty)$,
(ii) *for each* K, $\sup_{t_\nu' \leqq t \leqq K} |U_\nu(t) - \mu t| \to 0$ *in probability as* $\nu \to \infty$, *and*
(iii) *for each* K, $\sup_{t_\nu' \leqq t \leqq K} |V_\nu(t) - \theta| \to 0$ *in probability as* $\nu \to \infty$,

then $(X_\nu + U_\nu)V_\nu I_\nu \Rightarrow W_{(\lambda+\mu)\theta,\theta^2}$ *in* $D[0, \infty)$.

In applications of this lemma, we frequently have $U_\nu(t) = \hat{\mu}_n t$ and $V_\nu(t) = \hat{\theta}_n$ at $n = [n_\nu t]$ for some $\mathscr{B}_n$-measurable $\hat{\mu}_n$ and $\hat{\theta}_n$, where $\mathscr{B}_n \subset \mathscr{A}_n$, $\mathscr{B}_n \uparrow$, and

P_ν restricted to $\mathscr{B}_n$ is ν-free for every n. Then, if $n_\nu t_\nu' \to \infty$, and if $(\hat{\mu}_n, \hat{\theta}_n) \to (\mu, \theta)$ a.s., (ii) and (iii) hold.

Now introduce absorbing barriers $b < 0 < a$, and for any x in $D[0, \infty)$ write $T^a(x) = \inf\{t \mid x(t) \geqq a\}$, $T^b(x) = \inf\{t \mid x(t) \leqq b\}$ (or $+\infty$ if otherwise undefined), $T = \min(T^a, T^b)$, and $\phi(x) = I[T^a(x) < T^b(x)]$. Thus, T is the absorption time and ϕ the indicator of absorption in $[a, \infty)$.

The following lemma is necessary for sequential analysis applications: it assures that, if $S_\nu \Rightarrow W_{\delta,\sigma^2} \equiv Z$ in $D[0, \infty)$ with the J_1-topology, then $T(S_\nu) \Rightarrow T(Z)$ and $\phi(S_\nu) \Rightarrow \phi(Z)$; that is, stopping times converge weakly, decision variables (and their expectations—the *OC-function*) converge weakly, $P_\nu(T(S_\nu) < \infty) \to 1$, etc.; see Hall (1975). (Convergence of expected stopping times (ASN's) requires extra assumptions, however.)

LEMMA A2. *T^a, T^b, T and ϕ are measurable in $C[0, \infty)$ and $D[0, \infty)$ and are continuous almost surely with respect to Wiener measure with drift.*

The continuity of T^a may be established by an argument similar to that on page 232 of Billingsley (1968); the rest is standard. We could replace "$\geqq$" by "$>$" in the definition of T^a, etc., and exactly the same results hold. Whitt (1971) has shown that such a T^a is a continuous functional everywhere, but he used the weaker M_1-topology.

REFERENCES

[1] BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] HÁJEK, JAROSLAV and ŠIDÁK, ZBYNEK (1967). *Theory of Rank Tests.* Academic Press, New York.

[3] HALL, W. J. (1973). Two asymptotically efficient sequential t-tests. *Proceedings of the Prague Symposium on Asymptotic Statistics,* **2** 89–108 (Jaroslav Hájek, ed.). Charles Univ., Prague.

[4] HALL, W. J. (1975). On construction of asymptotic sequential tests. In preparation.

[5] HALL, W. J. and LOYNES, R. M. (1977). On the concept of contiguity. *Ann. Probability* **5** (April).

[6] HALL, W. J. and LOYNES, R. M. (1975). Sequential linear rank tests—asymptotic theory. In preparation.

[7] LE CAM, L. (1960). Local asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.* **3** 37–98.

[8] LOYNES, R. M. (1976). A criterion for tightness for a sequence of martingales. *Ann. Probability* **4** 859–862.

[9] ROUSSAS, GEORGE G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge Univ. Press.

[10] STONE, CHARLES (1964). Weak convergence of stochastic processes defined on semi-infinite time intervals. *Proc. Amer. Math Soc.* **14** 694–696.

[11] WHITT, WARD (1971). Weak convergence of first passage time processes. *J. Appl. Probability* **8** 417–422.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NEW YORK 14627

DEPARTMENT OF PROBABILITY AND STATISTICS
UNIVERSITY OF SHEFFIELD
SHEFFIELD S3 7RH, UK