EXAMPLES OF EXPONENTIALLY BOUNDED STOPPING TIME OF INVARIANT SEQUENTIAL PROBABILITY RATIO TESTS WHEN THE MODEL MAY BE FALSE

R. A. WIJSMAN
UNIVERSITY OF ILLINOIS

1. Summary and introduction

Two examples are presented (one of them the sequential χ^2 test) of parametric models in which the invariant Sequential Probability Ratio Test has exponentially bounded stopping time N, that is, satisfies $P(N > n) < \rho^n$ for some $\rho < 1$, where the true distribution P may be completely arbitrary except for the exclusion of a certain class of degenerate distributions. Another example demonstrates the existence of P under which N is not exponentially bounded, but even for those P we have $P(N < \infty) = 1$. In the last section a proof is given of the representation (2.1), (2.2) of the probability ratio R_n as a ratio of two integrals over the group G if G consists of linear transformations and translations.

Let $Z_1, Z_2, \cdots$ be independent, identically distributed (i.i.d.) random vectors which take their values in d dimensional Euclidean space E^d and possess distribution P. The symbol P will also be used for the probability of an event that depends on all the Z_i. Let Θ be an index set (parameter space) such that for each $\theta \in \Theta$, P_θ is a probability distribution on E^d. We shall say "the model is true" if the true distribution P is one of the P_θ, $\theta \in \Theta$, but it should be kept in mind throughout that we shall also consider the possibility that the model is false, that is, that P is not one of the P_θ. In the latter case we shall also speak of P being outside the model as opposed to P being in the model. Let Θ_1, Θ_2 be two disjoint subsets of Θ. It is not assumed that their union is Θ. The problem is to test sequentially H_1 versus H_2, where H_j is the hypothesis: $P = P_\theta$ for some $\theta \in \Theta_j$, $j = 1, 2$.

If the hypotheses H_j are simple, that is, $\Theta_j = \{\theta_j\}$, $j = 1, 2$, then Wald's Sequential Probability Ratio Test (SPRT) [23] computes the sequence of probability ratios

$$R_n = \prod_{i=1}^{n} \frac{p_2(Z_i)}{p_1(Z_i)}, \qquad n = 1, 2, \cdots, \tag{1.1}$$

Research supported in part by National Science Foundation Grant GP-23835.

p_j being the density of P_{θ_j} with respect to some common dominating measure; stopping bounds $B < A$ are chosen and sampling continues until the first n when R_n is either $\geqq A$ (accept H_2) or $\leqq B$ (accept H_1). More convenient for our purpose is to consider $L_n = \log R_n$. Put $\log B = \ell_1$, $\log A = \ell_2$, N the random stopping time, then N is the first n when

$$\ell_1 < L_n < \ell_2 \tag{1.2}$$

is violated.

If we put $Y_i = \log [p_2(Z_i)/p_1(Z_i)]$ it follows from (1.1) that $L_n = \Sigma_1^n Y_i$, that is, L_n is a random walk on the real line, starting from zero. Wald [22] used this to prove $P(N < \infty) = 1$ for every P except if $P(Y_1 = 0) = 1$, which can also be written

$$P\{p_1(Z_1) = p_2(Z_1)\} = 1. \tag{1.3}$$

Under the same assumption $P(Y_1 = 0) < 1$, Stein [21] obtained a much stronger result

$$P(N > n) < \rho^n, \qquad n \geqq r, \tag{1.4}$$

in which $0 < \rho < 1$ and r is some suitably chosen integer. Stein's proof relies on the demonstration that there exists an integer r and $p > 0$ such that

$$P\{L_{n+r} \leqq \ell_1 \quad \text{or} \quad \geqq \ell_2 | \ell_1 < L_n < \ell_2\} > p, \qquad n = 1, 2, \cdots. \tag{1.5}$$

The property $P(N < \infty) = 1$ is usually referred to as "termination with probability one." For short, the property of N expressed by (1.4) will be called exponential boundedness of N (it would of course be more correct to call it exponential boundedness of the distribution of N; the shorter terminology follows Berk [3], [4]). The nondegeneracy condition that P not satisfy (1.3) is obviously not only sufficient but also necessary for N to be exponentially bounded. Thus, Stein's result is the most complete general result possible for the Wald SPRT.

If the hypotheses H_j are composite, a SPRT can be formulated only if somehow the composite hypotheses can be reduced to simple ones. In this paper we shall consider only the reduction that results from applying the principle of invariance (for the theory of invariance see, for example, [15]; for an extensive discussion of weight function and invariance reduction SPRT and their possible relation see Berk [4]). Suppose $X^{(n)}$ is the sample space of $(Z_1, \cdots, Z_n)$. Let G be a group of invariance transformations on $X^{(n)} \times \Theta$. The principle of invariance restricts decisions at the nth stage of sampling to those that depend on $x \in X^{(n)}$ only through its orbit Gx. At the same time, the distribution of orbits in $X^{(n)}$ depends on $\theta \in \Theta$ only through the orbit $G\theta$ of θ. (An alternative way of expressing these facts is by means of so called maximal invariants.) Suppose now that for $j = 1, 2$, Θ_j is a single orbit, that is, there exists $\theta_j \in \Theta_j$ such that $\Theta_j = G\theta_j$. Then as far as the distribution of orbits in $X^{(n)}$ is concerned, the hypotheses H_j are simple.

The probability ratio R_n at the nth stage may then be taken as the ratio of (random) orbit densities, one for θ_1 and one for θ_2. The resulting sequence $\{R_n: n = 1, 2, \cdots\}$ is used for testing in the same way as in the Wald SPRT. Such a test will be called an *invariant* SPRT. Many well known sequential tests of composite hypotheses are of this form: sequential t test, F test, and so forth.

The questions of termination with probability one and of exponential boundedness of N can also be asked of invariant SPRT but are much harder to answer. The results obtained so far are much less complete than for the Wald SPRT. In fact, the history of these endeavors has had a very modest beginning with only the question of termination with probability one being attacked and then only for P in the parametric model: Johnson [13], David and Kruskal [7], Ray [18], Jackson and Bradley [12], Wirjosudirjo [30], Ifram [9]. Berk [2] was the first explicitly to consider P outside the model. This was followed by Wijsman in [28] where the question of termination with probability one was also treated for a class of distributions larger than given by the model.

Study of the exponential boundedness of N started much later. Ifram [11] still restricted P within the parametric model. Later studies of parametric models by Wijsman [29], Abu-Salih [1] and Berk [4] allowed P outside the model. The first nonparametric model (sequential two sample rank order test) was treated by Savage and Sethuraman [19] and sharper results were obtained by Sethuraman [20]. In both [19] and [20] P was allowed outside the model.

The only result in the literature to date for invariant SPRT that approximates the completeness of Stein's result for the Wald SPRT is Sethuraman's [20] since he shows for the particular rank order test in question that N is exponentially bounded for every P, except for P in a certain class of degenerate distributions (but whether the P in this class actually misbehave, that is, cause N not to be exponentially bounded, is not known). In all parametric problems studied the results have been considerably less complete. In order to obtain termination with probability one a minimal assumption, made by all authors, has been that under P certain random variables (the X_i of (2.4)) have first moments, and for exponential boundedness of N the existence of their moment generating functions. In addition, nearly all authors have been plagued by exceptional P (where certain moments have certain exceptional values, not to be confused with degenerate P) under which only weaker results or even no results could be obtained (see, for example, the discussion in [29] Section 1). This is not typical for parametric as opposed to nonparametric problems, for Savage and Sethuraman [19] encountered the same difficulty. The only authors whose methods have made it possible to avoid having to exclude exceptional distributions are Berk [4] and Sethuraman [20]. The latter author revived the Stein type of proof in which (1.5) has to be established (this is of course much harder if L_n is not a random walk). Berk [4] followed suit and used a modification of Stein's method to prove exponential boundedness of N. Judging from these examples it is not unreasonable to guess that for sharpest results it is necessary to use the Stein type of proof, centering around (1.5). (However, the absence of exceptional distributions in

Berk's work must be due to his method [3] and assumptions in [3], [4], since his proof of termination with probability one has nothing to do with Stein's type of proof.)

There is a certain amount of overlap in results and in assumptions between [4] on one hand and [28], [29] on the other. There is a discussion of this in [4] Section 2. However, there are also differences. Berk obtains the strongest results whenever his method (relying on [3]) applies. Thus, he obtains exponential boundedness of N in the sequential t test also for those exceptional P for which only weaker results could be obtained in [29]. Moreover, [28] and [29] apply only to invariant SPRT and assume a multivariate normal model whereas Berk's method applies to a wider class of models and to weight function SPRT that are not necessarily invariant SPRT. On the other hand there are some intrinsic limitations in Berk's method that prevent the applicability of [3], [4] to certain cases that are amenable to [28], [29]. Specifically, Assumption 2.4(d) in [4] (which already assumes the existence of $E_P X_1$) implies in our notation (see (2.3), (2.4)) that $\lambda'(\theta) E_P X_1 + \log c(\theta)$ has a unique maximum as θ varies through $\Theta_j, j = 1, 2$. This condition is violated in Example 2, Section 5, if P is such that $E_P Z_1 Z_1'$ is a constant times the identity matrix because then $\lambda'(\theta) E_P X_1 + \log c(\theta)$ is constant on each Θ_j. The same happens in Example 3, Section 6, if $E_P Z_1 = 0$.

The main purpose of this paper is to provide two examples (2 and 3), one of them being the sequential χ^2 test, in which N is shown to be exponentially bounded for every P, excluding a certain class of degenerate P. In particular, it is not assumed that any of the random variables has finite expectation under P. We rely heavily on the Stein type of proof. To the best of our knowledge this is the first time such examples have been given in parametric models. These examples give hope that in the future other (perhaps all?) invariant SPRT may be shown to have exponentially bounded N for all but certain degenerate P.

In Section 3 a general discussion is given on distributions P for which N is not exponentially bounded. Such P will be called *obstructive*. That such P do indeed exist is illustrated in Example 1, Section 4. This example is nontrivial but so simple that the behavior of $\{L_n\}$ can be studied easily. It seems to be the first example of any nontrivial invariant SPRT—parametric or nonparametric—where the existence of obstructive P has been demonstrated. Berk [4], Section 5, gives examples of weight function SPRT where the tests do not even terminate with probability one for certain P. However, those tests do not seem to be invariant SPRT and fall therefore outside the scope of this paper.

2. Representation of R_n and approximation of L_n by Φ_n

We assume that for each $\theta \in \Theta$, P_θ has a density p_θ with respect to Lebesgue measure on E^d. Let G be a group of invariance transformations $Z_i \to gZ_i$, $i = 1, 2, \cdots$, where to each $g \in G$ is associated a d^2 nonsingular matrix C and a vector $b \in E^d$, and the action is defined by $gZ_i = C(Z_i + b)$. The precise assumptions on G are set forth in Section 7, Theorem 7.1. The hypotheses H_j for

$j = 1, 2$, are of the form: P is one of the P_θ with $\theta \in G\theta_j$, where θ_1 and θ_2 are not on the same orbit so that their orbits $G\theta_j$ are disjoint. It will be proved in Section 7 that the probability ratio R_n at the nth stage for the invariant SPRT can be represented in the form

$$R_n = J_n(\theta_2)/J_n(\theta_1), \qquad n = d+1, d+2, \cdots, \tag{2.1}$$

in which

$$J_n(\theta) = \int \prod_{i=1}^{n} p_{g\theta}(Z_i)\nu_G(dg). \tag{2.2}$$

In (2.2) ν_G is right Haar measure on G and we assume $n \geqq d + 1$. Note that the right invariance of ν_G guarantees that (2.2) is constant on each orbit in Θ, so that R_n does not depend on the particular choice of θ_1 and θ_2. The representation (2.1), (2.2) bears a strong resemblance to (3) in [27], but in [27] the group G acted linearly whereas in the present paper the action of G may include translations. Furthermore, (2.2) is written in terms of G acting on Θ, whereas (3) in [27] was expressed in terms of G acting on the sample space.

From now on we shall assume that the model is exponential, that is, the density p_θ of the Z_i is expressible as

$$p_\theta(z) = c(\theta)e^{\lambda'(\theta)s(z)}h(z), \tag{2.3}$$

in which $\lambda(\theta), s(z) \in E^k$ for some $k \geqq 1$, $c(\theta), h(z) > 0$ and prime denotes transposition. It is convenient to denote

$$X_i = s(Z_i), \qquad i = 1, 2, \cdots \tag{2.4}$$

so that $X_1, X_2, \cdots$ are i.i.d. vectors in E^k. Furthermore we introduce the function

$$\psi_\theta(g, x) = \lambda'(g\theta)x + \log c(g\theta), \qquad x \in E^k, \tag{2.5}$$

and write $\bar{X}_n$ for $(1/n)\ \Sigma_1^n\ X_i$. After substitution of (2.3) to (2.5) into (2.2) we obtain

$$J_n(\theta) = \int \exp\{n\psi_\theta(g, \bar{X}_n)\}\nu_G(dg). \tag{2.6}$$

If in (2.6) the integration were over a Euclidean space instead of over a group, the method of Laplace (see, for example, [6]) could be used to approximate $J_n(\theta)$ for large n. Assuming that Laplace's method is applicable even if the integration is over a group (provided certain conditions are fulfilled) we expect the main contribution to the integral to come from values of g that come close to maximizing ψ_θ. This suggests putting

$$\phi_\theta(x) = \max_{g \in G} \psi_\theta(g, x), \tag{2.7}$$

assuming the maximum exists, and putting

$$\Phi(x) = \phi_{\theta_2}(x) - \phi_{\theta_1}(x). \tag{2.8}$$

Then we may expect that asymptotically $R_n \sim \exp\{n\Phi(\bar{X}_n)\}$. Taking log and writing

$$\Phi_n = n\Phi(\bar{X}_n) \tag{2.9}$$

we may expect then that Φ_n is in some sense an approximation to $L_n = \log R_n$. This approach was used in [28], [29].

Suppose now that the approximation of L_n by Φ_n has the following form: there is a constant B such that for all n

$$|L_n - \Phi_n| < B. \tag{2.10}$$

Let $\ell'_1 = \ell_1 - B, \ell'_2 = \ell_2 + B$, where ℓ_1 and ℓ_2 are the stopping bounds on L_n (see (1.2)). In analogy to N defined in (1.2), define N' to be the first n for which $\ell'_1 < \Phi_n < \ell'_2$ is violated. Then clearly $N \leqq N'$ so that if the latter is exponentially bounded, so is the former. Since our aim is to prove exponential boundedness of N for all choices of ℓ_1, ℓ_2, the problem then becomes to prove exponential boundedness of N' for all choices of ℓ'_1, ℓ'_2. This shifts therefore the original problem to a similar one, the only difference being that the original sequence $\{L_n\}$ has been replaced by a new sequence $\{\Phi_n\}$. The advantage is that $\{\Phi_n\}$ is usually much more tractable than $\{L_n\}$. In Examples 2 and 3 it is indeed possible first to show (2.10) and then to prove exponential boundedness of N'. In the proofs the primes on ℓ_1, ℓ_2 and N will be dropped for notational convenience.

3. Obstructive distributions

For the purpose of this paper we shall call a distribution P *obstructive* if for some stopping bounds (1.4) is not true, that is, N is not exponentially bounded when the true distribution is P. In Wald's SPRT Stein's result shows that P is obstructive if and only if it satisfies (1.3). In the case of invariant SPRT we know much less about the characterization of obstructive P. In fact, there is not even one example where this has been carried out. All that has been achieved so far is that in a few isolated examples classes of P that contain the obstructive P have been characterized. For instance, Sethuraman [20] shows in a sequential two sample rank test that the obstructive P satisfy (in his notation) $P(V(X_1, Y_1) = 0) = 1$, but it is not known whether all P in this class are obstructive. In Examples 2 and 3, Sections 5 and 6, certain classes of P will be characterized and shown to include the obstructive P. Again it is not known whether the inclusion is proper but this seems likely in the light of Example 1.

In the exponential model there is some reason to conjecture that any obstructive P satisfies

$$P(v'X_1 = \text{constant}) = 1 \qquad \text{for some } v \in E^k, \qquad v \neq 0, \tag{3.1}$$

X_1 being defined in (2.4). That is, X_1 is confined to a $(k - 1)$ dimensional hyperplane, a.e. P. However, very likely the class defined by (3.1) is usually too big.

At least in Examples 2 and 3 the class of P defined by (3.1) is larger than the class of obstructive P. A smaller class than defined by (3.1) is the one defined by Berk in his Assumption 2.4(c) of [4] (in different notation):

$$P\{p_{\theta'}(Z_1) = p_{\theta''}(Z_1)\} = 1 \qquad \text{for some } \theta' \in \Theta_1,\ \theta'' \in \Theta_2. \tag{3.2}$$

This is the composite hypotheses analogue to (1.3). It is immediately seen that in the exponential model (3.2) implies (3.1). In Examples 2 and 3 it is true that the obstructive P, if any, are contained in the class defined by (3.2), but again the latter is too big. It turns out in those examples that obstructive P satisfy (3.2) only for certain pairs of (θ', θ''). One can make a conjecture that the pairs (θ', θ'') for which (3.2) may give obstructive P are those pairs that minimize the Kullback-Leibler "divergence between $p_{\theta'}$ and $p_{\theta''}$" (see (2.6) in [14]). It should be mentioned that in Examples 2 and 3 no P are actually demonstrated to be obstructive.

Example 1, Section 4 demonstrates for an invariant SPRT the existence of obstructive distributions P. They do satisfy (3.2), with the further restriction to pairs (θ', θ'') that minimize the Kullback-Leibler divergence; but even in this smaller class not all P are obstructive. In Example 1 we would have a complete characterization of obstructive P if we would know that every P outside the class (3.2) is not obstructive. Unfortunately, this is not known at present.

There is a qualitative difference between obstructive P for Wald's SPRT and obstructive P for invariant SPRT, at least for those of Example 1. In Wald's SPRT P is obstructive if and only if (1.3) is satisfied and in that case $L_n = 0$ for all n a.e. P so that not only is N not exponentially bounded but $N = \infty$ a.e. P. The obstructive P in Example 1 behave much more mildly: $\{L_n\}$ is truly random and $P(N < \infty) = 1$. The only thing that goes wrong is that the distribution of N does not have nice properties. Not only is N not exponentially bounded but the stopping bounds of the test can be chosen so that not even the first moment of N is finite.

4. Example 1

Let $Z_1, Z_2, \cdots$ be i.i.d. normal with mean ζ variance σ^2, both unknown. H_j: $\sigma = \sigma_j$, $j = 1, 2$, where the σ_j are given and distinct. Thus, ζ is a nuisance parameter. The problem is invariant under the transformation $Z_i \to Z_i + b$, $i = 1, 2, \cdots$, $\zeta \to \zeta + b$, $-\infty < b < \infty$, $\sigma \to \sigma$. The group G is therefore the group of reals b under addition, and right (= left) Haar measure $\nu_G(dg)$ can be taken simply as db. The points θ_1 and θ_2 on the orbits Θ_j can be chosen as $\theta_j = (0, \sigma_j)$, $j = 1, 2$. Then

$$P^{\theta}_{g_j}(Z_i) = \sqrt{2\pi}\sigma_j)^{-1} \exp\{-(2\sigma_j^2)^{-1}(Z_i - b)^2\}, \tag{4.1}$$

where b corresponds to g. Substituting this into (2.1) and (2.2), integrating with respect to b, and taking log yields

$$L_n = [(2\sigma_1^2)^{-1} - (2\sigma_2^2)^{-1}] \sum_{i=1}^{n} (Z_i - \bar{Z}_n)^2 + (n-1) \log \frac{\sigma_1}{\sigma_2} \tag{4.2}$$

in which $\bar{Z}_n = (1/n)\Sigma_1^n Z_i$.

Now for convenience choose $\sigma_1 < \sigma_2$ in such a way that

$$\sigma_1^{-2} + 2 \log \sigma_1 = \sigma_2^{-2} + 2 \log \sigma_2. \tag{4.3}$$

It is easily checked that with this choice the two normal densities with mean zero and variance σ_1^2, σ_2^2, respectively, are equal at ± 1, that is,

$$p_{\theta_1}(\pm 1) = p_{\theta_2}(\pm 1). \tag{4.4}$$

Take the distribution P as follows:

$$P(Z_1 = -1) = P(Z_1 = 1) = \tfrac{1}{2}, \tag{4.5}$$

so that $\Sigma_1^n Z_i^2 = n$ a.e. P. Substituting this into (4.2) and using (4.3) we get

$$L_n = \left[\log \frac{\sigma_2}{\sigma_1}\right](1 - n\bar{Z}_n^2) \qquad \text{a.e. } P. \tag{4.6}$$

Suppose ℓ_2 is chosen $> \log (\sigma_2/\sigma_1)$ then $L_n < \ell_2$ for all n a.e. P. The double inequality (1.2) reduces then to the single inequality $\ell_1 < L_n$, which can be put in the form

$$\left|\sum_{i=1}^{n} Z_i\right| < a\sqrt{n} \tag{4.7}$$

for some $a > 0$ depending on σ_1, σ_2 and ℓ_1. The law of the iterated logarithm guarantees that for every $a > 0$, we have $P(N < \infty) = 1$, but if $a \geqq 1$ then $E_P N = \infty$ [5].

It follows from (4.4) and (4.5) that P satisfies (3.2) for the couple $(\theta', \theta'') = (\theta_1, \theta_2)$, where $\theta_j = (0, \sigma_j)$. It is clear from (4.2) that the behavior of L_n is unchanged if P puts probability $\frac{1}{2}$ on $c \pm 1$ for arbitrary c, so that such a P is also obstructive. It can be shown that this class of P, with $-\infty < c < \infty$, constitutes the class of obstructive P among all P satisfying (3.2) (or even among all P for which Z_1^2 has finite moment generating function). At the same time, a P with support on $c \pm 1$ satisfies (3.2) with

$$\theta' = (c, \sigma_1), \qquad \theta'' = (c, \sigma_2), \tag{4.8}$$

and a simple computation shows that the pairs (θ', θ'') given by (4.8) are the pairs that minimize the Kullback-Leibler divergence. But note that not even all P satisfying (3.2) with $\theta', \theta'')$ satisfying (4.8) are obstructive, since for this property it is not only required that P is supported on $c \pm 1$ but also that the probabilities in these two points are $\frac{1}{2}$.

5. Example 2

Let $Z_1, Z_2, \cdots$ be i.i.d. bivariate normal with mean 0 and covariance matrix Σ. Put $Z_i = (x_i, y_i)'$. We want to test the hypotheses $H_j, j = 1, 2$, that the characteristic roots of Σ^{-1} are σ_j, τ_j, with $\sigma_j > \tau_j$, where the σ_j and τ_j are given positive numbers such that $(\sigma_1, \tau_1) \neq (\sigma_2, \tau_2)$. Consider the invariance transformation $Z_i \to \Omega Z_i, i = 1, 2, \cdots, \Sigma \to \Omega\Sigma\Omega'$, with Ω 2^2 orthogonal. It is sufficient and more convenient to restrict G to all Ω having determinant $+1$. Then we may take Ω in the form

$$\Omega = \begin{bmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{bmatrix}, \qquad 0 \leqq \omega < 2\pi, \tag{5.1}$$

and right (= left) Haar measure $\nu_G(d\Omega)$ can be taken as $d\omega$. The density p_θ of Z_i with respect to Lebesgue measure in E^2 is

$$p_\theta(Z_i) = (2\pi)^{-1}|\Sigma|^{-1/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} Z_i Z_i'\}. \tag{5.2}$$

We may identify θ with Σ^{-1} and take $\theta_j = \operatorname{diag}(\sigma_j, \tau_j), j = 1, 2$. In order to evaluate R_n, using (2.1), (2.2), we have to replace in (5.2) Σ^{-1} by $\Omega \operatorname{diag}(\sigma_j, \tau_j)\Omega'$, $j = 1, 2$, substitute into (2.1), (2.2), and integrate. The integrals involve expressions of the form

$$\tfrac{1}{2}\pi \int_0^{2\pi} \exp\{x \cos t\}\, dt = I_0(x), \tag{5.3}$$

where I_0 is one of the Bessel functions of imaginary argument ([25], p. 77). The random variables $Z_i = (x_i, y_i)'$ enter into the result only in the following combinations:

$$U_n = \sum_{i=1}^{n} (x_i^2 + y_i^2), \tag{5.4}$$

$$V_n = \left[\left(\sum_1^n x_i^2 - \sum_1^n y_i^2\right)^2 + \left(2\sum_1^n x_i y_i\right)^2\right]^{1/2}. \tag{5.5}$$

In this notation we obtain for $L_n = \log R_n$ the result

$$\begin{aligned} L_n = {} & \tfrac{1}{4}(\sigma_1 + \tau_1 - \sigma_2 - \tau_2)U_n + \frac{n}{2}\log\frac{\sigma_2\tau_2}{\sigma_1\tau_1} \\ & + \log I_0[\tfrac{1}{4}(\sigma_2 - \tau_2)V_n] - \log I_0[\tfrac{1}{4}(\sigma_1 - \tau_1)V_n]. \end{aligned} \tag{5.6}$$

On the other hand, performing the maximization in (2.7) and substituting into (2.8) we find

$$\begin{aligned} \Phi_n = {} & \tfrac{1}{4}(\sigma_1 + \tau_1 - \sigma_2 - \tau_2)U_n + \frac{n}{2}\log\frac{\sigma_2\tau_2}{\sigma_1\tau_1} \\ & + \tfrac{1}{4}(\sigma_2 - \tau_2 - \sigma_1 + \tau_1)V_n. \end{aligned} \tag{5.7}$$

From the continuity of I_0 and its asymptotic behavior,

$$\lim_{x \to \infty} (2\pi x)^{1/2} \exp\{-x\} I_0(x) = 1 \tag{5.8}$$

(see [25], p. 203, but this also follows directly from the integral representation of $I_0(x)$ and Laplace's method), we deduce that for any given $\alpha, \beta > 0$ there exists B such that for all $x \geqq 0$

$$|\log I_0(\alpha x) - \log I_0(\beta x) - (\alpha - \beta)x| < B. \tag{5.9}$$

Using this to compare (5.6) and (5.7) we see that (2.10) is satisfied. Hence for the question of exponential boundedness of N we may replace L_n by Φ_n, as discussed in Section 2.

It turns out that the distributions P, for which we cannot prove exponential boundedness of N, satisfy (3.2) with the pairs (θ', θ'') of the form $(g\theta_1, g\theta_2)$, $g \in G$ (and these are precisely the pairs that minimize the Kullback-Leibler divergence). That is, the two values of Σ^{-1} (under θ' and θ'') are $\Omega \operatorname{diag}(\sigma_1, \tau_1)\Omega'$ and $\Omega \operatorname{diag}(\sigma_2, \tau_2)\Omega'$, where Ω is any matrix of the form (5.1). Substituting (5.2) for Z_1 into (3.2), with the above values of Σ^{-1} and putting $\cos 2\omega = -u_1$, $\sin 2\omega = -u_2$, the negation of (3.2) then has the form

$$\begin{aligned} P\{\tfrac{1}{2}(\sigma_2 - \tau_2 - \sigma_1 + \tau_1)[u_1(x_1^2 - y_1^2) + u_2(2x_1y_1)] \\ + \tfrac{1}{2}(\sigma_1 + \tau_1 - \sigma_2 - \tau_2)(x_1^2 + y_1^2) + \log\frac{\sigma_2\tau_2}{\sigma_1\tau_1} = 0\} < 1. \end{aligned} \tag{5.10}$$

Suppose first $\sigma_2 - \tau_2 - \sigma_1 + \tau_1 = 0$. Using (5.7) and (5.4) it is seen now that Φ_n is of the form $\Phi_n = \Sigma_1^n W_i$, with $W_1, W_2, \cdots$ i.i.d. and

$$W_i = \tfrac{1}{4}(\sigma_1 + \tau_1 - \sigma_2 - \tau_2)(x_i^2 + y_i^2) + \tfrac{1}{2}\log(\sigma_2\tau_2/\sigma_1\tau_1), \tag{5.11}$$

and thus by (5.10) $P(W_1 = 0) < 1$. Exponential boundedness follows from [21].

Assuming from now on that $\sigma_2 - \tau_2 - \sigma_1 + \tau_1 \neq 0$ we shall put

$$a = \frac{\sigma_1 + \tau_1 - \sigma_2 - \tau_2}{\sigma_2 - \tau_2 - \sigma_1 + \tau_1}, \qquad b = \frac{2}{\sigma_2 - \tau_2 - \sigma_1 + \tau_1}\log\frac{\sigma_2\tau_2}{\sigma_1\tau_1}, \tag{5.12}$$

so that (5.10) can be written

$$P\{u_1(x_1^2 - y_1^2) + u_2(2x_1y_1) + a(x_1^2 + y_1^2) + b = 0\} < 1. \tag{5.13}$$

For the purpose of proving exponential boundedness of N we may divide Φ_n by any nonzero constant. Dividing in (5.7) by the coefficient of V_n we get

$$\Phi_n = V_n + aU_n + bn. \tag{5.14}$$

A further simplification in notation will be made. Consider the vectors $Y_1, Y_2, \cdots$, where $Y_i = (x_i^2 - y_i^2, 2x_iy_i)'$, then $x_i^2 + y_i^2 = \|Y_i\|$ and (5.13), (5.14) can be put in the form

$$P\{u'Y_1 + a\|Y_1\| + b = 0\} < 1 \tag{5.15}$$

for every $u \in U$,

$$\Phi_n = \|\sum_1^n Y_i\| + a\sum_1^n \|Y_i\| + bn. \tag{5.16}$$

in which $U = \{u \in E^2 : \|u\| = 1\}$. In (5.15), (5.16) there are some restrictions on a and b that follow from (5.12) and $\sigma_j \geqq \tau_j, j = 1, 2$. For instance, if $b = 0$ we must have $-1 < a < 0$. However, these restrictions are of no help in the proof of exponential boundedness of N and we shall ignore them. Also, the fact that the dimension of the sample space of the Y_i is two plays no role. We shall give the proof for $Y_i \in E^k$, for any positive integer k. This necessitates a redefinition of the set of unit vectors

$$U = \{u \in E^k : \|u\| = 1\}. \tag{5.17}$$

LEMMA 5.1. *Let $Y_1, Y_2, \cdots$ be i.i.d. random vectors taking their values in E^k for some $k \geqq 1$, their common distribution P satisfying* (5.15) *with U defined in* (5.17). *Let N be the first integer n such that $\ell_1 < \Phi_n < \ell_2$ is violated, where Φ_n is given by* (5.16) *and ℓ_1, ℓ_2 are arbitrary real numbers. Then N is exponentially bounded, that is, satisfies* (1.4).

Before proving Lemma 5.1 it is convenient to prove

LEMMA 5.2. *Let $\{X_t, t \in T\}$ be a family of real valued random variables, T a compact index set, such that for every $t \in T$,* (i) $P(X_t < 0) > 0$, *and* (ii) $X_s \to X_t$ *in law as $s \to t$. Then there exists $\delta, \varepsilon > 0$ such that $P(X_t < -\delta) > \varepsilon$ for every $t \in T$.*

PROOF. Let F_t be the distribution function of X_t, and $x_t < 0$ a continuity point of F_t such that $F_t(x_t) = 2\varepsilon_t$ for some $\varepsilon_t > 0$ (this can be done by assumption (i)). For any $t \in T$ let $s \to t$ then $F_s(x) \to F_t(x)$ for every continuity point x of F_t, using assumption (ii). In particular, $F_s(x_t) \to F_t(x_t) = 2\varepsilon_t$ so that there exists a neighborhood U_t of t such that $s \in U_t$ implies $F_s(x_t) > \varepsilon_t$. Since T is compact, we can cover T with a finite number of such U_t, say $U_{t_1}, \cdots, U_{t_n}$. Put $\varepsilon = \min(\varepsilon_{t_i}, i = 1, \cdots, n)$, $\delta = \min(-x_{t_i}, i = 1, \cdots, n)$ so that $\varepsilon, \delta > 0$. Take an arbitrary $t \in T$ then t is in one of the U_{t_i}, say in U_{t_j}. We have then $F_t(-\delta) \geqq F_t(x_{t_j}) > \varepsilon_{t_j} \geqq \varepsilon$.

PROOF OF LEMMA 5.1. The case $P(Y_1 = 0) = 1$ is trivial for then by (5.15), (5.16) $\Phi_n = bn$, $b \neq 0$ so that N is constant. Henceforth we shall assume $P(Y_1 = 0) < 1$. We shall distinguish two cases.

Case 1. There exists $u \in U$ (defined in (5.17)) such that

$$P\{u'Y_1 + a\|Y_1\| + b \geqq 0\} = 1. \tag{5.18}$$

We define

$$W_i = u'Y_i + a\|Y_i\| + b \tag{5.19}$$

so that $W_1, W_2, \cdots$ are i.i.d. real valued random variables. Equation (5.18) states that $P(W_1 \geqq 0) = 1$, and (5.15) in addition implies that $P(W_1 > 0) > 0$.

Thus, the random walk $T_n = \Sigma_1^n W_i$ takes only nonnegative steps, and the steps are positive with positive probability. Let N' be the first integer n such that $\ell_1 < T_n < \ell_2$ is violated, then N' is exponentially bounded according to [21]. Define

$$S_n = \sum_{i=1}^{n} Y_i \tag{5.20}$$

then $\|S_n\| \geqq u'S_n$ and therefore, using (5.16) and (5.19),

$$\Phi_n \geqq u'S_n + a \sum_1^n \|Y_i\| + bn = \sum_1^n W_i = T_n. \tag{5.21}$$

This, together with the fact that T_n is nondecreasing, implies $N \leqq N'$ so that N is also exponentially bounded.

Case 2. For all $u \in U$, $P(u'Y_1 + a\|Y_1\| + b \geqq 0) < 1$, which can also be written in the form

$$P\{u'Y_1 + a\|Y_1\| + b < 0\} > 0 \tag{5.22}$$

for every $u \in U$. We shall prove that there exists a positive integer r and $p > 0$ such that

$$P\{\ell_1 < \Phi_{n+i} < \ell_2,\ i = 1, \cdots, r | \ell_1 < \Phi_n < \ell_2\} < 1 - p, \qquad n = 1, 2, \cdots. \tag{5.23}$$

This statement is weaker than (1.5) but exponential boundedness of N follows from (5.23) in the same way as it does from (1.5).

We apply Lemma 5.2 to $X_t = t'Y_1 + a\|Y_1\| + b$, $t \in U$. Obviously, U is compact. Assumption (i) in Lemma 5.2 is (5.22) and assumption (ii) follows immediately from the fact that $X_s \to X_t$ everywhere as $s \to t$. Using Lemma 5.2, there exists $\delta_1, \varepsilon_1 > 0$ such that

$$P\{u'Y_1 + a\|Y_1\| + b < -2\delta_1\} > 2\varepsilon_1 \tag{5.24}$$

for every $u \in U$. Now take B_1 so large that $P(\|Y_1\| \geqq B_1) < \varepsilon_1$ and combine with (5.24), then

$$P\{\|Y_1\| < B_1, u'Y_1 + a\|Y_1\| + b < -2\delta_1\} > \varepsilon_1 \tag{5.25}$$

for every $u \in U$. For any vectors $s, y \in E^k$ we compute

$$\|s + y\| - \|s\| \leqq \|s\|^{-1}(s'y + \tfrac{1}{2}\|y\|^2). \tag{5.26}$$

With S_n defined in (5.20) define $u_n = S_n/\|S_n\|$ if $S_n \neq 0$, and if $S_n = 0$ take u_n equal to any fixed vector in U. Let i be any positive integer. Since $S_{n+i} = S_{n+i-1} + Y_{n+i}$, (5.26) states that

$$\|S_{n+i}\| - \|S_{n+i-1}\| \leqq u'_{n+i-1}Y_{n+i} + \frac{\|Y_{n+i}\|^2}{2\|S_{n+i-1}\|}. \tag{5.27}$$

From (5.16) we obtain

(5.28) $$\Phi_{n+i} - \Phi_{n+i-1} = \|S_{n+i}\| - \|S_{n+i-1}\| + a\|Y_{n+i}\| + b.$$

Substitution of (5.27) into (5.28) gives

(5.29) $$\Phi_{n+i} - \Phi_{n+i-1} \leqq u'_{n+i-1} Y_{n+i} + a\|Y_{n+i}\| + b + (2\|S_{n+i-1}\|)^{-1}\|Y_{n+i}\|^2.$$

Let B_2 be so large that $B_1^2/(2B_2) < \delta_1$, then

(5.30) $$[\|Y_{n+i}\| < B_i, \|S_{n+i-1}\| > B_2] \Rightarrow [(2\|S_{n+i-1}\|)^{-1}\|Y_{n+i}\|^2 < \delta_1].$$

Together with (5.29) this gives

(5.31) $$[\|Y_{n+i}\| < B_1, \|S_{n+i-1}\| > B_2, u'_{n+i-1}Y_{n+i} + a\|Y_{n+i}\| + b < -2\delta_1] \Rightarrow [\Phi_{n+1} - \Phi_{n+i-1} < -\delta_1].$$

Choose integer n_1 so that $n_1\delta_1 > d = \ell_2 - \ell_1$ and put $B = B_2 + n_1B_1$. Then

(5.32) $$[\|S_n\| > B, \|Y_{n+i}\| < B_1, i = 1, \cdots, n_1] \Rightarrow [\|S_{n+i-1}\| > B_2, i = 1, \cdots, n_1].$$

Define the events

(5.33) $$A_i = [\|Y_i\| < B_1, u'_{i-1}Y_i + a\|Y_i\| + b < -2\delta_1], \qquad i = 1, 2, \cdots,$$

then given $\|S_n\| > B, A_{n+i}$ implies the left side of (5.31) for $i = 1, \cdots, n_1$ (making use of (5.32)) and therefore

(5.34) $$\left[\|S_n\| > B, \bigcap_{i=1}^{n_1} A_{n+i}\right] \Rightarrow [\Phi_{n+i} - \Phi_{n+i-1} < -\delta_1, i = 1, \cdots, n_1] \Rightarrow [\Phi_{n+n_1} - \Phi_n < -n_1\delta_1] \Rightarrow [\Phi_{n+n_1} - \Phi_n < -d].$$

From (5.25) and the independence of the Y_i it follows that $P(A_i|S_{i-1}) > \varepsilon_1$ for every given value of S_{i-1}. (Note that the A_i are not independent, but A_i depends on $Y_1, \cdots, Y_{i-1}$ only through u_{i-1}, that is, through S_{i-1}.) Therefore, given $\|S_n\| > B$, the extreme left side of (5.34) has probability $\geqq \varepsilon_1^{n_1}$, and consequently

(5.35) $$P\{\Phi_{n+n_1} - \Phi_n < -d\,|\,S_n\} > \varepsilon_1^{n_1} \qquad \text{if} \quad \|S_n\| > B.$$

Since $P(Y_1 = 0) < 1$, we can choose $u \in U$ so that $P(u'Y_1 > 0) > 0$. Then there exists $\delta_2, \varepsilon_2 > 0$ such that $P(u'Y_1 > \delta_2) > \varepsilon_2$. Choose integer n_2 such that $n_2\delta_2 > 2B$. Then

(5.36) $$[\|S_n\| \leqq B, u'Y_{n+i} > \delta_2, i = 1, \cdots, n_2] \Rightarrow [\|S_{n+n_2}\| > B].$$

Therefore

(5.37) $$P\{\|S_{n+n_2}\| > B|S_n\} > \varepsilon_2^{n_2} \qquad \text{if} \quad \|S_n\| \leqq B.$$

Now put $r = n_1 + n_2$ and consider

$$P\{\Phi_{n+i} \leqq \ell_1 \quad \text{for some} \quad i = 1, \cdots, r \,|\, \ell_1 < \Phi_n < \ell_2, S_n\}. \tag{5.38}$$

If $\|S_n\| > B$ it follows from (5.35) that (5.38) is $> \varepsilon_1^{n_1}$. If $\|S_n\| \leqq B$ it follows from (5.37) and from (5.35) with n replaced by $n + n_2$, that (5.38) is $> \varepsilon_2^{n_2}\varepsilon_1^{n_1}$. Taking $p = \varepsilon_2^{n_2}\varepsilon_1^{n_1}$, we have then, whether $\|S_n\|$ is $> B$ or $\leqq B$,

$$P\{\Phi_{n+i} \leqq \ell_1 \quad \text{for some} \quad i = 1, \cdots, r \,|\, \ell_1 < \Phi_n < \ell_2\} > p \tag{5.39}$$

and (5.24) is an immediate consequence.

6. Example 3: sequential χ^2 test

Let $Z_1, Z_2, \cdots$ be i.i.d. d-variate normal with identity covariance matrix and unknown mean vector ζ. We want to test the hypotheses H_j that $\|\zeta\| = \gamma_j$, where $\gamma_1 \neq \gamma_2$ are given positive numbers. (Note that we exclude the null hypothesis $\zeta = 0$.) Let G be the group of all d^2 orthogonal matrices Ω and consider the invariance transformations $Z_i \to \Omega Z_i, i = 1, 2, \cdots, \zeta \to \Omega\zeta$. The parameter θ in Section 2 may be identified with ζ and we may take $\theta_j = \gamma_j u, j = 1, 2$, where u is any fixed element of U defined in (5.17) with $k = d$. The density of Z_i is given by

$$p_\theta(Z_i) = (2\pi)^{-d/2} \exp\{-\tfrac{1}{2}(Z_i - \zeta)'(Z_i - \zeta)\}. \tag{6.1}$$

Replacing in (6.1) ζ by $\Omega\theta_j = \gamma_j\Omega u, j = 1, 2$, substituting into (2.1), (2.2) and taking log results in

$$L_n = \log f(\gamma_2\|S_n\|) - \log f(\gamma_1\|S_n\|) + (n/2)(\gamma_1^2 - \gamma_2^2), \tag{6.2}$$

in which

$$S_n = \sum_{i=1}^{n} Z_i \tag{6.3}$$

and

$$f(x) = \int e^{xu'\Omega u}\, \mu(d\Omega), \qquad [x \geqq 0, u \in U]. \tag{6.4}$$

In (6.4) μ is Haar measure on G. Clearly, f does not depend on the particular choice of $u \in U$. The function f satisfies the differential equation $f''(x) + (d - 1)x^{-1}f'(x) = f(x)$ and can be related to the hypergeometric function ${}_0F_1$ [24], [10]: $f(x) = {}_0F_1[\frac{1}{2}d(\frac{1}{2}x)^2]$ provided μ is normalized. This permits writing down the asymptotic behavior of f; we have $\lim_{x\to\infty} x^{(d-1)/2}e^{-x}f(x) = \text{constant}$. We can obtain this also directly from (6.4) using Laplace's method after writing $f(x)$ in the form $c\int_0^\pi \exp\{x\cos\omega\}(\sin\omega)^{d-2}\,d\omega$. A third way is provided by verifying that $f(x) = cx^{-\nu}I_\nu(x)$, with $\nu = \frac{1}{2}d - 1$ and consulting [25], p. 203 for the asymptotic expansion of I_ν. It follows that for any $\alpha, \beta > 0$ there exists B such that

$$|\log f(\alpha x) - \log f(\beta x) - (\alpha - \beta)x| < B. \tag{6.5}$$

Maximizing the integrand in (6.4) in order to obtain Φ_n defined in (2.9) we get

$$\Phi_n = (\gamma_2 - \gamma_1)\|S_n\| + (n/2)(\gamma_1^2 - \gamma_2^2) \tag{6.6}$$

and comparison between (6.2) and (6.6), using (6.5), shows that (2.10) is satisfied so that we may replace L_n by Φ_n in order to investigate exponential boundedness of N.

As in Example 2, in order to prove exponential boundedness of N it suffices to exclude distributions P satisfying (3.2) with (θ', θ'') of the form $(g\theta_1, g\theta_2)$, $g \in G$ (again, these are the pairs that minimize the Kullback-Leibler divergence). Equating $p_\theta(Z_1)$ for these pairs, using (6.1), yields $2(\gamma_2 - \gamma_1)u'Z_1 + \gamma_1^2 - \gamma_2^2 = 0$, with $u \in U$. If this is not to hold a.e. P then P must satisfy

$$P(u'Z_1 + b = 0) < 1 \tag{6.7}$$

for every $u \in U$ in which $b = -\frac{1}{2}(\gamma_1 + \gamma_2)$. On the other hand, (6.6), after dividing by the immaterial factor $\gamma_2 - \gamma_1$, may be written

$$\Phi_n = \|S_n\| + bn. \tag{6.8}$$

It is seen that (6.7), (6.8) is a special case of (5.15), (5.16), with $a = 0$, and therefore exponential boundedness of N follows from Lemma 5.1.

7. Representation of orbit density ratio as ratio of integrals over the group G

Suppose X is a Euclidean space carrying a family of probability densities $p_\theta(x)$ with respect to Lebesgue measure in X, $\theta \in \Theta$. Suppose G is a group of invariance transformations, implying (among other things) that for every $g \in G$, $\theta \in \Theta$ and measurable set $A \subset X$

$$\int_{gA} p_{g\theta}(x)\,dx = \int_A p_\theta(x)\,dx \tag{7.1}$$

in which dx is Lebesgue measure in X. Let Q_θ be the distribution on the orbit space X/G and assume it has a density q_θ with respect to some sigma-finite measure (since we shall need q_θ only for two values of θ, this assumption is justified), where $q_\theta(x)$ depends on x only through its orbit Gx. Let θ_1, θ_2 have distinct orbits $G\theta_1$, $G\theta_2$, and consider the orbit density ratio $R(x) = q_{\theta_2}(x)/q_{\theta_1}(x)$. If G acts on X linearly, it was shown in [27], Equation (3) that for all x, except possibly for x in a set of Lebesgue measure zero, $R(x)$ can be expressed as

$$R(x) = J(\theta_1, x)/J(\theta_2, x) \tag{7.2}$$

with

$$J(\theta, x) = \int p_\theta(gx)|g|\mu_G(dg), \tag{7.3}$$

in which $|g|$ is the Jacobian of the transformation $x \to gx$ and μ_G is left Haar measure on G. Using (7.1) it is easy to establish that

$$p_\theta(gx)|g| = p_{g^{-1}\theta}(x) \tag{7.4}$$

and substitution of (7.4) into (7.3) gives

$$J(\theta, x) = \int p_{g^{-1}\theta}(x)\mu_G(dg). \tag{7.5}$$

The measure ν_G defined by $\nu_G(dg) = \mu_G(dg^{-1})$ is right Haar measure on G and therefore (7.5) can be put in the form

$$J(\theta, x) = \int p_{g\theta}(x)\nu_G(dg). \tag{7.6}$$

This leads immediately to (2.1), (2.2) provided (7.2), (7.3) can be justified also for G whose action on X consists not only of linear transformations but also of translations. The derivation of (3) in [27] depended on the nature of the action of G only insofar the existence of a flat local cross section at almost every x had been shown only for linear Cartan G spaces [26], Theorem 2. We shall extend this result now and show the existence of a flat local cross section at almost all x when G consists of linear transformations as well as translations. Basic to the proof are certain results of Palais [17]. Definitions in [17] and [26] will be used freely. No use will be made of Lemma 5 in [26].

THEOREM 7.1. *Let $X_1, \cdots, X_n$ be copies of E^d, where $n > d \geqq 1$, and let $X = \prod_1^n X_i$. Let $G = LH$, in which L is a Lie group of linear transformations on E^d (with the usual topology as a matrix group) and H is the group of all translations of a subspace $B \subset E^d$ such that $LB = B$. Let the action of G on E^d be defined as follows: if $g = \ell h$, $\ell \in L$ corresponding to the d^2 nonsingular matrix C, $h \in H$ corresponding to the vector $b \in B$, then $gz = C(z + b)$, $z \in E^d$. Let the action of G on X be defined by $g(x_1, \cdots, x_n) = (gx_1, \cdots, gx_n)$, $x_i \in X_i$, $i = 1, \cdots, n$. Then there is an open subset X_0 of X, invariant under G and $X - X_0$ having zero Lebesgue measure, such that each point $x \in X_0$ admits a flat local cross section.*

Before proving the theorem we shall prove a lemma. Also, the definition of G needs discussion. As an analytic manifold G is defined as $L \times H$. Now there are two ways of defining the multiplication in order to make G into a group. The first one of these is consistent with the way we have defined the action of G on E^d (that is, $gz = C(z + b)$) and is given by

$$(C_1, b_1)(C_2, b_2) = (C_1C_2, b_2 + C_2^{-1}b_1), \tag{7.7}$$

$C_1, C_2 \in L$, $b_1, b_2 \in B$, where it should be observed that $b_2 + C_2^{-1}b_1 \in B$ by virtue of the assumption $LB = B$. From (7.7) follows that the elements of G of the form $(C, 0)$ form a subgroup isomorphic to L, and the elements of the form (I, b) (I = identity matrix) form a subgroup isomorphic to H. Furthermore, also from (7.7), $(C, 0)(I, b) = (C, b)$ so that the group multiplication (7.7) is consistent with writing the elements $g \in G$ in a unique way as $g = \ell h$, $\ell \in L$, $h \in H$. This suggests the notation $G = LH$. It is also easily verified that $g^{-1}hg \in H$, $g \in G$, $h \in H$, so that H is normal in G. The second way of defining the group multiplication, which also results in H being normal and which is consistent with the group action $gz = Cz + b$, is given by $(C_1, b_1)(C_2, b_2) = (C_1C_2, b_1 + C_1b_2)$. This is consistent with writing $g \in G$ in a unique way as $g = h\ell$, $h \in H$, $\ell \in L$, and we would have denoted $G = HL$. Either way of defining G from L and H is called a *semidirect product* (see (2.6), (6.20) in [8] and 30D in [16]). It is immaterial which of the two ways one chooses, as long as the action of G on E^d is defined in accordance. We have arbitrarily chosen the first way. After the proof of Theorem 7.1 we shall deal with Haar measure on G.

LEMMA 7.1. *Let $V_1, \cdots, V_d$ be copies of E^d and let L be a Lie group of d^2 matrices (with the usual topology) acting linearly on E^d. Let $V = \Pi_1^d V_i$, consider the points $v \in V$ as d^2 matrices $v = [v_1, \cdots, v_d]$, $v_i \in V_i$ being a $d \times 1$ column vector, and let $V_0 = \{v \in V: v \text{ is a nonsingular matrix}\}$. Then V_0 is a linear Cartan L space.*

PROOF. Since L is contained in the group $GL(d, R)$ of all real d^2 nonsingular matrices and L has the relative topology of $GL(d, R)$, the conclusion of the lemma is true for L if it is true for $GL(d, R)$. We shall therefore proceed to prove the lemma for $L = GL(d, R)$. Note that V_0 is a copy of L. Hence the task is to prove that L is a linear Cartan L space.

In order to show that every $\ell \in L$ has a thin neighborhood it is sufficient to do this for the d^2 identity matrix I. Let E be the d^2-dimensional Euclidean space in which L is embedded and observe that L has the relative topology of E. Define $U_0 = \{M \in E: \operatorname{tr} MM' \leqq c\}$ in which $c > 0$ is chosen so that $I + M$ is nonsingular for every $M \in U_0$ (it can be shown that any $c < 1$ will do). Then $U = I + U_0 \subset L$ is a compact neighborhood of I. It follows from the continuity of inversion and group multiplication that UU^{-1} is compact. Let $A = \{\ell \in L: \ell U \cap U \neq \emptyset\}$ then $A \subset UU^{-1}$ so that A has compact closure and therefore U is a thin neighborhood of I. *Q.E.D.*

PROOF OF THEOREM 7.1. Choose any n^2 orthogonal matrix Ω whose last column has all its entries equal to $n^{-1/2}$. Write $X = [X_1, \cdots, X_n]$ and let $W = [W_1, \cdots, W_n]$ be defined by $W = X\Omega$. It suffices to prove the conclusion of the theorem for W instead of X. The action of G on W is determined by the action of G on X. An easy calculation shows that L acts on the W_i in the same way as on the X_i. On the other hand, H acts trivially on $W_1, \cdots, W_{n-1}$ whereas the action on W_n is given by $hw_n = w_n + bn^{1/2}$ if h corresponds to $b \in B$. Write $W_n = W_n^{(1)} \times W_n^{(2)}$, where $W_n^{(1)} = B$ and $W_n^{(2)}$ its orthogonal complement in W_n. Thus, $LW_n^{(1)} = W_n^{(1)}$ and H is transitive on $W_n^{(1)}$ and acts trivially on $W_n^{(2)}$. It is easily established from the invariance of $W_n^{(1)}$ under L that the action of L on W_n induces an action of L on $W_n^{(2)}$ so that $W_n^{(2)}$ is a linear L space. Put $V = \Pi_1^d W_i$ and let V_0 be the subspace of V as defined in Lemma 7.1 so that $V - V_0$ is an invariant nullset. Write Y for $(\Pi_{d+1}^{n-1} W_i) \times W_n^{(2)}$, or $Y = W_n^{(2)}$ if $n = d + 1$, and Z for $W_n^{(1)}$, then Y is a linear L space and

$$W_0 = V_0 \times Y \times Z \tag{7.8}$$

differs from W by an invariant nullset. Since by Lemma 7.1 V_0 is a Cartan L

space and Y is also an L space, it follows from [17] Proposition 1.3.3 that $V_0 \times Y$ is a Cartan L space (it is also linear but that is not being used). Using the fact that H is a group of translations of Z it is elementary to establish that W_0 is a Cartan G space. Furthermore, from the nature of L acting on V_0 it is obvious that the isotropy group is trivial at each point of V_0 and it follows that G acting on W_0 has trivial isotropy group G_x at each $x \in W_0$. Using [17] Lemma 2.2 there is at $x \in W_0$ a flat near slice (take as S^* in that lemma a translate through x of any linear complement to $(Gx)_x$, taking into account that G_x is trivial). By [17] Proposition 2.1.7 this near slice at x contains a slice S at x. Using [26] Lemma 3 and the fact that G_s is trivial for every $s \in S$, we conclude that S is a local cross section. This concludes the proof of Theorem 7.1.

We conclude this section by showing how Haar measure on $G = LH$ can be obtained from the Haar measures on L and H. Our G is a special case of the following general semidirect product. Given Lie groups L and H such that for each $\ell \in L$ there is an automorphism σ_ℓ of H satisfying $\sigma_{\ell_1\ell_2} = \sigma_{\ell_2}\sigma_{\ell_1}, \ell_1, \ell_2 \in L$, (in our G, σ_ℓ is the automorphism $b \to C^{-1}b$ if C corresponds to ℓ), and moreover the function $(\ell, h) \to \sigma_\ell(h)$ is an analytic mapping of $L \times H$ onto H. Then G is defined as the analytic manifold $L \times H$ with the group multiplication $(\ell_1, h_1)(\ell_2, h_2) = (\ell_1\ell_2, \sigma_{\ell_2}(h_1)h_2)$ (see 30D in [16]). (With this group multiplication and the obvious identification of L and H with subgroups of G it is readily verified that $\sigma_\ell(h) = \ell^{-1}h\ell$.) Let μ_L, μ_H be left Haar measure on L, H, respectively. Then the product measure $\mu_L \times \mu_H$ on $L \times H$ is left Haar measure on G(see 30D, E in [16]). This can also be expressed as follows: let ν_L, ν_H be right Haar measure on L, H, respectively, and let A be any measurable set in G, then the measure $\mu_G(\nu_G)$ defined by

$$\mu_G A = \iint_{\ell h \in A} \mu_L(d\ell)\mu_H(dh), \tag{7.9}$$

$$\nu_G A = \iint_{h\ell \in A} \nu_L(d\ell)\nu_H(dh) \tag{7.10}$$

is left (right) Haar measure on G. Here (7.10) follows from (7.9) using the familiar fact that for any group K, if μ_K is a left invariant measure on K then ν_K defined by $\nu_K A = \mu_K A^{-1}$ is right invariant. We can write (7.9), (7.10) also in the following form: let f be integrable on G with respect to $\mu_G(\nu_G)$, then

$$\int f(g)\mu_G(dg) = \iint f(\ell h)\mu_L(d\ell)\mu_H(dh), \tag{7.11}$$

$$\int f(g)\nu_G(dg) = \iint f(h\ell)\nu_L(d\ell)\nu_H(dh). \tag{7.12}$$

The fact that the integral on the right side of (7.11) (of (7.12)) is unchanged if $f(g)$ is replaced by $f(g_1g)$ (by $f(gg_1)$), for any $g_1 \in G$, can of course also be checked directly (as in (15.29)(a) of [8]).

APPLICATIONS. Using (7.11), the function $\psi(h)$ in Equation (10) of [27] could have been omitted by writing on the right side $p(hgx_0)$ instead of $p(ghx_0)$.

Using (7.12) and observing that right (= left) Haar measure on H can simply be taken as Lebesgue measure db on B, we can write (2.2) now in the form

$$J_n(\theta) = \iint \prod_{i=1}^{n} p_{h\ell\theta}(Z_i)\nu_L(d\ell)\, db \tag{7.13}$$

and (2.6) in the form

$$J_n(\theta) = \iint \exp\{n\psi_\theta(h\ell, \bar{X}_n)\}\nu_L(d\ell)\, db. \tag{7.14}$$

◇ ◇ ◇ ◇ ◇

I would like to thank R. H. Berk for several valuable comments.

REFERENCES

[1] M. S. ABU-SALIH, "On termination with probability one and bounds on sample size distribution of sequential probability ratio tests where the underlying model is an exponential family," unpublished Ph.D. thesis, University of Illinois, 1969.

[2] R. H. BERK, "Asymptotic properties of sequential probability ratio tests," unpublished Ph.D. thesis, Harvard University, 1964.

[3] ———, "Consistency a posteriori," *Ann. Math. Statist.*, Vol. 41 (1970), pp. 894–906.

[4] ———, "Stopping time of SPRTS based on exchangeable models," *Ann. Math. Statist.*, Vol. 41 (1970), pp. 979–990.

[5] D. BLACKWELL and D. FREEDMAN, "A remark on the coin tossing game," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 1345–1347.

[6] N. G. DE BRUIJN, "*Asymptotic methods in analysis*," Amsterdam, North-Holland Publishing Co., 1958.

[7] H. T. DAVID and W. H. KRUSKAL, "The WAGR sequential t-test reaches a decision with probability one," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 797–805 and Vol. 29 (1958), p. 936.

[8] E. HEWITT and K. A. ROSS, *Abstract Harmonic Analysis*, Berlin-Heidelberg-New York, Springer, 1963.

[9] A. F. IFRAM, "On the asymptotic behavior of densities with applications to sequential analysis," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 615–637.

[10] ———, "Hypergeometric functions in sequential analysis," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 1870–1872.

[11] ———, "On the sample size and simplification of a class of sequential probability ratio tests," *Ann. Math. Statist.*, Vol. 37 (1966), pp. 425–434.

[12] J. E. JACKSON and R. A. BRADLEY, "Sequential χ^2- and T^2-tests," *Ann. Math. Statist.*, Vol. 32 (1961), pp. 1063–1077.

[13] N. L. JOHNSON, "Some notes on the application of sequential methods in the analysis of variance," *Ann. Math. Statist.*, Vol. 24 (1953), pp. 614–623.

[14] S. KULLBACK and R. A. LEIBLER, "On information and sufficiency," *Ann. Math. Statist.*, Vol. 22 (1951), pp. 79–86.

[15] E. L. LEHMANN, *Testing Statistical Hypotheses*, New York, Wiley, 1959.

[16] L. H. LOOMIS, *An Introduction to Abstract Harmonic Analysis*, New York, Van Nostrand, 1953.

[17] R. S. PALAIS, "On the existence of slices for actions of non-compact Lie groups," *Ann. Math.*, Vol. 73 (1961), pp. 295–323.

[18] W. D. RAY, "A proof that the sequential probability ratio test (S.R.P.T.) of the general linear hypothesis terminates with probability unity," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 521–523.

[19] 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 and Vol. 38 (1967), p. 1309.

[20] 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.

[21] C. STEIN, "A note on cumulative sums," *Ann. Math. Statist.* Vol. 17 (1946), pp. 498–499.

[22] A. WALD, "On cumulative sums of random variables," *Ann. Math. Statist.*, Vol. 15 (1944), pp. 283–296.

[23] ———, *Sequential Analysis*, New York, Wiley, 1947.

[24] G. N. WATSON, "Asymptotic expansions of hypergeometric functions," *Cambridge Phil. Soc. Trans.*, Vol. 22 (1912), pp. 277–308.

[25] ———, *A Treatise on the Theory of Bessel Functions*, Cambridge, Cambridge University Press, 1922.

[26] R. A. WIJSMAN, "Existence of local cross-sections in linear Cartan G-spaces under the action of noncompact groups," *Proc. Amer. Math. Soc.*, Vol. 17 (1966), pp. 295–301.

[27] ———, "Cross-sections of orbits and their application to densities of maximal invariants," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 1, pp. 389–400.

[28] ———, "General proof of termination with probability one of invariant sequential probability ratio tests based on multivariate normal observations," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 8–24.

[29] ———, "Bounds on the sample size distribution for a class of invariant sequential probability ratio tests," *Ann. Math. Statist.*, Vol. 39 (1968), pp. 1048–1056.

[30] S. WIRJOSUDIRJO, "Limiting behavior of a sequence of density ratios," unpublished Ph.D. thesis, University of Illinois, 1961. (Abstract in *Ann. Math. Statist.*, Vol. 33 (1962), pp. 296–297.)