GENERAL PROOF OF TERMINATION WITH PROBABILITY ONE OF INVARIANT SEQUENTIAL PROBABILITY RATIO TESTS BASED ON MULTIVARIATE NORMAL OBSERVATIONS[1]

By R. A. Wijsman

University of Illinois[2]

0. Summary. $Z_1, Z_2, \cdots$ is a sequence of iid k-vectors with common distribution P. G^* is a group of transformations $Z_n \to CZ_n + b$, $C \varepsilon G$, where G is a Lie group of k^2 matrices, dim $G \geqq 1$, G closed in the group of all nonsingular k^2 matrices, and the totality of translation vectors b is a subspace of k-space invariant under G. Let $\mathfrak{N}$ be all $N(\mu, \Sigma)$ distributions, with Σ k^2 nonsingular. Let $U = (U_1, U_2, \cdots)$ be a maximal invariant under G^* in the sample space, $\gamma = \gamma(\theta)$ a maximal invariant in $\mathfrak{N}$, where $\theta = (\mu, \Sigma)$. For given $\theta_1, \theta_2 \varepsilon \mathfrak{N}$ such that $\gamma(\theta_1) \neq \gamma(\theta_2)$ let R_n be the probability ratio of $(U_1, \cdots, U_n)$. The limiting behavior of R_n is studied under the assumption that the actual distribution P belongs to a family $\mathfrak{F} \supset \mathfrak{N}$, defined as follows: the components of Z_1 have finite 4th moments, and there is no relation $Z_1'AZ_1 + b'Z_1 =$ constant a.e. P, with A symmetric, unless $A = 0, b = 0$. It is proved that $\mathfrak{F}$ can be partitioned into 3 subfamilies, and for every P in the first subfamily $\lim R_n = \infty$ a.e. P, in the second $\lim R_n = 0$ a.e. P, and in the third $\lim\sup R_n = \infty$ a.e. P or $\lim\inf R_n = 0$ a.e. P. This implies that any SPRT based on $\{R_n\}$ terminates with probability one for every $P \varepsilon \mathfrak{F}$.

1. Introduction. There are many testing problems where it is possible to eliminate nuisance parameters by invoking the principle of invariance. Among the parametric problems it is especially in problems involving normally distributed variables that the success of invariance has been spectacular. Application of the principle of invariance to nonparametric problems is no less important but will fall entirely without the scope of this paper. We shall restrict ourselves to sequential tests of composite hypotheses, based on a normal model, where nuisance parameters can be eliminated by using the principle of invariance. More specifically, we shall investigate the problem whether a sequential probability ratio test (SPRT) based on a maximal invariant sequence terminates with probability one. Throughout, it should be kept in mind that for the study of the behavior of any SPRT one needs three distributions: two to define the sequence of probability ratios, whereas the third one is the "actual" distribution of the observations and need not belong to the model. In fact, the wider this last class of pos-

Received 12 April 1966.

[1] Research was supported in part by the National Science Foundation under Grant G-21507.

[2] Most of the research was done at Cornell University during a sabbatical leave from the University of Illinois. I greatly appreciate the use of facilities that were made available to me by the Department of Mathematics at Cornell University.

sible actual distributions is relative to which termination with probability one can be proved, the better.

Let $U = (U_1, U_2, \cdots)$ be a sequence of random variables, and, for two given distributions of U, let R_n be the probability ratio of $(U_1, \cdots, U_n)$. In Wald's SPRT [13] of a simple hypothesis against a simple alternative the U_n are independent and identically distributed (iid) and termination with probability one is easy to prove for *arbitrary* actual common distribution of the U_n (excepting only the distribution according to which the probability ratio of one observation equals 1 with probability one). However, if the sequence U arises from the application of the principle of invariance, the U_n are usually not iid and the termination proof is incomparably more difficult. For several well-known sequential tests of composite hypotheses individual proofs have been given; e.g. for the sequential t-test by David and Kruskal [4]; for special cases of the sequential F-test by Johnson [9] and Ray [11]; for the sequential χ^2- and T^2-test by Jackson and Bradley [8]. Results of greater generality, making various assumptions, have been obtained by Wirjosudirjo [15], Ifram [7] and Berk [1].

This paper offers a rather general termination proof, including the above mentioned tests as special cases. The sequence $\{R_n\}$ of probability ratios is computed under the assumption that the observations are iid multivariate normal and the invariance group is a Lie group of affine transformations subject to some weak restrictions (Assumption A, Section 2). The actual common distribution of the observations, under which the behavior of $\{R_n\}$ is studied, may be almost completely arbitrary (family $\mathfrak{F}$, Section 2) although we have to exclude certain distributions for which termination cannot be proved with the present methods. This is somewhat unfortunate from an aesthetical point of view, even though it probably matters little for applications. Note that also in the existing termination proofs, mentioned above, certain restrictions have to be placed on the actual underlying distribution. This is true also for the recent proof, by Savage and Sethuraman [12], of termination with probability one of an invariant SPRT in a nonparametric problem. It may be conjectured that termination with probability one is valid for a wider family than $\mathfrak{F}$, but this remains to be investigated. In any case, the restrictions are very mild. Certainly every nonsingular multivariate normal distribution belongs to $\mathfrak{F}$.

The main tool used in the proof is the possibility of writing the probability ratio of a maximal invariant as the ratio of two integrals with respect to Haar measure on the group of transformations. This method seems to have been introduced into statistics by C. M. Stein (for references see [14]). Whereas in the case of a known group of transformations there are usually other methods available for obtaining the probability ratio, in our case the group is almost completely unknown, subject only to Assumption A, and the method of "integration over the group" seems the only possible way to get a handle on the probability ratio. In the ratio of integrals the asymptotic behavior of numerator and denominator is studied, applying what amounts to the idea of Laplace's method to integration over a Lie group. It is Ifram's success with Laplace's method in [7] that led us to believe that it would work also in the present more general problem where integration takes place over a Lie group.

2. The theorem: statement and examples. Throughout the random variables $Z_1, Z_2, \cdots$ are assumed to be iid k-vectors with common distribution P. The joint distribution of $Z_1, Z_2, \cdots$ will also be denoted P. Let $\bar{Z}_n = (1/n) \sum_{i=1}^n Z_i$ and $S_n = (1/n) \sum_{i=1}^n (Z_i - \bar{Z}_n)(Z_i - \bar{Z}_n)'$ be the sample mean and covariance matrix of $Z_1, \cdots, Z_n$. We consider two families of distributions to which P may belong. The first, denoted $\mathfrak{N}$, is the family of all k-variate normal distributions $N(\mu, \Sigma)$, Σ nonsingular, indexed by $\theta = (\mu, \Sigma)$. The second is defined as follows:

DEFINITION. $\mathfrak{F}$ is the family of all distributions P such that the components of Z_1 have finite 4th moments and, if A is a k^2 symmetric matrix and b a k-vector, then $P(Z_1'AZ_1 + b'Z_1 = \text{constant}) = 1$ implies $A = 0$, $b = 0$.

For any $P \varepsilon \mathfrak{F}$ we denote by μ the mean and by Σ the covariance matrix of Z_1. As in the case of the family $\mathfrak{N}$ we put $\theta = (\mu, \Sigma)$. If it is important to stress the dependence of θ on P we shall write $\theta(P)$. It follows from the definition of $\mathfrak{F}$, by taking $A = 0$, that Σ is nonsingular. It is also easy to see that any nonsingular k-variate normal distribution satisfies the conditions of $P \varepsilon \mathfrak{F}$, so that $\mathfrak{F} \supset \mathfrak{N}$.

Concerning the group G^* of invariance transformations we make the following assumption:

ASSUMPTION A. $G^* = GH$, where (i) G is a Lie subgroup of $GL(k, R)$ (i.e. the general linear group of all real nonsingular k^2 matrices); (ii) G is closed in $GL(k, R)$ and of dimension $\geqq 1$; (iii) H is a group of translations of k-space with k-vectors b, the totality of vectors b constituting a subspace invariant under G; (iv) each transformation $g^* = (C, b)$, $C \varepsilon G$, $b \varepsilon H$, transforms $(Z_1, Z_2, \cdots)$ according to $Z_n \to CZ_n + b$, $n = 1, 2, \cdots$.

Under the group G^* let $U = (U_1, U_2, \cdots)$ be a maximal invariant in the sample space, $\gamma = \gamma(\theta)$ a maximal invariant in $\mathfrak{N}$. If $P \varepsilon \mathfrak{N}$, the distribution of U depends on θ only through γ [10]. Let $\theta_1, \theta_2 \varepsilon \mathfrak{N}$ be such that $\gamma_1 \neq \gamma_2$, where $\gamma_i = \gamma(\theta_i)$, $i = 1, 2$. Denote $U^n = (U_1, \cdots, U_n)$ and let p'_{in} be its density under γ_i, $i = 1, 2$, with respect to some common sigma-finite measure. Denote $r_n = p'_{2n}/p'_{1n}$ and

$$R_n = r_n(U^n), \tag{2.1}$$

then R_n is the probability ratio at the nth stage of sampling based on the maximal invariant U. A SPRT based on $\{R_n\}$ continues sampling as long as R_n is between two fixed stopping bounds, taking the appropriate decision the first time one of the bounds is exceeded. We shall call a SPRT based on $\{R_n\}$ also an *invariant* SPRT.

THEOREM. *Let* $Z_1, Z_2, \cdots$ *be* iid *with common distribution* $P \varepsilon \mathfrak{F}$, G^* *satisfy Assumption* A, *and* R_n *be defined by* (2.1). *Then* $\lim R_n = \infty$ a.e. P *or* $= 0$ a.e. P *according as* $\Phi(\theta) > 0$ *or* < 0, *where* $\theta = \theta(P)$ *and* Φ *is defined in* (3.9). *For* P

such that $\Phi(\theta) = 0$, $\limsup R_n = \infty$ a.e. P *or* $\liminf R_n = 0$ a.e. P. *Consequently, any invariant* SPRT *terminates with P-probability one if* $P \varepsilon \mathfrak{F}$.

Before proving this theorem it may be helpful to see how some of the well-known sequential tests fit into this scheme.

EXAMPLE 1. (sequential t-test). $Z_1, Z_2, \cdots$ are iid normal (univariate) with unknown mean μ and unknown standard deviation σ. Suppose the two hypotheses are $\mu/\sigma = \gamma_1$ and $\mu/\sigma = \gamma_2$. The group of invariance transformations transforms $Z_n \rightarrow cZ_n$, $c > 0$, so that H is trivial and G consists of the multiplicative group of positive reals, which is a one-dimensional subgroup of $GL(1, R)$ and closed in $GL(1, R)$. The conclusion of the Theorem applies then to the one-sided sequential t-test, whose R_n equals the probability ratio of student's t-statistic computed from $Z_1, \cdots, Z_n$. The same conclusion holds for the two-sided sequential t-test if we allow c to be <0 as well as >0, i.e. $G = GL(1, R)$.

EXAMPLE 2 (sequential T^2-test). The k-vectors $Z_1, Z_2, \cdots$ are iid $N(\mu, \Sigma)$ and $\gamma = \mu'\Sigma^{-1}\mu$. Here H is trivial and $G = GL(k, R)$. The conclusion of the theorem applies then to the sequential test whose R_n equals the probability ratio of Hotelling's T^2-statistic computed from $Z_1, \cdots, Z_n$.

EXAMPLE 3 (sequential multiple correlation coefficient test). The k-vectors $Z_1, Z_2, \cdots$ are iid $N(\mu, \Sigma)$ and $\gamma = \sigma_{11}^{-\frac{1}{2}}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{\frac{1}{2}}$ is the multiple correlation coefficient between the first and the $k - 1$ remaining variates, where Σ is partitioned in the usual way. Here G consists of all matrices C that have the form diag (c, C_{22}), c real $\neq 0$ and C_{22} $(k - 1)^2$ nonsingular. H consists of all translations of k-space, so that the subspace of Assumption A (iii) is all of k-space and is clearly invariant under G. The conclusion of the Theorem applies then to the sequential multiple correlation coefficient test [7] whose R_n is the probability ratio of the multiple correlation computed from $Z_1, \cdots, Z_n$.

EXAMPLE 4 (sequential F-test). The k-vectors $Z_1, Z_2, \cdots$ are iid $N(\mu, \sigma^2 I_k)$; each Z_n is partitioned into Z_{n1}, Z_{n2}, Z_{n3}, where Z_{ni} is a k_i-vector, $k_1 + k_2 + k_3 = k$, with μ partitioned similarly; μ_2 is known to be 0 whereas σ, μ_1 and μ_3 are unknown, and $\gamma = \mu_1'\mu_1/\sigma^2$. The group of invariance transformations contains all translations $Z_{n3} \rightarrow Z_{n3} + b_3$, i.e. H consists of all translation vectors b with $b_1 = b_2 = 0$. The group of linear transformations G consists of all matrices $c\Omega$, $c > 0$ and $\Omega = \text{diag}(\Omega_{11}, \Omega_{22}, \Omega_{33})$ with Ω_{ii} being k_i^2 orthogonal. If k-space E^k is written $E_1 \times E_2 \times E_3$, with E_i being k_i-space, then G leaves each E_i invariant. In particular G leaves invariant E_3, which is the totality of all translation vectors b under H. Thus, Assumption A is satisfied and the conclusion of the theorem applies to the sequential F-test, whose R_n is the probability ratio of the usual F-statistic computed from $Z_1, \cdots, Z_n$.

3. Proof of the Theorem. Several lemmas are needed whose proofs are rather long. Instead of proving the lemmas first and then the theorem, we shall turn it around and state each lemma without proof the first time it is needed, while coming back to the proofs of the lemmas in Sections 4, 5 and 6. Furthermore, in order to avoid tiresome repetition, the phrase "a.e. P" will be suppressed whenever there is question of the asymptotic behavior of a sequence of random variables.

Since we are concerned with the derivation of an expression for r_n, the term "sufficient statistic" is used in the following relative to the family $\mathfrak{N}$ of multivariate normal distributions. For each n, let T_n be sufficient for U^n, then the probability ratio r_n of U^n equals the probability ratio of T_n so that we may denote the latter also by r_n. The sequence $T_1, T_2, \cdots$ (called an *invariantly sufficient sequence* in [6]), which was obtained by first applying invariance, then sufficiency, may also be obtained by applying sufficiency first and then invariance (this is proved in [6]). That is, for each n let V_n be sufficient for $Z_1, \cdots, Z_n$, then T_n is a maximal invariant obtained by applying the group of invariance transformations to V_n. For V_n we choose the sufficient statistic $V_n = (\bar{Z}_n, S_n)$, defined in Section 2. Then the transformation $Z_n \to CZ_n + b$ for each n, induces the following transformation of $V_n : \bar{Z}_n \to C\bar{Z}_n + b$, $S_n \to CS_nC'$. It is readily checked that part (iii) of Assumption A guarantees that H is a normal subgroup of G^*. This implies that a maximal invariant can be obtained by first applying H and then G [10]. Let the invariant subspace corresponding to H (Assumption A (iii)) be E_2 and write $E^k = E_1 \times E_2$, where E_1 has dimension l, $0 \leqq l \leqq k$. Without loss of generality we may assume the coordinate system chosen so that the first l coordinate axes span E_1 and the remaining span E_2. Let all vectors and matrices be partitioned according to E_1, E_2. The invariance of E_2 under G implies that for each $C \varepsilon G$ the partitioned C has $C_{12} = 0$, i.e. C has the form

$$C = \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix}. \tag{3.1}$$

The transformations of H are of the form $Z_{n2} \to Z_{n2} + b_2$, which induces on V_n the transformations $S_n \to S_n$, $\bar{Z}_{n2} \to \bar{Z}_{n2} + b_2$, $b_2 \varepsilon E_2$. A maximal invariant under H is $X_n = (\bar{Z}_{n1}, S_n)$ and the transformation of X_n induced by $C \varepsilon G$ is $\bar{Z}_{n1} \to C_{11}\bar{Z}_{n1}$, $S_n \to CS_nC'$. A maximal invariant under this group of transformations is what we have called T_n, and the first object is to find an expression for its probability ratio.

X_n takes its values in a space $\mathfrak{X}$ of points $x = (z, s)$, where z is an l-vector and s a k^2 positive definite matrix, so that $\mathfrak{X}$ is an open subset of E^q, where $q = l + (k(k+1)/2)$. Put $\zeta = \mu_1$ ($=$ projection of μ on E_1) and rename θ by putting $\theta = (\zeta, \Sigma)$, then if $P \varepsilon \mathfrak{N}$ the distribution of X_n is determined by θ. In fact, letting from now on $n \geqq k + 1$ and denoting by $p_{\theta n}$ the density of X_n with respect to Lebesgue measure in $\mathfrak{X}$, we have

$$\begin{aligned} p_{\theta n}(x) = {} & c_n|\Sigma|^{-(n-1)/2}|\Sigma_{11}|^{-\frac{1}{2}}|s|^{(n-k-2)/2} \\ & \cdot \exp\left[-(n/2) \operatorname{tr} \Sigma^{-1}s - (n/2)(z-\zeta)'\Sigma_{11}^{-1}(z-\zeta)\right], \end{aligned} \tag{3.2}$$

in which vertical bars around a matrix denote the absolute value of the determinant, and c_n is a constant depending on n whose value does not interest us since it will drop out when we form the probability ratio.

The group G will temporarily be considered a group of matrices g of linear transformations of $\mathfrak{X}$ onto intself according to $x \to gx$. Let μ_G be left Haar measure on G. Then it is proved in [14], Theorems 2 and 4, that the probability ratio of a maximal invariant under G can be written as

$$(3.3) \qquad r_n(x) = \int p_{\theta_2 n}(gx)|g|\mu_G(dg)/\int p_{\theta_1 n}(gx)|g|\mu_G(dg).$$

Returning now to our concept of G as a group of k^2 matrices C, we observe that the connection between g and C is as follows: the transformation $x \to gx$ is given by $(z, s) \to (C_{11}z, CsC')$. From this we compute $|g| = |C|^{k+1}|C_{11}|$. Before substituting this and (3.2) into (3.3), it is convenient to introduce

$$(3.4) \quad \psi(x, \theta, C) = -\tfrac{1}{2}\mathrm{tr}\, \Sigma^{-1}CsC' - \tfrac{1}{2}(C_{11}z - \zeta)'\Sigma_{11}^{-1}(C_{11}z - \zeta) + \ln |C| - \tfrac{1}{2}\ln |\Sigma|$$

and

$$(3.5) \qquad J_{n\theta}(x) = \int e^{n\psi(x,\theta,C)}|C|^{-1}|C_{11}|\mu_G(dC).$$

Then we can express r_n as

$$(3.6) \qquad r_n(x) = KJ_{n\theta_2}(x)/J_{n\theta_1}(x)$$

in which K is a constant depending only on θ_1 and θ_2.

LEMMA 1. *For every x and θ, ψ defined in* (3.4) *has a maximum when C varies over G.*

Define

$$(3.7) \qquad \varphi(x, \theta) = \max_{C\varepsilon G}\psi(x, \theta, C).$$

LEMMA 2. *For every fixed θ, $\varphi(\cdot, \theta)$ given by* (3.7) *is continuous on $\mathfrak{X}$.*

LEMMA 3. *Let N be an open subset of $\mathfrak{X}$, such that* tr s, tr s^{-1} *and $\|z\|$ are bounded on N. Let d be the dimension of G. There exist $K_1(\theta)$, $K_2(\theta)$, both >0 (and dependent on N) such that for $x \varepsilon N$*

$$(3.8) \qquad K_1(\theta)n^{-d}e^{n\varphi(x,\theta)} < J_{n\theta}(x) < K_2(\theta)e^{n\varphi(x,\theta)}.$$

Substituting (3.8) into (3.6) we find

$$(3.9) \qquad K_1n^{-d}e^{n\Phi(x)} < r_n(x) < K_2n^de^{n\Phi(x)}, \qquad x \varepsilon N,$$

in which K_1, K_2 are some positive constants (depending on N), and

$$(3.10) \qquad \Phi(x) = \varphi(x, \theta_2) - \varphi(x, \theta_1).$$

R_n, defined by (2.1), also equals $r_n(X_n)$. The asymptotic behavior of $\{R_n\}$ will follow from (3.9) and the strong law of large numbers applied to $\{X_n\}$. We have, for $P \varepsilon \mathfrak{F}$, $\bar{Z}_{n1} \to \zeta$ and $S_n \to \Sigma$, so that $X_n \to \theta$. Choose the set N in (3.9) to be a neighborhood of θ then $X_n \varepsilon N$ eventually. We have from (3.9) for sufficiently large n:

$$(3.11) \quad n^{-\frac{1}{2}}\ln (K_1n^{-d}) + n^{\frac{1}{2}}\Phi(X_n) < n^{-\frac{1}{2}}\ln R_n < n^{-\frac{1}{2}}\ln (K_2n^d) + n^{\frac{1}{2}}\Phi(X_n).$$

Letting $n \to \infty$ we get

$$\limsup n^{-\frac{1}{2}} \ln R_n = \limsup n^{\frac{1}{2}} \Phi(X_n),$$
$$\liminf n^{-\frac{1}{2}} \ln R_n = \liminf n^{\frac{1}{2}} \Phi(X_n). \tag{3.12}$$

In the following it should be kept in mind that Φ, defined in (3.10), is continuous by virtue of Lemma 2, and that $X_n \to \theta$. We distinguish three cases, according as $\Phi(\theta) > 0$, <0, or $=0$.

CASE 1. $\Phi(\theta) > 0$. Then the right hand sides of both equations (3.12) are ∞, so that $R_n \to \infty$.

CASE 2. $\Phi(\theta) < 0$. Then the right hand sides of both equations (3.12) are $-\infty$, so that $\ln R_n \to -\infty$ and therefore $R_n \to 0$.

CASE 3. $\Phi(\theta) = 0$. We shall show $\limsup n^{\frac{1}{2}}\Phi(X_n) = \infty$ or $\liminf n^{\frac{1}{2}}\Phi(X_n) = -\infty$, so that $\limsup R_n = \infty$ or $\liminf R_n = 0$. To prove this, several more lemmas are needed.

LEMMA 4. *The directional derivative* $\Phi'(x_0, x)$ *defined by*

$$\Phi'(x_0, x) = \lim_{t \downarrow 0} t^{-1}[\Phi(x_0 + tx) - \Phi(x_0)] \tag{3.13}$$

exists for all $x_0 \varepsilon \mathfrak{X}$, $x \varepsilon E^q$, *and* $\gamma(\theta_1) \neq \gamma(\theta_2)$ *implies that for every* x_0, $\Phi'(x_0, \cdot)$ *is not identically equal* 0.

LEMMA 5. *For any set of vectors* $a_1, \cdots, a_m \varepsilon E^q$ *such that the convex cone* $C_0 = \{x: \min_i a_i'x > 0\}$ *has positive Lebesgue measure, we have*

$$\limsup \min_i a_i' n^{\frac{1}{2}}(X_n - \theta) = \infty$$

if $P \varepsilon \mathfrak{F}$.

LEMMA 6. *If a real-valued function f on* $\mathfrak{X}$ *has the property that there exists a cone* C_0 *as in Lemma* 6, *and constants* $a, b > 0$ *such that* $f(x) \geqq b\,\|x\|$ *whenever* $x \varepsilon C_0$ *and* $\|x\| \leqq a$, *then* $\limsup n^{\frac{1}{2}} f(X_n - \theta) = \infty$ *if* $P \varepsilon \mathfrak{F}$.

LEMMA 7. *If* $\Phi(x_0) = 0$ *and* $\Phi'(x_0, \cdot) \not\equiv 0$, *and f is defined by* $f(x) = \Phi(x_0 + x)$, *then f or* $-f$ *satisfies the condition on the function f of Lemma* 6.

Case 3 can now be completed, using Lemmas 4 through 7. Remember that the θ_i used in forming the probability ratio (3.3) are chosen such that $\gamma(\theta_1) = \gamma_1 \neq \gamma_2 = \gamma(\theta_2)$, so that Lemma 4 applies. In Lemma 7 take $x_0 = \theta$, so that $f(x) = \Phi(\theta + x)$, then the conclusion of Lemma 6 holds for f or $-f$. That is, $\limsup n^{\frac{1}{2}}\Phi(X_n) = \infty$ or $\limsup -n^{\frac{1}{2}}\Phi(X_n) = \infty$. This finishes the proof of the theorem.

REMARKS. 1. Case 3 is by far the hardest case to prove, i.e. there may be certain exceptional values of θ (or rather of γ: this follows from the invariance of Φ) for which the termination proof is much more difficult. This difficulty with exceptional parameter values is typical of all termination proofs that have been given in specific cases (e.g. [4] [8]) as well as in the more general treatments [7] [15]. In [15] there is the possibility of existence of an exceptional value of γ for which termination with probability one cannot be proved. In [7], in order to assure termination at the exceptional value stronger assumptions have to be made than would be necessary otherwise. In the present paper, if it were not for Case 3 the definition of $\mathfrak{F}$ (Section 2) could have been widened to include all P with finite Σ (singular Σ permitted). The further restrictions on $P \varepsilon \mathfrak{F}$ are only used to also cover Case 3 by ensuring asymptotic normality of $n^{\frac{1}{2}}(X_n - \theta)$ (see proof of Lemma 5).

2. The existence of exceptional parameter values occurs even in a problem as simple as proving termination with probability one of Wald's SPRT based on iid observations. There, $\ln R_n = Y_1 + \cdots + Y_n$, where $Y_1, Y_2, \cdots$ are iid real valued random variables; i.e. $\ln R_n$ performs a random walk on the real line with iid steps. It follows that $\ln R_n \to \infty$ or $-\infty$ according as $E_\theta Y_1 > 0$ or <0. If $E_\theta Y_1 = 0$ for some exceptional parameter value θ, then $\ln R_n$ does not converge at all. Instead, for such θ, $\limsup \ln R_n = \infty$ and $\liminf \ln R_n = -\infty$. This same behavior of $\ln R_n$ at an exceptional parameter value has also been demonstrated in tests like the sequential t-test, etc. [7] [15]. Unfortunately, in the present paper that result is not quite achieved, since we have been able to prove only that at an exceptional θ (where $\Phi(\theta) = 0$), $\limsup \ln R_n = \infty$ *or* $\liminf \ln R_n = -\infty$. It is not known at the present whether in general the "or" can be replaced by "and".

3. Since the proof of the theorem (including the proofs of the lemmas) is rather long, it may be helpful to indicate some of the ideas in the proof. By (3.6) $r_n(x)$ is essentially the ratio of two integrals given by the right hand side of (3.5), written down for θ_1 and θ_2. If this were integration in Euclidean space, the asymptotic behavior of such an integral would be as $\exp [n \max_C \psi]$ times a factor depending only on n (Laplace's method). It turns out that this is still essentially correct if the integration takes place in a Lie group. Since, by (3.7), $\max_C \psi = \varphi$, we get $r_n(x) \sim \exp [n\Phi(x)]$, using (3.10). Thus, in cases 1 and 2, $r_n(\theta) \to \infty$ or 0 according as $\Phi(\theta) > 0$ or <0. The same conclusion holds then for R_n since $X_n \to \theta$ and Φ is continuous.

In order to sketch the idea behind the treatment of Case 3, imagine the simple situation where X_n and θ are real valued. The family of directional derivatives (indexed by x) of Lemma 4 has then only two members: the left and the right hand derivative of Φ at x_0. In Case 3, $\Phi(\theta) = 0$. According to Lemma 4, one of the one-sided derivatives at θ is $\neq 0$, say the right hand derivative equals $a \neq 0$. Since $X_n \to \theta$, the asymptotic behavior of $n^{\frac{1}{2}}\Phi(X_n)$ is as $a\, n^{\frac{1}{2}}(X_n - \theta)$ whenever $X_n > \theta$. It is not hard to see that $\limsup n^{\frac{1}{2}}(X_n - \theta) = \infty$, so that $\limsup n^{\frac{1}{2}}\Phi(X_n) = \infty$ if $a > 0$, and $\liminf n^{\frac{1}{2}}\Phi(X_n) = -\infty$ if $a < 0$.

4. Proofs of Lemmas 1, 2 and 3. First we prove a useful matrix lemma, which will be used repeatedly in the sequel.

LEMMA 8. *If A and B are k^2 matrices, A positive definite and B nonnegative definite, then*

$$\operatorname{tr} B/\operatorname{tr} A^{-1} \leqq \operatorname{tr} AB \leqq \operatorname{tr} A \operatorname{tr} B. \tag{4.1}$$

PROOF. To prove the right hand inequality it is sufficient to prove

$\operatorname{tr} C'AC \leqq \operatorname{tr} A \operatorname{tr} CC'$ for any k^2 matrix C (for then we can take $C = B^{\frac{1}{2}}$). Let $x_1, \cdots, x_k$ be the columns of C, then $\operatorname{tr} C'AC = \sum_{i=1}^k x_i'Ax_i$. Now $x_i'Ax_i \leqq \lambda_{\max}\|x_i\|^2 \leqq \operatorname{tr} A \|x_i\|^2$, where $\lambda_{\max}$ is the largest eigenvalue of A. Summing over i and noting $\sum_{i=1}^k \|x_i\|^2 = \operatorname{tr} CC'$, the result follows. The left hand side inequality is a consequence: $\operatorname{tr} B = \operatorname{tr} A^{-1}(A^{\frac{1}{2}}BA^{\frac{1}{2}}) \leqq \operatorname{tr} A^{-1} \operatorname{tr} (A^{\frac{1}{2}}BA^{\frac{1}{2}})$, using the right hand side inequality.

PROOF OF LEMMAS 1 AND 2. In the following it should be kept in mind that the topology in G is the relative topology of G as a subset of $GL(k, R)$: this follows from Assumption A (ii) that G be closed in $GL(k, R)$ [2] [3]. In other words, considering the k^2 matrices C as vectors in E^{k^2}, the topology of G is the relative Euclidean topology of E^{k^2}. It should also be noted that, since $GL(k, R)$ is not closed in E^{k^2}, a subset of G that is closed in G is not necessarily closed in E^{k^2}. Lastly, in the proof of Lemmas 1 and 2, with ψ defined in (3.4), it is sufficient to give the proof with Σ set equal to I_k, if we replace $\Sigma^{-\frac{1}{2}}C\Sigma^{\frac{1}{2}}$ by C, $\Sigma^{-\frac{1}{2}}s\Sigma^{-\frac{1}{2}}$ by s and $\Sigma_{11}^{-\frac{1}{2}}z$ by z (the group $\Sigma^{-\frac{1}{2}}G\Sigma^{\frac{1}{2}}$ is isomorphic to G). Multiplying the right hand side of (3.4) by 2, we put $f(x, C) = -\operatorname{tr} CsC' - \|C_{11}z - \zeta\|^2 + 2 \ln |C|$. Let $x_0 \varepsilon \mathfrak{X}$ be fixed and let N be a neighborhood of x_0 such that $\operatorname{tr} s$, $\operatorname{tr} s^{-1}$ and $\|z\|$ are bounded above on N. Take any $C \varepsilon G$ then, for this C, $\operatorname{tr} CsC'$ is bounded above on N, using the right hand inequality (4.1), and so is $\|C_{11}z - \zeta\|^2$. Hence, for this C, $f(x, C)$ is bounded below on N. It follows that there is a finite constant K such that $\sup_{C\varepsilon G} f(x, C) \geqq K$, $x \varepsilon N$. Let $G_1 = \{C \varepsilon G: f(x, C) \geqq K \text{ for all } x \varepsilon N\}$, then $G_1 \neq \emptyset$ and G_1 is closed in G (because $f(x, \cdot)$ is continuous on G) so that G_1 is closed in $GL(k, R)$. Let $\{G_i\}$ be any sequence of matrices in G_1 converging to a k^2 matrix C. Since for $x \varepsilon N$, $2 \ln |C_i| \geqq f(x, C_i) \geqq K$, we have $2 \ln |C| \geqq K$ so that $|C| > 0$ and consequently $C \varepsilon GL(k, R)$, and therefore $C \varepsilon G_1$ since G_1 is closed in $GL(k, R)$. It follows that G_1 is closed in E^{k^2}. Next, we shall show G_1 bounded. Let $1/a$ be an upper bound for $\operatorname{tr} s^{-1}$, $x \varepsilon N$, then, by the left hand inequality (4.1) (applied to $A = s$, $B = CC'$) we have $a \operatorname{tr} CC' \leqq \operatorname{tr} CsC'$ if $x \varepsilon N$. Therefore, if $x \varepsilon N$, $-a \operatorname{tr} CC' + \ln |CC'| \geqq f(x, C)$, and since $f(x, C) \geqq K$ if $x \varepsilon N$, $C \varepsilon G_1$, we have that $-a \operatorname{tr} CC' + \ln |CC'| \geqq K$ for $C \varepsilon G_1$. Now $|CC'|$ is the product of the positive, real eigenvalues of CC', and each eigenvalue is less than the sum of the eigenvalues, i.e. $< \operatorname{tr} CC'$. Hence $|CC'| < (\operatorname{tr} CC')^k$, so that $-a \operatorname{tr} CC' + k \ln \operatorname{tr} CC' \geqq K$ for $C \varepsilon G_1$. This implies that $\operatorname{tr} CC'$ is bounded on G_1. Since $\operatorname{tr} CC'$ is precisely the squared Euclidean norm of C as a vector in E^{k^2}, we have proved that the closed set G_1 is bounded, and therefore compact. It follows that, for $x \varepsilon N$, the supremum of $f(x, C)$ over all $C \varepsilon G_1$ is a maximum, proving Lemma 1. Lemma 2 now follows by a standard argument, taking N to be compact and observing that f is continuous on $N \times G_1$, and therefore uniformly continuous.

PROOF OF LEMMA 3. As in the proof of Lemmas 1 and 2 we may suppose $\Sigma = I_k$. Let N be as in the hypothesis of Lemma 3, then N is included in a compact set, so that $\varphi(\cdot, \theta)$ is bounded on N by virtue of Lemma 2. First we shall prove the right hand inequality in (3.8). Using (3.5) and (3.7), we have $J_{n\theta}(x) \leqq \exp[(n - k - 1)\varphi(x, \theta)] \int \exp[(k + 1)\psi(x, \theta, C)]|C|^{-1} |C_{11}| \mu_G(dC) \leqq$

$\exp[-(k+1)\varphi(x,\theta)]\exp[n\varphi(x,\theta)]\int\exp[-\frac{1}{2}(k+1)\operatorname{tr}CsC']|C|^k|C_{11}|\,\mu_G(dC)$. Since φ is bounded on N, the factor $\exp[-(k+1)\varphi(x,\theta)]$ is bounded on N by a constant, which is part of $K_2(\theta)$ in (3.8). It remains to be shown that the integral is bounded by a constant. Using the left hand inequality in (4.1) and the boundedness of $\operatorname{tr} s^{-1}$ on N, there is a positive constant a such that $\exp[-\frac{1}{2}(k+1)\cdot\operatorname{tr}CsC'] \leqq \exp[-a\operatorname{tr}CC']$ on N. Furthermore, $|C_{11}| \leqq (\operatorname{tr}C_{11}C_{11}')^{l/2}$ by the argument given in the proof of Lemmas 1 and 2. Since $\operatorname{tr}C_{11}C_{11}' \leqq \operatorname{tr}CC'$, we have $|C_{11}| \leqq (\operatorname{tr}CC')^{l/2}$. We shall show now that for any $a > 0$, $m \geqq 0$,

$$\int e^{-a\operatorname{tr}CC'}(\operatorname{tr}CC')^m|C|^k\mu_G(dC) < \infty. \tag{4.2}$$

(Remark: If G were given to be the group, say G', of all nonsingular matrices of the form (3.1), then the Haar measure μ_G in (4.2) could be written down explicitly and (4.2) proved directly. However, G is only known to be a subgroup of G' so that μ_G cannot be given explicitly. To prove (4.2) anyway, a trick will be used.) First, we claim that for any $a > 0$, $m \geqq 0$,

$$\int e^{-a\operatorname{tr}AA'}(\operatorname{tr}AA')^m\prod da_{ij} < \infty \tag{4.3}$$

where the integration is over all k^2 matrices A, with elements a_{ij}, and the product in the integral runs over all i, j from 1 to k. It is easy to show (4.3), e.g. by introducing polar coordinates, and the proof will be omitted. Let d be the dimension of G. If $d = k^2$ then G must contain the component of the identity of $GL(k, R)$. The integral in (4.2) would therefore not be decreased if the integration is taken over all of $GL(k, R)$, which equals the integral over all k^2 matrices. Then (4.2) follows from (4.3) (with m replaced by $m + (k^2/2)$) after observing that $|C|^k \leqq (\operatorname{tr}CC')^{k^2/2}$. Suppose now that $d < k^2$. Using the theory of Lie groups and Assumption A (ii) that G is closed in $GL(k, R)$, we can assert the existence at the identity (i.e. I_k) of a local cross-section ([2], p. 110, [3], Theorem 6.5.2), say $\mathcal{B}$, of the left cosets $GL(k, R)/G$. The matrices in $\mathcal{B}$ will be denoted by B, and we may choose $\mathcal{B}$ so that $\operatorname{tr}BB'$ and $\operatorname{tr}(BB')^{-1}$ are both bounded above on $\mathcal{B}$. Let $\mathcal{A} = \{A = CB : C \varepsilon G, B \varepsilon \mathcal{B}\}$. Since B is the unique intersection of $\mathcal{B}$ and the left coset GB, every $A \varepsilon \mathcal{A}$ has a unique representation $A = CB$, $C \varepsilon G$, $B \varepsilon \mathcal{B}$. In other words, there is a 1-1 correspondence between $\mathcal{A}$ and $G \times \mathcal{B}$. The set $\mathcal{A}$ is a subset of $GL(k, R)$ of positive k^2-dimensional Lebesgue measure. Any integral over $\mathcal{A}$ with respect to Lebesgue measure can be evaluated as an integral over $G \times \mathcal{B}$ after observing that $\prod da_{ij} = |C|^k\mu_G(dC)\nu(dB)$, with ν some measure on $\mathcal{B}$ such that $\nu(\mathcal{B}) > 0$ (this formula for the volume element can be derived by making the transformation $A \to C_0A$ with any fixed $C_0 \varepsilon G$, which transforms $\prod da_{ij} \to |C_0|^k\prod da_{ij}$, but leaves $\mu_G(dC)$ and $\nu(dB)$ invariant; cf. [14], proof of Theorem 3). Let b be a common upper bound for $\operatorname{tr}BB'$ and $\operatorname{tr}(BB')^{-1}$, $B \varepsilon \mathcal{B}$. Then if $A = CB$, $B \varepsilon \mathcal{B}$, we derive from Lemma 8 the following bounds: $\operatorname{tr}CC' \leqq b\operatorname{tr}AA'$ and $\exp[-a\operatorname{tr}CC'] \leqq \exp[-(a/b)\operatorname{tr}AA']$. In order to prove (4.2) it suffices to multiply the left hand side of (4.2) first by $\nu(\mathcal{B}) = \int_{\mathcal{B}}\nu(dB)$ and to show that the result, which now may be considered as an integral over $G \times \mathcal{B}$, is finite. Writing the resulting integral now as an integral over $\mathcal{A}$, and observing the above derived bounds for the integrand, we obtain the following bound for the integral:

$$b^m \int_G e^{-(a/b)\operatorname{tr} AA'} (\operatorname{tr} AA')^m \prod da_{ij}.$$

This integral can only be increased by extending the integral over all k^2 matrices A, and then we know by (4.3) that the result is finite. This concludes the proof of the right hand inequality in (3.8).

Next, we shall prove the left hand inequality in (3.8). In the following, θ will be held fixed and the dependence of various functions on θ will be suppressed (e.g. $\psi(x, C)$ instead of $\psi(x, \theta, C)$). For any $x \varepsilon \mathfrak{X}$, denote by C_x any matrix $C \varepsilon G$ that maximizes $\psi(x, C)$, the existence of C_x being guaranteed by Lemma 1. Furthermore, denote $x_m(x) = (z_m(x), s_m(x))$, with

$$(4.4) \qquad z_m(x) = C_{x11}z, \qquad s_m(x) = C_x s C_x'.$$

Putting, temporarily, $C^* = CC_x^{-1}$ (the dependence of C^* on x has been suppressed), we can write $\psi(x, C) = -\frac{1}{2} \operatorname{tr} C^* s_m(x) C^{*\prime} - \frac{1}{2} \|C_{11}^* z_m(x) - \zeta\|^2 + \ln |C^*| + \ln |C_x|$. In the integrand on the right hand side of (3.5), $|C|^{-1} |C_{11}| = |C^*|^{-1} |C_{11}^*| \, |C_x|^{-1} |C_{x11}|$ and $\mu_G(dC) = \mu_G(dC^* C_x) = m(C_x)\mu_G(dC^*)$ in which $m(\cdot)$ is the modular function. Now for $x \varepsilon N$ we have necessarily $C_x \varepsilon G_1$, where G_1 was defined in the proof of Lemmas 1 and 2, and shown to be compact. Since also $G_1 \subset GL(k, R)$, $|C_x|$ and $|C_{x11}|$ as well as their reciprocals are bounded above on G_1, so that $|C_x|^{-1} |C_{x11}|$ is bounded below by a positive constant, for $x \varepsilon N$. The same is true for $m(C_x)$, m being continuous. The lower bound for $|C_x|^{-1} |C_{x11}| \, m(C_x)$ may be absorbed in the constant $K_1(\theta)$ on the left hand side in (3.8). We integrate now over C^*, but since C runs through G, so does C^*. We now drop the asterisk and define

$$(4.5) \quad \psi_m(x, C) = -\tfrac{1}{2} \operatorname{tr} C s_m(x) C' - \tfrac{1}{2} \|C_{11} z_m(x) - \zeta\|^2 + \ln |C| + \ln |C_x|.$$

Then $\psi_m(x, I_k) = \varphi(x)$, defined in (3.7), and we are seeking a lower bound for

$$(4.6) \qquad \int e^{n\psi_m(x,C)} |C|^{-1} |C_{11}| \, \mu_G(dC), \qquad x \varepsilon N.$$

Let V be a compact neighborhood of the identity I_k in G on which there is a *chart* [3], i.e. a set of coordinates $u_1, \cdots, u_d$ such that the elements of $C \varepsilon V$ are analytic functions of the u_i (remembering that $d = \dim G$, and $d \geqq 1$ by Assumption A (ii)). Put $u = (u_1, \cdots, u_d)$ and assume, without loss of generality, that $u = 0$ at $C = I_k$. On V define $f(x, \cdot)$ by $f(x, u) = \psi_m(x, C)$, so that $f(x, 0) = \varphi(x)$ is the maximum of $f(x, \cdot)$. V compact implies that $\operatorname{tr} CC'$ and $(\operatorname{tr} CC')^{-1}$ are bounded on V, which, in turn, implies that $|C|^{-1} |C_{11}|$ is bounded below by a positive constant on V. Furthermore, there exists $b > 0$ such that on V, $\mu_G(dC) > b \prod du_i$. Absorbing all positive constants into $K_1(\theta)$ in (3.8) and restricting the integration in (4.6) to V, we are seeking a lower bound for $\int_V \exp [nf(x, u)] \prod du_i$. Define $N_m = \{x_m(x) : x \varepsilon N\}$, $x_m(x)$ given by (4.4), then it follows from Lemma 8 and the boundedness of $\operatorname{tr} s$, $\operatorname{tr} s^{-1}$ and $\|z\|$ on N that $\operatorname{tr} s_m(x)$, $\operatorname{tr} s_m(x)^{-1}$ and $\|z_m(x)\|$ are bounded on N so that N_m has compact closure. The function $f(x, u)$, indirectly defined by (4.5) and depending on x through x_m, on u through C, is clearly analytic jointly in (x_m, u), $x_m \varepsilon N_m$, $u \varepsilon V$. The radial derivative of f with respect to u is then an analytic function of (x_m, u) on the set $N_m \times V$ which has compact closure. Consequently, this directional derivative is bounded below, say by $-a$, $a > 0$. It follows that $f(x, u) \geqq \varphi(x) - a\|u\|$, $x \varepsilon N$. Thus, we get a lower bound for our integral:

$$\int_V e^{nf(x,u)} \prod du_i \geqq e^{n\varphi(x)} \int_V e^{-na\|u\|} \prod du_i = e^{n\varphi(x)} n^{-d} \int_{nV} e^{-a\|u\|} \prod du_i .$$

The latter integral is $> \int_V e^{-a\|u\|} \prod du_i$, which is a positive constant. This establishes the left hand inequality in (3.8) and concludes the proof of Lemma 3.

REMARK. We could have bounded f little sharper by $f(x, u) \geqq \varphi(x) - a\|u\|^2$, for some $a > 0$. Then the factor n^{-d} on the left hand side in (3.8) would have been $n^{-d/2}$, which gives a slightly better bound. However, it makes no difference in the proof of the theorem.

5. Proof of Lemma 4. First we prove the following auxiliary lemma.

LEMMA 9. *Let A be a subset of some topological space, and let, for each $\alpha \varepsilon A$, f_α be a function defined on the real line by $f_\alpha(t) = -a(\alpha) + b(\alpha)t - c(\alpha)t^2$, in which a, b, and c are real valued functions on A with the following properties:* (i) *a and c are nonnegative;* (ii) *there exist disjoint sets A_1, A_2 with $A_1 \cup A_2 = A$ and A_1 compact, such that a and b are continuous on A_1 and $\min_{\alpha\varepsilon A_1} a(\alpha) = 0$;* (iii) *there exists $a_0 > 0$ and $r < \infty$ such that for all $\alpha \varepsilon A_2$, $a(\alpha) \geqq a_0$ and $b(\alpha)/a(\alpha) \leqq r$. Define $A_0 = \{\alpha: a(\alpha) = 0\}$ and put $b_0 = \max_{\alpha\varepsilon A_0} b(\alpha)$. Then $L \equiv \lim_{t\downarrow 0} t^{-1} \cdot \sup_{\alpha\varepsilon A} f_\alpha(t) = b_0$.*

PROOF. We shall first reduce the proof to the case $c \equiv 0$. Denote temporarily $f_\alpha^*(t) = -a(\alpha) + b(\alpha)t$, so that $f_\alpha^* \geqq f_\alpha$. Note that A_0 is nonempty and compact. Let $\alpha_0 \varepsilon A_0$ be such that $b(\alpha_0) = b_0$. Suppose the lemma proved for $\{f_\alpha^*\}$. Then, on the one hand, $\limsup_{t\downarrow 0} t^{-1} \sup_\alpha f_\alpha(t) \leqq \lim_{t\downarrow 0} t^{-1} \sup_\alpha f_\alpha^*(t) = b_0$, and, on the other hand, $\liminf_{t\downarrow 0} t^{-1} \sup_\alpha f_\alpha(t) \geqq \liminf_{t\downarrow 0} t^{-1} f_{\alpha_0}(t) = \lim_{t\downarrow 0} (b(\alpha_0) - c(\alpha_0)t) = b(\alpha_0) = b_0$. Thus, the limit in the definition of L indeed exists and equals b_0.

Dropping the asterisk, we shall in the remainder of the proof take $f_\alpha(t) = -a(\alpha) + b(\alpha)t$. Since $\sup_\alpha f_\alpha$ is convex, its one-sided derivative at $t = 0$ exists. In the first part of the proof it was shown that $L \geqq b_0$, so it remains to show that $L \leqq b_0$. Take $M > 0$ arbitrarily. With a_0 and r defined in (iii) of the hypothesis of the lemma, if $\alpha \varepsilon A_2$ and $t^{-1} > r + (M/a_0)$, then $t^{-1} f_\alpha(t) < -M$. Hence, $\limsup_{t\downarrow 0} t^{-1} \sup_{\alpha\varepsilon A_2} f_\alpha(t) \leqq -M$, and since M was arbitrary, $\lim_{t\downarrow 0} t^{-1} \cdot \sup_{\alpha\varepsilon A_2} f_\alpha(t) = -\infty$. It follows that $L = \lim_{t\downarrow 0} \sup_{\alpha\varepsilon A_1} f_\alpha(t)$ (here the sup is really a max, by (ii) of the hypothesis). Let $\{t_n\}$ be any sequence such that $t_n \downarrow 0$, and, for each n, let $\alpha_n \varepsilon A_1$ maximize $f_\alpha(t_n)$. By the compactness of A_1 we may assume that $\alpha_n \to \alpha^*$, say, with $\alpha^* \varepsilon A_1$, so that $a(\alpha_n) \to a(\alpha^*)$, $b(\alpha_n) \to b(\alpha^*)$. We have now $L = \lim_{n\to\infty} t_n^{-1} f_{\alpha_n}(t_n) = \lim_{n\to\infty} [-t_n^{-1} a(\alpha_n) + b(\alpha_n)] = b(\alpha^*) - \lim_{n\to\infty} t_n^{-1} a(\alpha_n)$. If $\lim a(\alpha_n) = a(\alpha^*)$ were >0, then L would be $-\infty$, contradicting $L \geqq b_0$. Therefore, $a(\alpha^*) = 0$ so that $\alpha^* \varepsilon A_0$. It follows that $L \leqq b(\alpha^*) \leqq \max_{\alpha\varepsilon A_0} b(\alpha) = b_0$, which finishes the proof of Lemma 9.

Proof of Lemma 4. For any $x_0 \in \mathfrak{X}$, $x \in E^q$, define

$$\varphi_\theta'(x_0, x) = \lim_{t \downarrow 0} t^{-1}[\varphi(x_0 + tx, \theta) - \varphi(x_0, \theta)]. \tag{5.1}$$

In the course of the proof we shall show that the limit on the right hand side in (5.1) exists. In that case, using (3.10), the limit on the right hand side in (3.13) also exists, and

$$\Phi'(x_0, x) = \varphi_{\theta_2}'(x_0, x) - \varphi_{\theta_1}'(x_0, x). \tag{5.2}$$

Using the definition (3.7) of φ, we compute

$$\varphi(x_0 + tx, \theta) - \varphi(x_0, \theta) = \max_{C \in G}[-a_\theta(C) + b_\theta(C)t - c_\theta(C)t^2], \tag{5.3}$$

in which (with ψ defined in (3.4))

$$-a_\theta(C) = \psi(x_0, \theta, C) - \max_{C^* \in G} \psi(x_0, \theta, C^*) \tag{5.4}$$

$$b_\theta(C) = -\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} C s C' - (C_{11} z_0 - \zeta)' \Sigma_{11}^{-1} C_{11} z \tag{5.5}$$

$$c_\theta(C) = \tfrac{1}{2} z' C_{11}' \Sigma_{11}^{-1} C_{11} z. \tag{5.6}$$

From (5.1) and (5.3) we have then

$$\varphi_\theta'(x_0, x) = \lim_{t \downarrow 0} t^{-1} \max_{C \in G}[-a_\theta(C) + b_\theta(C)t - c_\theta(C)t^2]. \tag{5.7}$$

Suppressing the dependence of the functions a_θ, b_θ and c_θ on θ, we shall show that these functions satisfy the conditions of Lemma 9. Condition (i) is obviously fulfilled. Choose $a_0 > 0$ arbitrarily and define $A_1 = \{C \in G: a(C) \leqq a_0\}$, $A_2 = G - A_1$. The set A_1 is the set on which $\psi(x_0, \theta, \cdot) \geqq$ constant, and it was shown in the proof of Lemmas 1 and 2, section 4, that this is a compact set. All functions are obviously continuous and $\min_{C \in A_1} a(C) = 0$. Thus, condition (ii) has been verified. The only part of (iii) that is left to verify is the boundedness of $b(C)/a(C)$, $C \in A_2$. Since this ratio is continuous, and the denominator bounded away from 0, we only have to check that $b(C)/a(C)$ remains bounded as $\operatorname{tr} CC' \to \infty$. In order to obtain bounds, Lemma 8 will be used repeatedly. We compute $b(C) < -(C_{11} z_0 - \zeta)' \Sigma_{11}^{-1} C_{11} z \leqq \|z_0\| \, \|z\| \operatorname{tr} \Sigma^{-1} \operatorname{tr} C_{11} C_{11}' + \|\zeta\| \, \|z\| \cdot \operatorname{tr} \Sigma^{-1} (\operatorname{tr} C_{11} C_{11}')^{\frac{1}{2}} \leqq k_1 \operatorname{tr} CC'$ for some finite k_1. It remains to be shown that $a(C) \geqq k_2 \operatorname{tr} CC'$ for some $k_2 > 0$. For this purpose, consulting (5.4) and (3.4), it is sufficient to show that $\operatorname{tr} \Sigma^{-1} C s C' - 2 \ln |C| \geqq k_3 \operatorname{tr} CC'$ for some $k_3 > 0$ for sufficiently large $\operatorname{tr} CC'$. This follows from $\operatorname{tr} \Sigma^{-1} C s C' \geqq (\operatorname{tr} \Sigma)^{-1} (\operatorname{tr} s^{-1})^{-1} \operatorname{tr} CC'$ and $2 \ln |C| \leqq k \ln \operatorname{tr} CC'$ (see the derivation of $|CC'| \leqq (\operatorname{tr} CC')^k$ in the proof of Lemmas 1 and 2). Thus, Lemma 9 applies, from which it follows that the limit on the right hand side in (5.7) exists. This concludes the proof of the first part of Lemma 4.

Applying the conclusion of Lemma 9 to (5.7), we obtain

$$\varphi_\theta'[x_0, x) = \max_{C \in G_\theta} b_\theta(C) \tag{5.8}$$

in which $b_\theta(C)$ is given by (5.5), and

$$G_\theta = \{C \in G: \psi(x_0, \theta, C) = \max_{C^* \in G} \psi(x_0, \theta, C^*)\}. \tag{5.9}$$

Observing that $\text{tr}\,\Sigma^{-1}CsC' = \text{tr}\,C'\Sigma^{-1}Cs = \sum_{ij}(C'\Sigma^{-1}C)_{ij}s_{ij} = \sum_i (C'\Sigma^{-1}C)_{ii}s_{ii} + 2\sum_{i<j}(C'\Sigma^{-1}C)_{ij}s_{ij}$, we may write $b_\theta(C) = v_\theta(C)'x$, where the components of the vector $v_\theta(C)\ \varepsilon\ E^q$ consists of the components of $-C_{11}'\Sigma_{11}^{-1}(C_{11}z_0 - \zeta)$ and the elements of $-\frac{1}{2}C'\Sigma^{-1}C$ on and above the diagonal (the latter counted twice). Thus, according to (5.8), $\varphi_\theta'(x_0, x) = \max_{C\varepsilon G_\theta} v_\theta(C)'x$. This may be considered as the value at x of the support function of the set $V_\theta = \{v_\theta(C): C\ \varepsilon\ G_\theta\}$ in E^q [5]. Since v_θ is continuous on the compact set G_θ, V_θ is compact. We shall prove the second part of Lemma 4 by contradiction. That is, we shall assume $\Phi'(x_0, \cdot) \equiv 0$ and then show that $\gamma(\theta_1) = \gamma(\theta_2)$. For convenience we replace subscript θ_i wherever it occurs by i, $i = 1, 2$. If $\Phi'(x_0, \cdot) \equiv 0$ then, according to (5.2), $\varphi_1'(x_0, \cdot) \equiv \varphi_2'(x_0, \cdot)$. This is equivalent to saying that the two compact sets V_1 and V_2 have the same convex hull. It is not hard to show that, under those circumstances, V_1 and V_2 must have a point in common. Thus, there exist $C_i\ \varepsilon\ G_i$, $i = 1, 2$, such that $v_1(C_1) = v_2(C_2)$. In view of the definition of $v_\theta(C)$, this equality becomes the two equalities

$$C_1'\Sigma_1^{-1}C_1 = C_2'\Sigma_2^{-1}C_2, \tag{5.10}$$

$$C_{1,11}'\Sigma_{1,11}^{-1}(C_{1,11}z_0 - \zeta_1) = C_{2,11}'\Sigma_{2,11}^{-1}(C_{2,11}z_0 - \zeta_2) \tag{5.11}$$

in which $\theta_i = (\zeta_i, \Sigma_i)$, $i = 1, 2$. Throughout, it should be kept in mind that, for $i = 1, 2$, C_i has the form (3.1). Then (5.10) implies

$$C_{1,11}'\Sigma_{1,11}^{-1}C_{1,11} = C_{2,11}'\Sigma_{2,11}^{-1}C_{2,11} \tag{5.12}$$

and, using (5.12), (5.11) reduces to

$$C_{1,11}^{-1}\zeta_1 = C_{2,11}^{-1}\zeta_2\,. \tag{5.13}$$

Putting $C_3 = C_2C_1^{-1}$, which is in G since C_1 and C_2 are, it is easily verified that (5.10) and (5.13) are equivalent to $\Sigma_2 = C_3\Sigma_1C_3'$ and $\zeta_2 = C_{3,11}\zeta_1$. In other words, if g is the transformation $\theta \to g\theta$ corresponding to C_3, then $\theta_2 = g\theta_1$ so that θ_1 and θ_2 are on the same orbit; i.e. $\gamma(\theta_1) = \gamma(\theta_2)$. Q.E.D.

6. Proofs of Lemmas 5, 6, and 7.

PROOF OF LEMMA 5. Throughout this proof it will be assumed that $P\ \varepsilon\ \mathfrak{F}$. First we investigate the limiting distribution of $n^{\frac{1}{2}}(X_n - \theta) = n^{\frac{1}{2}}(\bar{Z}_{n1} - \zeta, S_n - \Sigma)$. We write $n^{\frac{1}{2}}S_n = n^{-\frac{1}{2}}\sum_1^n (Z_i - \mu)(Z_i - \mu)' - n^{\frac{1}{2}}(\bar{Z}_n - \mu)(\bar{Z}_n - \mu)'$. Since $n^{\frac{1}{2}}(\bar{Z}_n - \mu)$ has a limiting (multivariate normal) distribution, $n^{\frac{1}{2}}(\bar{Z}_n - \mu)\cdot(\bar{Z}n - \mu)' \to 0$ in probability. Therefore, the limiting distribution of $n^{\frac{1}{2}}(X_n - \theta)$ (still to be shown to exist) is the same if we replace nS_n by $\sum_1^n (Z_i - \mu)(Z_i - \mu)'$. Furthermore, $n\bar{Z}_{n1} = \sum_1^n Z_{i1}$, where Z_{i1} is the vector of first l components of Z_i. Let Y_i be the q-vector ($q = l + k(k + 1)/2$) consisting of the first l components of Z_i and the $k(k + 1)/2$ elements of $(Z_i - \mu)(Z_i - \mu)'$. Then the limiting distribution of $n^{\frac{1}{2}}(X_n - \theta)$ is the same as the limiting distribution of $n^{-\frac{1}{2}}\sum_1^n (Y_i - \theta)$. Now $Y_1, Y_2, \cdots$ are iid with mean θ, and it follows from the definition of $\mathfrak{F}$ (Section 2) that the covariance matrix of Y_1 is finite and nonsingular. Consequently, the limiting distribution exists and is nonsingular multivariate normal, say Q.

Define, for any real c, $A_c = \{\min_i a_i' n^{\frac{1}{2}}(X_n - \theta) > c$ for infinitely many $n\}$. Clearly, A_c does not depend on any finite number of the Z_i, and since $Z_1, Z_2, \cdots$ are independent, $PA_c = 0$ or 1, by the zero-one law. We shall show that $PA_c = 0$ is impossible. Define $A_{cn} = \{\min_i a_i' m^{\frac{1}{2}}(X_m - \theta) > c$ for some $m \geqq n\}$ and $B_{cn} = \{\min_i a_i' n^{\frac{1}{2}}(X_n - \theta) > c\}$, then $B_{cn} \subset A_{cn} \downarrow A_c$. If we would have $PA_c = 0$, then $PB_{cn} \to 0$. But by hypothesis $C_0 = \{x: \min_i a_i' x > 0\}$ has positive q-dimensional Lebesgue measure, and then so has $C_c = \{x: \min_i a_i' x > c\}$ for every c. It follows that $PB_{cn} \to QC_c > 0$, contradicting $PB_{cn} \to 0$. Hence, $PA_c = 1$ for every c, implying the conclusion of Lemma 5.

Proof of Lemma 6. Denote $X_n - \theta$ by Y_n, then we have for any real c, using $Y_n \to 0$ and Lemma 5

$$(6.1) \quad P\{\text{for infinitely many } n, \quad \|Y_n\| \leqq a \quad \text{and} \quad \min_i a_i' n^{\frac{1}{2}} Y_n > c\} = 1.$$

Let $e = \max_i \|a_i\|^{-1}$, then $\min_i a_i' x > c \Rightarrow \|x\| > ec$. From now on we shall take $c > 0$. Observe that $\min_i a_i' n^{\frac{1}{2}} x > 0 \Rightarrow \min_i a_i' x > 0 \Rightarrow x \varepsilon C_0$. Consider now, for each n, the string of implications $[\|Y_n\| \leqq a$ and $\min_i a_i' n^{\frac{1}{2}} Y_n > c] \Rightarrow [\|Y_n\| \leqq a$ and $Y_n \varepsilon C_0$ and $\|n^{\frac{1}{2}} Y_n\| > ec] \Rightarrow [f(Y_n) \geqq b \|Y_n\|$ and $\|n^{\frac{1}{2}} Y_n\| > ec] \Rightarrow [n^{\frac{1}{2}} f(Y_n) > bec]$. Using (6.1), we have then $P\{n^{\frac{1}{2}} f(Y_n) > bec$ for infinitely many $n\} = 1$. Since $be > 0$ and $c > 0$ arbitrary, the conclusion of Lemma 6 is proved.

Proof of Lemma 7. We shall first show that given $\epsilon > 0$ and any compact set $K \subset E^q$ there exists $t_0 > 0$ (t_0 may depend on θ) such that

$$(6.2) \quad |\varphi_\theta'(x_0, x) - t^{-1}[\varphi(x_0 + tx, \theta) - \varphi(x_0, \theta)]| < \epsilon \quad \text{if} \quad 0 < t \leqq t_0, \quad x \varepsilon K.$$

We shall use (5.3) and we shall again suppress the dependence on θ. It should kept in mind, however, that b and c are also functions of x.

First, recall from the proof of Lemma 4 that on the right hand side in (5.7) the maximum may be taken over all C in the compact set A_1 defined in the proof of Lemma 4. In the following all maxima will be taken over $C \varepsilon A_1$. Second, recall from the proof of Lemma 9 that $\lim_{t \downarrow 0} t^{-1} \max_C [-a(C) + b(C)t - c(C)t^2] = \lim_{t \downarrow 0} t^{-1} \max_C [-a(C) + b(C)t]$. Using (5.7), we have then that $\lim_{t \downarrow 0} t^{-1} \cdot \max_C [-a(C) + b(C)t] = \varphi'(x_0, x)$. Since $\max_C [-a(C) + b(C)t]$ is a convex function of t, equal to 0 at $t = 0$, the convergence is downward. Furthermore, $\varphi'(x_0, \cdot)$ is continuous: this follows from (5.8), (5.5) and the compactness of G_θ. By Dini's theorem, the convergence is then uniform for $x \varepsilon K$ so that there exists $t_1 > 0$ such that

$$(6.3) \quad |\varphi'(x_0, x) - t^{-1} \max_C [-a(C) + b(C)t]| < \epsilon/2 \quad \text{if} \quad 0 < t \leqq t_1, \quad x \varepsilon K.$$

Third, since the function c of (5.6), as a function on $G \times E^q$, is bounded on the compact $A_1 \times K$, there exists $t_2 > 0$ such that $c(C)t \leqq \epsilon/2$ if $0 < t \leqq t_2$, $C \varepsilon A_1$ and $x \varepsilon K$. We have then

$$(6.4) \quad |t^{-1} \max_C [-a(C) + b(C)t] - t^{-1} \max_C [-a(C) + b(C)t - c(C)t^2]| < \epsilon/2 \quad \text{if} \quad 0 < t \leqq t_2, \quad x \varepsilon K.$$

Combining (6.3) and (6.4), using (5.3) and restoring the dependence on θ, (6.2) follows when we take $t_0 = \min(t_1, t_2)$.

For simplicity in notation we shall put $h_i(x) = \varphi(x_0 + x, \theta_i) - \varphi(x_0, \theta_i)$, $h_i'(x) = \varphi_{\theta_i}'(x_0, x)$. Let $\delta > 0$ and K compact, to be determined later, be given. From (6.2) it follows that there exists $t_0 > 0$ such that

$$(6.5) \qquad |h_i'(x) - t^{-1}h_i(tx)| < \delta/2, \qquad i = 1, 2, \quad \text{if} \quad 0 < t \leqq t_0, \quad x \varepsilon K.$$

Using the hypothesis $\Phi(x_0) = 0$ of the lemma, i.e. $\varphi(x_0, \theta_1) = \varphi(x_0, \theta_2)$, we compute $f(x) = \Phi(x_0 + x) = h_2(x) - h_1(x)$. Note that $f(0) = 0$. We shall denote $\Phi'(x_0, x) = h_2'(x) - h_1'(x)$ by $f'(x)$. Since $\varphi_\theta'(x_0, \cdot)$ is continuous for each θ, $f'(\cdot)$ is continuous. By hypothesis, $f'(\cdot) \not\equiv 0$, and it follows that there is a compact neighborhood $K \subset E^q$, which we may take to exclude the origin, such that $f' > 0$ or <0 on K. Suppose the former, the latter yielding the same conclusion for $-f$. Then there is $\delta > 0$ such that $f'(x) \geqq 2\delta$ for all $x \varepsilon K$. That is,

$$(6.6) \qquad h_2'(x) - h_1'(x) \geqq 2\delta, \qquad x \varepsilon K.$$

It is this K and δ that is to be used in (6.5). Combining (6.5) and (6.6) we get $t^{-1}[h_2(tx) - h_1(tx)] > \delta$, or

$$(6.7) \qquad t^{-1}f(tx) > \delta, \qquad 0 < t \leqq t_0, \qquad x \varepsilon K.$$

Now put $b = \delta(\max_{x \varepsilon K} \|x\|)^{-1} > 0$, then (6.7) can be written in the form

$$(6.8) \qquad f(tx) \geqq b\,\|tx\|, \qquad 0 \leqq t \leqq t_0, \qquad x \varepsilon K.$$

The cone $\{tx : x \varepsilon K, 0 \leqq t < \infty\}$ contains a cone C_0 as in Lemma 5, for some a_i. Put $a = t_0 \min_{x \varepsilon K} \|x\| > 0$. If $x \varepsilon C_0$, then $x = tx_1$ for some $x_1 \varepsilon K$ and $t \geqq 0$, so that $\|x\| \geqq t \min_{x^* \varepsilon K} \|x^*\| = ta/t_0$. If, at the same time, $\|x\| \leqq a$, then we have $ta/t_0 \leqq a$, so that $0 \leqq t \leqq t_0$. That is, $x \varepsilon C_0$ and $\|x\| \leqq a$ together imply $x = tx_1$, with $0 \leqq t \leqq t_0$, $x_1 \varepsilon K$. Replacing in (6.8) x by x_1 and then tx_1 by x, we have that $x \varepsilon C_0$ and $\|x\| \leqq a$ together imply $f(x) \geqq b\,\|x\|$, which is the conclusion of the lemma.

REFERENCES

[1] Berk, Robert H. (1964). Asymptotic properties of sequential probability ratio tests. Ph.D. thesis (unpublished), Harvard Univ.

[2] Chevalley, Claude (1946). *Theory of Lie Groups I*. Princeton Univ. Press.

[3] Cohn, P. M. (1957). *Lie Groups*, Cambridge Univ. Press.

[4] David, Herbert T. and Kruskal, William H. (1956). The WAGR sequential t-test reaches a decision with probability one. *Ann. Math. Statist.* **27** 797–805 and **29** 936.

[5] Eggleston, H. G. (1958). *Convexity*, Cambridge University Press.

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

[7] Ifram, Adnan F. (1965). On the asymptotic behavior of densities with applications to sequential analysis. *Ann. Math. Statist.* **36** 615–637.

[8] Jackson, J. Edward and Bradley, Ralph A. (1961). Sequential χ^2- and T^2-tests. *Ann. Math. Statist.* **32** 1063–1077.

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

[10] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

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

[12] SAVAGE, I. RICHARD and J. SETHURAMAN (1966). Stopping time of a rank-order sequential probability ratio test based on Lehmann alternatives. *Ann. Math. Statist.* **37** 1154–1160.

[13] WALD, ABRAHAM (1947). *Sequential Analysis,* Wiley, New York.

[14] WIJSMAN, R. A. (1965). Cross-sections of orbits and their application to densities of maximal invariants. *Proc. 5th Berkeley Symp. Math. Statist. and Prob.* Univ. of California Press.

[15] WIRJOSUDIRJO, SUNARDI (1961). Limiting behavior of a sequence of density ratios. Ph.D. thesis (unpublished), Univ. of Illinois. Abstract in *Ann. Math. Statist.* **33** 296–297.