ON SUFFICIENCY AND INVARIANCE
WITH APPLICATIONS IN SEQUENTIAL ANALYSIS

by

William Jackson Hall

University of North Carolina

This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 228
June, 1959

ON SUFFICIENCY AND INVARIANCE WITH APPLICATIONS IN SEQUENTIAL ANALYSIS[1]

By Wm. Jackson Hall

University of North Carolina

1. Introduction. We investigate in what sense sufficiency properties are preserved under the invariance principle, and thereby obtain an interpretation of the sufficiency of a statistic in the presence of nuisance parameters - an interpretation which facilitates the derivation of some sequential tests.

Suppose we consider a family of distributions indexed by a parameter θ, and some group of transformations on the sample space (e.g., changes in sign, location, scale, or order). When one invokes invariance[2] under the transformations -- that is, requires that his decision procedures not be affected by the transformations -- then all invariant functions have distributions depending only on some function, say γ, of θ. We show that, loosely speaking, if a statistic s contains all relevant information about θ, then the maximal invariant function of s contains all the relevant information about γ that is available in any invariant function. We call such functions invariantly sufficient; they are sufficient for the family of distributions of any invariant function on the sample space.

1. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-261. Reproduction in whole or in part is permitted for any purpose of the United States Government.

2. For an explanation of the invariance principle, see [11], [12], [6], [18], [13].

Invoking invariance can frequently be considered as a way of handling nuisance parameters.[3] For example, consider a normal distribution with mean θ and known variance. This can be considered as a two-parameter distribution, one parameter being the magnitude of the mean and the other its sign. By invoking invariance under changes in sign, we shall see that one can obtain the magnitude of the sample mean as an invariantly sufficient statistic -- sufficient for a family of distributions indexed by $|\theta|$ -- the sign of θ being a nuisance parameter. Similarly, with normal mean and variance unknown, the sample variance is invariantly sufficient for the population variance under changes in location, and the t-statistic is invariantly sufficient for the mean in standard deviation units under changes in scale; with two normal populations with equal and known variances, the difference between the sample means is invariantly sufficient for the difference between population means under changes in location, and so on. The precise meaning of these statements will be made clear in what follows.

In each instance, the invoking of invariance in problems of inference insures that the error probabilities or risk functions will be independent of the nuisance parameters; however, it may be possible to reduce the error probabilities or risk functions by using non-invariant decision procedures, as shown by various examples in the literature (e.g., [5], [16]).[4]

3. D. A. S. Fraser [9] has considered a generalized sufficiency definition in the presence of nuisance parameters, differing from this approach through invariance.
4. W. L. Smith has called the author's attention to several other examples in which all invariant procedures are inadmissible, from an unpublished paper by Charles Stein.

These results, to be stated explicitly here, seem to be implicitly assumed by other authors. Thus, E. L. Lehmann [11] considers the maximal invariants of sufficient statistics, and derives most of the common significance tests about normal random variables as most powerful invariant tests based on these statistics. We felt the need for clarification of the sense in which these statistics remain sufficient, particularly with regard to an application in sequential analysis.

D. R. Cox [7] has described a method of obtaining sequential tests for certain kinds of composite hypotheses. The method essentially consists of applying a Wald sequential probability ratio test (SPRT) for simple hypotheses about γ to what we shall call an <u>invariantly sufficient sequence</u> of random variables -- or more generally, to any sequence of statistics whose distribution depends only on γ. He has given conditions under which the joint density of m terms in this sequence factors conveniently. Our main result already mentioned clarifies the motivation of these tests, reinterpreting them through the principle of invariance, and also constitutes a major simplification of Cox's factorization theorem. We shall illustrate this method by deriving a two-sided sequential t-test, a slight extension of the usual one [14].

In section 2 we give a not completely rigorous development of these ideas, with simple examples in section 3 and application to sequential tests in sections 4 and 5. A rigorous parallel development in terms of subfields rather than statistics is given in section 6, based on the work of R. R. Bahadur on statistics and subfields [1], [2], [4].

It is trivial to show that (bounded) completeness of a family of distributions is preserved under invariance reductions; in such cases, minimality (or necessity) properties of sufficiency reductions are also preserved (see Theorem 3 in [3]). Investigation of general conditions under which minimality is preserved under the invariance principle is in progress.

2. Invariantly sufficient statistics. We represent a random variable, or probability space, by

$$X_\theta = (\mathcal{X}, \underline{X}, P_\theta)$$

where x represents a point in the sample space $\mathcal{X}$, $\underline{X}$ the sample field of subsets of $\mathcal{X}$, and P_θ a probability measure on $\underline{X}$; we represent a class of random variables indexed by a parameter θ by $X_\Theta = \{X_\theta : \theta \epsilon \Theta\}$ (Θ may be abstract). We consider a group $\mathcal{G}$, with elements g, of measurable [$\underline{X}$, Θ] 1-1 transformations of $\mathcal{X}$ into itself and $\bar{\mathcal{G}}$ the induced group of transformations $\bar{g}$ on Θ, defined by

$$P_\theta(g^{-1}A) = P_{\bar{g}\theta}(A), \; A \epsilon \underline{X}, \; g \epsilon \mathcal{G}, \theta \epsilon \Theta$$

where $g^{-1}A = \{x : gx \epsilon A\}$, i.e., $P_\theta g^{-1} = P_{\bar{g}\theta}$. In what follows, all functions are assumed measurable, and many equalities are to be interpreted as holding except perhaps on a set of measure zero, with the appropriate (class of) probability measure(s); these lapses of rigor will not appear in section 6.

We are concerned with two kinds of reduction principles and their order of application: invariance and sufficiency. We treat invariance following Lehmann [11] and assume the Halmos-Savage concept of sufficiency.

The group $\mathcal{G}$ partitions $\mathcal{X}$ into equivalence classes which we call *orbits* (following Wesler [18]); thus, one can "move freely within an orbit", but not "among orbits", by applying transformations $g \in \mathcal{G}$. A function constant on an orbit is (strictly) *invariant*;[5] if it assumes a different value on each orbit, it is a *maximal invariant*:

DEFINITION: A function u *on* $\mathcal{X}$ *is a maximal invariant on* $\mathcal{X}$ *under* $\mathcal{G}$ *if*

(i) u *is invariant under* $\mathcal{G}$: $u(gx) = u(x)$ *for all* g,x; *and*

(ii) $u(x') = u(x'')$ *implies there exists a* $g \in \mathcal{G}$ *such that* $x' = gx''$.

If s *is a function on* $\mathcal{X}$ *with range in* $\mathcal{S}$ *, then a function* v *on* $\mathcal{X}$ *through* $\mathcal{S}$ *is a maximal invariant on* $\mathcal{X}$ *through* $\mathcal{S}$ (*or simply on* $\mathcal{S}$) *under* $\mathcal{G}$ *if*

(i) v *is invariant under* $\mathcal{G}$; *and*

(ii) $v(x') = v(x'')$ *implies there exists a* $g \in \mathcal{G}$ *such that* $s(gx') = s(x'')$.

A well-known property of maximal invariant functions on $\mathcal{X}$ is that any function t on $\mathcal{X}$ is invariant if and only if it is a function of x only through the maximal invariant; maximal invariants on $\mathcal{S}$ have a comparable property. (A step-wise method of finding a maximal invariant function is treated in [11].)

Any function t on $\mathcal{X}$ induces a new probability space (random variable or *statistic*) which we denote as

---

5. "Strictly" implies "for all x" and not just "for almost all x".

$$T_\theta = (\mathcal{T}, \underline{T}, P_\theta^t)$$

with similar notation for other functions. Here, $P_\theta^t = P_\theta t^{-1}$ is the induced probability measure on $\underline{T}$, the induced field of subsets of $\mathcal{T}$, the induced sample space of the function t. If t is invariant under $\mathcal{G}$, then its distribution depends only on a maximal invariant function, say γ, on Θ under $\mathcal{G}$ (e.g., see [11]); we thus denote an invariant statistic by

$$T_\gamma = (\mathcal{T}, \underline{T}, P_\gamma^t),\ \gamma \in \Gamma = \{\gamma(\theta) : \theta \in \Theta\}.$$

<u>DEFINITION</u>: <u>A function</u> v <u>on</u> $\mathcal{X}$ <u>is invariantly sufficient for</u> X_Θ <u>under</u> $\mathcal{G}$ <u>if</u>

(i) v <u>is invariant under</u> $\mathcal{G}$, <u>and</u>

(ii) v is sufficient for T_Γ for every function t on $\mathcal{X}$ which is invariant under $\mathcal{G}$.

Clearly, an equivalent requirement to (ii) is that v be sufficient for U_Γ where u is a maximal invariant on $\mathcal{X}$ under $\mathcal{G}$.

Consider the following schematic diagram, in which vertical arrows indicate maximal invariance reductions and horizontal arrows sufficiency reductions:

$$\begin{array}{c|ccc} \theta & X_\theta & \longrightarrow & S_\theta \\ \downarrow & \downarrow & & \downarrow \\ \gamma & U_\gamma & \longrightarrow & V_\gamma \end{array}$$

We shall justify these implications, that V_γ may be obtained by either route from X_θ;

THEOREM: If s is sufficient for X_Θ and v is a maximal invariant on $\mathcal{S}$ under $\mathcal{G}$, then v is invariantly sufficient for X_Θ under $\mathcal{G}$.

A sketch of the proof follows: v is invariant by assumption; we must verify that it is sufficient for U_Γ, that is, that the conditional probability of $u \in B$ (a set in U) given v is independent of γ. This probability can be written as

$$P^u_{\gamma(\Theta)}(B|v) = P_\Theta u^{-1}(B|v) = P_\Theta(u^{-1}B|v) = \mathcal{E}_{S_\Theta}[(P_\Theta(u^{-1}B|s)|v],$$

the last statement being a variation of the well-known fact that "expectation = expectation of conditional expectation" (a rigorous proof follows easily from the definitions as Radon-Nikodyn derivatives). We first note that the conditional probability on the RHS is independent of Θ by the sufficiency of s, and secondly (see below) that it is an invariant function on $\mathcal{S}$. Since v is a maximal invariant on $\mathcal{S}$, $P(u^{-1}B|s)$ is a function of v only, so that the conditional expectation operation is applied to a constant and is thus free of Θ, proving that the LHS is also free of Θ (and γ). For the invariance proof, we have for any g

$$P_\Theta(u^{-1}B|s(gx)=s) = P_{g\Theta}(gx \in u^{-1}B|s(x)=s) = P_{g\Theta}(x \in u^{-1}B|s(x)=s)$$

since u is invariant; but all terms are independent of Θ, so that $P(u^{-1}B|s)$ is invariant.

A satisfactory converse to this theorem is not as yet available.

3. Examples. As a simple illustrative example, consider the following (we omit the parentheses in the functional notation):

$x = (x_1, x_2)$ where the x_i's are independent $N(\theta, 1)$, $\Theta = R_1$;

$\mathcal{G} = (g^+, g^-)$ where $g^+ x = x = (x_1, x_2)$, $g^- x = -x = (-x_1, -x_2)$;

$\bar{\mathcal{G}} = (\bar{g}^+, \bar{g}^-)$ where $\bar{g}^+ \theta = \theta$, $\bar{g}^- \theta = -\theta$;

$sx = x_1 + x_2$, which is sufficient for X_Θ ;

$vs = |s|$, $vsx = |x_1 + x_2|$;

$ux = (u_1 x, u_2 x, u_3 x)$ where $u_1 x = |x_1|$, $u_2 x = |x_2|$, and

$$u_3 x = \begin{cases} 1 & \text{if } \operatorname{sgn} x_1 = \operatorname{sgn} x_2 \\ 0 & \text{otherwise;} \end{cases}$$

$\gamma\theta = |\theta|$.

That γ, v, and u are all maximal invariants is readily verified; we prove it for u only: (i) $ug^{\pm}x = ux$; (ii) $ux' = ux''$ implies $x_i' = \pm x_i''$ with the same sign for i=1,2, that is, $x' = g^+x''$ or $x' = g^-x''$. Thus, the absolute values of each coordinate x_i together with knowledge of which pair of diagonally opposite quadrants contains (x_1, x_2), form a maximal invariant u (a diagram may be helpful).

The theorem states that the distribution of any invariant function t -- that is, any function for which $t(x) = t(-x)$ -- given v is independent of θ (or $|\theta|$); i.e., among invariant functions, $|x_1 + x_2|$ contains all the available information about $|\theta|$. For example, the distribution of $|x_1|$, or of $(|x_1|, |x_2|)$, given $|x_1 + x_2|$ is independent of $|\theta|$.

For a second example, suppose $x = (x_1, \ldots, x_n)$, where the x_i's are independent $N(\theta_1, \theta_2^2)$. The sample mean and sum of squared deviations provide a sufficient statistic, and Student's t-statistic is a maximal invariant under changes in scale; θ_1/θ_2 is a maximal invariant on the parameter space. Thus, Student's t is invariantly sufficient under changes in scale. Similarly, $|t|$ is invariantly sufficient under changes in scale and sign (see section 5). See [11] for details of these and other examples, including the F-ratio and Hotelling's T^2 statistic.

4. Application to sequential tests of composite hypotheses. Suppose $x_1, x_2, \ldots$ is a sequence of independent observations on X_θ, θ fixed but unknown, and it is desired to make some inference concerning θ. Bahadur [1] has shown that a sequential decision rule need depend only a sufficient sequence (assuming regularity or transitivity), a sequence $s_1, s_2, \ldots$ of sufficient statistics where $s_m = s_m(x_1 \ldots, x_m)$ is sufficient for the family of distributions of $x_{(m)} = (x_1, \ldots, x_m)$ -- i.e., for the class of m-fold products $X_\theta^m = X_\theta x \ldots x X_\theta$.

Now consider a group of transformations on the sequential sample space, and let $u_1, u_2, \ldots$ be a sequence of maximal invariants, where $u_m = u_m(x_{(m)})$ is invariant under transformations on the m-fold sample space. Invoking invariance in the sequential problem requires replacement of the original sequence $x_1, x_2, \ldots$ by such a sequence $u_1, u_2, \ldots$. A sufficiency reduction in the invariant problem leads to a new sequence $v_1, v_2, \ldots$ where v_m is sufficient for $u_{(m)} = (u_1, \ldots, u_m)$; a maximal invariance reduction applied to the sufficient sequence $s_1, s_2, \ldots$ leads also (by the theorem) to $v_1, v_2, \ldots$. We call such a sequence an invariantly sufficient sequence, and the results of Bahadur [1] on sufficient sequences, together with the invariance principle, imply that invariant sequential decision rules need depend only on invariantly sufficient sequences.

D. R. Cox [7] (whose paper is in the nuisance parameter terminology) has suggested that sequential tests of simple hypotheses about γ, which are composite hypotheses about Θ, may be obtained by applying a SPRT to a sequence of statistics whose distributions depend only on γ; such a sequence is $v_1, v_2, \ldots$. We see now that this is simply an instance of invoking the invariance principle. Moreover, since v_m is sufficient for $u_{(m)}$ (by our theorem), it is clearly sufficient for $v_{(m)} = (v_1,\ldots,v_m)$, a function of $u_{(m)}$. Thus, the density (having assumed a fixed dominating measure) of $v_{(m)}$ factors according to the Fisher-Neyman factorization theorem for sufficient statistics. The ratio of densities of $v_{(m)}$ at γ_1 and γ_0, say -- on which the SPRT is based -- thus reduces to the ratio of densities of v_m at γ_1 and γ_0. This is the essence of Cox's theorem. However, instead of verifying his conditions (i) to (iv), one need only verify the sufficiency of s_m and the maximal invariance of v_m and γ. And moreover, the justification for confining attention to the sequence $v_1,v_2,\ldots$ is made precise.

However, since the v_m's are not independent and indentically distributed, no optimal properties of the SPRT applied to them are known; it is only known that the test will possess (approximately) the prescribed error probabilities, using Wald's boundaries. (Note: It seems to be a common fallacy (e.g., [7], [8]) to claim that, in order to use Wald's boundaries for a SPRT, one must prove that the test terminates with probability one; undoubtedly, it is desirable to prove this when possible, but it is not necessary for the applicability of Wald's boundaries giving approximate prescribed error probabilities. This can be seen by referring to Wald's original development beginning on page 41 of [17]; each time that $1-\beta$ or $1-\alpha$ appears in the text on page 41, it can be preceded by "is less than or equal to" without assuming termination with certainty. His inequalities (3:12), (3:13), and (3:27), justifying use of his boundaries thus remain valid.)

It would be tempting to replace the sequence $x_1, x_2, \ldots$ by a sequence of independent and identically distributed invariant functions $t_1, t_2, \ldots$ by taking t_m as a maximal invariant function of x_m (rather than $x_{(m)}$). Then all properties of the common SPRT would hold. There is no other justification, however; this technique may be exceedingly extravagant, as is perhaps best seen by an example: in the example treated in section 5, $t_1, t_2, \ldots$ would be a sequence of constants.

As an example of a test based on an invariantly sufficient sequence, suppose x_i is $N(\theta,1)$, and consider the group with two elements g^+ and g^-, where g^+ is the identity transformation and g^- changes the signs of all x_i's. $s_m = x_1 + \ldots + x_m$, $m=1,2,\ldots$, provides a sufficient sequence, and $v_m = |s_m|$ is a maximal invariant on $\mathcal{S}^m$. $v_1, v_2, \ldots$ is then an invariantly sufficient sequence, each v_m being invariantly sufficient for the m-fold product $\mathcal{X}^m_{\Theta}$. A sequential test of $|\theta| = \gamma_0$ or γ_1 is provided by a SPRT on $v_1, v_2, \ldots$. The ratio of joint densities after m observations is the ratio of densities of v_m, which is readily obtained as:

$$\exp [-m(\gamma_1^2-\gamma_0^2)/2] \cdot [\cosh(v_m\gamma_1)/\cosh(v_m\gamma_0)].$$

Sampling is continued as long as this ratio remains between B and A, functions of the prescribed error probabilities.

A second example appears in section 5.

Note: If we wish to test $|\theta| \leq \gamma_0$ or $\geq \gamma_1$ $(0 \leq \gamma_0 < \gamma_1)$, instead of equalities, for examples such as those mentioned, we must show that the error probabilities are still valid for these sets of θ-values, which are now composite in γ $(=|\theta|)$. No general results are known by the author. Wald (pp. 204-207 of [17])claims in an example similar to those considered here that it is sufficient to prove that the ratio of densities is monotone in v_m; this can frequently be done, but it is not clear to this author that this is in fact sufficient.

5. A two-sided sequential t-test. Suppose $x = x_1, x_2, \ldots$ is a sequence of independent observations on $N(\mu, \sigma^2)$, and we wish to decide whether μ is close to some specified value μ_o (decision d_0) or appreciably different from μ_o (decision d_1), requiring that the probability of making d_i if $|(\mu-\mu_o)/\sigma| = \gamma_j$ (i,j=0,1 or 1,0, $0 \leq \gamma_0 < \gamma_1$) be at a prescribed level, given γ_0 and γ_1. Such a problem is commonly stated as that of testing $(\mu, \sigma) \in \omega_0$ or ω_1, where

$$\omega_0 = \{\mu, \sigma : |\frac{\mu - \mu_o}{\sigma}| \leq \gamma_0\} \text{ and } \omega_1 = \{\mu, \sigma : |\frac{\mu - \mu_o}{\sigma}| \geq \gamma_1\},$$

with specified strength (α,β). With $\gamma_0 = 0$, this reduces to the common two-sided sequential t-test problem treated in [14]. We feel, however, that it is more reasonable to assign a prescribed error probability for an interval around μ_o, rather than for $\mu=\mu_o$, a hypothesis frequently known to be false a priori. We shall take $\mu_o = 0$ for simplicity.

(With $\theta^* = \text{Prob}\ (X > \mu_o)$, an equivalent problem is one in terms of

$$\omega_0^* = \{\theta^* : |\theta^* - \tfrac{1}{2}| \leq \gamma_0^*\},\ \omega_1^* = \{\theta^* : |\theta^* - \tfrac{1}{2}| \geq \gamma_1^*\}.$$

As an alternative to a t-test, a sequential sign test can be developed for these hypotheses without any assumption of normality.)

Let $\mathcal{G}$ be the group of scale and sign changes; specifically, for each non-zero real number r there corresponds a transformation $g_r \in \mathcal{G}$ defined by

$$g_r x = g_r(x_1,x_2,\ldots) = (rx_1,rx_2,\ldots) = rx.$$

A sufficient sequence is $s_1,s_2,\ldots$ where $s_m = (\bar{x}_m, \Sigma_{i=1}^m (x_i-\bar{x}_m)^2)$ and $m\bar{x}_m = x_1 + \ldots + x_m$. Denoting

$$v_m = |\sqrt{m}\ \bar{x}_m / \sqrt{\Sigma_{i=1}^m (x_i - \bar{x}_m)^2}|,\ m > 1 \text{ (absolute Student's t)}$$

$$v_1 = \text{arbitrary constant},$$

then $v_1,v_2,\ldots$ is an invariantly sufficient sequence since v_m is easily verified as being a maximal invariant function of s_m. The induced group on the parameter space (half-plane) has elements $\bar{g}_r$ defined by $\bar{g}_r(\mu, \sigma^2) = (r\mu, r^2\sigma^2)$ and $\gamma = |\mu/\sigma|$ is a maximal invariant. To test $\gamma = \gamma_0$ against $\gamma = \gamma_1$, we can use the SPRT applied to the sequence of random variables $v_1, v_2, \ldots$.

By our theorem, v_m is sufficient for all invariant functions of $x_{(m)} = (x_1, \ldots, x_m)$ (whose distributions are indexed by γ) and thus, in particular, for $v_{(m)}$. The probability ratio after m observations is then, by the factorization theorem, the ratio of two non-central t^2-densities (having replaced v_m by $v_m{}^2$ for convenience). From the non-central t density, or from the non-central F, we obtain as the density of $z = t^2$:

$$[2^{m-2}\,\Gamma^2(\overline{m-1}/2)\,m\text{-}1\,\pi\, z]^{-1/2} \cdot \left(\frac{m-1}{m-1+z}\right)^{m/2} \cdot \exp\left(-\frac{1}{2}\,\frac{\overline{m-1}\; m\,\gamma^2}{m-1+z}\right)$$

$$\cdot \exp\left(-\frac{1}{2}\,\frac{m\,\gamma^2 z}{m-1+z}\right) \int_0^\infty w^{m-1}\, e^{-w^2/2} \cosh\left(\frac{m\,\gamma^2 z}{m-1+z}\, w^2\right)^{1/2} dw .$$

The probability ratio, with $y = mz/(m-1+z)$, thus becomes

$$e^{-m(\gamma_1^2 - \gamma_0^2)/2} \cdot \frac{\int_0^\infty w^{m-1}\, e^{-w^2/2} \cosh(\sqrt{y}\,\gamma_1\, w)\, dw}{\int_0^\infty w^{m-1}\, e^{-w^2/2} \cosh(\sqrt{y}\,\gamma_0\, w)\, dw} .$$

From the bottom of page vi in [14], we have the identity for all real a:

$$F\left(\frac{m}{2}, \frac{1}{2}; \frac{a^2}{2}\right) \cdot \int_0^\infty w^{m-1}\, e^{-w^2/2}\, dw = \int_0^\infty w^{m-1}\, e^{-w^2/2} \cosh(aw)\, dw$$

where F denotes the confluent hypergeometric function. Using this, the ratio becomes

$$R_m = e^{-m(\gamma_1^2 - \gamma_0^2)/2} \cdot \frac{F(\frac{m}{2},\frac{1}{2};\frac{y\gamma_1^2}{2})}{F(\frac{m}{2},\frac{1}{2};\frac{y\gamma_0^2}{2})}$$

where y is most simply computed as

$$y = y_m = \frac{m\, v_m^2}{m-1+v_m^2} = \frac{(\Sigma_{i=1}^m x_i)^2}{\Sigma_{i=1}^m x_i^2} .$$

The sequential test is then: after m observations, continue sampling if the ratio R_m lies between $B = \beta/(1-\alpha)$ and $A = (1-\beta)/\alpha$, and stop sampling otherwise; make the decision d_0 if the ratio $\leq B$ and d_1 if it is $\geq A$. The approximate error probabilities at $\gamma = \gamma_0$ and γ_1 (i.e., for any (μ,σ) for which $|\mu/\sigma| = \gamma_i$) are α and β, respectively. With $\gamma_0 = 0$, it is easily verified that this is the two-sided WAGR test in [14]. (Note: Using Wald's weight function method [17], one can obtain the same test with the modification of reducing m by 1 in the first argument of the confluent hypergeometric functions.)

Tables of the confluent hypergeometric function (e.g., [15]) or its logarithm may be used to facilitate the test. It might be noted that $\log R_m = L(m,\gamma_1,y) - L(m,\gamma_0,y)$, the L-function being defined, and its inverse tabled, in [14].

In order to use this test for $\gamma \leq \gamma_0$ against $\gamma \geq \gamma_1$, one should verify that the strength requirements are maintained. It would be sufficient to show that the operating characteristic function $L(\gamma)$ is nonotone in γ; this seems intuitively plausible, but the author has not been able to construct a general proof.

It is possible to prove that v_m has a monotone likelihood ratio. Since v_m is monotone in y_m, this is equivalent to proving the test ratio R_m is monotone in y_m, which is a consequence of the following lemma:

LEMMA: Let $f(x) = \int_0^\infty w^{m-1} e^{-w^2/2} \cosh(\gamma_1 x w)\, dw \Big/ \int_0^\infty w^{m-1} e^{-w^2/2} \cosh(\gamma_0 x w)\, dw$ where $0 \leq \gamma_0 < \gamma_1$ and m is a positive integer $(-\infty < x < +\infty)$. $f(x)$ is an increasing (decreasing) function of x for $x > 0$ $(x < 0)$.

A proof can be constructed (not given here) along the lines of an analogous result in [10]. As noted in section 4, Wald's argument would assert this to be a sufficient condition for the monotonicity of $L(\gamma)$; this argument is not clear to the present author.

It has not been verified that this test terminates with probability one; presumably, a proof along the lines of [8] accomplish it.

6. Invariance and subfields. We now give an alternative parallel development of section 2 in terms of subfields rather than statistics. The subfields can be thought of as those induced by the corresponding statistics. We thus avoid specific designation of values to be assumed by a statistic, these values being irrelevant. We shall utilize the terminology and some elementary results of [1].

We consider a class of probability spaces X_Θ and a group of transformations $\mathcal{G} = \{g\}$ as in section 2. Differing from section 2, we now let T_θ represent $(\mathcal{X}, \underline{T}, P_\theta)$ where $\underline{T}$ denotes a subfield of $\underline{X}$; this is an advantage of the subfield approach--no new range space or induced probability measures need be introduced.

A set $A \in \underline{X}$ is <u>invariant</u> <u>under</u> $\mathcal{G}$ if $x \in A$ implies $gx \in A$ for all $g \in \mathcal{G}$. A subfield $\underline{T} \subseteq \underline{X}$ is an <u>invariant</u> <u>subfield</u> <u>of</u> X <u>under</u> $\mathcal{G}$ if all sets $B \in \underline{T}$ are invariant under $\mathcal{G}$. A subfield $\underline{T} \subseteq \underline{X}$, defined by

$$\underline{U} = \{B : B \in \underline{X},\ B \text{ invariant under } \mathcal{G}\}$$

is the <u>maximal</u> <u>invariant</u> <u>subfield</u> <u>of</u> X <u>under</u> $\mathcal{G}$. (It is easily verified that $\underline{U}$ is a field.) Thus, $\underline{U}$ is the class of all invariant sets in $\underline{X}$, and equivalently can be defined as the subfield generated by the <u>orbits</u> in $\underline{X}$ under $\mathcal{G}$. If $\underline{X}_o$ is a subfield of $\underline{X}$, then (<u>maximal</u>) <u>invariant</u> <u>subfields</u> <u>of</u> $\underline{X}_o$ <u>under</u> $\mathcal{G}$ are defined analogously.

Following Bahadur [1], we define: a subfield $\underline{S} \subseteq \underline{X}$ is a <u>sufficient</u> <u>subfield</u> of $\underline{X}$ for X_{Θ} if, corresponding to each set $A \in \underline{X}$, there exists an $\underline{S}$-measurable function $\phi_A(x)$ such that

$$\phi_A(x) = \mathcal{E}_\theta(\chi_A(x) \mid \underline{S}) \quad [\underline{X}, \theta] \text{ for each } \theta \in \Theta$$

where χ_A is the indicator function for A.

A subfield $\underline{V} \subseteq \underline{X}$ is an <u>invariantly</u> <u>sufficient</u> <u>subfield</u> of $\underline{X}$ <u>for</u> X_{Θ} <u>under</u> $\mathcal{G}$ if $\underline{V}$ is a sufficient subfield of $\underline{U}$ for U_{Θ}, where $\underline{U}$ is the maximal invariant subfield for $\underline{X}$ under $\mathcal{G}$. Thus, corresponding to each invariant set $B \in \underline{X}$, there exists a $\underline{V}$-measurable function $\psi_B(x)$ such that

$$\psi_B(x) = \mathcal{E}_\theta(\chi_B(x) \mid \underline{V}) \quad [\underline{X}, \theta] \text{ for each } \theta \in \Theta .$$

Actually, if any set in $\underline{U}$ or $\underline{V}$ contains a subset of zero measure for all θ which is not invariant, this would not affect the sufficiency properties; we have defined them as strictly invariant for simplicity. We now give a series of lemmas; the first three are almost immediate consequences of the corresponding definitions.

LEMMA 1 (Lemma 4.8 in [1]): If $\underline{X}_{oo}$ and $\underline{X}_o$ are subfields for which $\underline{X}_{oo} \subsetneq \underline{X}_o \subseteq \underline{X}$, then, for every $A \in \underline{X}$,

$$\mathcal{E}_\theta(\chi_A(x) \mid \underline{X}_{oo}) = \mathcal{E}_\theta(\mathcal{E}_\theta(\chi_A(x) \mid \underline{X}_o) \mid \underline{X}_{oo}) \ [\underline{X}, \theta].$$

LEMMA 2: Let $f(x)$ denote a real-valued $\underline{X}$-measurable function and $\underline{U}$ the maximal invariant subfield of $\underline{X}$ under $\mathcal{G}$. Then $f(x)$ is (strictly) invariant under $\mathcal{G}$ if and only if $f(x)$ is $\underline{U}$-measurable.

Proof: $A_r = \{ x : f(x) < r \} \in \underline{X}$ for all real r. If f is invariant, $x \in A_r$ implies $f(x) < r$ implies $f(gx) < r$ implies $gx \in A_r$ implies $A_r \in \underline{U}$. The converse is also straightforward and is omitted.

LEMMA 3 (Lemma 4.6 in [1]): If $\underline{X}_o$ is a subfield of $\underline{X}$, and if $f(x)$ is a bounded real-valued $\underline{X}_o$-measurable function, then

$$\mathcal{E}_\theta(f(x) \mid \underline{X}_o) = f(x) \ [\underline{X}, \theta].$$

LEMMA 4: If B is an invariant set in $\underline{X}$ under $\mathcal{G}$, if $\underline{X}_o$ is a subfield of $\underline{X}$, and if $\phi_{B,\theta}(x)$ represents a fixed determination of $\mathcal{E}_\theta(\chi_B(x) \mid \underline{X}_o)$, then

$$\phi_{B\theta}(x) = \phi_{B,\bar{g}\theta}(gx) \ [\underline{X}, \theta]$$

for all $g \in \mathcal{G}$ and $\theta \in \Theta$.

Proof: For $A_o \in \underline{X}_o$,

$$\int_{A_o} \phi_{B,\bar{g}\theta}(x)\, dP_{\bar{g}\theta}(x) = P_{\bar{g}\theta}(A_o \cap B) \text{ by definition of } \phi$$

$$= P_\theta(g^{-1}A_o \cap g^{-1}B) = P_\theta(g^{-1}A_o \cap B) \text{ since B is invariant}$$

$$= \int_{g^{-1}A_o} \phi_{B,\theta}(x)\, dP_\theta(x) = \int_{A_o} \phi_{B,\theta}(g^{-1}y)\, dP_\theta g^{-1}(y) \text{ transforming } y=gx$$

$$= \int_{A_o} \phi_{B,\theta}(g^{-1}x)\, dP_{\bar{g}\theta}(x).$$

Thus, $\phi_{B,\bar{g}\theta}(x) = \phi_{B,\theta}(g^{-1}x)$ $[\underline{X}, \theta)$ for all g.

Remark: If $\underline{X}_o$ in Lemma 4 is sufficient for $\underline{X}_\Theta$, then $\phi_{B,\theta}(x)$ may be chosen to be free of θ; the lemma them asserts that $\phi_B(x)$ is invariant under $\mathcal{G}$ for almost all x, and, by redefinition if necessary, $\phi_B(x)$ can be taken to be strictly invariant.

THEOREM: If $\underline{S}$ is a sufficient subfield of $\underline{X}$ for X_Θ and $\underline{V}$ is the maximal invariant subfield of $\underline{S}$ under $\mathcal{G}$, then $\underline{V}$ is an invariantly sufficient subfield of $\underline{X}$ for X_Θ under $\mathcal{G}$.

Proof: We must prove that, corresponding to each $B \in \underline{U}$, there exists a $\underline{V}$-measurable function

$$\psi_B(x) = \mathcal{E}_\theta(\chi_B|\underline{V}) \quad [\underline{X}, \theta] \text{ for each } \theta \in \Theta .$$

Since $\underline{V} \subseteq \underline{S} \subseteq \underline{X}$, Lemma 1 implies

$$\mathcal{E}_\theta(\chi_B(x)|\underline{V}) = \mathcal{E}_\theta(\mathcal{E}_\theta(\chi_B(x)|\underline{S})|\underline{V}) \quad [\underline{X}, \theta].$$

By the sufficiency of S, we may replace the inner expectation on the RHS by an S-measurable function $\phi_B(x)$, which, by the remark above, we note is invariant under $\mathcal{G}$. By Lemma 2, with S and V replacing X and U, $\phi_B(x)$ is V-measurable, which, using Lemma 3, yields

$$\mathcal{E}_\theta(\chi_B(x)|\underline{V}) = \phi_B(x) \quad [\underline{X}, \theta];$$

with $\psi_B = \phi_B$, the proof is complete.

REFERENCES

1. R. R. Bahadur, "Sufficiency and statistical decision functions," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 423-461.

2. R. R. Bahadur, "Statistics and subfields," Annals of Mathematical Statistics, Vol. 26 (1955), pp. 490-497.

3. R. R. Bahadur, "On unbiased estimates of uniformly minimum variance," Sankhyā, Vol. 18 (1957), pp. 211-224.

4. R. R. Bahadur and E. L. Lehmann, "Two comments on sufficiency and statistical decision functions," Annals of Mathematical Statistics, Vol. 26 (1955), pp. 139-142.

5. David Blackwell, "On the translation parameter problem for discrete variables," Annals of Mathematical Statistics, Vol. 22 (1951), pp. 393-398.

6. David Blackwell and M. A. Girshick, Theory of Games and Statistical Decisions John Wiley and Sons, New York, 1954.

7. D. R. Cox, "Sequential tests for composite hypotheses," Proceedings of the Cambridge Philosophical Society. Vol. 48 (1952), pp. 290-299.

8. Herbert T. David and William H. Kruskal, "The WAGR sequential t-test reaches a decision with probability one," Annals of Mathematical Statistics, Vol. 27 (1956), pp. 797-805.

9. D. A. S. Fraser, "Sufficient statistics with nuisance parameters," Annals of Mathematical Statistics, Vol. 27 (1956), pp. 838-842.

10. William Kruskal, "The monotonicity of the ratio of two non-central t density functions," Annals of Mathematical Statistics, Vol. 25 (1954), pp. 162-165.

11. E. L. Lehmann, Notes on Theory of Testing Hypotheses, Chapter IV, mimeographed, Associated Students Store, Berkeley, California.

12. E. L. Lehmann, "Some principles of the theory of testing hypotheses," Annals of Mathematical Statistics, Vol. 21 (1950), pp. 1-26.

13. E. L. Lehmann and C. M. Stein, "The admissibility of certain invariant statistical tests involving a translation parameter," Annals of Mathematical Statistics, Vol. 24 (1953), pp. 473-479.

14. National Bureau of Standards, Tables to Facilitate Sequential t-Tests, Applied Mathematics Series, No. 7, Government Printing Office, Washington, 1951.

15. National Bureau of Standards, Tables of the Confluent Hypergeometric Function F(n/2, 1/2; x) and Related Functions, Applied Mathematics Series, No. 3, Government Printing Office, Washington, 1949.

16. Charles Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, Berkeley and Los Angeles, University of California Press, 1956, pp. 197-206.

17. Abraham Wald, Sequential Analysis, John Wiley and Sons, New York, 1947.

18. Oscar Wesler, "Invariance theory and a modified minimax principle," Annals of Mathematical Statistics, Vol. 30 (1959) pp. 1-20.