INVARIANCE, MINIMAX SEQUENTIAL ESTIMATION, AND CONTINUOUS TIME PROCESSES

By J. Kiefer[1]

Cornell University

1. Introduction and summary. The main purpose of this paper is to prove, by the method of invariance, that in certain sequential decision problems (discrete and continuous time) there exists a minimax procedure δ^* among the class of all sequential decision functions such that δ^* observes the process for a constant length of time. In the course of proving these results a general invariance theorem will be proved (Sec. 3) under conditions which are easy to verify in many important examples (Sec. 2). A brief history of the invariance theory will be recounted in the next paragraph. The theorem of Sec. 3 is to be viewed only as a generalization of one due to Peisakoff [1]; the more general setting here (see Sec. 2; the assumptions of [1] are discussed under Condition 2b) is convenient for many applications, and some of the conditions of Sec. 2 (and the proofs that they imply the assumptions) are new; but the method of proof used in Sec. 3 is only a slight modification of that of [1]. The form of this extension of [1] in Secs. 2 and 3, and the results of Secs. 4 and 5, are new as far as the author knows.

In 1939 Pitman [2] suggested on intuitive grounds the use of best invariant procedures in certain problems of estimation and testing hypotheses concerning scale and location parameters. In the same year Wald [3] had the idea that the theorem of Sec. 3 should be valid for certain nonsequential problems of estimating a location parameter; unfortunately, as Peisakoff points out, there seems to be a lacuna in Wald's proof. During the war Hunt and Stein [4] proved the theorem for certain problems in testing hypotheses in their famous unpublished paper whose results have been described by Lehmann in [5a], [5b]. Peisakoff's previously cited work [1] of 1950 contains a comprehensive and fairly general development of the theory and includes many topics such as questions of admissibility and consideration of vector-valued risk functions which will not be considered in the present paper (the latter could be included by using the devise of taking linear combinations of the components of the risk vector). Girshick and Savage [6] at about the same time gave a proof of the theorem for the location parameter case with squared error or bounded loss function. In their book [7], Blackwell and Girshick in the discrete case prove the theorem for location (or scale) parameters. The referee has called the author's attention to a paper by H. Kudō in the *Nat. Sci. Report of the Ochanomizu University* (1955), in which certain nonsequential invariant estimation problems are treated by extending the method of [7]. All of the results mentioned above are nonsequential. Peisakoff [1] mentions that sequential analysis can be considered in his development,

Received July 12, 1956; revised March 15, 1957.

[1] Research sponsored by the Office of Naval Research.

but (see Sec. 4) his considerations would not yield the results of the present paper.

A word should be said about the possible methods of proof. (The notation used here is that of Sec. 2 but will be familiar to readers of decision theory.) The method of Hunt and Stein, extended to problems other than testing hypotheses, is to consider for any decision function δ a sequence of decision functions $\{\delta_i\}$ defined by

$$\delta_i(x, \Delta) = \int_{G_n} \delta_i(gx, g\Delta)\mu(dg)/\mu(G_n)$$

where μ is left Haar measure on a group G of transformations leaving the problem invariant and $\{G_n\}$ is a sequence of subsets of G of finite μ-measure and such that $G_n \to G$ in some suitable sense. If G were compact, we could take $\mu(G) = 1$ and let $G_1 = G$; it would then be clear that δ_1 is invariant and that $\sup_F r_{\delta_1}(F) \leqq \sup_F r_\delta(F)$, yielding the conclusion of the theorem of Sec. 3. If G is not compact, an invariant procedure δ_0 which is the limit in some sense of the sequence $\{\delta_i\}$ must be obtained (this includes proving that, in Lehmann's terminology, suitable conditions imply that any almost invariant procedure is equivalent to an invariant one) and $\sup_F r_{\delta_0}(F) \leqq \sup_F r_\delta(F)$ must be proved. Peisakoff's method differs somewhat from this, in that for each δ one considered a family $\{\delta_g\}$ of procedures obtained in a natural way from δ, and shows that an average over G_n of the supremum risks of the δ_g does not exceed that of δ as $n \to \infty$; there is an obvious relationship between the two methods. Similarly, in [7] the average of $r_\delta(gF_0)$ for g in G_n and some F_0 is compared with that of an optimum invariant procedure (the latter can thus be seen to be Bayes in the wide sense); the method of [6] is in part similar. In some problems it is convenient (see Example iii and Remark 7 in Sec. 2) to apply the method of Hunt and Stein to a compact group as indicated above in conjunction with the use of Peisakoff's method for a group which is not compact. The possibility of having an unbounded weight function does not arise in the Hunt-Stein work. Peisakoff handles it by two methods, only one of which is used in the present paper, namely, to truncate the loss function. The other method (which also uses a different assumption from Assumption 5) is to truncate the region of integration in obtaining the risk function. Peisakoff gives several conditions (usually of symmetry or convexity) which imply Assumption 4 of Sec. 2 or the corresponding assumption for his second method of proof in the cases treated by him, but does not include Condition 4b or 4c of Sec. 2. Blackwell and Girshick use Condition 4b for a location parameter in the discrete case with W continuous and not depending on x, using a method of proof wherein it is the region of integration rather than the loss function which is truncated. (The proof in [6] is similar, using also the special form of W there.) It is Condition 4c which is pertinent for many common weight functions used in estimating a scale parameter, e.g., any positive power of relative error in the problem of estimating the standard deviation of a normal d.f.

The overlap of the results of Secs. 4 and 5 of the present paper with previous publications will now be described. There are now three known methods for proving the minimax character of decision functions. Wolfowitz [8] used the Bayes method for a great variety of weight functions for the case of sequential estimation of a normal distribution with unknown mean (see also [9]). Hodges and Lehmann [10] used their Cramér-Rao inequality method for a particular weight function in the case of the normal distribution with unknown mean and gamma distribution with unknown scale (as well as in some other cases not pertinent here) to obtain a slightly weaker minimax result (see the discussion in Sec. 6.1 of [12]) than that obtainable by the Bayes method. The Bayes method was used in the sequential case by Kiefer [11] in the case of a rectangular distribution with unknown scale or exponential distribution with unknown location, for a particular weight function. This method was used by Dvoretzky, Kiefer and Wolfowitz in [12] for discrete and continuous time sequential problems involving the Wiener, gamma, Poisson, and negative binomial processes, for particular classes of weight functions. The disadvantage of using the Cramér-Rao method is in the limitation of its applicability in weight function and in regularity conditions which must be satisfied, as well as in the weaker result it yields. The Bayes method has the disadvantage that, when a least favorable a priori distribution does not exist, computations become unpleasant in proving the existence (if there is one) of a constant-time minimax procedure unless an appropriate sequence of a priori distributions can be chosen in such a way that the a posteriori expected loss at each stage does not depend on the observations (this is also true in problems where we are restricted to a fixed experimentation time or size, but it is less of a complication there); thus, the weight functions considered in [12] for the gamma distribution were only those relative to which such sequences could be easily guessed, while the proof in [11] is made messy by the author's inability to guess such a sequence, and even in [8] the computations become more involved in the case where an unsymmetric weight function is treated. (If, e.g., $\mathfrak{F}$ is isomorphic to G, the sequence of a priori distributions obtained by truncating μ to G_n in the previous paragraph would often be convenient for proving the minimax character by the Bayes method if it were not for the complication just noted.) The third method, that of invariance, has the obvious shortcoming of yielding little unless the group G is large enough and/or there exists a simple sequence of sufficient statistics; however, when it applies to the extent that it does in the examples of Secs. 4 and 5, it reduces the minimax problem to a trivial problem of minimization.

Several other sequential problems treated in Section 4 seem never to have been treated previously by any method or for any weight function; some of these involve both an unknown scale and unknown location parameter. A multivariate example is also treated in Sec. 4. In example xv of Sec. 4 will be found some remarks which indicate when the method used there can or cannot be applied successfully.

In Sec. 5, in addition to treating continuous time sequential problems in a manner similar to that of Sec. 4, we consider another type of problem where the group G acts on the *time* parameter of the process rather than on the values of the sample function.

2. Assumptions, conditions, examples, and counterexamples. We use the set-up and notation of a fixed sample-size decision problem (the inclusion of the sequential case will be described in Secs. 4 and 5). A random variable X takes on values in $\mathfrak{X}$, which we may think of as being the underlying sample space with Borel field $B_{\mathfrak{X}}$. The family $\mathfrak{F}$ (possible states of nature) is a class of probability measures on $(\mathfrak{X}, B_{\mathfrak{X}})$. We write $P_F\{\ \}$ and $E_F\{\ \}$ to mean "probability of" and "expected value of" when X has probability measure F. The decision spade D has a Borel field B_D associated with it. The weight function W we take to be extended real (possibly $+\infty$) and nonnegative (this could be generalized) on $\mathfrak{F} \times \mathfrak{X} \times D$, jointly measurable in its last two arguments. $\mathfrak{D}$ is the class of decision functions δ from $\mathfrak{X} \times B_D$ into the unit interval which are available to the statistician (not necessarily all possible δ); each such δ is measurable in its first argument and a probability measure in its second one. For fixed F and δ, a probability measure $m_{F,\delta}$ on $\mathfrak{X} \times D$ is defined by its values on rectangles being given by $m_{F,\delta}(Q \times R) = E_F\{\chi_Q(X)\delta(X, R)\}$ where χ_Q is the characteristic function of Q. The risk function of δ is given by $r_\delta(F) = \int W(F, x, s)m_{F,\delta}(dx, ds)$. We define $\bar{r}_\delta = \sup_{F \varepsilon \mathfrak{F}} r_\delta(F)$.

Let G be a group of transformations on $\mathfrak{F} \times \mathfrak{X} \times D$ which operates componentwise; i.e., each $g \varepsilon G$ can be written $g = (g_1, g_2, g_3)$ where g_1, g_2, g_3 are transformations on $\mathfrak{F}$, $\mathfrak{X}$, D, respectively, and where $g(F, x, d) = (g_1F, g_2x, g_3d)$ for all F, x, d. For simplicity of notation we shall write gF, gx, gd in place of g_1x, g_2x, g_3d; this will never be ambiguous. G will be a group (not necessarily the largest) which leaves the problem invariant; i.e., for each $g \varepsilon G$, the probability measure of gX is gF when that of X is F, and $W(gF, gx, gd) = W(F, x, d)$ for all F, x, d. Of course, it is necessary to impose some measurability restrictions on G: the elements of G should be measurable transformations on $\mathfrak{X} \times D$ (thus, gX is a andom variable); moreover, we assume G to be a measurable group; i.e., there s a σ-ring S (closed under differencing and countable intersection but not necesrarily containing G) and a measure μ on (G, S) such that $g \varepsilon G$, $A \varepsilon S$ implies i$A \varepsilon S$ and $\mu(gA) = \mu(A)$ and such that the transformation t of $G \times G$ onto stself defined by $t(g, h) = (g, gh)$ is $S \times S$ measurable. The reader is referred to g13] for a detailed discussion. We mention here the fundamental existence and uniqueness theorem, which states that every locally compact Hausdorff group has such a μ (left Haar measure) on (G, S) where $S =$ Borel sets of G, such that iμ is finite on compacta, positive on non-empty open sets, unique to within multi[plicative constant, and regular. We also impose on W a measurability restriction which will make such integrals as

$$\int_{\mathfrak{X}} \int_H \int_D W^b(F, x, gr)\delta(g^{-1}x, dr)\mu(dg)F(dx)$$

meaningful in Sec. 3, where $\mu(H) < \infty$ and we define $W^b = \max(W, b)$ for each positive number b. We also define r_δ^b to be the risk function of δ when W^b is the weight function, and $\bar{r}_\delta^b = \sup_F r_\delta^b(F)$. We note that assumptions of measurability and invariance are unaltered when W is replaced by W^b. (It is worth noting that any nondecreasing sequence of measurable invariant functions W^{*b} for which $W^{*b} \leqq b$ and $\lim_{b\to\infty} W^{*b} = W$ could be used in place of the W^b throughout this paper. Thus, in some sequential problems where W is a sum of experimental cost and loss due to incorrect decisions, it may be more convenient to use a W^{*b} reflecting separate truncation of these two components than to use W^b which truncates their sum.)

A decision function δ is said to be invariant if $\delta(gx, g\Delta) = \delta(x, \Delta)$ for all $g \varepsilon G$, $x \varepsilon \mathfrak{X}$, $\Delta \varepsilon B_D$. We denote the class of all invariant decision functions in $\mathfrak{D}$ by $\mathfrak{D}_I$.

Let $\mathfrak{F} = \cup_\beta J_\beta$ where β ranges over some index set and the J_β are equivalence classes of $\mathfrak{F}$ under the equivalence $F_1 \sim F_2$ if $F_1 = gF_2$ for some $g \varepsilon G$. Similarly, let $\mathfrak{X} = \cup_\alpha K_\alpha$ where the K_α are equivalence classes under $x_1 \sim x_2$ if $x_1 = gx_2$ for some $g \varepsilon G$. The number of elements in each J_β (or K_α) need not be the same, nor need there be the same number of J_β as K_α, etc. We hereafter denote by F_β a fixed member of J_β.

Remark 1. If $\delta \varepsilon \mathfrak{D}_I$, clearly $r\delta$ is constant on each J_β.

We now list our five assumptions and examples of conditions which imply them.

Assumption 1. *For each δ in $\mathfrak{D}$ there is a function γ_δ from $\mathfrak{X}$ into G such that, writing $\gamma_\delta(x) = g_x$ and $g_x^{-1}x = x^*$ (we shall hereafter not display the allowed dependence on δ), we have $x^* = \bar{x}^* \varepsilon K_\alpha$ if $x, \bar{x} \varepsilon K_\alpha$, and such that for each g in G the function δ_g defined by*

$$\delta_g(x, \Delta) = \delta(gx^*, gg_x^{-1}\Delta) \tag{2.1}$$

is in $\mathfrak{D}$. (We shall sometimes write x_α for the constant value of x^*, $x \varepsilon K_\alpha$.)

It may help the reader to see what δ_g looks like in a simple example. Suppose $\mathfrak{X} = D = G = R^1$ (additive group of reals), so that there is one K_α and we take $x_\alpha = 0$ and $g_x u = x + u$. If δ is a nonrandomized estimator, which we may think of as being a function t from $\mathfrak{X}$ into D, the corresponding δ_g (g a real number) is the function t_g defined by $t_g(x) = x + t(g) - g$.

Remark 2. The measurability portion of Assumption 1 is usually trivial. One must take care to ascertain that $\mathfrak{D}$ is large enough to satisfy the remainder of the assumption. For example, if $\mathfrak{D}$ were taken to be tests of some specified size γ (or $\leqq \gamma$) in a problem of testing hypotheses, δ_g might have size $<\gamma$ (or $> \gamma$) and would not be in $\mathfrak{D}$. This situation is easily handled as noted in Condition 2a below. Counterexample B at the end of this section considers another case where Assumption 1 may be violated.

Assumption 2. *For every δ in $\mathfrak{D}$, h in G, d in D, and x,*

$$g_{hx}^{-1}hd = g_x^{-1} d. \tag{2.2}$$

Remark 3. Since $hx \varepsilon K_\alpha$ if $x \varepsilon K_\alpha$, (2.1) and (2.2) imply

$$\delta_g(hx, h\Delta) = \delta(gx^*, gg_{hx}^{-1}h\Delta) = \delta(gx^*, gg_x^{-1}\Delta) = \delta_g(x, \Delta),$$

so that $\delta_g \varepsilon \mathfrak{D}_I$ for every g. We thus also note, putting g = identity in (2.1) and $g = g_x^{-1}$ in the definition of invariant decision function, that a necessary and sufficient condition for $\delta \varepsilon \mathfrak{D}_I$ is that $\delta \varepsilon \mathfrak{D}$ and $\delta(x, \Delta) = \delta(x^*, g_x^{-1}\Delta)$.

CONDITION 2a. (Testing hypotheses.) Let ω be a non-empty proper subset of $\mathfrak{F}$ and suppose G leaves both ω and also $\mathfrak{F} - \omega$ invariant. Let D consist of two elements d_1, d_2, and suppose $W(F, x, d_2) = c$ if $F \varepsilon \omega$, $W(F, x, d_1) = 1$ if $F \varepsilon \mathfrak{F} - \omega$, and $W = 0$ otherwise. If G is such that gX has probability measure gF when X has F, then G leaves the problem invariant, where G acts trivially[2] on D (i.e., $gd_i = d_i$). Hence, Assumption 2 is automatically satisfied. Let $\mathfrak{D}$ be the class of all tests. It is easy to see that as we let c vary from 0 to ∞ the class of minimax procedures (assuming they exist) for the above problem will yield procedures which maximize the minimum power on $\mathfrak{F} - \omega$ among all tests of size γ (or $\leqq \gamma$) for $0 < \gamma < 1$. An analogous result holds for problems of testing with general invariant W. In particular, the problem of finding a most stringent test of size γ falls within our framework (see *e.g.*, [4]; [5] for discussion). (Our use of the term "size α" does not entail similarity.)

The above condition can obviously be generalized to include k-decision problems where $\mathfrak{F} = \sum_{i=1}^{k} \omega_i$ and G leaves each ω_i invariant. (The problem might be to find a procedure which maximizes the minimum probability of making a correct decision. In some examples such as ranking problems, G may also permute the ω_i.)

CONDITION 2b. For each α, K_α is a homogeneous space G/M_α, M_α being the subgroup of G which leaves x_α fixed (see, e.g., [14]), where M_α acts trivially on D. (A particular important instance of this condition, hereafter denoted Assumption 2b′, is that where $\mathfrak{X} = Y \times Z$, Y being a homogeneous space G/M where M is the subgroup of G leaving some element x_0 of $\mathfrak{X}$ fixed, M acts trivially on D, and G acts trivially on Z. In this case we can write $gx = g(y, z) = (gy, z)$, and we can identify the index α with values $z \varepsilon Z$ since G is transitive on Y and trivial on Z. Some examples where this condition is satisfied will be considered at the and of this section.) To see that Condition 2b implies that (2.2) is satisfied, we note that $x \varepsilon K_\alpha$ implies that $q = g_{hx}^{-1}h$ takes x into x_α, so that qg_x leaves x_α fixed and is thus some element m_α of M_α. Hence, $qd = g_x^{-1}m_\alpha d = g_x^{-1}d$, whichis (2.2).

REMARK 4. Peisakoff assumes, in the notation of Condition 2b′, that Y is isomorphic to G and that $\mathfrak{F}$ consists of the possible probability measures of gX for $g \varepsilon G$ when X has a given probability measure F_0 (thus, we may think of G as being the "parameter space," too). This special case of Condition 2b′ we hereafter refer to as Condition 2bp (see also Example iv below.) Note that in Condition 2b(2b′), $M_\alpha(M)$ need not be normal in G, so $K_\alpha(Y)$ need not be a subgroup of G. Of course, G might be either "larger" or "smaller" than $\mathfrak{F}$, which will be partly reflected by the J_β.

REMARK 5. It is convenient at this point to discuss the question of whether or not it is necessary to consider, as we have, randomized decision functions.

[2] Throughout this paper we shall say that G acts trivially on D or a factor of D if the appropriate component of every g in G is the identity transformation.

We discuss this without consideration of questions of atomicity, our interest here being in the relationship of G and $\mathfrak{F}$ to randomization. Suppose, for example, that the following condition were satisfied:

CONDITION NR. *G is transitive on $\mathfrak{F}$.*

Let $\bar{\alpha}$ be defined by $\bar{\alpha} = \alpha$ when $X \varepsilon K_\alpha$. Define X^* by $X^* = x^*$ if $X = x$. It will usually be a trivial measurability verification to see that $\bar{\alpha}$ and X^* are random variables. If Assumptions 1 and 2 and Condition NR are satisfied, $\delta \varepsilon \mathfrak{D}_I$ implies (see Remark 1) that r_δ is constant and (F_0 being any fixed member of $\mathfrak{F}$) equal to

$$\begin{aligned} r_\delta(F_0) &= E_{F_0} E_{F_0} \left\{ \int_D W(F_0, X, s)\delta(X, ds) \mid \bar{\alpha} \right\} \\ &= E_{F_0} E_{F_0} \left\{ \int_D W(F_0, X, g_X s)\delta(X^*, ds) \mid \bar{\alpha} \right\} \\ &\geqq E_{F_0} \inf_s E_{F_0}\{ W(F_0, X, g_X s) \mid \bar{\alpha} \}, \end{aligned}$$

where the invariance of δ has been used in passing from the second expression to the third. Thus, whether or not the infimum in the last expression is attained, there clearly exists a function s^* of α into D such that, if δ^* is the nonrandomized decision function defined by $\delta^*(x, g_x, s^*(\alpha)) = 1$ when $x \varepsilon K_\alpha$ (we can think of δ^* as a function from $\mathfrak{X}$ into D which takes on the value $g_x s^*(\alpha)$ when $x \varepsilon K_\alpha$), then $r_{\delta^*}(F_0) \leqq r_\delta(F_0)$, provided only that there are no measurability difficulties in defining the function s^*. We shall not go into the last provision, remarking only that mild semi-continuity restrictions on W would suffice and that one could even avoid any measurability considerations by defining risk as an outer integral for "nonmeasurable decision functions." In order to show that, for minimax considerations, one can do as well with the nonrandomized members of $\mathfrak{D}_I$ as with all of $\mathfrak{D}_I$, it remains to show that $\delta^* \varepsilon \mathfrak{D}_I$; this follows at once upon noting that δ^* satisfies the condition given in the last sentence of Remark 3.

In [1] and [7], the authors restrict their consideration to nonrandomized decision functions; we note that Condition NR is satisfied in [1] and [7]. In general, one can not dispense with randomization, as can be seen from many examples where G is not transitive on $\mathfrak{F}$. For example, in estimating the mean θ of a binomial distribution ($0 < \theta < 1$) with $W(\theta, x, d) = |\theta - d|^\rho$ with $0 < \rho < 1$, the only minimax procedures are randomized (see [10]); G consists of two elements here. In many discrete problems of testing hypotheses randomization will also be necessary.

We note that a δ^* formed from an s^* which achieves the infimum (w.p.1 under F_0) above is obviously a uniformly minimum risk decision function among members of $\mathfrak{D}_I$: thus, if Condition NR is satisfied in addition to Assumptions 1 to 5, this gives a prescription for explicitly writing down a minimax procedure. A similar remark applies to ϵ-minimax procedures if the infimum is not attained.

ASSUMPTION 3. *For each b there is a subset Γ_b of $\mathfrak{F}$ with $\Gamma_b \supset \{F_\beta\}$, and a family*

S_b of probability measures on Γ_b which includes each measure giving probability one to a single element of Γ_b, such that

$$\inf_{\delta \varepsilon \mathfrak{D}_I} \sup_{\xi \varepsilon S_b} r_\delta^b(\xi) = \sup_{\xi \varepsilon S_b} \inf_{\delta \varepsilon \mathfrak{D}_I} r_\delta^b(\xi), \tag{2.3}$$

where $r_\delta^b(\xi)$ is the expected value of r_δ^b with respect to the probability measure ξ on $\mathfrak{F}$.

Remark 6. Whenever $\{F_\beta\}$ is finite (e.g., if G is transitive on ω and $\mathfrak{F} - \omega$ or on each ω_i in the case of Condition 2a or is transitive on $\mathfrak{F}$ in 2b; see also Example vi below), if also $\mathfrak{D}$ (or merely $\mathfrak{D}_I$) is convex, (2.3) is trivial. In many other cases it may suffice to let S_b be a family of totally atomic (discrete) measures, so that no measurability difficulty arises in defining $r_\delta^b(\xi)$ (see, e.g., [15], [16]). If one tries to verify (2.3) using an S_b containing more general measures with respect to some Borel field on $\mathfrak{F}$, one must also make sure that conditions implying the existence of the integral $r_\delta^b(\xi)$ are satisfied for these ξ.

It is not clear how essential Assumption 3 is to the validity of the theorem of Sec. 3; it will be seen there that it is used because (3.3) can not in general be verified if integration with respect to ξ is replaced by a supremum over Γ_b there (Counterexamples A to D at the end of this section show that none of our other four assumptions can be entirely dispensed with). The reason for not necessarily putting $\Gamma_b = \{\mathfrak{F}_\beta\}$ is that (2.3) may sometimes be more obvious for a larger Γ_b than for $\{F_\beta\}$.

Assumption 4. *We assume that*

$$\lim_{b \to \infty} \inf_{\delta \varepsilon \mathfrak{D}_I} \bar{r}_\delta^b = \inf_{\delta \varepsilon \mathfrak{D}_I} \bar{r}_\delta . \tag{2.4}$$

(By monotone convergence, the right side of (2.4) is equal to the left with the operations of limit and infimum interchanged.)

Condition 4a. If W is bounded, (2.4) is trivial. This condition will usually be satisfied in problems of testing hypotheses and interval estimation.

Condition 4b. The following set of conditions is not the most general possible of this type, but covers many important cases such as the examples of this section and Sec. 4 for many commonly employed unbounded W. We assume that G is a topological group satisfying Condition 2bp, that $D = G$, and (writing $\mathfrak{F}$ as G) that $W(g, (y, z), d)$ does not depend on y and may hence be written $W(g^{-1}d, z)$. We also assume for each z the existence of an increasing sequence $\{U_r^z\}$ of compact sets whose limit is G and such that every compact subset of G is in some U_r^z, that $W(h, z)$ is bounded in h in each U_r^z and tends to ∞ uniformly in h for $h \not\varepsilon U_r^z$ as $r \to \infty$, and such that, for each r_0 and r_1, the set $U_{r_0}^z(G - U_n^z)$ (group multiplication) is disjoint from $U_{r_1}^z$ for all sufficiently large n. We also assume regularity conditions on W of the type mentioned in the discussion of Condition NR and that there exist (as there will if $(\mathfrak{X}, B_\mathfrak{X})$ is Euclidean with the Borel sets or a countable product of such spaces) conditional probability measures on Y given the Z-coordinate of X. Let F_0 denote the probability measure of X, and let $\bar{F}_0$ denote the probability measure of the Z-coördinate of X, when the element g of $\mathfrak{F}$ is the identity; and let $F_0(A \mid z)$ denote a version of the conditional probability measure on Y, evaluated at the set $A \subset Y$, given that the Z-coördinate of X is z. Our final assumption of Condition 4b is that the compact and sequentially compact subsets of G coincide (this is clearly removable if the next phrase is appropriately re-stated) and that, for each $g_0 \varepsilon G$, $g_i \rightarrow g_0$ implies $\liminf W(yg_i, z) \geqq W(yg_0, z)$ w.p.1 under F_0.

The above condition is not as complicated as it may first seem: for example, if G is the additive group R^m and W is for each z bounded on bounded sets and ∞ at ∞, we can take U_r^z to be the sphere of radius r centered at the origin.

We now verify that Condition 4b implies Assumption 4. If Condition 4b is satisfied, so is Condition NR, and we can restrict ourselves to nonrandomized members of $\mathfrak{D}_I$ in computing either side of (2.4). According to Remark 3 and the discussion of Condition 2b′, these are functions from $\mathfrak{X}$ onto D of the form $t^1(y, z) = yt(z)$ where t is an arbitrary measurable function from Z into D. We hereafter label nonrandomized members of $\mathfrak{D}_I$ by t in place of δ. Since $r_\delta(F) = r_\delta(F_0)$ for $\delta \varepsilon \mathfrak{D}_I$, we have

$$\bar{r}_t^b = \int_Z \int_Y W^b(yt(z), z) F_0(dy \mid z) \bar{F}_0(dz),$$

the same equation holding with no superscript b. Thus, if we show that

$$\lim_{b \to \infty} \inf_{t \varepsilon Y} \int_Y W^b(yt, z) F_0(dy \mid z) = \inf_t \int_Y W(yt, z) F_0(dy \mid z)$$

for each fixed z, (2.4) will follow from monotone convergence. Thus, we may neglect a set of $\bar{F}_0$-measure 0 and delete the z, and it will then suffice to prove that, if W is a nonnegative function on G, satisfying the conditions assumed above for each z in a set of $\bar{F}_0$-measure one, and Q is a fixed probability measure on $Y = G$, and if $\epsilon > 0$ and

$$0 \leqq q < \inf_t \int_Y W(yt) Q(dy),$$

then there is a B such that $b > B$ implies

$$\int_Y W^b(yt) Q(dy) > q(1 - \epsilon) \text{ for all } t \varepsilon G.$$

We hereafter denote integration with respect to Q (over y) by E_Q. First let U_0' be a compact subset of Y with $Q(U_0) > 1 - \epsilon$, and let $V_b = \{y \mid W(y) \leqq b\}$. Thus, the closure of V_b is compact. By our assumption, there is a compact set U_{r_1} of $\{U_r\}$ such that $y \varepsilon U_0'$ and $t \not\varepsilon U_{r_1}$ imply $yt \not\varepsilon V_q$. Hence, $t \not\varepsilon U_{r_1}$ and $b > q$ imply that $E_Q W^b(yt) > q(1 - \epsilon)$, and it remains to show that $t \varepsilon U_{r_1}$ and $b > B'$ for some B' imply the same result. Let $t_b \varepsilon U_{r_1}$ be chosen so that $E_Q W^b(yt_b) < b^{-1} + \inf_{t \varepsilon U_{r_1}} E_Q W^b(yt)$ and let $\{t_{b_i}, i = 1, 2, \cdots\}$ be a subsequence of $\{t_b\}$ with limit t' (say). Then, for each r, since $U_r U_{r_1}$ (group multiplication) is compact,

$$\liminf_{b\to\infty}\, \inf_{t \,\varepsilon\, U_{r1}} E_Q W^b(yt) = \lim_{b\to\infty} E_Q W^b(yt_b) \geqq \liminf_{i\to\infty} \int_{U_r} W^{b_i}(yt_{b_i})Q\,(dy)$$

$$= \liminf_{i\to\infty} \int_{U_r} W(yt_{b_i})Q\,(dy) \geqq \int_{U_r} W(yt')Q\,(dy).$$

Letting $r \to \infty$, the last member must tend to a value $\geqq q$, completing the proof that Condition 4b implies (2.4).

Condition 4c. For brevity we state this for the case where $\mathfrak{F}$, $\mathfrak{X}$, D are as in Condition 4b with G = additive group R^1, but it is easily generalized to versions for other groups. Writing the group operation as addition, we again assume W to be of the form $W(d - g, z)$, but now assume for each z that $W(y, z) < L_z < \infty$ if $y \leqq e_z$, that $W(y, z) \to c_z$ as $y \to -\infty$, and that, for $y \geqq e_z$, $W(y, z)$ is finite and nondecreasing and $\to \infty$ as $y \to \infty$. We also assume as before that, for each real t, $t_i \to t$ implies $\liminf W(y + t_i, z) \geqq W(y + t, z)$ w.p. 1 under F_0. Finally, we assume that there exists at least one member δ_0 of $\mathfrak{D}_I$ for which $r_{\delta_0} < \infty$.

As in the consideration of Condition 4b, by neglecting a set of $\bar{F}_0$-measure zero, we can reduce the problem to proving that $\inf_t E_Q W^b(y + t) \to \inf_t E_Q W(y + t)$ where $W(y) \to c$ as $y \to -\infty$, $W(y) < L < \infty$ for $y \leqq e$, $W(y)$ is nondecreasing for $y \geqq e$ and $\to \infty$ as $y \to \infty$, $t_i \to t$ implies $\liminf W(t_i + y) \geqq W(t + y)$ w.p.1 under Q, and $\inf_t E_Q W(y + t) = q < \infty$. Clearly, for some B_1 and T_1, $b > B_1$ and $t > T_1$ imply $E_Q W^b(y + t) > q$. Also, letting t_0 be any value for which $E_Q W(y + t_0) < \infty$, since $W(y + t) < L + W(y + t_0)$ for $t < t_0$ and all y, we obtain $E_Q W(y + t) \to c$ as $t \to -\infty$ by bounded convergence. Thus, $q \leqq c$. Obviously, for $\epsilon > 0$, $b > L$ and $t < T_2(\epsilon)$ imply $E_Q W^b(y + t) > c - \epsilon \geqq q - \epsilon$. To summarize then, it remains to prove that

$$\lim_{b\to\infty} \inf_{T_2 \leqq t \leqq T_1} E_Q W^b(y + t) \geqq q,$$

where T_1 and T_2 are finite. This case is treated in the same way the case $t \,\varepsilon\, U_{r_1}$ was treated in Condition 4b. Thus, Condition 4c implies (2.4).

The form of our next assumption is Peisakoff's; he calls it "weak boundedness." As usual, we denote $(A - B) \cup (B - A)$ by $A \Delta B$.

Assumption 5. *There exists a sequence* $\{G_n\}$ *of measurable subsets of* G *with* $0 < \mu(G_n) < \infty$ *and such that, for each* g *in* G,

$$\lim_{n\to\infty} \mu(gG_n \Delta G_n)/\mu(G_n) = 0. \tag{2.5}$$

Condition 5a. If G is compact Hausdorff, we can take $G_n = G$ and Assumption 5 is satisfied.

Condition 5b. Peisakoff [1] also gives the following examples of groups satisfying Assumption 5:

(1) G = additive group of R^n (take G_n to be the cube of side n, centered at 0).

(2) G = real affine group; here an element of G is a pair (b, c) with b positive and c real and $(b, c)\,(b', c') = (bb', bc' + c)$, and $d\mu = db\,dc/b^2$; in [1], (2.5) is verified directly if G_n is taken to be the set where $|\, c/b \,| \leqq e^{n^2}$ and $e^{-n} \leqq b \leqq e^n$. A less computational verification can be obtained using Condition 5d below.

Peisakoff attempts to show that the full linear group $GL(n)$ also satisfies (2.5), but his proof seems to be incorrect (see also Counterexample D cited below).

CONDITION 5c. G satisfies Assumption 5 if it is the direct product of two groups satisfying Assumption 5. We omit the obvious proof.

Condition 5c can be used, for example, if G is a direct product of real affine groups. Another example (see Example iv below) is that where G is the direct product of the multiplicative group of positive numbers (scale group) and the orthogonal group $O(n)$ on R^n. In connection with this last example, note that the two factors which generate the group, considered as subgroups of G, of course commute; it is instructive to contrast this or the proof of Condition 5d below with the difficulty one encounters if one tries to verify Assumption 5 for $GL(n)$ by representing an element of the group as (for example) Q_1P or Q_1DQ_2 where Q_1, Q_2 are orthogonal, D is diagonal, and P is positive definite. We next prove that G satisfies (2.5) if a slight strengthening of (2.5) is satisfied for a normal subgroup and factor group of G. This can be used in examples such as that of Condition 5b(2), Example vi, etc.

CONDITION 5d. Suppose a locally compact G has a closed normal subgroup $G^{(1)}$ with factor group $G^{(2)} = G/G^{(1)}$; that for $i = 1, 2$ there is an increasing sequence $\{Q_m^{(i)}\}$ of sets whose union is $G^{(i)}$ and such that $Q_m^{(1)}$ has compact closure and any compact subset of $G^{(1)}$ is in some $Q_m^{(1)}$; that there is a sequence $\{G_m^{(i)}\}$ of measurable subsets of $G^{(i)}$ such that $G_m^{(2)}$ has compact closure; and that $m > n$ and $g^{(i)} \varepsilon Q_n^{(i)}$ imply $\mu^{(i)}(g^{(i)}G_m^{(i)} \cap G_m^{(i)}) > (1 - \epsilon_n)\mu^{(i)}(G_m^{(i)})$ for some sequence $\{\epsilon_n\}$ with $\lim_n \epsilon_n = 0$, where $\mu^{(i)}$ is a left Haar measure on $G^{(i)}$. Under these conditions we shall show that G satisfies Assumption 5. Let $\nu(m) > m$ be such that $\tau^{-1}g^{(1)}\tau \varepsilon Q_{\nu(m)}^{(1)}$ if $g^{(1)} \varepsilon Q_m^{(1)}$ and $\tau \varepsilon G_m^{(2)}$ (the set of all such $\tau^{-1}g^{(1)}\tau$ is contained in a compact set). We shall show that $G_m = G_m^{(2)}G_{\nu(m)}^{(1)}$ satisfies Assumption 5. For let $\varepsilon > 0$ and let $g = g^{(1)}g^{(2)}$ be an arbitrary element of G. Choose n so that $g^{(i)} \varepsilon Q_n^{(i)}$ and $(1 - \epsilon_n)(1 - \epsilon_{\nu(n)}) \geqq 1 - \epsilon$. Since (see Sec. 63 of [13] and the references cited there) $\mu(E) = \int \mu^{(1)}([\tau^{-1}E] \cap G^{(1)})\mu^{(2)}(d\tau)$, and since $(\tau^{-1}g^{(1)}\tau)(\tau^{-1}g^{(2)}G_m^{(2)})G_{\nu(m)}^{(1)} \cap G^{(1)} = \tau^{-1}g^{(1)}\tau G_{\nu(m)}^{(1)}$ if $\tau^{-1}g^{(2)}G_m^{(2)}$ contains the identity and = the empty set otherwise (where $\tau \varepsilon G^{(2)}$), we have, for $m > n$,

$$\begin{aligned}\mu(gG_m \cap G_m) &= \mu(g^{(1)}g^{(2)}G_m^{(2)}G_{\nu(m)}^{(1)} \cap G_m^{(2)}G_{\nu(m)}^{(1)}) \\ &= \int_{g^{(2)}G_m^{(2)} \cap G_m^{(2)}} \mu^{(1)}(\tau^{-1}g^{(1)}\tau G_{\nu(m)}^{(1)} \cap G_{\nu(m)}^{(1)})\mu^{(2)}(d\tau) \\ &\geqq (1 - \epsilon_n)(1 - \epsilon_{\nu(n)})\mu^{(1)}(G_{\nu(m)}^{(1)})\mu^{(2)}(G_{(m)}^{(2)}) \geqq (1 - \epsilon)\mu(G_m),\end{aligned}$$

proving our assertion. (It is easy to extend Condition 5d to more factors.)

EXAMPLES. We list briefly a few examples (of estimation except for Example vi) to illustrate some of the concepts of this section. In each case W will be assumed to satisfy appropriate conditions which will be obvious, and the possible choices of D will be evident if not stated.

(i) (Location parameter) $\mathfrak{X} = R^n$ and, ϵ denoting the n-vector $(1, \cdots, 1)$, $X = (X_1, \cdots, X_n)$ has c.d.f. $F_0(x - \theta\epsilon)$ for some $\theta \varepsilon R^1$ (identified with $\mathfrak{F}$), the form of F_0 being known. Here Condition 2bp is satisfied with $Y = R^1$ = space of X_1 and $Z = R^{n-1}$ = space of $X_2 - X_1, \cdots, X_n - X_1$.

(i′) (Scale parameter) Let R^{*n} be the subset of R^n where no coördinate is zero. For simplicity we assume $R^n - R^{*n}$ has probability zero according to every element of $\mathfrak{F}$, so that we can take $\mathfrak{X} = R^{*n}$. Here $\mathfrak{F}$ is identified with the positive reals and $X = (X_1, \cdots, X_n)$ has c.d.f. $F_0(x/\theta)$ for some $\theta > 0$. Letting $X_1' = \log |X_i|$, $t_i = \operatorname{sgn} X_i$, $t = (t_1, \cdots, t_n)$ and $\theta' = \log \theta$, this problem can be transformed to that considered in Example 1 with the trivial and inessential modification that the sample space is $R^n \times T$ where T, the space of 2^n possible values of t, is acted on trivially by G. (The case where $R^n - R^{*n}$ has positive probability is handled similarly by considering $\mathfrak{X}$ to be the union of subspaces $\mathfrak{X}_i^*$ ($0 \leqq i \leqq n$), where $X_1 = \cdots = X_i = 0$ and $X_{i+1} \neq 0$ in $\mathfrak{X}_i^*$. A similar remark applies in other examples.)

(ii) (Scale and location parameters). Let R^{**n} be the subset of R^n where no two coördinates are equal and $n \geqq 2$ (see also Example v). All elements of $\mathfrak{F}$ will give probability one to R^{**n}, which we take to be $\mathfrak{X}$. $\mathfrak{F}$ will be identified with G = real affine group, and $X = (X_1, \cdots, X_n)$ has d.f. $F_0((x - \theta_1\epsilon)/\theta_2)$ for some $\theta_2 > 0$ and real θ_1. Condition 2bp is satisfied if we take Y to be the space of $(X_1, |X_1 - X_2|)$ and Z to be the space of sgn $(X_1 - X_2)$ and $(X_1 - X_i) / |X_1 - X_2|$, $3 \leqq i \leqq n$.

In the above examples, if F_0 and W have additional symmetry properties, a larger group might leave the problem invariant. Our next two examples illustrate this possibility.

(iii) Consider the setup of Example i with $D = R^1$, $F_0(x)$ symmetric about 0, and W a symmetric function of $\theta - d$ satisfying Assumption 4. As in Example i, the group $G^{(1)}$ = additive group of reals leaves the problem invariant; but so does the larger group G^* = direct product of $G^{(1)}$ and $G^{(2)}$ where $G^{(2)}$ consists of the identity element and an element which takes x, θ, and d into their negatives. We cannot apply Condition 2b here with $G = G^*$ since $G^{(2)}$ does not act trivially on D. However, we can apply Condition 2b (or even 2bp) with $G = G^{(1)}$ and then make a trivial application of the Hunt and Stein method in order to assert that, if $\delta^*(x, \Delta) = [\delta(x, \Delta) + \delta(-x, -\Delta)]$, then $\bar{r}_{\delta^*} \leqq \bar{r}_\delta$; thus, we can conclude that the conclusion of the theorem of Section 3 holds with $G = G^*$. Note that we cannot conclude that there will be a G^*-invariant minimax (or ϵ-minimax) *non-randomized* procedure, since Assumption 2 is violated for G^*; indeed, without some monotonicity restriction on the density of F_0 and on W (which *would* yield this result) this conclusion is false, as can be seen from consideration of the weight function $W = 0$ if $2 < |\theta - d| < 3$ and $W = 1$ otherwise when $F_0(x)$ is normal with mean 0 and variance 1.

The advantage of obtaining the conclusion of the theorem of Section 3 for $G = G^*$ instead of merely $G = G^{(1)}$ in examples of the above variety is, of course, that there are fewer G^*-invariant procedures than $G^{(1)}$-invariant procedures among which we must search for a minimax procedure. Although we would therefore usually like to take G as large as possible, the above example illustrates that the apparent reduction obtained in using G^* in place of a smaller $G^{(1)}$ may in some cases only be illusory, since we may lose the reduction to nonrandomized procedures in passing from $G^{(1)}$ to G^*. However, the example might suggest that the method of Hunt and Stein, used *ab initio*, would result in a simpler treatment. Counter-example C below shows, though, that the use of that method also could not avoid the verification of something like Assumption 2 for the non-compact factor of G. In the following remark we summarize the general result obtained by using the Hunt and Stein method as in Example iii.

REMARK 7. *If G^* is the direct product of $G^{(1)}$ and $G^{(2)}$ where $G^{(2)}$ is compact Hausdorff and where the conclusion of the theorem of Sec. 3 is valid for $G = G^{(1)}$, then that conclusion is valid for $G = G^*$.*

(iv) $\mathfrak{X}$ is $R^{*n} = R^n$-origin of R^n, while $\mathfrak{F}$ is the set of c.d.f.'s $F_0(x / \theta)$ for $\theta > 0$ where under $\theta = 1$ the X_i are independent and normal with mean 0 and variance 1. D is the set of positive reals and, e.g., W is a function of d/θ. We can take G to be the group cited as the second example under Condition 5c. Here $\mathfrak{X} = G / O(n - 1)$ (we can think of $O(n - 1)$ as leaving the point $(1, 0, \cdots, 0)$ fixed) and $O(n - 1)$ (in fact, $O(n)$) acts trivially on D. Thus, Condition 2b′ is satisfied. Since Condition NR is also satisfied, our search for a minimax procedure is reduced to considering nonrandomized estimators of the form $c \sum X_i^2$ where the constant c is chosen to minimize the risk.

Note that Condition 2bp cannot be satisfied for the G used in the above example. If we had treated the example as a case of Example i′ so as to use Condition 2bp, we would have ended up searching through a much larger class of procedures unless we invoke some further principle such as that of sufficiency (in a manner similar to that of Sec. 4). We remark that Peisakoff indicated another method which could be used in some examples such as this one when one wants to use Condition 2bp: Let Q be a random variable independent of X and uniformly distributed on the component $O^+(n)$ of the identity of $O(n)$, and apply Condition 2bp to the G considered above on the sample space $R^{*n} \times O^+(n)$ of $X' = (X, Q)$. The disadvantage of using this technique, where it is possible to do so, is that in some examples further considerations may be required to reduce the class of invariant procedures to that which would have been obtained if Condition 2b′ had been used directly. Note that the technique used here is really related to that of Remark 7, which would give the desired result more directly here, but which would still be inferior to the direct use of Condition 2b′ which does not require the technique of Remark 7 in the present example.

(v) $\mathfrak{X}$ and G are as in Example iii, but with $n = 1$. $D = R^1$, the object being to estimate θ_1. The weight function is, e.g., a function of $(d - \theta_1) / \theta_2$, which we hereafter take to be the argument of W. There is one K_α, and if we try to verify Condition 2b′ we run into trouble. For example, take $x_\alpha = 0$ so that M is the multiplicative group of reals (not normal in G) and $\mathfrak{X} = G/M$; M does not act trivially on D, so Condition 2b′ is not satisfied. If we consider this example as a case of Example i (i.e., let G be the smaller group used there), we obtain for $\mathfrak{D}_I$ the class C of procedures δ for which $\delta(x, \Delta + x) = \delta(0, \Delta)$. If the conclusion of the theorem of Sec. 3 were valid for G = affine group, we could restrict ourselves to those members of C for which $\delta(x, \Delta) = \delta(ax, a\Delta)$ for all $a > 0$; putting $x = 0$, this means $\delta(0, \Delta) = \delta(0, a\Delta)$ for all $a > 0$; taking Δ to be the interval $(-1, 1)$, this means $\delta(0, 0) = 1$; noting the equation defining C, this means that there is only one invariant procedure δ^* under the affine group, the nonrandomized estimator $t(x) = x$. One would like to conclude that this estimator is minimax. If $\liminf_{t\to 0} W(t + X) \geqq W(X)$ w.p.1 when $(\theta_1, \theta_2) = (0, 1)$, an application of Fatou's lemma to the equation

$$r_\delta(\theta_1, \theta_2) = \int \int W\left(u + \frac{r}{\theta_2}\right) F_0(du)\delta(0, dr) \text{ if } \delta \varepsilon C$$

yields the fact that $\liminf_{\theta_2\to\infty} r_\delta(\theta_1, \theta_2) \geqq r_{\delta^*}(=$ constant) for $\delta \varepsilon C$, and the conclusion that δ^* is minimax is justified. However, Counterexample C below shows that without some such additional assumption as the one made here on W, *this conclusion is false*: δ^* need not be minimax and we can only conclude that there is a $\delta \varepsilon C$ which is minimax (or ϵ-minimax).

(vi) The univariate general linear hypothesis (GLH) is discussed in detail in many places. If γ is the parameter on which the power function of the usual F-test of specified size ϵ depends, it is easily proved (see, e.g., [5a]) that this test is uniformly most powerful invariant of size ϵ of the GLH $\gamma = 0$ (against $\gamma > 0$). There are several ways to apply the theorem of the next section to conclude, *e.g.*, that this test is most stringent of size ϵ (first proved by Hunt and Stein). One is to consider for fixed $\gamma_0 > 0$ the problem of testing $\gamma = 0$ against $\gamma = \gamma_0$, to note that G is transitive on ω and on $\mathfrak{F} - \omega$ in this case so that Assumption 3 is satisfied (as are the other assumptions), and thus to conclude that the above test is most stringent of size ϵ; since this is true for every $\gamma_0 > 0$, it follows that the test is most stringent for the original GLH. Another method (better than the above in other problems where such a property uniform in γ_0 may not hold) is to verify Assumption 3 directly for GLH; we can do this easily by applying the theory of [15] to the present case. Alternatively, (2.3) can be verified by considering, on the right side of (2.3), a ξ assigning probability one to the set consisting of one point in ω and one point at which the power function of the F-test differs most from the envelope power function.

COUNTEREXAMPLES. We now list briefly four counterexamples to the conclusion of the theorem of Sec. 3, only the third of which is new, in order to indicate that Assumptions 1, 2, 4, and 5 cannot be entirely dispensed with.

(A) In [6] and also in [7] are given examples which show that the conclusion of the theorem of Sec. 3 is false if (in terms of the present treatment) Assumption 4 is violated. We note here also that if, in the notation of p. 313 of [7], the weight function is altered to $f(s) = 1$ if s is an integer and $f(s) = \max(s, 0)$ otherwise, then there exist invariant procedures with finite risk $(=1)$, but the conclusion of the theorem is still false; thus, we see that if in Condition 4c the condition that $W(y, z) \to c_z$ as $y \to -\infty$ and $\to\infty$ monotonically as $y \to \infty$ were dropped while maintaining the condition that a $\delta_0 \varepsilon \mathfrak{D}_I$ with finite risk exists, Assumption 4 would not be implied.

(B) As Peisakoff has pointed out, the invariance theory applies to the general sequential case only if we restrict $\mathfrak{D}$ to consist of procedures which take at least a first observation with probability one. In Section 4 we shall discuss this in more detail (there are cases where this restriction of $\mathfrak{D}$ is not necessary); for the moment, we give an example to demonstrate that the conclusion of the theorem of the next section would not generally be true without such a restriction. Suppose we are limited to taking a single observation or else no observation on a random variable whose distribution depends only on a location parameter θ which we desire to estimate (see Example (i)), the loss from estimating θ incorrectly being bounded by 1 and the cost of experimentation being 2 or 0 depending on whether or not we take an observation. Any minimax procedure in $\mathfrak{D}$ must clearly take no observation with probability $\geqq \frac{1}{2}$ (a similar remark applying for ϵ-minimax procedures); however, the only invariant procedures take a first observation with probability one (see Sec. 4 for further discussion). The difficulty here is that Assumption 1 is violated, since g_x must depend on the observation and thus, for a δ which requires no observations, the δ_g of (2.1) would require no observations but would depend on the observation, and would thus not be a legitimate decision function.

(C) As an example which shows that Assumption 2 cannot be entirely dispensed with, consider the setup of Example v with $F_0(x) = 0$ if $x < 0$ and $=1$ if $x \geqq 0$, and let $W = 1$ if $d = \theta$ and $=0$ otherwise. This is essentially a game where one player says "don't you name the real number I name" and then names a real number, while the only affine-invariant procedure for the other player is, on hearing the number, to name the same number. The procedure δ^* of Example v is in fact uniformly worst and is clearly not minimax, while there exist many minimax procedures in the class C. This example can be made into one where all members of $\mathfrak{F}$ have densities with respect to a fixed σ-finite measure by restricting $\mathfrak{X}$, D, and G to the rationals (of course, this changes μ, and Condition 5b(2) is no longer applicable), and can be made more probabilistic by letting F_0 assign probability $\frac{1}{3}$ to each of the values $-1, 0, 1$; but the phenomenon persists. See Example v for an example of a condition which eliminates the phenomenon encountered in Counterexample C.

(D) Stein [17] has announced an example in testing hypotheses where all our assumptions except Assumption 5 are satisfied and where the conclusion of the theorem is false. This example shows that the real projective group and $GL(2)$ do not satisfy Assumption 5.

3. Proof of invariance theorem. We now use a modification of the method of proof used in [1] under Condition 2bp and Assumptions 4 and 5, in order to prove the following theorem (see also Remark 7 of Sec. 2):

THEOREM. *If G leaves the problem invariant and if Assumptions 1 to 5 are satisfied, then for any $\delta \varepsilon \mathfrak{D}$ and $\epsilon > 0$ there is a $\delta' \varepsilon \mathfrak{D}_I$ such that $\bar{r}_{\delta'} \leqq \epsilon + \bar{r}_\delta$. In* *particular, if δ^* is minimax among procedures in $\mathfrak{D}_I$, then it is minimax among procedures in $\mathfrak{D}$.*

PROOF. Our first step is to prove (3.5) below. Denote right invariant measure on G by μ^{-1}; i.e., $\mu^{-1}(E) = \mu(E^{-1})$. Fix b and $\delta \varepsilon \mathfrak{D}$ and let $\{G_n\}$ be a sequence satisfying Assumption 5, and define

$$H_{F,x}(g) = \int_D W^b(F, x, gr)\delta(g^{-1}x, dr). \tag{3.1}$$

Then, for $\gamma \varepsilon G$,

$$\begin{aligned} \lim_{n\to\infty} \left| \int_{G_n^{-1}} [H_{F,x}(g^{-1}) - H_{F,x}(\gamma g^{-1})]\mu^{-1}(dg)/\mu(G_n) \right| \\ = \lim_{n\to\infty} \left| \int_{G_n} [H_{F,x}(h) - H_{F,x}(\gamma h)]\mu(dh)/\mu(G_n) \right| \\ \leqq \lim_{n\to\infty} 2b\mu(\gamma G_n \Delta G_n)/\mu(G_n) = 0, \end{aligned} \tag{3.2}$$

by Assumption 5. Using (3.2) with $\gamma = g_x$ and bounded convergence, we obtain, for any fixed $\xi \varepsilon S_b$,

$$\lim_{n\to\infty} \int_{\Gamma_b} \xi(dF) \int_{\mathfrak{X}} F(dx) \int_{G_n^{-1}} [H_{F,x}(g^{-1}) - H_{F,x}(g_x g^{-1})]\mu^{-1}(dg)/\mu(G_n) = 0. \tag{3.3}$$

It will simplify notation if we define the operation L by

$$L = \liminf_{n\to\infty} \int_{\Gamma_b} \xi(dF) \int_{\mathfrak{X}} F(dx) \int_{G_n^{-1}} \mu^{-1}(dg)[\mu(G_n)]^{-1} \int_D . \tag{3.4}$$

Using (3.3), a change of variables, and (2.1), we obtain

$$\begin{aligned} LW^b(F, x, g^{-1}r)\delta(gx, dr) &= LW^b(F, x, g_x g^{-1}r)\delta([g_x g^{-1}]^{-1}x, dr) \\ &= LW^b(F, x, u)\delta(gx^*, d_u g g_x^{-1}u) \\ &= LW^b(F, x, u)\delta_g(x, du). \end{aligned} \tag{3.5}$$

Let $\delta \varepsilon \mathfrak{D}$. Using the fact (Assumption 3) that S_b includes every measure giving probability one to a single element of $\Gamma_b \supset \{F_\beta\}$ and that gX has probability measure gF when X has measure F, we have for any fixed $\delta \varepsilon \mathfrak{D}$,

$$\begin{aligned} \bar{r}_\delta &= \sup_{F \varepsilon \mathfrak{F}} \sup_b \int_{\mathfrak{X}} F(dx) \int_D W^b(F, x, r)\delta(x, dr) \\ &= \sup_b \sup_{\xi \varepsilon S_b} \sup_{g \varepsilon G} \int_{\Gamma_b} \xi(dF) \int_{\mathfrak{X}} F(dx) \int_D W^b(gF, gx, r)\delta(gx, dr) \\ &\geqq \sup_b \sup_{\xi \varepsilon S_b} \liminf_{n\to\infty} \int_{G_n^{-1}} \mu^{-1}(dg)[\mu(G_n)]^{-1} \\ &\qquad \cdot \int_{\Gamma_b} \xi(dF) \int_{\mathfrak{X}} F(dx) \int_D W^b(gF, gx, r)\delta(gx, dr), \end{aligned} \tag{3.6}$$

where the inequality follows from the fact that an average is no greater than a supremum. Using Fubini's theorem ($\mu^{-1}(dg)$ on G_n^{-1} and $\xi(dF)F(dx)$ on $\Gamma_b \times \mathfrak{X}$ are both finite) and the invariance of W (i.e., $W^b(gF, gx, r) = W^b(F, x, g^{-1}r)$), we see that the last member of (3.6) is equal to the supremum with respect to b and ξ of the first member of (3.5). On the other hand, again using Fubini's theorem, the supremum with respect to b and ξ of the last member of (3.5) is equal to

$$(3.7) \qquad \sup_b \sup_{\xi \varepsilon S_b} \liminf_{n\to\infty} \int_{G_n^{-1}} \mu^{-1}(dg) r^b_{\delta_g}(\xi)/\mu(G_n) \geqq \sup_b \sup_{\xi \varepsilon S_b} \inf_{\delta \varepsilon \mathfrak{D}_I} r^b_\delta(\xi),$$

the inequality following from the fact that an average is no less than an infimum and that $\delta_g \varepsilon \mathfrak{D}_I$ for $g \varepsilon G$ (see Remark 3). Using first Assumption 3 and the fact that $\sup_{\xi \varepsilon S_b} r^b_\delta(\xi) = \bar{r}^b_\delta$ if $\delta \varepsilon \mathfrak{D}_I$, and then using Assumption 4, we see that the right side of (3.7) is equal to

$$(3.8) \qquad \sup_b \inf_{\delta \varepsilon \mathfrak{D}_I} \bar{r}^b_\delta = \inf_{\delta \varepsilon \mathfrak{D}_I} \bar{r}_\delta .$$

Thus, for each $\delta \varepsilon \mathfrak{D}$, the first member of (3.6) is no less than the last member of (3.8), proving the theorem.

The above theorem does not, of course, treat the question of whether or not a minimax procedure exists, i.e., whether $\inf_{\delta \varepsilon \mathfrak{D}} \bar{r}_\delta$ is attained. Conditions for this may be found, e.g., in [15] and [16]; the same conditions will usually apply for both $\mathfrak{D}$ and $\mathfrak{D}_I$, so that the conclusion of our theorem can be strengthened by the additional remark that a minimax procedure exists in $\mathfrak{D}_I$ if one exists in $\mathfrak{D}$. Various conditions for the attainment of $\inf_\delta \bar{r}_\delta$ are also given in [1] and [4] (see [5a]). Of course, for suitably simple W one can often write down an explicit formula for a minimax invariant procedure in the manner discussed under Condition NR of Sec. 2; for example, by now this formula is well known in the case studied in [6].

It is of interest to note an observation of Peisakoff to the effect that his proof (under Condition 2bp) will go through in many cases where the elements of $\mathfrak{F}$ are not *all* the distributions gF_0 for $g \varepsilon G$, but only a suitably large subset of these: e.g., in Example i of Sec. 2, the restriction $\theta \geqq 0$ might be imposed. This extension can also be carried out under our assumptions in certain cases where the restricted class of elements g for which $gF_\beta \varepsilon \mathfrak{F}$ is not compact.

4. The sequential case. Our setup in this section is that of Secs. 2 and 3 with certain interpretations. For simplicity our description is specialized to handle the examples stated at the end of this section, although a more general setup is obvious. The space $\mathfrak{X}$ is a product space $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \cdots$ with denumerably many factors or a trivial modification of such a space as in Example ii or iv of Sec. 2, and we write a point of $\mathfrak{X}$ as $x = (x_1, x_2, \cdots)$ and the random variable X as $(X_1, X_2, \cdots)$. In the examples we treat, the $\mathfrak{X}_i$ will be copies of the same Euclidean space and the X_i will be independent and identically distributed according to each $F \varepsilon \mathfrak{F}$. The group G will act componentwise on $\mathfrak{X}$, so we may write $gx = (gx_1, gx_2, \cdots)$. The space D will be a product space $D_1 \times E$ where the "terminal decision space" D_1 has the role the space D had in fixed sample-size problems and $B_D = B_{D_1} \times B_E$ where (B_{D_1}, D_1) is the Borel sets on a subset of a Euclidean space and B_E contains at least the countable subsets of E. The "experimental decision space" E consists of all ordered k-tuples of (not necessarily distinct) positive integers for $k = 0, 1, 2, \cdots$, as well as infinite sequences of positive integers; we represent an element of E by $e_k = (a_1, a_2, \cdots, a_k)$, such a k-tuple representing an experiment carried out in k stages, the ith of which consisted of a_i "observations," namely, on $X_{s_{i-1}+1}, \cdots, X_{s_i}$, where we write $s(e)$ for the sum of the integers in e and $s_k = s((a_1, \cdots, a_k))$; e_0 represents the taking of no observations, and we write $e_\infty = (a_1, a_2, \cdots)$ for an e where experimentation never ceases. The group G acts trivially on E, so that we may write $gd = g(d_1, e) = (gd_1, e)$ in the sequel. The weight function W can depend on F, d_1, and e; for simplicity of exposition, in this section the weight function W will be a sum of two non-negative parts:

$$W(F, x, (d_1, e)) = W_1(F, d_1) + W_2(e), \tag{4.1}$$

although the more general form $W(F, d_1, e)$ can be treated in similar fashion. Thus, W_1 takes the place of the W of the fixed sample-size case and must satisfy the invariance condition $W_1(gF, gd_1) = W_1(F, d_1)$ for all F, d_1 and g. The cost of experimentation $W_2(e_k)$ we assume to be non-negative and finite if $k < \infty$ and infinite if $k = \infty$ (the cases where $W_2(e_k)$ is permitted to be infinite for $k < \infty$ in some treatments of decision theory to reflect upper limits on sampling will be covered by restricting $\mathfrak{D}$ as indicated in Remark 8 below), and we assume the existence of a finite number q and a real nondecreasing function h tending to infinity with its non-negative argument and such that, for all $k < \infty$ and e,

$$\begin{aligned} &W_2((a_1, \cdots, a_k, 1)) &&- W_2((a_1, \cdots, a_k)) < q, \\ &W_2(e) &&> h(s(e)), \\ &W_2((a_1, \cdots, a_k, a_{k+1})) &&\geqq W_2((a_1, \cdots, a_k)); \end{aligned} \tag{4.2}$$

in other words, the cost of taking *one* additional observation at any stage is bounded, for any finite number M only finitely many different e's cost less than M, and additional observations always have non-negative cost. One often imposes on W_2 practical restrictions such as $W_2((a_1 + a_2)) \leqq W((a_1, a_2))$, but this is inessential for our considerations. Typical specializations of W_2 often encountered in practice are $W_2((a_1, \cdots, a_k)) = \sum_1^k W_2((a_i))$ or $= W_2((\sum_1^k a_i))$ the latter case with $W_2((t)) = ct$ being especially important.

Denote by B_n the Borel field of members of $B_{\mathfrak{X}}$ which are cylinder sets with base in $\mathfrak{X}_1 \times \cdots \times \mathfrak{X}_n$; i.e., a B_n-measurable real function of x is one which depends on x only through $(x_1, \cdots, x_n)$, the only B_0-measurable functions being constants. We denote by $\mathfrak{D}^0$ the class of all sequential decision functions δ, i.e., functions δ on $\mathfrak{X} \times B_D$ which are probability measures on D for each x (see also the discussion of the paragraph containing (4.3) below for interpretation) where, in addition to the measurability requirements of Section 2, each $\delta \in \mathfrak{D}^0$ is assumed to satisfy the restriction that if $e = (a_1, \cdots, a_k)$ with $s(e) = r$ and if $Q_{e,a}$ is the set of all elements e_∞ or e_k of E of the form $e_\infty = (a_1, \cdots, a_k, a, \cdots)$ or $e_j = (a_1, \cdots, a_k, a, a_{k+2}, \cdots, a_j)$ for all $j \geqq k + 1$ and all $a_{k+2}, \cdots$, then $\delta(x, \Delta_1 \times e)$ (for each $\Delta_1 \varepsilon B_{D_1}$) and $\delta(x, D_1 \times Q_{e,a})$ are B_r-measurable in x; that is, the decision to stop taking observations or to take a particular number of observations at the next stage depends only on observations which have already been taken. Let $\mathfrak{D}^i$ denote the class of all δ in $\mathfrak{D}^0$ for which $\delta(x, D_1 \times e) = 0$ whenever $s(e) < i$; i.e., which for each x observe at least $x_1, \cdots x_i$ w.p.1. For $i \geqq 0$, let $\mathfrak{D}_I^i$ denote the invariant procedures in $\mathfrak{D}^i$; of course, δ is invariant if $\delta(gx, g\Delta_1 \times e) = \delta(x, \Delta_1 \times e)$ for all g, x, Δ_1, e. We have already seen in Counterexample B of Sec. 2 that the theorem of Sec. 3 will not generally be true if $\mathfrak{D} = \mathfrak{D}^0$ because not all of the δ_g of Assumption 1 will be decision functions. Of course, if G were compact we could use the method of [4] directly as outlined in Sec. 1, without any difficulty. For the examples treated at the end of this section it will suffice to take $\mathfrak{D} = \mathfrak{D}^1$ or $\mathfrak{D}^2$. (The sequential considerations of [1] consist of briefly pointing out an example of the sequential setup of $\mathfrak{D}$ and the necessity of not taking $\mathfrak{D} = \mathfrak{D}^0$.)

The question arises, how much do we lose by restricting $\mathfrak{D}$ to be $\mathfrak{D}^1$ or $\mathfrak{D}^2$ rather than $\mathfrak{D}^0$? The answer will usually be easy to verify. For example, suppose D_1, G, $\mathfrak{F}$, and the X_i are as in Example i (or i′) of Sec. 2 (Examples vii to x of the present section) and that W_1, which we may think of as a function of $\theta - d_1$, tends to its supremum w (say) when its argument tends to ∞ (or, similarly, $-\infty$). Then any procedure δ which requires 0 observations w.p.1 clearly has $\bar{r}_\delta = w$. Since any member of $\mathfrak{D}^0$ can be written as a probability mixture of a procedure in $\mathfrak{D}^1$ and a procedure which requires 0 observations w.p.1, it is evident that either every procedure requiring 0 observations w.p.1 is minimax, or else there is a $\delta \varepsilon \mathfrak{D}^1$ which is minimax. Which of these is the case will be easy to verify in most practical examples. In particular, if $w = \infty$, the second is always the case.

The function δ as given above is (with a different notation) the function p defined in Eq. (1.3) of [15]; $\delta(x, \Delta_1 \times e)$ is the probability, when δ is used and $X = x$, that the experiment will terminate with experimental decision e and terminal decision an element of the subset Δ_1 of D_1. The usual representation of a sequential decision function is obtained by letting $\bar{D}$ be the union of D_1 with the space L of positive integers and writing, for each element e of E and subset $\bar{\Delta}$ of $\bar{D}$,

$$\delta(\bar{\Delta} \mid x, e) = \frac{\delta(x, \bar{\Delta}_1 \times e) + \delta(x, D_1 \times Q_e')}{\delta(x, D_1 \times Q_e)}, \tag{4.3}$$

where Q_e is the set of all elements of the form $e_\infty = (a_1, \cdots, a_k, \cdots)$ or $e_j = (a_1, \cdots, a_k, \cdots, a_j)$ of E for all $j \geqq k$, when $e = (a_1, \cdots, a_k)$ (thus, Q_e is the union of all $Q_{e,a}$ for $a > 0$, together with e), while Q_e' is the union over $a \varepsilon \bar{\Delta} \cap L$ of the sets $Q_{e,a}$, and we let $\bar{\Delta}_1 = \bar{\Delta} \cap D_1$. If the denominator of the right side of (4.3) is 0, define $\delta(\bar{\Delta} \mid x, e) = 1$ or 0 according to whether or not $1 \in \bar{\Delta} \cap L$; the definition in this case is only for definiteness and could be made in many other ways. The left side of (4.3) represents the conditional probability, when δ is used and given that $X = x$ and that the experiment has already proceeded (if $e = (a_1, \cdots, a_k)$) through k stages of experimentation as represented by e, that a terminal decision in $\bar{\Delta}_1$ is made or that the next stage of the experiment consists of a number of observations in $\bar{\Delta}_2 = \bar{\Delta} \cap L$. Clearly, $\delta(\bar{\Delta} \mid x, e)$ is B_r-measurable in x if $s(e) = r$, and the functions $\delta(\bar{\Delta} \mid x, e)$ on $B_{\bar{D}} \times \mathfrak{X} \times E$ satisfying obvious restrictions are in 1-to-1 correspondence with the functions $\delta(x, \Delta)$ on $\mathfrak{X} \times B_D$ as described originally ($B_{\bar{D}}$ consists of every union of a set in B_{D_1} and a set in B_E). Moreover, in terms of our later description, δ is invariant if $\delta(\bar{\Delta} \mid x, e) = \delta(g\bar{\Delta} \mid gx, e)$, where $g\bar{\Delta} = g(\bar{\Delta}_1 \cup \bar{\Delta}_2) = (g\Delta_1) \cup \Delta_2$. We shall use this representation of $\mathfrak{D}_I^i$ below.

The problems we are going to consider are ones in which the difficulty encountered in Counterexample B can be avoided as indicated above, and in which there is a very simple sufficient sequence $\{T_i\}$ of functions on $\mathfrak{X}$, T_i being B_i-measurable (the range space of T_i is immaterial). If one does not employ the principle of sufficiency in the manner of this section the theorem of Sec. 3 will only yield the dependence of the stopping rule on $x_\alpha(=(x_2 - x_1, \cdots, x_n - x_1)$ after n observations in Example vii, for example), nothing like the result we obtain. Specifically, we assume (see Example xv for further remarks)

ASSUMPTION 6. *For some positive integer m, Assumptions 1 and 2 are satisfied for $\mathfrak{D} = \mathfrak{D}^m$ with g_x a B_m-measurable function of x. There exists a sequence $\{T_i\}$ of functions with T_i a B_i-measurable sufficient statistic for $[(X_1, \cdots, X_i), \mathfrak{F}]$, such that there exist conditional probability* d.f.'s

$$F_r(y_1, \cdots, y_r \mid t_r) = P\{g_x^{-1}(X_1, \cdots, X_r) \leqq (y_1, \cdots, y_r) \mid T_r(X) = t_r\}$$

for $r \geqq m$ with the property

$$F_r(y_1, \cdots, y_r \mid t_r) \text{ does not depend on } t_r. \tag{4.4}$$

It will aid understanding to consider an example at this point, Example vii of this section. The X_i are normal with unknown mean and known variance, and $\mathfrak{X}_i = G = D_1 = R^1$. We also identify $\mathfrak{F}$ with R^1 in obvious fashion. We can let $n = 1$ and $g_x u = u + x_1$ for u in $\mathfrak{X}_i$ or D_1, and identify the indices α with sequences $x_\alpha = g_x^{-1}x = (0, x_2 - x_1, x_3 - x_1, \cdots)$. Let $T_i = \sum_{j=1}^{i} X_j$. Since $g_x^{-1}X_1 = 0$ and $g_x^{-1}(X_2, \cdots, X_r) = (X_2 - X_1, \cdots, X_r - X_1)$, the distribution of $g_x^{-1}(X_1, \cdots, X_i)$ given that $T_i(X) = t_i$ is multivariate normal with means and covariances independent of t_i, so that Assumption 6 is satisfied. Similarly, in Example xi with G the affine group and $\mathfrak{X}_i = R^1$, we put $g_x^{-1}x_i = (x_i - x_1)/(x_2 - x_1)$, etc.

Assumption 6 is related to a property cited in [5a] as being proved in [4] in certain regular cases, to the effect that we lose nothing in the validity of the theorem of Sec. 3 for problems considered in [4] if we first use the principle of sufficiency and then apply the invariance principle to the space of a correctly chosen sufficient statistic. Assumption 6 also includes an additional strong property in that (4.4) is obviously not implied by this result of [4] (see also Example xv below). This assumption is easily verified in Examples vii to xiv.

Denote by $Q(s)$ the infimum of $\bar{r}_\delta - W_2(e_1^{(s)})$ over all δ with $\delta(x, D_1 \times e_1^{(s)}) = 1$, where $e_1^{(s)} = (s)$, and by $Q_I(s)$ the infimum when δ is also restricted to be invariant; thus, $Q(s)$ and $Q_I(s)$ are the values of $\inf_\delta \bar{r}_\delta$ over all δ or all invariant δ for the fixed sample-size problem with sample size s when the weight function is W_1. We assume

ASSUMPTION 7. *Either $\bar{r}_\delta = \infty$ for all $\delta \varepsilon \mathfrak{D}^m$ or else there is an integer m' with $Q(m') < \infty$.*

This assumption is easy to verify in practical cases for the examples considered in this section, where one will usually know $Q_I(j) < \infty$ for some j. The assumption can be shown, in fact, to be implied by our other assumptions under mild regularity conditions, although for the sake of brevity we forego such considerations here.

The main remaining difficulty in applying the theorem of Sec. 2 to the present problem is the verification of Assumption 4, which would usually be difficult to verify directly in sequential problems. Our form of the theorem which follows reduces this verification to the much simpler nonsequential one of Sec. 2.

THEOREM. *If G leaves the problem invariant and Assumptions* 3, NR, 5, 6, 7, (4.1), *and* (4.2), *as well as Assumption* 4 *for W_1 in each fixed sample-size problem with sample size $\geqq m$, are satisfied, and if $\mathfrak{D} = \mathfrak{D}^m$, then for each $\epsilon > 0$ there exists a fixed sample-size invariant procedure δ^* (the sample perhaps being taken according to some grouping) with $\bar{r}_{\delta^*} \leqq \epsilon + \inf_{\delta\varepsilon\mathfrak{D}} \bar{r}_\delta$. Thus, if $Q_I(s(e)) + W_2(e)$ is minimized over $s(e) \geqq m$ by $e = e'$ and if δ^* is a minimax invariant procedure for the fixed sample-size problem with sample size $s(e')$ ignoring W_2, then a minimax procedure for the sequential problem is to take $s(e')$ observations according to the grouping e' (which minimizes $W_2(e)$ over e satisfying $s(e) = s(e')$) and then to use δ^*.*

REMARK 8. Before proving the theorem we remark that the first paragraph of the proof below can easily be altered to handle the case where $\mathfrak{D}$ is further restricted in some way such as bounding k or the a_i or $s(e_k)$ in $e_k = (a_1, \cdots, a_k)$, etc. We have already noted the fact that it will usually be easy to verify whether a minimax procedure of $\mathfrak{D}^m$ or a more trivial procedure is minimax in $\mathfrak{D}^0$. We also note that one can think of G as acting on T_r for $r \geqq m$ in the examples treated by us, so that the conclusion of the theorem could be phrased in terms of invariant functions of T_r.

PROOF OF THEOREM. We may assume $\rho = \inf_{\delta\varepsilon\mathfrak{D}} \bar{r}_\delta < \infty$, the theorem being trivial otherwise. By Assumption 7 there is an m' and a procedure δ^0 with $\delta^0(x, D_1 \times e_1^{(m')}) = 1$ and $\bar{r}_{\delta^0} - W_2(e_1^{(m')}) = C < \infty$. Since the X_i are independent and identically distributed we can clearly assume $m' \geqq m$. Let ϵ be a positive number. The second line of (4.2) implies the existence of a number $N' > m$ such that any procedure $\delta \varepsilon \mathfrak{D}^m$ with $\bar{r}_\delta < \rho + \epsilon$ must require fewer than N' observations with probability $>1 - \epsilon$ for all $F \varepsilon \mathfrak{F}$. For any such δ define the procedure δ' as one which proceeds like δ except that whenever experimentation has reached a stage e (including e_0) where $s(e) < N' \leqq s(e) + t$ for some t with $\delta(x, e_1^{(t)} \mid e) > 0$, δ' assigns the probability $\delta(x, e_1^{(t)} \mid e)$ which δ assigned to the taking of t observations at the next stage (there may of course be several such t) to the taking of exactly m' additional observations one-by-one and, if these observations are taken, uses δ^0 on these last m' observations to reach a terminal decision. Since the X_i are independent and identically distributed, by the first and last lines of (4.2) we clearly have $\bar{r}_{\delta'} < \bar{r}_\delta + \epsilon(C + qm')$ and $\delta' \varepsilon \mathfrak{D}^m$. Since $\epsilon > 0$ is arbitrary, we conclude that our theorem will be proved if we prove it for the case where $\mathfrak{D}$ is restricted to the class $\mathfrak{D}^{m,N}$ of procedures in $\mathfrak{D}^m$ for which $\delta(x, D_1 \times E_N) = 1$, where N is a fixed integer and E_k is the set of e for which $s(e) < k$. We hereafter assume $\mathfrak{D} = \mathfrak{D}^{m,N}$.

In order to apply the theorem of Sec. 3 to the present case, it remains to verify Assumption 4 when $\mathfrak{D} = \mathfrak{D}^{m,N}$. Let $Q_I^b(s)$ be the value of $Q_I(s)$ when W_1 is replaced by W_1^b. By Assumption 4 in the fixed sample-size case, we have

$$\lim_{b\to\infty} [W_2(e) + Q_I^b(s(e))] = W_2(e) + Q_I(s(e))$$

for each fixed e with $s(e) \geqq m$. Since there are only finitely many e with $m \leqq s(e) < N$, we obtain, for $\mathfrak{D} = \mathfrak{D}^{m,N}$,

$$\begin{aligned}\lim_{b\to\infty} \inf_{\delta\,\varepsilon\,\mathfrak{D}_I} \bar{r}_\delta^b &= \lim_{b\to\infty} \min_{m\leqq s(e)<N} [W_2(e) + Q_I^b(s(e))] \\ &= \min_{m\leqq s(e)<N} [W_2(e) + Q_I(s(e))] = \inf_{\delta\,\varepsilon\,\mathfrak{D}_I} \bar{r}_\delta,\end{aligned}$$

which is Assumption 4 for the present problem.

Applying, then, the theorem of Sec. 3, we obtain for any $\delta \,\varepsilon\, \mathfrak{D}^{m,N}$ and $\epsilon > 0$ an invariant procedure δ' with $\bar{r}_{\delta'} \leqq \bar{r}_\delta + \epsilon$. Since δ' is invariant, we have

$$\begin{aligned}\delta'(x, \Delta \mid e) &= \delta'(g_x^{-1}x, g_x^{-1}\Delta \mid e) \\ &= \delta'(g_x^{-1}x, g_x^{-1}\Delta_1 \mid e) + \delta'(g_x^{-1}x, \Delta \cap L \mid e).\end{aligned}$$

Define the procedure δ'' by

$$\begin{aligned}\delta''(x, \Delta \mid e) &= E\{\delta'(g_X^{-1}X, g_X^{-1}\Delta_1 \mid e) \mid T_{s(e)} = T_{s(e)}(x)\} \\ &\quad + \int \delta'(y, \Delta \cap L \mid e) F_{s(e)}(dy_1, \cdots, dy_{s(e)} \mid T_{s(e)}(x)).\end{aligned} \tag{4.5}$$

Since B_{D_1} = Borel sets on a Euclidean set, this defines a decision function for some version of the conditional expected value (see, e.g., [18]). Clearly $\delta''(x, D_1 \times E_k) = 0$ for $k = m$ and $= 1$ for $k = N$, so $\delta'' \,\varepsilon\, \mathfrak{D}^{m,N}$. Since $\{T_i\}$ is sufficient, $r_{\delta'} = r_{\delta''}$. But for each e and $\Delta_2 \subset L$, Assumption 6 implies that $\delta''(x, \Delta_2 \mid e)$ is a constant. Hence δ'' can be considered to be a member of the class ϕ of probability mixtures of fixed sample-size procedures of sample-sizes $s(e)$ with $m \leqq s(e) < N$, where the sample may be taken according to some grouping e (independent of X). It is easy to see that, under our assumptions, the result of the previous paragraph remains true if $\mathfrak{D}^{m,N}$ is replaced by ϕ and that Assumptions 1, 2, 3, and 5 remain satisfied; thus, the theorem of Sec. 3 is valid for $\mathfrak{D} = \phi$, so that there is a $\delta^* \varepsilon \phi_I \subset \mathfrak{D}_I^{m,N}$ with $\bar{r}_{\delta^*} \leqq \bar{r}_{\delta''} + \epsilon \leqq \bar{r}_\delta + 2\epsilon$. This completes the proof of the theorem, since condition NR implies the constancy of r_{δ^*} and hence the existence of a fixed sample-size $\delta^\# \varepsilon \phi_I$ with $r_{\delta^\#} \leqq r_{\delta^*}$.

We note that δ'' in the preceding paragraph can be proved invariant in our examples, for an appropriate version of the first term on the right in (4.5), but the proof as given seems just as short. The lack of dependence of W on x in (4.1) is of course used in invoking sufficiency.

Examples. We shall use the following notation in our examples, where z and θ_1 are real and θ_2 and γ are positive:

$$f_1(z; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi}\theta_2} e^{-(z-\theta_1)^2/2\theta_2^2},$$

$$f_2(z; \theta_1, \theta_2) = \begin{cases} 1/2\theta_2 & \text{if } |z - \theta_1| < \theta_2 \\ 0 & \text{otherwise,} \end{cases}$$

$$f_{3\gamma}(z; \theta_1, \theta_2) = \begin{cases} (z - \theta_1)^{\gamma-1} e^{-(z-\theta_1)/\theta_2}/\theta_2^\gamma \Gamma(\gamma) & \text{if } z > \theta_1 \\ 0 & \text{otherwise.} \end{cases}$$

In all the examples except xiv, the X_i will be independent real random variables whose common Lebesgue density will be assumed to be in some class of the above densities, which class we identify with $\mathfrak{F}$.

(vii) $\mathfrak{F}$ consists of the densities f_1 for $-\infty < \theta_1 < \infty$ with θ_2 assumed known and θ_1 to be estimated and hence $G = D_1 = \mathfrak{F} =$ additive group of R^1, W_1 being a function of $\theta_1 - d_1$. Note that in most practical examples Assumption 4 can be verified by applying Condition 4a or 4b or 4c, and the question of whether to use a procedure requiring no observations or one in $\mathfrak{D}^1$ will be easy to settle. Of course, we can take $T_i = \sum_{j=1}^{i} X_j$, $m = 1$, and $g_x u = u + x_1$, as previously mentioned. Thus, the conclusion of the theorem will be satisfied for most W_1 and W_2 encountered in practice. Of course, $Q_I(s)$ is easily computed in this case to be given by

$$Q_I(s) = \inf_h \int_{-\infty}^{\infty} W_1(h + u) f_1(u; 0, \theta_2 s^{-\frac{1}{2}})\, du; \tag{4.6}$$

and, if h_s achieves the minimum, a nonrandomized sequential minimax estimator will be given by taking $s(e')$ observations according to the grouping e' described in the statement of the theorem and then estimating θ_1 by $s(e')^{-1} T_{s(e')} + h_{s(e')}$.

(vii+) We mention several extensions of Example vii: (1) The form of the minimax (or an analogous ϵ-minimax) estimator above depends on θ_2 in such a way that if it were only known that θ_2 belonged to some set B (not necessarily the set of all positive numbers) and if W_1 were a function of $(d_1 - \theta_1)\theta_2^{-1}$ instead of $(d - \theta_1)$, then the estimator of the previous problem vii for the case $\theta_2 = 1$ would be minimax (or ϵ-minimax) here. (2) A second extension is to note that, for the original setup of Example vii, if W_1 is symmetric we can also apply the group of reflections as in Example iii of Sec. 2. If in addition W_1 is nondecreasing in $|\theta_1 - d_1|$, we obtain the sample mean ($h_s = 0$ in vii) as minimax estimator, a result first obtained in [8], with a special case in [9]. Note that the question of whether a procedure in $\mathfrak{D}^1$ or one requiring no observations is minimax is trivial in this case.

(viii) Same as vii, except that the possible distributions are the $f_{3\gamma}$ with $\gamma = 1$ and θ_2 known and θ_1 unknown with $-\infty < \theta_1 < \infty$. In this case $T_i = \min(X_1, \cdots, X_i)$ and the considerations and conclusions are as in vii with f_1 replaced by $f_{31}(u; 0, \theta_2 s^{-1})$ in (4.6), the minimax estimator being $T_{s(e)} + h_{s(e)}$. A very special case of this was obtained tediously in [11].

(ix) $\mathfrak{F}$ consists of the densities f_2 with θ_1 known and θ_2 unknown, $0 < \theta_2 < \infty$. Here θ_2 is to be estimated, so $D_1 = \mathfrak{F} = G^{(1)}$ = multiplicative group of positive reals. The weight function is a function only of θ_2/d_1. We can either take $G = G^{(1)}$ or can think of θ_1 is being 0 and let G = direct product of $G^{(1)}$ and $G^{(2)}$ where $G^{(2)}$ contains the identity and an element which multiplies X_i by -1 and leaves $\mathfrak{F}$ and D fixed. We have

$$m = 1 \quad \text{and} \quad T_i = \max(|X_1 - \theta_1|, \cdots, |X_i - \theta_1|)$$

and $g_x^{-1}u = u/(x_1 - \theta_1)$ if $G = G^{(1)}$, with an obvious modification if we let $G = G^{(1)} \times G^{(2)}$. In either case, Assumption 6 is satisfied. Of course, this problem is really the same as that of estimating θ_2 when X_i has density $1/\theta_2$ for $0 < x_i < \theta_2$ and 0 otherwise (put $X_1' = |X_i - \theta_1|$ above), and in this form the problem may be reduced to that of viii by a logarithmic transformation as in Example i′. The form of the analogue of (4.6) and of the minimax procedure are obvious. The special case $W_2 = (\theta_2 - d_1)^2/\theta_2^2$ was considered in [11]; Condition 4c is satisfied there.

(x) $\mathfrak{F}$ consists of the densities $f_{3\gamma}$ where γ and θ_1 are known and θ_2 is unknown, $0 < \theta_2 < \infty$. This is a scale parameter problem with G = the $G^{(1)}$ of ix, and we need only remark that the theorem applies with $T_i = \sum_{j=1}^{i} X_j$, the analogues of (4.6) and the form of the minimax procedure being obvious. This problem was treated for a particular γ and weight function in [10] and for a special class of weight functions in [12].

(x′) If X_i has symmetric density about known θ_1, the density of $|X_i - \theta_1|$ being that of Example x, the same considerations apply, using also the $G^{(2)}$ of ix. Similarly, the problem of estimating θ_2 when f_1 is the density and θ_1 is known can obviously be reduced to that of Example x.

The next three examples are similar in that, in each, there is both an unknown location parameter θ_1 and also an unknown scale parameter θ_2 with $-\infty < \theta_1 < \infty$ and $0 < \theta_2 < \infty$. In each case $m = 2$, G is the real affine group (see Example ii), and $g_x^{-1}x_i = (x_i - x_1)/(x_2 - x_1)$. There are three main types of problems in each example: (1) estimation of both θ_1 and θ_2, so that $D_1 = G$, $d_1 = (d_{11}, d_{12})$, W_1 is a function of $(\theta_1 - d_{11})/\theta_2$ and d_{12}/θ_2, and

$$g_x^{-1}\, d_1 = ((d_{11} - x_1)/(x_2 - x_1),\, d_{12}/(x_2 - x_1));$$

(2) estimation of θ_1, where $D_1 = R_1$, W_1 is a function of $(\theta_1 - d_1)/\theta_2$, $g_x^{-1}\, d_1 = (d_1 - x_1)/(x_2 - x_1)$; (3) estimation of θ_2, where D_1 = positive reals, W_1 is a func-

tion of d_1/θ_2, $g_x^{-1} d_1 = d_1/(x_2 - x_1)$; of course, (2) and (3) can really be considered as special cases of (1) where W_1 only depends on one of its two arguments. For each type and example it is simple to write down an analogue to (4.6) and the corresponding form of a minimax procedure. In each case the conditions of the theorem are easily verified for many commonly used W, and the verification of whether one should use a procedure in $\mathfrak{D}^2$ or one requiring one or no observations is also easy. The use of the Bayes method in these examples would of course be much more complicated than that in [8], [11], and [12].

(xi) $\mathfrak{F}$ consists of all densities f_1. Putting $\bar{X}^{(i)} = i^{-1} \sum_{j=1}^{i} X_j$, we have $T_i = (\bar{X}^{(1)}, \sum_{j=1}^{i}(X_j - \bar{X}^{(i)})^2)$ for $i \geqq 2$. Note that the problem of estimating θ_2, even for the appropriate weight function, cannot be obtained by the method of [10] without some modification, because of the nature of the Cramér-Rao bound.

(xii) $\mathfrak{F}$ consists of all densities f_2. Putting $U_i = \min(X_1, \cdots, X_i)$ and $V_i = \max(X_1, \cdots, X_i)$, we can take $T_i = (U_i, V_i)$ or $((U_i + V_i)/2, (V_i - U_i))$, for $i \geqq 2$. (The second form of T_i here and in the next example are pertinent to remarks made below in Example xv.)

(xiii) $\mathfrak{F}$ consists of all densities f_{31} (i.e., γ is known to be 1), and in the notation of the previous two examples we can take $T_i = (U_i, \bar{X}^{(i)})$ or $(U_i, \bar{X}^{(i)} - U_i)$ for $i \geqq 2$.

(xiv) As an example of a multivariate nature, suppose $\mathfrak{X}_i = R^J$ for some positive integer J, the X_i again being independent and identically distributed. Here $X_i = (X_{i1}, \cdots, X_{iJ})$, and we assume X_i has a multivariate normal distribution with the identity convariance matrix and unknown mean $\theta = (\theta_1, \cdots, \theta_J) \varepsilon R^J$. The problem is to estimate θ, so that $D_1 = \mathfrak{F} = G =$ additive group of R^J and W_1 is a function of the difference between the vectors d_1 and θ_1. Taking $m = 1$ and $T_i = \bar{X}^{(i)}$ and $g_x^{-1}u = u - x_1$ for $u \varepsilon R^J$, the theorem is applicable for many common weight functions. (Examples viii to xiii have similar multivariate analogues.)

(xiv+) We can extend Example xiv in the manner of vii+. In particular, if W_1 is an increasing function of the usual Euclidean *distance* between d_1 and θ, it is easy to see that $\bar{X}^{(s(e'))}$ is a minimax sequential estimator. The orthogonal group also leaves the problem invariant in this case, but this fact need not be used in obtaining the above form of the minimax estimator, it sufficing to apply a result of [19]. It is interesting to note that it is shown in [20] that, when W_1 is the squared length of the distance and $J > 2$, this estimator is not admissible.

(xv) As an example which illustrates the fact that the method of this section yields little if no T_i satisfy Assumption 6, consider the problem of estimating θ_1 when the X_i have density f_2 and θ_2 is known. This problem is considered for certain W in [15] and [21], and the minimax procedures obtained there are not fixed sample-size. As in Example xii, (U_i, V_i) is a minimal sufficient statistic. Assumption 6 cannot be satisfied for any sufficient T_i. The application of our method in this example would yield the form of the estimator obtained in [15] and [21], but would only yield the fact that the minimax stopping rule depends on $U_i - V_i$ at the ith stage; the stationary form of the minimax stopping rule seems to depend strongly on the particular nature of f_2. It will be noted that the previous examples differ from this one in that in the former, but not in the latter, *there is a natural version of* T_i *for* $i \geqq m$ *whose range is* G *and such that the problem in terms of the* T_i *is left invariant by the natural operation of* G *on the range of* T_i. This is the essence of the examples where the method of this section yields the conclusion of the theorem, although we have seen that G may be modified somewhat from what this statement indicates (see Examples vii+, ix, x′, and xiv+) to the case where the *range of* T_i *is a subgroup or homogeneous space of* G. We may add that, in most sequential *testing* problems, the invariance principle yields little, for reasons similar to those present in Example xv.

REMARK 9. We end this section with a remark about other versions of the statistical problem, such as that of minimaxing the W_1 component of the risk subject to a bound on the W_2 component or vice versa. This includes such problems as the problem of finding optimum sequential estimators of bounded relative error of the scale parameters in Examples ix to xiv (in [7] there is some discussion of this problem but our results are not obtained) and that of obtaining optimum sequential interval estimators of prescribed length and confidence coefficient for the location parameters in Examples vii and viii. The latter problem is considered in [8] and [9] in the case of Example vii, while [8] considers also the problem of minimaxing one component of risk subject to inequalities on *two* others, etc. The discussion of [8], [21], and [12] shows at once on application of our theorem that results of all these types hold for appropriate fixed sample-size procedures, or probability mixtures thereof, in Examples vii to xiv.

5. Sequential problems with continuous time. In this section we will use the method developed in Secs. 3 and 4 to obtain certain sequential minimax results for decision problems concerned with stochastic processes with continuous time parameter. Two types of problems will be considered: in Part I of this section we treat problems where the invariance is present in the same form as in Sec. 4, while in Part II the invariance has to do with the time parameter.

I. *Extension of Section 4 to continuous time.* The problems we consider here will be continuous time analogues of certain of the problems of Sec. 4 (in fact, those of Sec. 4 can be considered as special cases of those here, in the manner of [12]). Since the proofs are essentially identical to those of Sec. 4, we shall not give them. In fact, rather than to state a general theorem, we shall merely list three examples. In each of these the separable process $\{X(t),\ t \geqq 0\}$ is one of independent and stationary increments which can be taken to be continuous on the right, and $X(T)$ is sufficient for $\{X(t),\ 0 \leqq t \leqq T\}$. As in Sec. 4, W can be a function of $\theta^{-1} d$ (θ being the unknown parameter) and of the experimentation decision, but for convenience of exposition we discuss the case where it is a sum $W_1 + W_2$. The cost of experimentation W_2 may either be taken to be of the form $W_2(T)$ if the process is observed continuously up to time T, or else the cost may be allowed to depend on the number and spacing of the instants at which the process is observed; a description of this and other modifications (such as the problem of having to give an estimate continuously), as well as a more detailed discussion of the nature of sequential decision functions in the case of continuous time, and of the processes considered, will be found in [12]. In all of the examples, assumptions on W_2 can be treated as in Sec. 4. The analogue here of the restriction to $\mathfrak{D}^1$ in Sec. 4 is that we must restrict ourselves to the union over all $\epsilon > 0$ of the classes $\mathfrak{D}^\epsilon$ of procedures which observe the process for at least $0 \leqq t \leqq \epsilon$ w.p.1 for all $F \varepsilon \mathfrak{F}$. When we consider $\mathfrak{D}^\epsilon$, the g_x is a function of $X(\epsilon)$. As in Sec. 4, it will be easy in most practical cases to decide whether there will be a minimax procedure in $\mathfrak{D}^\epsilon$ for some $\epsilon > 0$ or a minimax procedure which does not observe the process at all.

In each of the three examples, our result is, under assumptions on W like those of Sec. 4, that *there exists an invariant minimax or ϵ-minimax procedure which observes the process for a constant length of time* w.p.1 (or a minimax procedure which does not observe the process at all). Formulas for computing the minimax procedure can be given as in Sec. 4 or [12], and Remark 9 of Sec. 4 applies also to these examples.

(xvi) The process is the one-dimensional Wiener process with known variance per unit time and with $EX(t) = \theta_1 t$, the object being to estimate θ_1. Thus, G, $\mathfrak{F}$, D, and the form of W_1 are the same as in Example vii. In particular, in the special case of a symmetric monotone W_2, we obtain the result of Sec. 5 of [12].

(xvi′) For the Wiener process with unknown scale or unknown location and scale, it has been shown in [12] that the scale parameter can be estimated with arbitrarily high accuracy in arbitrarily short time. Hence, the only new practical problems that arise when the scale parameter is unknown do so because W_2 reflects the number of instants at which the process is observed. In this case, as indicated in [12], we obtain problems analogous to Example xi with G the affine group, or to Example x′ (see also the next example below). In either of these problems there will be an invariant minimax procedure which observes the process at a certain set of instants specified in advance of the experiment.

(xvii) The process is the Gamma process; i.e., $X(0) = 0$ and $X(1)$ has density function $f_{3\gamma}$ of Sec. 4 with $\theta_1 = 0$ and γ known, the object being to estimate the scale parameter θ_2. Here $\mathfrak{F}$, D, G, and W_1 are the same as in Example x of Sec. 4.

(xviii) Consider the J-variate Wiener process $X(t) = (X_1(t), \cdots, X_J(t))$ where the $X_i(t)$ are independent with known scale factors and $EX_i(t) = \theta_i t$, the θ_i being unknown, $-\infty < \theta_i < \infty$. This is the continuous time analogue of Example xiv, and the considerations there and in xiv+ carry over to the present example.

II. *Invariance in time.* We now consider a process $\{X(t), t \geqq 0\}$ with unknown parameter $\theta > 0$ and with the property that, if $\{X(t), t \geqq 0\}$ has probability law labeled θ, then the process $\{X_c(t), t \geqq 0\}$, defined by $X_c(t) = X(ct)$ where $c > 0$, has probability law labeled $c\theta$. The most familiar process of this kind is the Poisson process. Another such process is the gamma process with θ_2 known and γ unknown.

Suppose the weight function (for estimating θ) is a function only of d_1/θ and

T/θ, where d_1 is the terminal decision and T is the length of time experimentation is carried on (modifications of the type mentioned earlier in this section and discussed in [12] are also possible). Then clearly the multiplicative group of positive reals leaves the problem invariant, where we define $g(\{X(t)\}, \theta, (d_1, T)) = (\{X(gt)\}, g\theta, (gd_1, g^{-1}T))$, the group operation being ordinary multiplication. The difference here from previous problems is that G acts on the process by shifting the time argument of a sample function by a scale factor rather than by operating on the values of the sample function, and that G acts nontrivially on the experimental decision. The reason for allowing this last action and the accompanying dependence of W on T/θ rather than on T lies in the form of the result which this setup yields when one applies the invariance theorem and examines the invariant procedures.

The details here are slightly more delicate and lengthy than those in Part I, so we shall be content with sketching the main idea. Consider the Poisson process with right continuous sample functions. $X(\tau)$ is sufficient for $\{X(t), 0 \leqq t \leqq \tau\}$. Suppose we have a nonrandomized stopping function which depends on the sufficient statistic, i.e., a nonnegative functional T of the process with the property that the event $t_1 < T \leqq t_2$ is measurable with respect to the Borel field generated by $\{X(t), t_1 < t \leqq t_2\}$. For such a T to be invariant we must have $T(x) = cT(x_c)$ for all $c > 0$ and all sample functions x, where x_c is the sample function of X_c when x is the sample function of X. It is easy to see that such a stopping function as $T(x) = $ constant is not invariant, while $T_r(x) =$ first time t that $x(t) = r$, where r is a fixed positive integer, is. In the present problem we must restrict $\mathfrak{D}$ to decision functions which observe the process until at least the first time $X(t) = 1$ (that time gives g_x^{-1}). Under fairly general conditions one can verify whether or not a minimax procedure should observe the process at all and that, if it does, a stopping rule of the type T_r is minimax. Of course, an invariant nonrandomized estimator will be of the form constant$/T_r$. A special case of this result thus shows that the procedure suggested in Sec. 3 of [22] and which was asserted there to be minimax among all procedures using a particular stopping rule T_r (analogous to a fixed sample-size problem) actually has an optimum property among all sequential procedures: e.g., among all procedures which give at least the prescribed accuracy of estimation, this one minimaxes $E_\theta T/\theta$.

REFERENCES

[1] M. P. Peisakoff, "Transformation Parameters," Thesis, Princeton University, 1950.

[2] E. J. G. Pitman, "The estimation of location and scale parameters of a continuous population of any given form," *Biometrika*, Vol. 30 (1939); "Tests of hypotheses concerning location and scale parameters," *Biometrika*, Vol. 31 (1939), pp. 200–215.

[3] A. Wald, "Contributions to the theory of statistical estimation and testing hypotheses," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 299–326.

[4] G. A. Hunt and C. Stein, "Most stringent tests of statistical hypotheses," unpublished.

[5a] E. L. Lehmann, Notes on testing hypotheses, University of California, Berkeley, 1949.

[5b] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1-26.

[6] M. A. GIRSHICK AND L. J. SAVAGE, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–74.

[7] D. BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decision Functions*, John Wiley and Sons, New York, 1954.

[8] J. WOLFOWITZ, "Minimax estimates of the mean of a normal distribution with known variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 218–230.

[9] C. STEIN AND A. WALD, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 427–433.

[10] J. L. HODGES, JR., AND E. LEHMANN, "Some applications of the Cramér-Rao inequality," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 13–22.

[11] J. KIEFER, "Sequential minimax estimation for the rectangular distribution with unknown range," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 586–593.

[12] A. DVORETZKY, J. KIEFER, AND J. WOLFOWITZ, "Sequential decision problems for processes with continuous time parameter. Problems of estimation," *Ann. Math. Stat.*, Vol. 24, 1953, pp. 403–415.

[13] P. R. HALMOS, *Measure Theory*, D. Van Nostrand Co., Inc., New York, 1950.

[14] C. CHEVALLY, *Theory of Lie Groups*, Princeton University Press, Princeton, 1946.

[15] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

[16] L. LECAM, "An extension of Wald's theory of statistical decision functions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 69–81.

[17] C. STEIN, "On tests of certain hypotheses invariant under the full linear group" (abstract), *Ann. Math. Stat.*, Vol. 26 (1955), p. 769.

[18] R. R. BAHADUR, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.

[19] T. W. ANDERSON, "The integral of a symmetric unimodal function over a convex set and some probability inequalities," *Proc. Amer. Math. Soc.*, Vol. 6 (1955), pp. 170–176.

[20] C. STEIN, "Inadmissibility of the usual estimate for the mean of a multivariate normal distribution" (abstract), *Ann. Math. Stat.*, Vol. 26 (1955), p. 157.

[21] C. R. BLYTH, "On minimax statistical decision problems and their admissibility," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 22–42.

[22] M. A. GIRSHICK, H. RUBIN, AND R. SITGREAVES, "Estimates of bounded relative error in particle counting," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 276–285.