ASYMPTOTICALLY OPTIMUM SEQUENTIAL INFERENCE AND DESIGN

By J. Kiefer[1] and J. Sacks[2]

Cornell University and Northwestern University

0. Summary. In recent years the study of sequential procedures which are asymptotically optimum in an appropriate sense as the cost c per observation goes to zero has received considerable attention.

On the one hand, Schwarz (1962) has recently given an interesting theory of the asymptotic shape, as $c \to 0$, of the Bayes stopping region relative to an a priori distribution F, for testing sequentially between two composite hypotheses $\theta \leqq \theta_1$ and $\theta \geqq \theta_2$ concerning the real parameter θ of a distribution of exponential (Koopman-Darmois) type, with indifference region the open interval (θ_1, θ_2). (An example of Schwarz's considerations is described in connection with Figure 4.) One aim of the present paper is to generalize Schwarz's results to the case where (with or without indifference regions) the distributions have arbitrary form and there can be more than two decisions (Sections 2, 3, 4). In this general setting we obtain, under mild assumptions, a family $\{\delta_c\}$ of procedures whose integrated risk is asymptotically the same as the Bayes risk. (In fact, extending Schwarz's result, a family $\{\delta_c'\}$ can be constructed so as to possess this asymptotic Bayes property relative to all a priori distributions with the same support as F, or even with smaller indifference region support than F.) Procedures like our $\{\delta_c\}$ have already been suggested by Wald (1947) for use in tests of composite hypotheses (e.g., the sequential t-test), but his concern was differently inspired.

At the same time, we show how such multiple decision problems can be treated by using simultaneously a number of sequential tests for associated two-decision problems.

A second aim is to extend, strengthen, and somewhat simplify the asymptotic sequential design considerations originated by Chernoff (1959) and further developed by Albert (1961) and Bessler (1960) (Section 5). Our point of departure here is a device utilized by Wald (1951) in a simpler estimation setting, and which in the present setting amounts to taking a preliminary sample with predesignated choice of designs and such that, as $c \to 0$, the size of this preliminary sample tends to infinity, while its ratio to the total expected sample size tends to zero. The preliminary sample can then be used to "guess the true state of nature" and thus to choose the future design pattern once and for all rather than to have to reexamine the choice of design after subsequent observations. (In Wald's setting the only "design" problem was to pick the size of the second sample of his two-sample procedure.) The properties of the resulting procedure can then be inferred

Received August 8, 1962; revised March 5, 1963.

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

[2] Research sponsored by the Office of Naval Research and by NSF Grant G24500.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

from the considerations of Sections 2, 3, and 4, where there is no design problem but where most of the work in this paper is done; using Wald's idea, we thereby obtain procedures for the design problem fairly easily, once we have the (non-design) sequential inference structure to build upon. The family $\{\delta_c^*\}$ so obtained has the same asymptotic Bayes property as that described above for the family $\{\delta_c\}$ of the non-design problem. Furthermore, a family $\{\delta_c^{**}\}$ can be constructed so that, like $\{\delta_c'\}$ in the non-design problem, it is asymptotically Bayes for all a priori distributions with the same support. The value of the asymptotic Bayes risk of such a family is closely related to the lower bound which was obtained by Chernoff et al for the risk function of certain procedures, and which gives another form for the optimality statement.

The role of the sequential procedures considered by Donnelly (1957) and Anderson (1960) for hypothesis testing with an indifference region is indicated at the end of Section 1. Asymptotic solutions to the problem of Kiefer and Weiss (1957) are given.

An Appendix contains proofs of certain results on fluctuations of partial sums of independent random variables, which are used in the body of the paper.

1. Introduction and comments on related work. Chernoff (1959), whose work initiated the design investigations mentioned in the summary, has also given an introductory heuristic discussion which motivates these considerations. We therefore omit such a discussion, mentioning only the differences between the present and previous work.

As mentioned in the summary, most of the effort in the present paper is devoted to generalizing Schwarz's results. Since we no longer have his exponential type structure, it is not possible to obtain the concise proofs and elegant characterizations that he obtains. It should be noted that Chernoff's and Albert's papers can be regarded as considering, inter alia, this (nondesign) problem without indifference region, if one allows only one design in their treatments. The first of these considers finitely many states, while the latter obtains ϵ-optimum families of procedures for tests of hypotheses containing infinitely many states. (Albert mentions an indifference region in the early part of his paper, but does not consider it in the domain of his risk function in the later part, so that his procedures are really ϵ-optimum only for the problem without indifference region.) Bessler also considers finitely many states, but infinitely many experiments.

Since we cannot invoke the monotone likelihood ratio structure to obtain Schwarz's simple asymptotic reduction of a hypothesis like $\theta \leqq \theta_1$ to one like $\theta = \theta_1$, we use compactness and appropriate continuity in Section 2 to reduce the problem to one involving finitely many states. The final argument used to obtain optimality (Theorem 1) reduces the consideration to that of testing between two simple hypotheses, and a comparison with known properties of the Sequential Probability Ratio Test (SPRT). When we introduce an indifference region in Section 3, this comparison must be made with a Bayes sequential test between two simple hypotheses with a single indifference state.

In the absence of compactness of the space of states of nature, Remarks 1, 2, 3 and 5 of Section 2 state the compactification assumptions we require. Albert also uses a compactification device in this case, but his is associated with the type of maximum likelihood (ML) argument development used in ML consistency proofs by Wald (1949) and by Kiefer and Wolfowitz (1956). Our development, which is in terms of Bayes procedures (as is Schwarz's), seems somewhat simpler to us; Albert remarks on the possible complexity of procedures using ML estimates in this way, on page 798 of his paper. (Asymptotic relations between Bayes and ML procedures in simpler contexts are well known.)

The optimality results of Chernoff, Albert, and Bessler are all stated in the non-Bayesian language of our Corollary to Theorem 1, while Schwarz's results are obtained in the Bayesian terms of Theorem 1 itself. (Although Schwarz's discussion is mainly in terms of shapes of stopping regions, an earlier, mimeographed version of his paper described his result also in terms of the Bayes risk, itself; in our case the difference between the two descriptions is greater, and is essentially the difference between Lemma 4 and Theorem 1.) Schwarz mentions briefly (page 234) the relationship between these two forms of the optimality result in his case, and Theorem 2 (and its analogues in Theorems 4, 5, and 6) make the result precise in our case: *for a given set of possible states of nature, we obtain a family of procedures which is asymptotically Bayes for every a priori distribution whose support is the given set.* (When there are only finitely many states, Chernoff's procedure clearly achieves this.) One can easily find examples where the hypothesis of the Corollary to Theorem 1 is violated by an asymptotically Bayes family, for example, when the family is asymptotically Bayes relative to an a priori distribution whose support is a proper subset of the given set; the violation of hypothesis and conclusion are only to be expected in such a case, since one can not generally find a family which is simultaneously asymptotically Bayes against a priori distributions with different support. It would be interesting to obtain the conclusion to Theorem 2 for the procedure of Theorem 1, but we are unable to prove such a result.

In the presence of an indifference region I, the family of procedures we obtain has the additional property that *it is also asymptotically Bayes relative to every a priori distribution whose support* S' *satisfies* $S' = (S - I) \cup A$ *with* $A \subset I$, *where* S *is the given supporting set* $(S \supset I)$.

Turning to our design considerations, the Summary and Section 5 describe the spirit of these. As for the results themselves, our considerations extend those of the previous papers in this area by considering both infinite sets of possible states and infinite sets of possible experiments, by including multiple decisions, indifference regions, and semi-indifference regions, by eliminating the ϵ in Albert's ϵ-optimality result (by use of a trivial device which could also be applied in his treatment), and by obtaining procedures with the strong uniform Bayes property mentioned in the two previous paragraphs. However, we regard it as more important than these extensions, that one sees that asymptotically optimum designs can be described and used in the simpler manner evolving from our extension of

Wald's method (wherein design prescriptions are decided at the outset and at one later stage, rather than after every stage), and that their properties can be verified largely by reference to the nondesign results. While no asymptotic theory of this type can contain any strict admissibility results, an examination of various nonasymptotic design problems leads us to a preference in applications and where c is not too small, for certain nonrandomized design choices over randomized ones which are demonstrably worse. Chernoff had indicated in his paper that such nonrandomized choices can be used, but the formal demonstrations of optimality in previous papers refer only to the random method (mentioned in Section 5) for choosing designs.

Most of our notation is similar to that of previous papers in the area, as is our model of independent, identically distributed observations in the nondesign work, and of identical possible choices of experiment at each stage in the design work. The reader is referred to Wald (1947) and Chernoff (1959) for asymptotic properties of the SPRT which we use. The fundamental role of the information numbers (whose use, as Albert mentions, was described by Wald (1947), although they subsequently came to be known as the Kullback-Leibler numbers) is well described in Chernoff's introductory comments.

Simultaneous tests. We continue this introduction by describing briefly the treatment of multiple decision problems through the use of simultaneous tests. This device has long been used by practical people, and was treated in detail by Lehmann (1957a, 1957b) in questions of nonasymptotic, nonsequential admissibility. In our asymptotic sequential context, considerations are easier, due to the fact that even a fairly large change in the stopping boundary does not alter asymptotic optimality. As an example of this last fact, we plot, in a three-state, three-decision problem, the asymptotically (as $c \to 0$) optimum boundaries given in Section 4 for the three-decision analogue of the SPRT; they are the lines

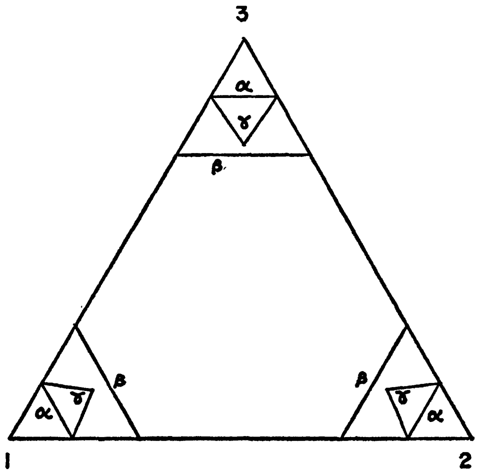

FIG. 1

α of Figure 1, namely (for $0 - 1$ loss) $\xi_{i,n} = 1 - c$ ($i = 1, 2, 3$), where $\xi_{i,n}$ is the a posteriori probability of hypothesis i after n observations. However, if these bounds are placed twice as far from the vertices, i.e., at $\xi_{i,n} = 1 - 2c$ (lines β), we change the expected sample size by a relative amount $\log 2/|\log c| = o(1)$, and still keep the probabilities of error to be of a smaller order of magnitude than the expected sample size. Thus, the second set of bounds is also asymptotically optimum. As Chernoff pointed out, the expected sample size is the overwhelming part of the asymptotic risk in these problems. As a consequence, the asymptotically optimum stopping bounds can vary greatly, and the almost impossible non-asymptotic problem of computing the best bounds disappears.

Now suppose that the statistician tried, in the classical tradition of solving new problems by using a conglomeration of old techniques, to carry on three SPRT's simultaneously, one between each of the three pairs of hypotheses and each with bounds $(1 - c)/c$ and $c/(1 - c)$ on the probability ratio, and that he stopped as soon as any two tests simultaneously dictated acceptance of the same hypothesis. Then his stopping bounds would be the broken lines γ, and these are again easily seen to be asymptotically optimum. It is not difficult to verify that the modification of this procedure which allows the tests to stop at different times (each one, as soon as possible) and makes a final decision as soon as two terminal decisions coincide or three are contradictory (in which case any decision can be made), is also asymptotically optimum, although this procedure cannot be represented so simply diagrammatically.

This technique of using simultaneous tests can also be used in more complicated cases. If one hypothesis consists of state 1 and a second consists of states 2 and 3, the procedure of Theorem 1 has the bounds $\alpha(\xi_{1,n} = 1 - c$ and $\xi_{2,n} + \xi_{3,n} = 1 - c)$ shown in Figure 2. If one instead tests 1 against 2 and 1 against 3 simultaneously with bounds $(1 - c)/c$ and $c/(1 - c)$ on the probability ratio, stopping as soon as both tests say to accept 1 or else either rejects 1, one obtains the bounds γ of Figure 2. The region where 1 is rejected is not even convex, but the procedure is asymptotically optimum; this again points up the extent to which small sample computational difficulties have become trivialized in the asymptotic theory. A modification like that indicated at the end of the previous paragraph is again possible.

As a final example, suppose we are testing between simple hypotheses 1 and 2, with 3 the simple indifference state. The procedure of Theorem 3 then stops as soon as either $\xi_{1,n}$ or $\xi_{2,n}$ is $<c$ (α, Figure 3). This rule can be described in terms of the tests of composite hypotheses without indifference region, of the previous paragraph: test $1 \cup 3$ against 2 and $2 \cup 3$ against 1, simultaneously. (If different critical constants were used, the lines α could be broken near the bottom line $\xi_{3n} = 0$.) Going all the way back to SPRT's, the boundaries γ of Figure 3, which are again asymptotically optimum, arise from using three simultaneous SPRT's and stopping as soon as either 1 or 2 is rejected by some test. The end of this section discusses these procedures further.

Further use of these ideas is contained in Sections 3 and 4. We summarize by

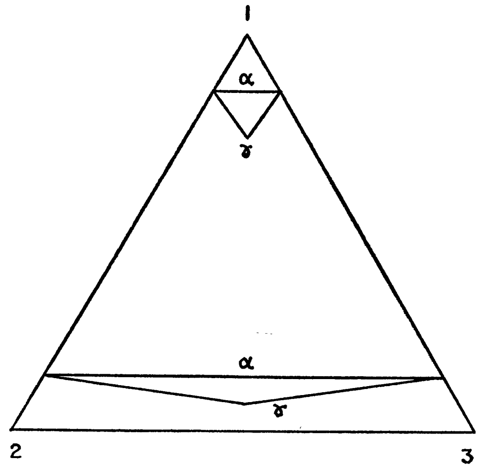

FIG. 2

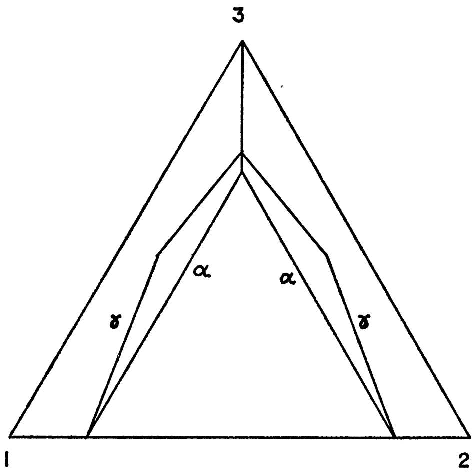

FIG. 3

remarking now that the whole development of asymptotically optimum multiple-decision procedures, with or without indifference and semi-indifference regions, can be carried out entirely in terms of such simultaneous tests. (If only SPRT's of *simple* hypotheses are used, this will, however, involve the use of appropriately fine coverings if the space of states of nature is infinite.)

The procedures of Donnelly, Anderson, and Schwarz. Let us consider the special application of the results of Section 3 to the problem where f_ω, h_α, and g_θ are normal with variance one and means -1, 0, and 1, respectively. In this case we can, with Schwarz, conveniently represent the asymptotic shape of procedures in the $(t, y) = (n/|\log c|, \sum_1^n X_i/|\log c|)$ plane. The solid lines β of Figure 4 indicate the stopping boundary corresponding to either α or γ of Figure 3 as $c \to 0$.

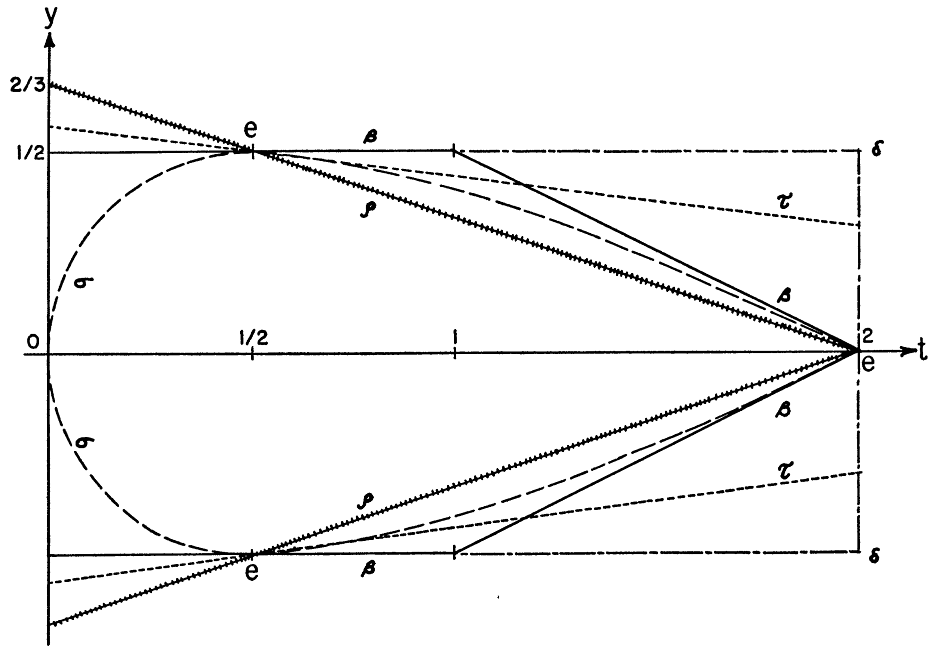

Fig. 4

This is Schwarz's famous pentagon. (The broken nature of β has nothing to do with that of γ, since α yields the same β; rather, the lower broken line β arises from the relation

$$\exp\left(\sum X_i - n/2\right)/\left[1 + \exp\left(\sum X_i - n/2\right) + \exp\left(-\sum X_i - n/2\right)\right] = c$$

which determines one half of α in Figure 3. It is the same shape which can be discerned more generally in the discussion of (3.16) of Section 3 in terms of coordinates $T_n/|\log c|$ and $S_n/|\log c|$ which correspond to $y - t$ and $-y - t$ in the present example.) The three points marked e are the intersection of the boundary with lines of "expected movement" under the three states ($\pm 45°$ lines under f_ω and g_θ, a horizontal line under h_α), and illustrate that the approximate expected sample size is $|\log c|/2$ under f_ω or g_θ and is $2\,|\log c|$ under h_α.

It is to be noted that, in the spirit of our discussion of simultaneous tests, this stopping rule can be obtained by drawing the well known parallel lines in terms of $(n, \sum X_i)$ for each of the three possible SPRT's with stopping constants c and $1/c$ for the probability ratio (it is impossible to reject h_α in both tests where it appears).

Other stopping boundaries of similar piecewise linear shape have been proposed by Donnelly (1957) and Anderson (1960). To obtain the same approximate sample sizes, such boundaries would also have to pass through the points e. One such boundary δ merely truncates, at $t = 2$, the SPRT of f_ω against g_θ. The idea of using such truncated sequential procedures goes back to Wald (1947). One can easily verify (e.g., Anderson, Equation (4.74), and an analogue of the parenthetical remark of the next paragraph regarding translation from continuous to discrete time) that this test is also asymptotically optimum. One might then wonder how Schwarz's test β can be optimum, since it superficially appears to stop sooner with presumably larger resulting probabilities of error. The explanation is that under f_ω it is so unlikely any path will reach a point P on the upper segment of β to the right of $t = 1$ having not gone out earlier, that nothing is lost (asymptotically) by stopping at P. Many other continuous boundaries passing through the three points e would also work, and we shall shortly discuss one such boundary, σ.

On the other hand, other boundaries formed from no more than three line segments which are symmetric in y and which go through the three points e, are *not* asymptotically optimum. (This conclusion also obviously applies to many other bounds through the three points, e.) For example, the "triangular" boundary consisting of the cross-hatched line segments $\rho(y = \pm\frac{2}{3} \mp t/3)$ and whose properties have been studied in detail by Anderson, has the right asymptotic expected sample sizes, but probabilities of error which are too large. This last can be verified from Equation (4.61) of Anderson, which gives probabilities of error of order $c^{8/9}$ for the corresponding Wiener process problem. To put it another way, if one wants symmetric triangular bounds with $P_{\pm 1}\{\text{error}\} = O(c)$ and $E_0N \sim 2 |\log c|$, then the bounds must be $y = \pm a \mp ta/2$ with $a \geqq 1$. For the best of these, $a = 1$, obtains $E_{\pm 1}N \sim \frac{2}{3} |\log c|$, efficiency $\frac{3}{4}$ compared with optimum bounds (like β or δ) for which $E_{\pm 1}N \sim \frac{1}{2} |\log c|$. (Translating the Wiener process results to the original discrete time normal problem, we see that by raising both lines slightly we obtain bounds ρ' for the Wiener process which still give order $c^{8/9+\epsilon}$ for probabilities of error, where $\epsilon < \frac{1}{9}$, and such that the probability that the corresponding discrete time process (obtained by examining the Wiener process at integral multiples of $t = 1/|\log c|$) with bounds ρ rejects f_ω when it is true, exceeds that for the Wiener process with bounds ρ' by $o(c)$; this last bound on the maximum difference between the two processes is standard.) A similar result holds for the procedure whose boundaries are the broken lines τ truncated at $t = 2$: we again obtain probabilities of error which are $c |\log c|/o(1)$, whereas the bounds β or δ yield a risk function of order $c |\log c|$.

We now discuss the particular boundary σ which is designated by a curved broken line in Figure 4 and whose equation is $y = \pm t \mp (2t)^{\frac{1}{2}}$ $(0 \leqq t \leqq 2)$. The boundary σ gives Schwarz's asymptotic shape of the Bayes stopping regions for the problem of testing $\theta \leqq -1$ against $\theta \geqq 1$ with indifference region $I = (-1, 1)$ when the support of the a priori distribution is the entire real line. For Schwarz's family of procedures (which is the same as the family $\{\delta_c^I\}$ of our Theorem 3) the asymptotic shape is given by σ when the a priori distribution is as described in the last sentence, and it is also easy to see that the same holds for the family $\{\delta_c'^{I}\}$ of our Theorem 4. Although $\{\delta_c^I\}$ and $\{\delta_c'^{I}\}$ have the same asymptotic shape, it is not clear that they enjoy the same asymptotic properties; e.g., in Theorem 4 we show (despite the lack of compactness in the parameter space, the methods of Section 3 will work because of the exponential structure of normal densities; the same will not generally be true in non-exponential situations) that $\{\delta_c'^I\}$, constructed for a given F, is asymptotically Bayes with respect to any G having the same support as F but we do not know if the same is true for $\{\delta_c^I\}$ (a discussion of this difference in the simpler context of Section 2 appears following the corollary to Theorem 1). The reason that the asymptotic shape does not seem to yield the asymptotic Bayes properties expressed in Theorem 4 is due to the necessity of knowing how P_θ {error} behaves asymptotically (for Theorem 4 we need P_θ {error} $= o(c |\log c|)$ uniformly for $\theta \not\in I$), and knowing the asymptotic shape of a family of procedures does not seem to yield precise enough knowledge about probabilities of error.

While knowledge of the boundary σ does not seem to yield the properties we obtain for $\{\delta_c'^I\}$, it is interesting to note that σ could be used to define a *new* family of procedures which will possess the desired properties. In fact, let $\bar{\delta}_c$ be the procedure which truncates at $2 |\log c|$ observations, continues observing when $n < 2 |\log c|$ and $n - (2 |\log c| n)^{\frac{1}{2}} < S_n < (2 |\log c| n)^{\frac{1}{2}} - n$, decides $\theta \geqq 1$ (resp. $\theta \leqq -1$) if S_n "goes out at the right" (resp. left) for $n < 2 |\log c|$, and, if $2 |\log c|$ observations are taken, $\bar{\delta}_c$ decides $\theta \geqq 1$ (resp. $\theta \leqq -1$) if $S_{2|\log c|} > 0$ (resp. < 0). By direct computation (which we omit), it can be shown that P_θ {wrong decision using $\bar{\delta}_c$} $= o(c |\log c|)$ uniformly for $|\theta| \geqq 1$, and that $E_\theta N(\bar{\delta}_c)$ has the "right" asymptotic value uniformly in θ, and this yields the result of Theorem 4. Thus, in this normal example, the family $\{\delta_c^I\}$ can be used to obtain an asymptotic shape σ from which we can go "back" and obtain a new family $\{\bar{\delta}_c\}$ (which is not the same as $\{\delta_c^I\}$) with asymptotic shape σ, which has the desired asymptotic Bayes properties, and which is simple to describe. For non-exponential problems we cannot generally expect to find simple asymptotic shapes which can be used to obtain such families $\{\bar{\delta}_c\}$.

Finally, we note that the $\{\delta_c'^I\}$ corresponding to the boundary σ (i.e., the $\{\delta_c'^I\}$ constructed for an F whose support is $(-\infty, \infty)$ is Bayes with respect to any G whose support contains the two points -1 and $+1$. That we can ignore that part of the support of the a priori distribution in $(-\infty, -1)$, or in $(1, \infty)$, is a consequence of the exponential structure of normal densities and is not, generally, possible in non-exponential situations. But, the reason we can "do away" with $(-1, 1)$ is generally true, and is stated in Theorem 4, where it is shown that we can replace the indifference region I by any subset I' of I and $\{\delta_c'^I\}$ remains asymptotically optimum. In particular, therefore, the $\{\delta_c'^I\}$ associated with σ is asymptotically optimum for our original three-state problem and is also asymptotically equivalent to the SPRT for testing between the simple hypotheses $\theta = -1$ and $\theta = +1$. Of course, the procedures which gave rise to the β and δ boundaries are also asymptotically optimum for testing $\theta = -1$ against $\theta = +1$ as well as for the original three-state problem; however, these procedures will not be asymptotically optimum in other problems for which the $\{\delta_c'^I\}$ corresponding to σ is asymptotically optimum.

Minimizing the maximum of EN subject to bounds on probabilities of error.

Kiefer and Weiss (1957) considered problems which, for the sake of brevity, we describe only in the special normal three-state setting just discussed. One problem is to minimize the maximum expected sample size, $\max(E_\omega N, E_\theta N, E_\alpha N)$, subject to specified upper bounds on probabilities of error when ω or θ is true. This is easily shown to be equivalent to minimizing

$$k_1 E_\omega N + k_2 E_\theta N + k_3 E_\alpha N + k_4 P_\omega \{\text{error}\} + k_5 P_\theta \{\text{error}\}$$

for some positive k_1, k_2, k_3, k_4, k_5. As the specified bounds on the probability of error go to zero, this problem can be seen to be equivalent to the asymptotic problem of Section 3. Thus, if we seek a family which asymptotically minimizes the maximum of EN subject to $P_\omega\{\text{error}\} \leqq q(c)$ and $P_\theta\{\text{error}\} \leqq q(c)$ where $q(c) = o(c\,|\log c|)$ and $q(c) = c/O(1)$ (for example, $q(c) = Kc$ where K is a positive constant), procedures corresponding to boundaries like the β and σ of Schwarz and the δ of Anderson (but not ρ or τ) are asymptotically optimum, as are many others. In fact, the fixed sample size procedure corresponding to the line $t = 2$ is asymptotically optimum for the present problem, although not for the previous one. (The normalization of q is for convenience.)

A problem auxiliary to the above, and of less intrinsic importance, is to minimize $E_\alpha N$ alone, subject to restrictions on $P_\theta\{\text{error}\}$ and $P_\omega\{\text{error}\}$. This problem really arises only because it is often true that its solution yields a solution to the more meaningful problem of the previous paragraph; nevertheless, we discuss it briefly here because of theoretical interest. This problem is equivalent to one of minimizing $k_3 E_\alpha N + k_4 P_\omega\{\text{error}\} + k_5 P_\theta\{\text{error}\}$, and is thus slightly different from that of the previous paragraph, so that the theorems of Section 3 do not lead immediately to an asymptotic solution. However from Lemma 5 of Section 3, we do obtain an asymptotic solution which is, in fact, the same one as for the first problem. Thus, in the normal example considered above, the procedures corresponding to β, σ, and δ, as well as many others, are asymptotically optimum in the sense of the present problem. It is interesting to note that, as a consequence of Lemma 5, any boundary which always guarantees stopping earlier than $t = 2$ cannot guarantee error probabilities of order $O(c)$.

L. Weiss has recently developed algorithms for the exact computation of procedures which solve this problem in certain cases (JASA, 1962). Actual computations by Weiss and D. Freeman indicate that the minimum $E_\alpha N$ must be very large before the asymptotic theory estimates it accurately.

Similar remarks apply to minimizing the maximum of EN over a larger indifference zone, etc. For example, in the normal example just discussed, the boundaries β, σ, and δ yield procedures which approximately minimize the maximum of EN over *all* normal distributions with unit variance, subject to the stated bounds on probabilities of error if the mean is $\leqq -1$ or $\geqq 1$. The corresponding k-decision problem can be treated in exactly the same manner.

2. Hypothesis testing without indifference region. In this section we prove the principal results for the two decision problem without indifference region; the latter modification is introduced in Section 3. The methods used in subsequent sections are very similar to those used here. Let Ω and Θ be two disjoint subsets of some Euclidean space (the restriction to Euclidean space is for convenience). Let $\{f_\omega ; \omega \varepsilon \Omega\}$, and $\{g_\theta ; \theta \varepsilon \Theta\}$ be two sets of densities, all absolutely continuous with respect to the same measure μ. The problem is to decide whether the true density is from Ω (Decision 1) or from Θ (Decision 2). Let $L_2(\omega)$ be the loss if $\omega \varepsilon \Omega$ is "true" and Decision 2 is made and let $L_1(\theta)$ be the loss if $\theta \varepsilon \Theta$ is "true" and Decision 1 is made. If the "correct" decision is made no loss is incurred. Let F be an a priori distribution on $\Omega \cup \Theta$ and denote that part of F which is concentrated on Ω by ξ and the part of F concentrated on Θ by η. We shall assume that $\Omega \cup \Theta$ is closed and that the support of F is $\Omega \cup \Theta$. This is for convenience in stating the results and the assumptions. To handle matters more generally all assumptions and results should be stated for the support of F (including, for example, the definitions in Assumption 2).

Independent and identically distributed observations may be taken sequentially with the cost of each observation being c with $0 < c < 1$. Let δ_c^* denote a Bayes solution with respect to F (the dependence on F of the various decision functions we define will be suppressed). Let δ_c be the decision function which, after n observations, stops and makes Decision 1 if

$$(2.1) \qquad \int f_\omega(x_1, \cdots, x_n)L_2(\omega)\xi(d\omega) > \frac{1}{c}\int g_\theta(x_1, \cdots, x_n)L_1(\theta)\eta(d\theta),$$

stops and makes Decision 2 if

$$(2.2) \qquad \int f_\omega(x_1, \cdots, x_n)L_2(\omega)\xi(d\omega) < c\int g_\theta(x_1, \cdots, x_n)L_1(\theta)\eta(d\theta),$$

and takes an $(n + 1)$th observation if neither (2.1) nor (2.2) holds. (An asymptotically equivalent procedure, which is in the form of the k-decision procedure of Section 4, is to stop when the a posteriori risk is $< c$. This form also exhibits an invariance under changes in monetary scale which is not present in (2.1) and (2.2), although its absence in these last is of course irrelevant in asymptotic considerations.) The first result we wish to establish is that, under certain restrictions, $\lim_{c\to 0} r_c(F, \delta_c)/r_c(F, \delta_c^*) = 1$, i.e., that $\{\delta_c\}$ is "asymptotically Bayes" as the cost of observation goes to 0; r_c, of course, denotes the risk when c is the cost of observation.

Assumption 1. Put $\alpha_1 = \inf_\theta L_1(\theta)$, $\alpha_2 = \inf_\omega L_2(\omega)$, $\beta_1 = \sup_\theta L_1(\theta)$, $\beta_2 = \sup_\omega L_2(\omega)$. We assume that α_1 and α_2 are strictly positive and that β_1 and β_2 are finite.

Assumption 2. Put $\lambda_1(\omega, \theta) = E_\omega[\log f_\omega(X) - \log g_\theta(X)]$ and put $\lambda_2(\omega, \theta) = E_\theta[\log g_\theta(X) - \log f_\omega(X)]$. The indicated expectations are assumed to exist. Furthermore, we assume

(a) $\inf_\omega \inf_\theta \lambda_1(\omega, \theta) = \lambda_1 > 0$
(b) $\inf_\omega \inf_\theta \lambda_2(\omega, \theta) = \lambda_2 > 0$
(c) $\lambda_1(\omega, \theta)$ and $\lambda_1(\omega) = \inf_\theta \lambda_1(\omega, \theta)$ are both continuous in ω.

(d) $\lambda_2(\omega, \theta)$ and $\lambda_2(\theta) = \inf_\omega \lambda_2(\omega, \theta)$ are both continuous in θ.

REMARK. Assumption 2 guarantees that Ω and Θ are "separated."

ASSUMPTION 3. For each ω, θ

(a) $E_\omega[\log f_\omega(X) - \log g_\theta(X)]^2 < \infty$
$E_\theta[\log g_\theta(X) - \log f_\omega(X)]^2 < \infty$

(b) $\lim_{\rho\downarrow 0} E_\omega[\log \sup_{|\theta'-\theta|\leqq\rho} g_{\theta'}(X) - \log g_\theta(X)]^2 = 0$
$\lim_{\rho\downarrow 0} E_\theta[\log \sup_{|\omega'-\omega|\leqq\rho} f_{\omega'}(X) - \log f_\omega(X)]^2 = 0$

(c) $\lim_{\omega'\to\omega} E_\omega[\log f_{\omega'}(X) - \log f_\omega(X)]^2 = 0$
$\lim_{\theta'\to\theta} E_\theta[\log g_{\theta'}(X) - \log g_\theta(X)]^2 = 0.$

REMARK. (c) is slightly stronger than actually required. If (c) holds with the second moment replaced by the first moment and if the second moment is finite in some neighborhood of ω(or θ) then the argument below will proceed without change.

The first thing we do is to estimate the loss due to incorrect decision when δ_c is used. Let $P_\omega(c)$ be the probability of making Decision 2 when ω is the true state of nature and when δ_c is the decision function used. Let $Q_\theta(c)$ be the probability of making Decision 1 when θ is the true state of nature and δ_c is used.

LEMMA 1. $\int L_2(\omega)P_\omega(c)\xi(d\omega) + \int L_1(\theta)Q_\theta(c)\eta(d\theta) \leqq (\beta_1 + \beta_2)c$.

PROOF. Let A_j be the set of $(x_1, \cdots, x_j)$ where δ_c says stop after j observations and make Decision 2. Then

$$P_\omega(c) = \sum_{j=1}^{\infty} \int_{A_j} f_\omega(x_1, \cdots, x_j)\, d\mu^j.$$

Using (2.2) and Assumption 1 we have

$$\int_\Omega L_2(\omega)P_\omega(c)\xi\,(d\omega) \leqq c \sum_{j=1}^{\infty} \int_{A_j} \int_\Theta g_\theta(x_1, \cdots, x_j)L_1(\theta)\eta\,(d\theta)\, d\mu^j$$
$$= c \int (1 - Q_\theta(c))L_1(\theta)\eta\,(d\theta) \leqq c\beta_1.$$

The rest of the argument is obvious.

Lemma 2 which we now prove is the heart of the matter before us.

LEMMA 2. *Put* $\lambda_1(\omega) = \inf_\theta \lambda_1(\omega, \theta)$ (*see Assumption* 2). *Let* Θ *be a compact (therefore bounded) set in some Euclidean space. Let* $N(c)$ *be the number of observations required by* δ_c *to terminate. For each* ω_0,

$$E_{\omega_0}N(c) \leqq [1 + o(1)]\,[|\log c|/\lambda_1(\omega_0)] \tag{2.3}$$

as $c \to 0$ (*the* $o(1)$ *term may depend on* ω_0; *this dependence is removed in Lemma* 3$'$).

PROOF. We will show (2.3) by proving it to be so for some random variable $N^*(c)$ (see (2.21) et seq) which is larger than $N(c)$. To obtain an $N^*(c)$ with which we can work we will suitably discretize Θ and alter (2.1). The first step is to obtain (2.8) below.

Fix ω_0 and put, for $\theta \varepsilon \Theta$,

$$M(\omega_0, \theta, \rho) = E_{\omega_0}[\log f_{\omega_0}(X) - \log \sup_{|\theta'-\theta|<\rho} g_{\theta'}(X)].$$

From Assumption 3, we obtain

(2.4) $\qquad M(\omega_0, \theta, \rho)$ increases to $\lambda_1(\omega_0, \theta)$ as ρ decreases to 0.

Let $\{\epsilon_c ; c > 0\}$ be a set of positive numbers with $\epsilon_c \to 0$ as $c \to 0$. From (2.4) and Assumption 3, there is, for each c and each θ, a $\rho_c(\theta)$ such that

$$M(\omega_0, \theta, \rho_c(\theta)) \geqq \lambda_1(\omega_0, \theta) - \epsilon_c, \tag{2.5}$$
$$E_{\omega_0}[\log f_{\omega_0}(X) - \log \sup_{|\theta'-\theta| \leqq \rho_c(\theta)} g_{\theta'}(X)]^2 < \infty.$$

Let $\mathrm{U}(\theta, \rho_c(\theta)) = \{\theta' \varepsilon \Theta \mid |\theta' - \theta| < \rho_c(\theta)\}$. Then $\{\mathrm{U}(\theta, \rho_c(\theta)), \theta \varepsilon \Theta\}$ is a family of open sets which covers Θ and, since Θ is compact, we can extract a finite subcovering. Let $\{\mathrm{U}(\theta_i, \rho_c(\theta_i)), i = 1, \cdots, T_c\}$ be such a finite subcovering of Θ and abbreviate $\mathrm{U}(\theta_i, \rho_c(\theta_i))$ by U_i. The θ_i will also depend on c but we suppress this dependence. By (2.5), for each $i = 1, \cdots, T_c$,

$$E_{\omega_0}[\log f_{\omega_0}(X) - \log \sup_{\theta \varepsilon \mathrm{U}_i} g_\theta(X)] \geqq \lambda_1(\omega_0, \theta_i) - \epsilon_c. \tag{2.6}$$

From (c) of Assumption 3, we know that there is a γ_c such that, when $|\omega - \omega_0| < \gamma_c$,

$$E_{\omega_0}[\log f_\omega(X) - \log f_{\omega_0}(X)] > -\epsilon_c. \tag{2.7}$$

Let $V_c = \{\omega \mid |\omega - \omega_0| < \gamma_c\}$. Then (2.6) and (2.7) yield

$$E_{\omega_0}[\log f_\omega(X) - \log \sup_{\mathrm{U}_i} g_\theta(X)] \geqq \lambda_1(\omega_0, \theta_i) - 2\epsilon_c \tag{2.8}$$

for all $\omega \varepsilon V_c$ and $i = 1, \cdots, T_c$. Since ω_0 is in the support of ξ (as is every point in Ω)

$$\xi(V_c) > 0 \quad \text{for all} \quad c > 0. \tag{2.9}$$

We are now ready to work towards the definition of $N^*(c)$. First note that if $N_1(c)$ is the first n such that (2.1) holds then $N(c) \leqq N_1(c)$. Let $N_2(c)$ be the first n such that

$$\int_{V_c} f_\omega(x_1, \cdots, x_n) L_2(\omega) \xi(d\omega) > \sup_\theta g_\theta(x_1, \cdots, x_n) \frac{\beta_1}{c}. \tag{2.10}$$

Clearly $N_2(c) \geqq N_1(c)$. Dividing both sides of (2.10) by $\xi(V_c)$ (which is positive by (2.9)) and taking logarithms we obtain an equivalent of (2.10), viz.,

$$\log \int_{V_c} f_\omega(x_1, \cdots, x_n) L_2(\omega) \frac{\xi(d\omega)}{\xi(V_c)} > |\log c| + \log \beta_1 \tag{2.11}$$
$$+ |\log \xi(V_c)| + \log \sup_\theta g_\theta(x_1, \cdots, x_n).$$

By the concavity of log and Jensen's Inequality

$$\begin{aligned} \log \int_{V_c} f_\omega(x_1, \cdots, x_n) L_2(\omega) \frac{\xi(d\omega)}{\xi(V_c)} &\geqq \int_{V_c} \log f_\omega(x_1, \cdots, x_n) \frac{\xi(d\omega)}{\xi(V_c)} \\ &+ \int_{V_c} \log L_2(\omega) \frac{\xi(d\omega)}{\xi(V_c)} = \sum_{j=1}^n \int_{V_c} \log f_\omega(x_j) \frac{\xi(d\omega)}{\xi(V_c)} \\ &+ \int_{V_c} \log L_2(\omega) \frac{\xi(d\omega)}{\xi(V_c)}. \end{aligned} \tag{2.12}$$

Recalling that $\{U_i\}$ covers θ,

$$\log \sup_\theta g_\theta(x_1, \cdots, x_n) = \log \sup_{i \leqq T_c} \sup_{U_i} g_\theta(x_1, \cdots, x_n)$$

$$(2.13) \qquad = \sup_i \sup_{U_i} \sum_{j=1}^n \log g_\theta(x_j) \leqq \sup_i \sum_{j=1}^n \sup_{U_i} \log g_\theta(x_j)$$

$$= \sup_i \sum_{j=1}^n \log \sup_{U_i} g_\theta(x_j).$$

Put

$$(2.14) \quad A_c = |\log \xi(V_c)| + \log \beta_1 + |\log c| - \int_{V_c} \log L_2(\omega) \frac{\xi(d\omega)}{\xi(V_c)}$$

$$(2.15) \quad Y_j = \int_{V_c} \log f_\omega(X_j) \frac{\xi(d\omega)}{\xi(V_c)} \qquad \text{for } j = 1, 2, \cdots$$

$$(2.16) \quad Z_j^i = \log \sup_{U_i} g_\theta(X_j) \qquad \text{for } i = 1, \cdots, T_c; j = 1, 2, \cdots.$$

Let $N_3(c)$ be the first n such that

$$(2.17) \qquad \sum_{j=1}^n Y_j - \sup_i \sum_{j=1}^n Z_j^i > A_c$$

or, equivalently, the first n such that

$$(2.18) \quad \sum_{j=1}^n [Y_j - \epsilon_c - Z_j^1] + \min_{1 \leqq i \leqq T_c} \sum_{j=1}^n [Z_j^1 + \epsilon_c - Z_j^i] > A_c.$$

(2.12)–(2.16) imply that $N_3(c) \geqq N_2(c)$. From (2.8) we have

$$(2.19) \qquad E_{\omega_0}[Y_j - \epsilon_c - Z_j^i] \geqq \lambda_1(\omega_0, \theta_i) - 3\epsilon_c \text{ for } i = 1, \cdots, T_c.$$

Suppose that $\{U_i\}$ are indexed so that the minimum (over i) of the left hand side of (2.19) occurs when $i = 1$. Then

$$(2.20) \quad \begin{aligned} &E_{\omega_0}[Y_j - \epsilon_c - Z_j^i] \geqq \lambda_1(\omega_0) - 3\epsilon_c; \\ &E_{\omega_0}[Z_j^1 + \epsilon_c - Z_j^i] \geqq \epsilon_c \quad \text{for } i = 1, \cdots, T_c. \end{aligned}$$

Let S_n denote the nth partial sum of the sequence $\{Y_j - \epsilon_c - Z_j^1\}$ and let B_n^i $(i = 1, \cdots, T_c)$ denote the nth partial sum of the sequence $\{Z_j^1 + \epsilon_c - Z_j^i\}$. Since $\{X_j\}$ is a sequence of independent and identically distributed random variables the same is true for each of the sequences in the last sentence. Put $B_n = \min_{i \leqq T_c} B_n^i$. Then (2.18) is equivalent to $S_n + B_n > A_c$. Let $N^*(c)$ be the first n such that, simultaneously,

$$(2.21) \qquad S_n > A_c \quad \text{and} \quad B_n \geqq 0.$$

It is obvious that $N^*(c) \geqq N_3(c)$. Our problem now is to show that the lemma holds for $N^*(c)$.

We hereafter assume, as we may, that $\lambda_1(\omega_0) - 3\epsilon_c > 0$. Let ν_1 be the first n such that $S_n > A_c$, ν_2 the second n such that $S_n > A_c$, etc. Let ϕ_t be the indicator function of the set where $B_{\nu_t} < 0$, $t = 1, 2, \cdots$. Then

$$N^*(c) = \nu_1 + \sum_{j=1}^{\infty} (\nu_{j+1} - \nu_j) \prod_{t=1}^{j} \phi_t .$$

Let $\nu^*_{j+1} - \nu_j$ be the first $m > 0$ such that $S_{m+\nu_j} - S_{\nu_j} > 0$. Since $S_{\nu_j} > A_c$, it follows that $S_{\nu^*_{j+1}} > A_c$ and, therefore, $\nu^*_{j+1} - \nu_j \geqq \nu_{j+1} - \nu_j$. Since $\nu^*_{j+1} - \nu_j$ depends on X's whose indices are greater than ν_j, it follows that $\nu^*_{j+1} - \nu_j$ is independent of $\phi_1, \cdots, \phi_j$. Consequently

$$E_{\omega_0} N^*(c) \leqq E_{\omega_0}\nu_1 + \sum_{j=1}^{\infty} E_{\omega_0}(\nu^*_{j+1} - \nu_j) E_{\omega_0} \prod_{t=1}^{j} \phi_t . \tag{2.22}$$

$\nu^*_{j+1} - \nu_j$ has the same distribution as the first n for which $S_n > 0$ so that, using (4.6) of Spitzer (1956), the Chebyshev inequality, (2.20), Assumption 3, and (2.5), we have

$$\begin{aligned} E_{\omega_0}(\nu^*_{j+1} - \nu_j) &= 1 + \sum_{k=1}^{\infty} P\{\max_{1 \leqq j \leqq k} S_j \leqq 0\} \\ &= \exp\left[\sum_{k=1}^{\infty} (1/k) P\{S_k \leqq 0\}\right] \\ &\leqq \exp\left\{[\mathrm{Var}_{\omega_0}(Y_1 - \epsilon_c - Z_1^1)/(\lambda_1(\omega_0) - 3\epsilon_c)^2] \sum_{k=1}^{\infty} (1/k^2)\right\} \\ &= D(c, \omega_0) \quad \text{(say)}. \end{aligned} \tag{2.23}$$

For the purpose of estimating $E_{\omega_0} \prod_{t=1}^{j} \phi_t$ we let $\sigma_i (i = 1, \cdots, T_c)$ be the *last* time $B_n^i < 0$ and then $\sigma = \max(\sigma_1, \cdots, \sigma_{T_c})$ is the last time $B_n < 0$. Observe that ν_j is at least as big as j so that

$$\begin{aligned} \sum_{j=1}^{\infty} E_{\omega_0} \prod_{t=1}^{j} \phi_t &= \sum_{j=1}^{\infty} P_{\omega_0}\{B_{\nu_1} < 0, \cdots, B_{\nu_j} < 0\} \leqq \sum_{j=1}^{\infty} P_{\omega_0}\{\sigma \geqq \nu_j\} \\ &\leqq \sum_{j=1}^{\infty} P_{\omega_0}\{\sigma \geqq j\} = E_{\omega_0}\sigma \leqq \sum_{i=1}^{T_c} E_{\omega_0}\sigma_i . \end{aligned} \tag{2.24}$$

We now apply Theorem D in the Appendix to $\{-B_k^i\}$ with $\mu_0 = \epsilon_c$ since, by (2.20), the summands in B_k^i have mean $\geqq \epsilon_c$, and, by (2.5), they have finite variance, so that

$$E_{\omega_0}\sigma_i \leqq -(2/\epsilon_c) E_{\omega_0}[\min_{k \geqq 0} B_k^i] + \sum_{k=1}^{\infty} P\{B_k^i - k\epsilon_c/2 < 0\}. \tag{2.25}$$

Since $B_k^i - k\epsilon_c/2$ is the kth partial sum of independent and identically distributed random variables whose mean is $\geqq \epsilon_c/2$ and whose variance is finite, we obtain from Erdös (1949) the fact that the last term in (2.25) is finite. The first term to the right of the inequality in (2.25) is finite because the summands in B_k^i have positive ($\geqq \epsilon_c$) mean and finite variance (this is well-known and can be seen in Theorem B of the Appendix). (2.24) and (2.25) yield

$$\sum_{j=1}^{\infty} E_{\omega_0} \prod_{t=1}^{j} \phi_t = h(c, \omega_0) \text{ (say)} < \infty. \tag{2.26}$$

The final estimate we need is the well-known fact from renewal theory that

$$E_{\omega_0} \nu_1 = \frac{[1 + o(1)]A_c}{E_{\omega_0}[Y_1 - \epsilon_c - Z_1^1]} \leqq \frac{[1 + o(1)]A_c}{\lambda_1(\omega_0) - 3\epsilon_c}. \tag{2.27}$$

Putting (2.23), (2.26), (2.27) into (2.22) we get

$$E_{\omega_0} N^*(c) \leqq [1 + o(1)]A_c/[\lambda_1(\omega_0) - 3\epsilon_c] + D(c, \omega_0)h(c, \omega_0).$$

The final step is to observe that ϵ_c may be chosen to go to 0 so slowly that the choices of V_c, $\{U_i\}$, etc. (which depend on c through ϵ_c) result in

$$|\log \xi(V_c)| = o(|\log c|), \qquad D(c, \omega_0)h(c, \omega_0) = o(|\log c|).$$

For, if this is the case,

$$A_c \leqq [1 + o(1)]\, |\log c| + \log \beta_1 + |\log \alpha_2| = [1 + o(1)]\, |\log c|$$

and then $E_{\omega_0} N^*(c) \leqq [1 + o(1)]\, |\log c|/\lambda_1(\omega_0)$, which completes the proof.

If we put $\lambda_2(\theta) = \inf_\omega \lambda_2(\omega, \theta)$ we can state

LEMMA 2′. *If Ω is compact, then for each θ_0*

$$E_{\theta_0} N(c) \leqq \frac{[1 + o(1)]\, |\log c|}{\lambda_2(\theta_0)} \tag{2.28}$$

as $c \to 0$ (the $o(1)$ term may depend on θ_0; see, however, Lemma 3′).

The proof of Lemma 2′ is, of course, the same as for Lemma 2.

REMARKS.

1. We may extend Lemmas 2 and 2′ to cases where Ω and Θ are not bounded subsets of a Euclidean space by using the following compactification assumption (in addition to assuming that our previously stated assumptions hold for compact subsets):

ASSUMPTION 4. There is for each ω a number r_ω such that

(a) $E_\omega[\log f_\omega(X) - \log \sup_{|\theta| \geq r_\omega} g_\theta(X)]^2 < \infty$ and

(b) $E_\omega[\log f_\omega(X) - \log \sup_{|\theta| \geq r_\omega} g_\theta(X)] \geqq \lambda_1(\omega)$.

The argument of Lemma 2 will go through quite easily by observing that we can take the $\{U_i\}$ there as a covering of $\Theta \cap \{\theta \mid |\theta| \leqq r_{\omega_0}\}$ and then we can take the remaining part of Θ as a neighborhood of infinity and add that to $\{U_i\}$ to give a covering of Θ. There is no problem in carrying out the remaining part of the argument. An analogous assumption will take care of Lemma 2′. These assumptions are satisfied in many common problems, for example, in ones of the univariate or multivariate exponential (*K-D*) or translation parameter type.

2. When, for any ω, we have

$$E_\omega[\log f_\omega(X) - \log \sup_{|\theta| \geq r} g_\theta(X)] < \lambda_1(\omega) \tag{*}$$

for all r then the argument of Lemma 2 breaks down unless we have

ASSUMPTION 4′. If ω is such that (*) holds then there is a sequence $\{r_n\}$ (which may depend on ω) such that $r_n \to +\infty$ and

(a) $E_\omega[\log f_\omega(X) - \log \sup_{|\theta| \geqq r_n} g_\theta(X)]^2 < \infty$ for each n (boundedness in n is *not* assumed);

(b) The limit as $r \to +\infty$ of the left side of (*) is equal to $\lambda_1(\omega)$. (The limit exists because the l.h.s. of (*) is increasing in r).

To treat Lemma 2 in the presence of Assumption 4′ observe that, for fixed ω_0, we can take r_n sufficiently large *depending on c* so that $\Theta \cap \{\theta \mid |\theta| \geqq r_n\}$ can be added to $\{U_i\}$ to form a covering of Θ. Turning to (2.20) with the possibility now that $Z_j^1 = \log \sup_{|\theta| \geqq r_n} g_\theta(X_j)$ we see that all is well provided r_n is also large enough so that the l.h.s. of (*) with r replaced by r_n is larger than $\lambda_1(\omega_0) - \epsilon_c$.

For generalizing Lemma 2 to unbounded Ω and Θ we make the obvious remark that it is only necessary to have each $\omega \varepsilon \Omega$ satisfy either (a) or (b) of Assumption 4 or (a) or (b) of Assumption 4′.

3. The situations not covered by Remarks 1 and 2 include, of course, those where the second moment in (a) of each of the two assumptions fails to exist in the right way. In most regular situations the second moment will either exist for any choice of r or never exist. Here and in our other assumptions, finiteness of second moments can be weakened; recent work by R. H. Farrell replaces t^2 by convex $h(t)$ with $\lim_{t\to\infty} h(t)/|t| = \infty$ for some of these considerations. The possible pathology in the behavior of the first moment is that, for some ω,

$$\inf_\theta \lambda_1(\omega, \theta) = \lambda_1(\omega) > \lim_{r\to\infty} E_\omega[\log f_\omega(X) - \log \sup_{|\theta| \geqq r} g_\theta(X)]$$

indicating a lack of continuity at a point of infinity in Θ.

ASSUMPTION 5.

(a) $E_\omega[\log f_\omega(X) - \log g_\theta(X)]^2$ and $E_\omega[\log \sup_{|\theta'-\theta| \leqq \rho} g_{\theta'}(X) - \log g_\theta(X)]^2$ are continuous in ω; $E_\omega[\log f_{\omega'}(X) - \log f_\omega(X)]^2$ is jointly continuous in ω and ω'.

(b) Interchange ω and θ, ω' and θ', f_ω and g_θ in (a) and the assumption is continuity in θ and joint continuity in θ, θ'.

LEMMA 3. *If Ω and Θ are compact and Assumptions* 1, 2, 3 *and* 5 *are satisfied, then there is a constant M such that*

$$E_\omega N(c) \leqq M\,|\log c|, \qquad E_\theta N(c) \leqq M\,|\log c|$$

for all ω and θ.

PROOF. We will only consider $E_\omega N(c)$. Observe by the first part of the proof of Lemma 2 that, for each $\omega \varepsilon \Omega$, there is a family $U_1, \cdots, U_{T(\omega)}$ of open sets covering Θ such that for $i = 1, \cdots, T(\omega)$

$$E_\omega[\log f_\omega(X) - \log \sup_{U_i} g_\theta(X)] \geqq \lambda_1/2$$

$$E_\omega[\log f_\omega(X) - \log \sup_{U_i} g_\theta(X)]^2 < G_i(\omega) \text{ (say) } < \infty.$$

By making use of Assumptions 2, 3, and 5 we can obtain an open set $V(\omega)$ containing ω such that for any $\omega' \varepsilon V(\omega)$ and $\omega'' \varepsilon V(\omega)$

$$E_{\omega'}[\log f_{\omega''}(X) - \log \sup_{U_i} g_\theta(X)] \geqq \lambda_1/4, \qquad i = 1, \cdots, T(\omega) \tag{2.29}$$

$$E_{\omega'}[\log f_{\omega''}(X) - \log \sup_{U_i} g_\theta(X)]^2 \leqq G_1'(\omega) < \infty, \qquad i = 1, \cdots, T(\omega) \tag{2.30}$$

$\{V(\omega), \omega \varepsilon \Omega\}$ covers Ω so we can extract a finite covering $V(\omega_1), \cdots, V(\omega_k)$. We will show that for each j there is a constant M_j such that

$$\sup_{\omega\varepsilon V(\omega_j)} E_\omega N(c) \leqq M_j |\log c| \tag{2.31}$$

and then take $M = M_1 + \cdots + M_k$.

We will establish (2.31) for $j = 1$. Fix $\omega_0 \varepsilon V(\omega_1)$. We can now emulate the proof of Lemma 2 starting at (2.14), replacing V_c by $V(\omega_1)$ and ϵ_c by $\lambda_1/8$, and using (2.29) and (2.30) to conclude that

$$E_{\omega_0} N(c) \leqq E_{\omega_0}\nu_1 + E_{\omega_0}\nu^* \sum_{i=1}^{T(\omega_1)} E_{\omega_0}\sigma_i \tag{2.32}$$

where ν_1 is the first time $S_n > A_c = |\log c| + \gamma$ (γ is some constant), S_n is the nth partial sum of independent and identically distributed random variables whose mean is $\geqq \lambda_1/8$ and whose variance is $\leqq \sum_{i=1}^{T(\omega_1)} G_i'(\omega_1) = G'$ (say) $< \infty$, ν^* has the same distribution as the first time $S_n > 0$, and σ_i is the last time $B_n^i < 0$ where B_n^i is the nth partial sum of independent and identically distributed random variables with mean larger than $\lambda_1/8$ and variance $\leqq 2G'$. From (2.23) it is immediate that

$$E_{\omega_0}\nu^* \leqq \exp\left\{G'(8/\lambda_1)^2 \sum_{k=1}^{\infty} (1/k^2)\right\} = \rho_1 \quad \text{(say)} \tag{2.33}$$

for *all* $\omega_0 \varepsilon V(\omega_1)$. From (2.25) we obtain

$$E_{\omega_0}\sigma_i \leqq (-16/\lambda_1) E_{\omega_0}[\min_{n\geqq 0} B_n^i] + \sum_{k=1}^{\infty} P_{\omega_0}\{B_k^i - k\lambda_1/16 < 0\}.$$

The mean of the summands in B_k^i is always greater than $\lambda_1/8$ and the variance is bounded as long as $\omega_0 \varepsilon V(\omega_1)$ so we can use Theorems A and B in the Appendix to conclude that

$$E_{\omega_0}\sigma_i \leqq \rho_2 \quad \text{(say)}, \qquad i = 1, \cdots, T(\omega_1), \text{ all } \omega_0 \varepsilon V(\omega_1). \tag{2.34}$$

Similarly, by Theorem C in the Appendix, we obtain

$$E_{\omega_0}\nu_1 \leqq \rho_3 |\log c|, \qquad \text{all } \omega_0 \varepsilon V(\omega_1). \tag{2.35}$$

(2.33), (2.34), (2.35) used in (2.32) give (2.31) for $j = 1$ and we are finished.

REMARK 4. A refinement of the proof of Lemma 3 using the full strength of Theorem C in the Appendix can be used to show that the $o(1)$ term in Lemma 2 (and Lemma 2′) is uniform in $\omega_0(\theta_0)$. This refinement is necessary in the proof of Theorem 2 (to obtain (2.57)), and we state it as

LEMMA 3′. *Inequalities* (2.3) *and* (2.28) *hold with the* $o(1)$ *term independent of* ω_0 *and* θ_0.

REMARK 5. It will sometimes be possible to establish the conclusion of Lemmas 3 and 3′ even when Ω and Θ are not compact. If it can be shown that

$\sup_\Omega E_\omega N(c) = \sup_{\Omega_0} E_\omega N(c)$ where Ω_0 is a compact subset of Ω and similarly for Θ then Lemma 3 is clearly true. When Ω, Θ are disjoint intervals on the real line and f_ω, g_θ are exponential densities with respect to the same measure then such a result can be established. This is the context of the work of G. Schwarz (1962). Other examples are easy to give.

The next lemma is essentially due to Schwarz who proved it in the case of exponential densities. It will be seen that Lemma 4 is not used in the proof of Theorem 1 although a proof of Theorem 1 can be constructed using Lemma 4. We state it here because of its intrinsic interest and because of the intuitive understanding which it lends to the whole subject.

Let $S(c)$ be the set of points in the sample space where δ_c says to stop. Let $B(c)$ be the set of points where a Bayes solution with respect to F (and with c the cost of observation) says to stop.

LEMMA 4. *If Ω and Θ are compact and Assumptions* 1, 2, 3 *and* 5 *are satisfied, then there are positive constants σ and τ such that $S(\sigma c) \subset B(c) \subset S(\tau c\,|\log c|)$ for all $c \leqq c_0(\tau)$ ($c_0(\tau)$ is some positive number).*

REMARK 6. What is needed here is the validity of the conclusions of Lemmas 2 and 2′ and the conclusion of Lemma 3.

PROOF. The first inclusion will follow from the fact that no Bayes solution would continue to take observations if the a posteriori loss is smaller than the cost of taking another observation. Indeed, if $x_1, \cdots, x_n$ have been observed, and we have

$$\int L_2(\omega) f_\omega(x_1, \cdots, x_n)\xi\,(d\omega) < \sigma c \int L_1(\theta) g_\theta(x_1, \cdots, x_n)\eta\,(d\theta)$$

then the a posteriori risk is smaller than

$$\frac{\sigma c \int L_1(\theta) g_\theta(x_1, \cdots, x_n)\eta\,(d\theta)}{\int g_\theta(x_1, \cdots, x_n)\eta\,(d\theta) + \int f_\omega(x_1, \cdots, x_n)\xi\,(d\theta)} \leqq \sigma c \beta_1$$

with a similar relation if δ_c stops "at the other hypothesis." How to choose σ is apparent.

The second inclusion requires further argument. If $\delta_{\tau c|\log c|}$ says continue after observing $x_1, \cdots, x_n$ we have (abbreviating $f_\omega(x_1, \cdots, x_n)$ by f_ω^n and similarly with g_θ^n)

$$\tau c\,|\log c| \int L_1(\theta) g_\theta^n \eta\,(d\theta) < \int L_2(\omega) f_\omega^n\,\xi\,(d\omega) < \frac{1}{\tau c\,|\log c|} \int L_1(\theta) g_\theta^n \eta\,(d\theta).$$

Thus, the a posteriori loss when c is the cost, after observing $x_1, \cdots, x_n$, is greater than

$$\frac{\int L_2(\omega) f_\omega^n\,\xi\,(d\omega)}{\int f_\omega^n\,\xi\,(d\omega) + \int g_\theta^n \eta\,(d\theta)} \min\,[1, \tau c\,|\log c|],$$

and is also greater than

$$\frac{\int L_1(\theta) g_\theta^n \eta\,(d\theta)}{\int f_\omega^n \xi\,(d\omega) + \int g_\theta^n \eta\,(d\theta)} \min\,[1, \tau c\,|\log c|].$$

Consequently, the a posteriori risk is larger than

$$\tfrac{1}{2} \min\,[\alpha_1, \alpha_2] \min\,[1, \tau c\,|\log c|] \geqq \tau' c\,|\log c| \tag{2.36}$$

for c small enough (say $c \leqq c'$).

We will now show, if no τ yields the stated result, how to obtain a procedure which is an improvement of the Bayes procedure, and this gives a contradiction. Suppose B' is the set of sample points where the Bayes procedure stops and $\delta_{\tau c|\log c|}$ says continue. If $P_\omega(B')(P_\theta(B'))$ is the measure of B' induced by $f_\omega(g_\theta)$ then we might as well assume that

$$\pi(B') = \int P_\omega(B')\xi\,(d\omega) + \int P_\theta(B')\eta\,(d\theta) > 0.$$

If $x \varepsilon B'$ and n is the time the Bayes procedure says stop let us continue to take observations until the first ν such that

$$\frac{\int L_2(\omega) f_\omega(x_1, \cdots, x_{n+\nu})\xi\,(d\omega)}{\int L_1(\theta) g_\theta(x_1, \cdots, x_{n+\nu})\eta\,(d\theta)} > \frac{1}{c} \quad \text{or} \quad < c. \tag{2.37}$$

If we put $\xi^n(d\omega) = f_\omega^n \xi(d\omega)/(\int f_\omega^n \xi(d\omega) + \int g_\theta^n \eta(d\theta))$ and $\eta^n(d\theta) = g_\theta^n \eta(d\theta)/(\int f_\omega^n \xi(d\omega) + \int g_\theta^n \eta(d\theta))$ then to calculate the "conditional" properties (given $(x_1, \cdots, x_n)$) of (2.37) we need only turn to Lemmas 1 and 2 and substitute ξ^n for ξ. Thus the "conditional" loss (given $(x_1, \cdots, x_n)$) is smaller than $(\beta_1 + \beta_2)c$ and the "conditional" expected number of observations is, for each ω,

$$E_\omega \bar{N}(c) \leqq [1 + o(1)]\,|\log c|/\lambda_1(\omega) \tag{2.38}$$

and, for each θ,

$$E_\theta \bar{N}(c) \leqq [1 + o(1)]\,|\log c|/\lambda_2(\theta) \tag{2.39}$$

We are assuming here that ξ^n has Ω as its support and that η^n has Θ as its sup port. This is inessential and (2.38) and (2.39) will hold in any case as can be seen by applying Lemma 2 to the support of ξ^n and the support of η^n and concluding (2.38) with $\lambda_1(\omega)$ replaced by $\inf \{\lambda_1(\omega, \theta) \mid \theta \epsilon \text{ support of } \eta^n\}$ which is larger than $\lambda_1(\omega)$.

Let $\epsilon_1(\omega, x_1, \cdots, x_n, c)$ denote the $o(1)$ term in (2.38) and let $\epsilon_2(\theta, x_1, \cdots, x_n, c)$ denote the $o(1)$ term in (2.39)—the dependence on $x_1, \cdots, x_n$ is, of course, through ξ^n and η^n. From Lemma 3 we know that there is an $M(x_1, \cdots, x_n)$ such that $|\epsilon_1(\omega, x_1, \cdots, x_n, c)| + |\epsilon_2(\theta, x_1, \cdots, x_n, c)| \leqq M(x_1, \cdots, x_n)$

for all ω and all θ. Thus there is a number M_0 and a set $B^* \subset B'$ with $\pi(B^*) > 0$ such that for $x \varepsilon B^*$

$$|\epsilon_1(\omega, x_1, \cdots, x_n, c)| + |\epsilon_2(\theta, x_1, \cdots, x_n, c)| \leqq M_0$$

for all ω and all θ.

Now consider the procedure of (2.37) only for $x \varepsilon B^*$. We conclude from (2.38) and (2.39) that, for $\epsilon > 0$, there is a c_0 (which is the same for all $x \varepsilon B^*$) such that for all $c \leqq c_0$

$$\int E_\omega \bar{N}(c)\xi^n (d\omega) + \int E_\theta \bar{N}(c)\eta^n (d\theta) \leqq [1 + \epsilon] |\log c|/\min (\lambda_1, \lambda_2).$$

Thus if τ is chosen so that $\tau' > (1 + \epsilon)/\min (\lambda_1, \lambda_2)$ (see (2.36)) we can conclude that, for any $x \varepsilon B^*$, the a posteriori risk when the Bayes procedure says stop is larger than for the modified procedure. This contradiction establishes Lemma 4.

THEOREM 1. *If Ω and Θ are compact and Assumptions 1, 2, 3, and 5 are satisfied, then $\{\delta_c\}$ is asymptotically Bayes, i.e.,* $\lim_{c\to 0} r_c(F, \delta_c^*)/r_c(F, \delta_c) = 1$ *where δ_c^* is a Bayes solution when F is the a priori distribution and c the cost per observation.*

REMARK 7. Referring to Remarks 1–6 it is possible to describe more general conditions under which Theorem 1 is true.

PROOF. Let $\gamma_1 = \int_\Omega \xi(d\omega)/\lambda_1(\omega)$, $\gamma_2 = \int_\Theta \eta(d\theta)/\lambda_2(\theta)$. From Lemma 2 we know that, for each $\omega \varepsilon \Omega$,

$$\limsup_{c\to 0} E_\omega N(c)/|\log c| \leqq 1/\lambda_1(\omega)$$

and from Lemma 3 we have $E_\omega N(c) \leqq M |\log c|$ for all $\omega \varepsilon \Omega$. Consequently

$$\limsup_{c\to 0} \frac{1}{|\log c|} \int_\Omega E_\omega N(c)\xi (d\omega) \leqq \int_\Omega \limsup_{c\to 0} \frac{E_\omega N(c)}{|\log c|} \xi (d\omega) \leqq \gamma_1 .$$

Doing the same for Θ and letting $E_F N(c) = \int_\Omega E_\omega N(c)\xi(d\omega) + \int_\Theta E_\theta N(c)\eta(d\theta)$ we obtain

$$\limsup_{c\to 0} E_F N(c)/|\log c| \leqq \gamma_1 + \gamma_2 .$$

This together with Lemma 1 yields

$$\limsup_{c\to 0} r_c(F, \delta_c)/c\, |\log c| \leqq \gamma_1 + \gamma_2 . \tag{2.40}$$

If $\{\delta_c\}$ is *not* asymptotically Bayes we would have $\liminf_{c\to 0} r_c(F, \delta_c^*)/r_c(F, \delta_c) < 1 - 2\epsilon$ for some positive ϵ which implies, due to (2.40), that there is a sequence $\{c_i\}$ with $c_i \to 0$ such that

$$r_{c_i}(F, \delta_{c_i}^*) < (1 - 2\epsilon)c_i\, |\log c_i|\, (\gamma_1 + \gamma_2) \tag{2.41}$$

and, consequently,

$$E_F N^*(c_i) < (1 - 2\epsilon)\, |\log c_i|\, (\gamma_1 + \gamma_2) \tag{2.42}$$

where $N^*(c)$ is the number of observations required when δ_c^* is used. (2.42) implies that either $\int E_\omega N^*(c_i)\xi(d\omega) < (1-2\epsilon)\,|\log c_i|\,\gamma_1$ or $\int E_\theta N^*(c_i)\eta(d\theta) < (1-2\epsilon)\,|\log c_i|\,\gamma_2$. Suppose the former. Then

$$\int \left[\frac{E_\omega N^*(c_i)}{|\log c_i|} - \frac{1-\epsilon}{\lambda_1(\omega)}\right]\xi(d\omega) \leqq -\epsilon\gamma_1;$$

while from Lemma 3, Assumption 2, and the compactness of Ω,

$$\sup_{\omega,i}\left|\frac{E_\omega N^*(c_i)}{|\log c_i|} - \frac{1-\epsilon}{\lambda_1(\omega)}\right| = M_1 \quad \text{(say)},$$

where $M_1 < \infty$. Now, letting $\Omega_i = \{\omega \mid E_\omega N^*(c_i) - (1-\epsilon)\,|\log c_i|/\lambda_1(\omega) < 0\}$, we have

$$-M_1\xi(\Omega_i) \leqq \int_{\Omega_i}\left[\frac{E_\omega N^*(c_i)}{|\log c_i|} - \frac{(1-\epsilon)}{\lambda_1(\omega)}\right]\xi(d\omega) \leqq -\epsilon\gamma_1,$$

which implies that $\xi(\Omega_i) \geqq \epsilon\gamma_1/M_1 = \epsilon_1$ (say). Let $P^*_\omega(c)$ and $Q^*_\theta(c)$ be the probabilities of wrong decision when, respectively, ω and θ are true and δ^*_c is the decision function used. From (2.41) we know that

$$\int \frac{P^*_\omega(c_i)}{c_i\,|\log c_i|}\,\xi(d\omega) \leqq \frac{(\gamma_1+\gamma_2)}{\alpha_2}$$

(α_2 is defined in Assumption 1). Let $\epsilon_2 > 0$ and $\epsilon_1 > \epsilon_2$. It is possible to find a number K_1 such that $A_i = \{\omega \mid P^*_\omega(c_i) \leqq K_1c_i|\log c_i|\}$ has the property that $\xi(A_i) \geqq \xi(\Omega) - \epsilon_2$; in fact, we can take $K_1 = (\gamma_1+\gamma_2)/(\alpha_2\epsilon_2)$. Let $B_i = \Omega_i \cap A_i$. Then $\xi(B_i) \geqq \epsilon_1 - \epsilon_2 > 0$ for each i. Since $\xi(\Omega) < \infty$ and the B_i have uniformly positive ξ-measure, it is easy to see that there is a subsequence $\{i_1, i_2, \cdots\}$ of the positive integers such that $B_{i_1} \cap B_{i_2} \cap \cdots \neq \phi$. Since (2.41) is valid for the subsequence $\{c_{i_j}\}$ of $\{c_i\}$ we might as well assume that $\{i_1, i_2, \cdots\}$ is the set of positive integers and, therefore, $\cap B_i \neq \phi$. Pick $\omega_0 \varepsilon \cap B_i$. By definition of Ω_i and A_i we have, for all i,

$$E_{\omega_0}N^*(c_i) \leqq (1-\epsilon)|\log c_i|/\lambda_1(\omega_0) \tag{2.43}$$

$$P^*_{\omega_0}(c_i) \leqq K_1c_i\,|\log c_i|. \tag{2.44}$$

Let $\Theta_1 = \{\theta \,\varepsilon\, \Theta \mid (1-\epsilon)/\lambda_1(\omega_0) < (1-\epsilon/2)/\lambda_1(\omega_0,\theta)\}$. By the continuity of $\lambda_1(\omega,\theta)$ in θ we have Θ_1 non-empty and open and consequently $\eta(\Theta_1) > 0$. Just as the sets A_i were obtained above, we can find numbers K_2 and K_3 and sets G_i ($i = 1, 2, \cdots$) such that $G_i \subset \Theta_1$, $\eta(G_i) \geqq \epsilon_3 > 0$ for all i, and $E_\theta N^*(c_i) \leqq K_2\,|\log c_i|$, $Q^*_\theta(c_i) \leqq K_3\,|\log c_i|$ for $\theta \,\varepsilon\, G_i$. Again, as with $\{B_i\}$, we can find a subsequence of $\{G_i\}$ which has non-empty intersection and, for simplicity, we can assume $\cap G_i \neq \phi$. Choose $\theta_0 \,\varepsilon\, G_i$ and we conclude that

$$E_{\theta_0}N^*(c_i) \leqq K_2\,|\log c_i| \tag{2.45}$$

$$Q^*_{\theta_0}(c_i) \leqq K_3\,|\log c_i| \tag{2.46}$$

and, furthermore, from the definition of Θ_1 and (2.43), we obtain

$$E_{\omega_0} N^*(c_i) < (1 - \epsilon/2)|\log c_i|/\lambda_1(\omega_0, \theta_0). \tag{2.47}$$

Consider the problem of testing ω_0 vs. θ_0 with cost of observation Hc_i ($H > 0$ and will be chosen below), 0-1 loss function, and a priori probabilities α and $1 - \alpha$ (to be chosen below) on ω_0 and θ_0 respectively. If δ_i' denotes the Bayes solution for this problem then, since δ_i' is a sequential probability ratio test (SPRT), it follows from the properties of SPRT's (which could be obtained from the previous results of this paper or from Wald (1947) or Chernoff (1959)) that

$$\begin{aligned} &\alpha r_{Hc_i}(\omega_0, \delta_i') + (1 - \alpha) r_{Hc_i}(\theta_0, \delta_i') \\ &\qquad \sim \alpha Hc_i |\log Hc_i|/\lambda_1(\omega_0, \theta_0) + (1 - \alpha) Hc_i |\log Hc_i|/\lambda_2(\omega_0, \theta_0). \end{aligned} \tag{2.48}$$

Considering $\delta_{c_i}^*$ as a test of these two simple hypotheses in the obvious way, we obtain from (2.44), (2.45), (2.46), and (2.47) that

$$\begin{aligned} \alpha r_{Hc_i}(\omega_0, \delta_{c_i}^*) + (1 - \alpha) r_{Hc_i}(\theta_0, \delta_{c_i}^*) &\leqq (\alpha K_1 + (1 - \alpha) K_3) c_i |\log c_i| \\ &+ Hc_i |\log c_i| [\alpha(1 - \epsilon/2)/\lambda_1(\omega_0, \theta_0) + (1 - \alpha) K_2]. \end{aligned} \tag{2.49}$$

Now choose H such that $K_1 + (1 - \epsilon/2)H/\lambda_1(\omega_0, \theta_0) < (1 - \epsilon/4)H/\lambda_1(\omega_0, \theta_0)$, and then choose $\alpha < 1$ so that $(1 - \alpha)K_3 + (1 - \alpha)K_2 H < (\alpha H\epsilon/8)/\lambda_1(\omega_0, \theta_0)$. This choice of H and α makes the right hand side of (2.49) smaller than

$$\frac{(1 - \epsilon/8)\alpha Hc_i |\log c_i|}{\lambda_1(\omega_0, \theta_0)} \sim \frac{(1 - \epsilon/8)\alpha Hc_i |\log Hc_i|}{\lambda_1(\omega_0, \theta_0)}$$

and this leads to a contradiction of (2.48).

COROLLARY. *If* $\{\delta_c'\}$ *is any family of procedures for which*

$$\sup_\omega r_c(\omega, \delta_c') + \sup_\theta r_c(\theta, \delta_c') \leqq Kc |\log c| \tag{2.50}$$

for some constant K, *then, letting* $N'(c)$ *be the number of observations required by* δ_c' *to terminate,*

$$E_\omega N'(c) \geqq [1 + o(1)]|\log c|/\lambda_1(\omega) \qquad \textit{for all } \omega, \tag{2.51}$$

$$E_\theta N'(c) \geqq [1 + o(1)]|\log c|/\lambda_2(\theta) \qquad \textit{for all } \theta, \tag{2.52}$$

which imply that

$$r_c(\omega, \delta_c') \geqq [1 + o(1)]|\log c|/\lambda_1(\omega) \tag{2.53}$$

and

$$r_c(\theta, \delta_c') \geqq [1 + o(1)]|\log c|/\lambda_2(\theta). \tag{2.54}$$

REMARK. 8. The uniformity in (2.50) is unnecessary; it is sufficient to have $r_c(\omega, \delta_c')$ and $r_c(\theta, \delta_c')$ of order $c |\log c|$ for each ω, θ.

PROOF. If (2.51) fails for ω_0 (say) then there is a positive number ϵ and a sequence $\{c_i\}$ with $c_i \to 0$ such that

$$(2.55) \qquad E_{\omega_0} N'(c_i) < (1 - \epsilon)|\log c_i|/\lambda_1(\omega_0).$$

We can now argue as in Theorem 1 since (2.44), (2.45), and (2.46) follow immediately from (2.50) and (2.47) follows from (2.55).

It is important to note that the family of procedures $\{\delta_c\}$ depends on F and that the asymptotic Bayes property of $\{\delta_c\}$ is with respect to the same F. It is easily seen that $\{\delta_c\}$ is also asymptotically Bayes relative to any G whose support is the same as that of F and for which dG/dF is bounded. However it would be desirable to be able to state that $\{\delta_c\}$ is asymptotically Bayes with respect to *every* G having the same support (namely, $\Omega \cup \theta$) as F, without any further restrictions. To be able to state such a result it is sufficient to prove, in view of Lemma 3′ and Theorem 1, that

$$(2.56) \qquad \sup_\omega P_\omega(c) + \sup_\theta Q_\theta(c) = o(c\,|\log c|).$$

Of course, if Ω and Θ are finite, then (2.56) follows trivially from Lemma 1, with $O(c)$ for $o(c\,|\log c|)$. Our methods do not seem sufficient to prove (2.56) in general, but we are able to achieve essentially the same end by obtaining the same result (see (2.59)) for a family $\{\delta_c'\}$ which still satisfies the conclusion of Lemma 3′ (see (2.60)) and which differs only slightly from $\{\delta_c\}$ (essentially by using slightly larger stopping bounds; Albert (p. 798, (c)) discusses a similar difficulty and device in his treatment). We require some slight further restrictions on $\{f_\omega\}$ and $\{g_\theta\}$ to obtain this result on the existence of a family $\{\delta_c'\}$ which satisfies this asymptotic Bayes property for all G having $\Omega \cup \Theta$ as its support. To this end let us assume

ASSUMPTION 6. For each $\omega \in \Omega$, $\theta \in \Theta$, $E_\omega[\sup_{|\theta'-\theta|<\rho} g_{\theta'}(X)/f_{\omega'}(X)]$ is finite for some $\rho > 0$ and all ω' in some neighborhood of ω, is continuous at $\rho = 0$, and is continuous in ω and ω'.

Since $E_\omega[g_\theta(X)/f_\omega(X)] = 1$ and $E_\omega[g_\theta(X)/f_\omega(X)]^h$ is convex and not constant in $h \in [0, 1]$, we know that for ϵ small enough there is an h_ϵ with $h_\epsilon \to 1$ as $\epsilon \to 0$ such that $E_\omega[g_\theta(X)/f_\omega(X)]^{h_\epsilon} \leqq 1 - 2\epsilon$. Then by Assumption 6 there is a ρ_ϵ and a neighborhood V_ϵ of ω such that

$$E_{\omega''}[\sup_{|\theta'-\theta|<\rho_\epsilon} g_{\theta'}(X)/f_{\omega'}(X)]^{h_\epsilon} \leqq 1 - \epsilon$$

for all $\omega' \in V_\epsilon$ and $\omega'' \in V_\epsilon$. Taking advantage of the compactness of Θ and Ω in the same way that we did in the proofs of Lemmas 2 and 3, we can find open sets $U_1, \cdots, U_{T_\epsilon}$ which cover Θ, and open sets $V_1, \cdots, V_{M_\epsilon}$ which cover Ω, and a number h_ϵ^* such that

$$\sup_{\omega' \in V_j, \omega'' \in V_j} E_{\omega''}[\sup_{U_i} g_\theta(X)/f_{\omega'}(X)]^{h_\epsilon^*} \leqq 1 - \epsilon$$

for $j = 1, \cdots, M_\epsilon$, $i = 1, \cdots, T_\epsilon$, and with $h_\epsilon^* \to 1$ as $\epsilon \to 0$.

If we define $\delta_{c,\epsilon}^* = \delta_{\bar{c}}$ where $\bar{c} = c^{1/h_\epsilon^*}$, then, from Lemma 3′, we have

$$E_\omega N(\delta^*_{c,\epsilon}) \leqq [1 + o(1)]|\log c|/\lambda_1(\omega)h^*_\epsilon \tag{2.57}$$

with the $o(1)$ term going to zero with c and uniformly in ω. A similar relation holds for $E_\theta N(\delta^*_{c,\epsilon})$. To calculate $P_\omega(\delta^*_{c,\epsilon})$ observe that, as in the proof of Lemma 2, we have, for all $\omega' \varepsilon V_j$,

$$P_{\omega'}(\delta^*_{c,\epsilon}) \leqq P_{\omega'}\left\{\sum_{k=1}^n \int_{V_j} \log f_\omega(X_k) \frac{\xi(d\omega)}{\xi(V_j)} - \sup_{i \leqq T_\epsilon} \sum_{k=1}^n \log \sup_{U_i} g_\theta(X_k) < \frac{\log c}{h^*_\epsilon} + |\log \xi(V_j)| \text{ for some } n\right\}.$$

Hence, putting

$$S_{in} = \sum_{k=1}^\infty \int_{V_j} [\log f_\omega(X_k) - \log \sup_{U_i} g_\theta(X_k)] \frac{\xi(d\omega)}{\xi(V_j)},$$

we have, for all $\omega' \varepsilon V_j$

$$\begin{aligned} P_{\omega'}(\delta^*_{c,\epsilon}) &\leqq \sum_{i=1}^{T_\epsilon} \sum_{n=1}^\infty P_{\omega'}\{S_{in} < \log c/h^*_\epsilon + |\log \xi(V_j)|\} \\ &= \sum_{i=1}^{T_\epsilon} \sum_{n=1}^\infty P_{\omega'}\{\exp(-h^*_\epsilon S_{in}) > (1/c) \exp[h^*_\epsilon \log \xi(V_j)]\}. \end{aligned} \tag{2.58}$$

Now,

$$E_{\omega'} \exp(-h^*_\epsilon S_{in}) = [E_{\omega'} \exp(-h^*_\epsilon S_{i1})]^n$$

and

$$\begin{aligned} E_{\omega'} \exp(-h^*_\epsilon S_{i1}) &\leqq E_{\omega'} \int_{V_j} \left[\frac{\sup_{U_i} g_\theta(X_1)}{f_\omega(X_1)}\right]^{h^*_\epsilon} \frac{\xi(d\omega)}{\xi(V_j)} \\ &\leqq \int_{V_j} (1 - \epsilon) \frac{\xi(d\omega)}{\xi(V_j)} = (1 - \epsilon); \end{aligned}$$

and this, used in (2.58) via the Chebyshev inequality, yields

$$P_{\omega'}(\delta^*_{c,\epsilon}) \leqq \sum_{i=1}^{T_\epsilon} \sum_{n=1}^\infty \{c/[\xi(V_j)]^{h^*_\epsilon}\}(1 - \epsilon)^n = T_\epsilon c/\epsilon[\xi(V_j)]^{h^*_\epsilon}$$

If we let ϵ depend on c (we will write ϵ_c) in such a way that $T_{\epsilon_c}/\epsilon_c[\xi(V_j)]^{h^*_{\epsilon_c}} = o(|\log c|)$ as $c \to 0$ and $\epsilon_c \to 0$ as $c \to 0$, then we obtain, for the procedures $\delta'_c = \delta^*_{c,\epsilon_c}$,

$$\sup_\omega P_\omega(\delta'_c) = o(c\,|\log c|); \tag{2.59}$$

and, since $h^*_{\epsilon_c} \to 1$ as $c \to 0$, we will also have from (2.57)

$$E_\omega N(\delta'_c) \leqq [1 + o(1)]|\log c|/\lambda_1(\omega) \tag{2.60}$$

with the $o(1)$ uniform in ω. A similar result for $Q_\theta(\delta'_c)$ and $E_\theta N(\delta'_c)$, when combined with the above, yield

THEOREM 2. *If Ω and Θ are compact and Assumptions* 1, 2, 3, 5, 6 *are satisfied, then, for a given a priori distribution F whose support is $\Omega \cup \Theta$, the family $\{\delta_c'\}$ just defined satisfies* (2.59), (2.60), *and their analogues in Θ, and is therefore asymptotically Bayes* (*as the cost $c \to 0$*) *with respect to every a priori distribution G having the same support as F.*

3. Indifference region, two decisions. With Ω, Θ as before let us introduce an indifference region I. For $\alpha \varepsilon I$ let h_α denote the density of an observation. Let F be an a priori distribution on $\Omega \cup \Theta \cup I$ and let ξ (resp. η, ψ) denote the restriction of F to Ω (resp. Θ, I). The loss function on Ω and Θ satisfies the assumptions of Section 2, and it is zero on I. (For the sake of notational simplicity, we will carry out the details below when the loss is zero-or-one, but the proof in general is the same except for obvious changes.) Let $\xi^{(n)} = \int_\Omega f_\omega(x_1, \cdots, x_n)\xi(d\omega)$, $\eta^{(n)} = \int_\Theta g_\theta(x_1, \cdots, x_n)\eta(d\theta)$, $\psi^{(n)} = \int_I h_\alpha(x_1, \cdots, x_n)\psi(d\alpha)$. Let $\lambda_{I1}(\alpha) = \inf_\omega E_\alpha[\log h_\alpha(X_1) - \log f_\omega(X_1)]$, $\lambda_{I2}(\alpha) = \inf_\theta E_\alpha[\log h_\alpha(X_1) - \log g_\theta(X_1)]$ and define $I_2 = \{\alpha \varepsilon I \mid \lambda_{I1}(\alpha) \geqq \lambda_{I2}(\alpha)\}$, $I_1 = \{\alpha \varepsilon I \mid \lambda_{I1}(\alpha) \leqq \lambda_{I2}(\alpha)\}$.

We may as well assume, and do, that F has $\Omega \cup \Theta \cup I$ as its support. Our regularity assumptions are the following:

ASSUMPTION IIIA. $\Omega \cup I_1$ and $\Theta \cup I_2$ are compact.

ASSUMPTION IIIB. The assumptions of Section 2 hold when (Ω, Θ) of Section 2 is replaced here by $(\Omega \cup I_1, \Theta)$ and also when (Ω, Θ) is replaced here by $(\Omega, \Theta \cup I_2)$. (In this replacement of Ω by $\Omega \cup I_1$, f_ω on Ω is replaced by f_ω on Ω together with h_α on I_1, etc.)

REMARK 9. As in Section 2, the compactness can often be weakened. The form of the assumptions as stated above is designed to include common cases where I is not closed but where its closure contains points of Ω and Θ. For example, in the case of normal random variables with variance one and mean μ, in testing $\Omega: a \leqq \mu \leqq -1$ against $\Theta: 1 \leqq \mu \leqq b$ with indifference region $-1 < \mu < 1$, we obtain $I_1 = \{-1 < \mu \leqq 0\}$ and $I_2 = \{0 \leqq \mu < 1\}$, neither of which is compact, but the above assumptions are satisfied.

Define δ_c^I as follows: take an $(n + 1)$th observation if

$$(3.1) \qquad \xi^{(n)}/[\xi^{(n)} + \eta^{(n)} + \psi^{(n)}] > c \quad \text{and} \quad \eta^{(n)}/[\xi^{(n)} + \eta^{(n)} + \psi^{(n)}] > c,$$

stop and choose Ω (resp., Θ) if

$$(3.2) \qquad \eta^{(n)}/[\xi^{(n)} + \eta^{(n)} + \psi^{n}] \leqq c \text{ (resp., } \xi^{(n)}/[\xi^{(n)} + \eta^{(n)} + \psi^{(n)}] \leqq c)$$

and $\xi^{(n)} > \eta^{(n)}$ (resp., $\xi^{(n)} < \eta^{(n)}$), randomizing if $\xi^{(n)} = \eta^{(n)}$. The symmetric form of (3.2) is not really necessary; when both inequalities of (3.2) are satisfied, either decision could be made. (The form of (3.1) and (3.2) can be altered to reflect a posteriori *loss*, as in Section 2; again, this is a trivial alteration which is not even needed in view of the identical asymptotic behavior of the present and altered forms.)

Since we are only concerned with probabilities of error when the true state of nature is either in Ω or Θ, Lemma 1 is easily verified. Let $N_I(c)$ be the number of observation for δ_c^I to terminate. When $\omega \varepsilon \Omega$ is true, $E_\omega N_I(c) \leqq E_\omega N(c)$ where $N(c)$ is the first time $\eta^{(n)}/[\xi^{(n)} + \eta^{(n)}] \leqq c$ or, equivalently, $\xi^{(n)}/\eta^{(n)} \geqq 1/c - 1$. Now we can use Lemma 2 (which will work despite the fact that $\xi(\Omega) + \eta(\Theta) < 1$; it is only the positivity of $\xi(\Omega)$ and $\eta(\Theta)$ which is relevant) to conclude that

$$E_\omega N_I(c) \leqq [1 + o(1)] \, |\log c|/\lambda_1(\omega)$$

where λ_1 is as in Section 2. Similarly we get

$$E_\theta N_I(c) \leqq [1 + o(1)] \, |\log c|/\lambda_2(\theta).$$

For $\alpha \varepsilon I_2$, we have $E_\alpha N_I(c) \leqq E_\alpha N'(c)$ where $N'(c)$ is the first n such that $\xi^{(n)}/[\xi^{(n)} + \psi_2^{(n)}] \leqq c$, where $\psi_2^{(n)} = \int_{I_2} h_\alpha(x_1, \cdots, x_n)\psi(d\alpha)$. Again we use Lemma 2 to conclude that

$$E_\alpha N_I(c) \leqq [1 + o(1)] \, |\log c|/\lambda_{I2}(\alpha), \qquad \alpha \varepsilon I_2$$

and, similarly,

$$E_\alpha N_I(c) \leqq [1 + o(1)] \, |\log c|/\lambda_{I1}(\alpha), \qquad \alpha \varepsilon I_1.$$

Lemma 3 follows as before, noting again that we are always dealing with compact sets; for example, if the true state of nature is $\alpha \varepsilon I_1$, we are dealing with $\Omega \cup I_1$ and Θ. Lemma 4 also follows as before (but, again, it will not be used in proving the Theorem).

In the required modification of the proof of Theorem 1, we must now consider both the contingencies of Section 2 and also the possibility of $\int_I E_\alpha N^*(c_i)\psi(d\alpha)$ being too small, where $N^*(c_i)$ denotes the number of observations required by the Bayes solution to stop when c_i is the cost. As for the former, since δ_c^I is a test of Ω vs. Θ, we can conclude from Section 2 that

$$\begin{aligned}(3.3) \qquad &\int_\Omega E_\omega N_I^*(c)\xi\,(d\omega) + \int_\Theta E_\theta N_I^*(c)\eta\,(d\theta) \\ &\qquad \geqq [1 + o(1)]\,|\log c| \left[\int \frac{\xi\,(d\omega)}{\lambda_1(\omega)} + \int \frac{\eta\,(d\theta)}{\lambda_2(\theta)}\right].\end{aligned}$$

As for the latter, to obtain the result

$$(3.4) \qquad \int_I E_\alpha N_I^*(c)\psi\,(d\alpha) \geqq [1 + o(1)]\,|\log c| \int_I \frac{\psi\,(d\alpha)}{\max\,(\lambda_{I1}(\alpha), \lambda_{I2}(\alpha))}$$

we require Lemma 5 below. This lemma, which obtains the required results in the case where each of Ω, Θ, and I consists of one element, is used to obtain (3.4) in the same way that the analogous results concerning the SPRT were used to obtain (3.3) in Section 2. Putting these results together as in Section 2, we obtain

THEOREM 3. *Under the assumptions of the present section, with* δ_c *replaced by* δ_c^I, *the conclusion of Theorem* 1 *holds.*

We also obtain

COROLLARY. *Under the assumptions of the present section, if* $\{\delta_c'\}$ *is any family of* *procedures for which* $\sup_\omega r_c(\omega, \delta_c') + \sup_\theta r_c(\theta, \delta_c') + \sup_\alpha r_c(\alpha, \delta_c') \leqq Kc\,|\log c|$ *for some constant* K, *then* (2.51) *and* (2.52) (*and thus* (2.53) *and* (2.54)) *hold, and also*

$$(3.5)\quad [1 + o(1)]\, c^{-1} \sup_\alpha r_c(\alpha, \delta_c') = \sup_\alpha E_\alpha N'(c) \geqq \frac{[1 + o(1)]\,|\log c|}{\max\,[\lambda_{I1}(\alpha), \lambda_{I2}(\alpha)]}.$$

In fact, all of the Corollary except for (3.5) follows from the Corollary to Theorem 1 (Section 2) and the fact that any test in the present context can be regarded as a test for the problem of Section 2. Lemma 5 yields (3.5) in the same way that properties of the SPRT yielded (2.51) and (2.52) in Section 2.

REMARK 10. The analogue of Remark 8 applies here.

We now turn to Lemma 5. We shall state and prove, for use in Section 4, a more general result than that needed in the present section, where we need consider only three states f, g, h. (We have dropped subscripts for convenience.) The specialization of Lemma 5 to the present case is then obtained by putting h for ω; f for θ_1, g for θ_2, "decision i" for the decision that θ_i is *not* the true state, and $m = k = 2$; Condition (3.7) is vacuous. We then obtain the desired conclusion that, for either δ_c^I or for the Bayes stopping rule,

$$(3.6)\qquad E_h N(c) \geqq [1 + o(1)]\,|\log c|/\max\,(\lambda_{I1}, \lambda_{I2}).$$

(The analogous results for f and g, which have already been stated in a more general context in the Corollary, followed from a comparison with the SPRT; they can also be obtained from Lemma 5, which generalizes to certain multiple decision problems the particular SPRT optimality results it states when $m = 1$, $k = 2$.)

LEMMA 5. *Consider any k-decision problem where the observations* $\{X_i\}$ *are taken sequentially and are independent and identically distributed. Let* $\omega, \theta_1, \cdots, \theta_m$ ($m \leqq k$) *be any* $m + 1$ (*not necessarily distinct*) *states of nature with corresponding densities* $f, g_1, \cdots, g_m$. *Suppose* $\lambda(\omega, \theta_i) = E_\omega[\log f(X) - \log g_i(X)] > 0$ *and* $E_\omega[\log f(X) - \log g_i(X)]^2 < \infty$ *for* $i = 1, \cdots, m$. *Put* $\lambda = \max_i \lambda(\omega, \theta_i)$. *Suppose* $\{\delta_j\,; j \geqq 1\}$ *is a sequence of decision functions satisfying*

$$(3.7)\qquad \sum_{t=m+1}^{k} P_\omega\{\delta_j \text{ makes decision } t\} = o(1) \qquad \text{as } j \to \infty$$

$$(3.8)\qquad P_{\theta_i}\{\delta_j \text{ makes decision } i\} \leqq A c_j\,|\log c_j|$$

for all i, *all* j, *where* A *is some positive constant and* $c_j \to 0$ *as* $j \to \infty$. ((3.7) *is empty if* $m = k$.) *Let* $N(j)$ *be the number of observations required by* δ_j *to terminate. Then* $E_\omega N(j) \geqq [1 + o(1)]\,|\log c_j|/\lambda$ as $j \to \infty$.

PROOF. Write N for $N(j)$ (there will be no confusion since j will be fixed for each calculation), and let $\epsilon > 0$. Put $S_{iN} = \sum_{r=1}^{N} [\log g_i(X_r) - \log f(X_r)]$, $D_i = \{S_{iN} \geqq -(1 - \epsilon)\,|\log c_j|\}$, $A_{in} = \{N = n, \delta_j \text{ selects decision } i\}$ ($i = 1, \cdots, m$). We are assuming here, merely for convenience, that each δ_j is nonrandomized. Put $B_{in} = D_i \cap A_{in}$. The existence of $\lambda(\omega, \theta_1)$ implies that on the set where g_i is 0, f must be 0 *a. e.* (μ) so that, following Lemma 4 of Chernoff, we have

$$\begin{aligned} P_\omega\{\delta_j \text{ makes decision } i, D_i\} &= \sum_{n=1}^{\infty} \int_{B_{in}} \prod_{j=1}^{n} f(x_j)\, d\mu^{(n)} \\ &= \sum_{n=1}^{\infty} \int_{B_{in}} \prod_{j=1}^{n} [f(x_j)/g_i(x_j)] g_i(x_j)\, d\mu^{(n)} \\ &= \sum_{n=1}^{\infty} \int_{B_{in}} e^{-S_{iN}} \prod g_i(x_j)\, d\mu^{(n)} \end{aligned}$$

$$\leqq c_j^{\epsilon-1} \sum_{n=1}^{\infty} P_{\theta_i}\{B_{in}\} \leqq c_j^{\epsilon-1} P_{\theta_i}\{\delta_j \text{ makes decision } i\} \leqq A c_j^{\epsilon} |\log c_j|$$

where the last inequality follows from (3.8). Thus

$$P_\omega\{\delta_j \text{ makes decision } i, D_i\} = o(1) \qquad \text{as } j \to \infty \tag{3.9}$$

We might as well assume that

$$\sum_{t=1}^{k} P_\omega\{\delta_j \text{ makes decision } t\} = 1, \tag{3.10}$$

since, otherwise, we would have $E_\omega N = +\infty$.

Now

$$\begin{aligned} P_\omega\{N < -(1 - 2\epsilon) \log c_j/\lambda\} \leqq \sum_{i=1}^{m} P_\omega\{N <- (1 - 2\epsilon) - \log c_j/\lambda; \\ S_{iN} < (1 - \epsilon) \log c_j\} + P_\omega\{D_1 \cap \cdots \cap D_m\}. \end{aligned} \tag{3.11}$$

From (3.10), (3.9), and (3.7),

$$\begin{aligned} P_\omega\{D_1 \cap \cdots \cap D_m\} \leqq \sum_{i=1}^{m} P_\omega\{D_i, \delta_j \text{ decides } i\} \\ + \sum_{t=m+1}^{k} P_\omega\{\delta_j \text{ decides } t\} = o(1) \end{aligned} \tag{3.12}$$

as $j \to \infty$. Also, for $r > 0$,

$$\begin{aligned} &P_\omega\{N < -(1 - \epsilon) \log c_j/\lambda, S_{iN} < (1 - \epsilon) \log c_j\} \\ &\quad \leqq P_\omega\{\min_{1 \leqq n \leqq r} S_{in} < (1 - \epsilon) \log c_j\} \\ &\quad\quad + P_\omega\{S_{iN}/N < -[(1 - \epsilon)/(1 - 2\epsilon)] (\lambda(\omega, \theta_i)); N > r\}. \end{aligned} \tag{3.13}$$

The last probability can be made arbitrarily small by taking r sufficiently large and using once more the result of Erdös (1949), while the first probability on the right side of (3.13) goes to zero for fixed r as $j \to \infty$. Hence,

$$P_\omega\{N < -[(1 - 2\epsilon)/\lambda] \log c_j ; S_{iN} < (1 - \epsilon) \log c_j\} = o(1) \tag{3.14}$$

as $j \to \infty$. Thus, using (3.14) and (3.12) in (3.11), we obtain

$$P_\omega\{N < -[(1 - 2\epsilon)/\lambda] \log c_j\} = o(1) \tag{3.15}$$

as $j \to \infty$, which proves Lemma 5.

The conclusion of Lemma 5 when $k = m = 2$ is related to a result of Hoeffding (1960) whose concern, in this situation, is to find lower bounds for $E_\omega N$, and one of his lower bounds (to be precise, (1.4) in Hoeffding's paper) has the asymptotic value given by Lemma 5 when (3.8) (of Lemma 5) holds. Hoeffding's conditions are stronger than ours so that his result does not include Lemma 5.

The reader may find it illuminating to consider the specialization of these results to the three-state problem discussed in Section 1 in connection with Figure 4. Writing $U_i = \log[f(X_i)/h(X_i)]$, $V_i = \log[g(X_i)/h(X_i)]$, $S_n = \sum_1^n U_i$, and $T_n = \sum_1^n V_i$ (with $S_0 = T_0 = 0$), the Bayes continuation region, by the analogue of Lemma 4, is easily seen to be

$$\max\{ |S_n - T_n|, -S_n, -T_n\} < |\log c| \; (1 + o(1)), \tag{3.16}$$

for any f, h, g satisfying our conditions. If f, h, g are of exponential type with parameter values increasing in the order f, h, g, then S_n, T_n and n are linearly related, and the unbounded polygonal region (3.16) becomes Schwarz's (bounded) pentagon as exemplified by β of Figure 4 in the symmetric normal case.

Again, as in Section 2, we are able to establish the analogue of Theorem 2 for a family $\{\delta_c'^I\}$ obtained by modifying $\{\delta_c^I\}$, rather than for the original family $\{\delta_c^I\}$. In fact, there is now an additional difficulty: we cannot immediately refer to the argument preceding Theorem 2 because I is not necessarily separated from $\Omega \cup \Theta$ (see the example cited earlier in this section). We get around this by defining δ_{1c} to be the test of Section 2 with the same Ω but Θ replaced by $\Theta \cup I_2$, and define δ_{2c} to be the test of Section 2 with the same Θ but with Ω replaced by $\Omega \cup I_1$. Let $\bar{\delta}_c^I$ be the procedure which (1) continues observing if both δ_{1c} and δ_{2c} continue, (2) stops and selects Ω if δ_{2c} stops and selects $\Omega \cup I_1$ while δ_{1c} continues or stops and selects Ω, (3) stops and selects Θ if δ_{1c} stops and selects $\Theta \cup I_2$ while δ_{2c} continues or stops and selects Θ, and (4) stops and randomizes in all other situations. (In the case of finitely many possible states of nature this "simultaneous test" is essentially like δ_c^I or the procedure described in Figure 3; in the nonseparated case, it will differ from δ_c^I.) It is now possible to modify $\bar{\delta}_c^I$ (by modifying δ_{1c} and δ_{2c} in exactly the way δ_c was modified in the argument leading to Theorem 2) to obtain $\delta_c'^I$ with the desired properties.

We also observe that the family $\{\delta_c'^I\}$ whose construction was just described for a given F with support $\Omega \cup \Theta \cup I$ has an asymptotic risk function which, on any subset I' of I, is identical to that for the analogous asymptotically Bayes family relative to an F' with support $\Omega \cup \Theta \cup I'$. Thus, *the family* $\{\delta_c'^I\}$ *is also asymptotically optimum for the problem where* I *is replaced by a smaller* I' (although of course $\{\delta_c'^{I'}\}$ need not be optimum for $\Omega \cup \Theta \cup I$). This optimality of the given family for problems with reduced indifference region, which was discussed in the example illustrated by Figure 4, is to be contrasted with the loss of optimality which occurs if Ω or Θ is changed, except in certain cases such as those of exponential families as treated by Schwarz, and for which one family of procedures is optimum whether $\Omega = \{\omega : \omega \leqq a\}$ or $\Omega = \{\omega : \omega = a\}$.

It follows from this remark on the effect of I that, in designing sequential procedures for large sample applications, we may as well take I to be so large as to include even the most remotely suspected indifferent possibilities, as long as our assumptions remain satisfied and computations remain tractable. "Robustness" can easily be built into indifference region performance from the outset, in the asymptotic theory. (Of course, the rate of approach to optimality may depend on the size of the chosen I.)

THEOREM 4. *The family* $\{\delta_c'^I\}$ *whose construction (with respect to any* F *with support* $\Omega \cup \Theta \cup I$*) was just described satisfies* P_ϕ {wrong decision} $= o(c\,|\log c|)$ *and*

$$E_\phi N'(c)/|\log c| = [1 + o(1)]r_c(\phi, \delta_c'^I)/c\,|\log c|$$

$$= \begin{cases} [1 + o(1)]/\lambda_1(\phi) & \text{if } \phi \in \Omega, \\ [1 + o(1)]/\lambda_2(\phi) & \text{if } \phi \in \Theta, \\ [1 + o(1)]/\max\,[\lambda_{I1}(\phi), \lambda_{I2}(\phi)] & \text{if } \phi \in I, \end{cases}$$

where the $o(1)$ *terms do not depend on* ϕ. *Hence,* $\{\delta_c'^I\}$ *is asymptotically Bayes relative to every a priori distribution* G *with support* $\Omega \cup \Theta \cup I'$ *where* $I' \subset I$.

4. k-decisions, with or without indifference and semi-indifference regions.

The generalization of Section 2 and 3 to a k-decision problem is quite straightforward. First assume no indifference region, and let Ω_i, $i = 1, \cdots, k$ be k disjoint compact subsets of a Euclidean space and for $\omega \in \bigcup_{i=1}^k \Omega_i$ let f_ω denote the density of an observation. Let F be an a priori distribution on $\bigcup_{i=1}^k \Omega_i$ and let ξ_i be its restriction to Ω_i. For the sake of simplicity we shall describe the results when the loss is 0-1, the modifications needed for the more general case being obvious upon reference to Section 2. Let $\xi_{i,n,x}$ be the a posteriori probability of Ω_i when $x_1, \cdots, x_n$ is observed. Let δ_c be the procedure which stops at the first n for which $\xi_{i,n,x} > 1 - c$ for some i and selects that Ω_i for which $\xi_{i,n,x} > 1 - c$ (or randomizes in any way if there are several such i's, which could only occur if $c > 1/k$). The appropriate assumptions are obtained from those of Section 2 by replacing Ω and Θ by Ω_i and $\bigcup_{j \neq i} \Omega_j$, respectively, when the true state is an element of Ω_i (in place of Ω), for each $i = 1, 2, \cdots, k$. (For example we define $\lambda(\omega) = \inf_{\theta \notin \Omega_i} E_\omega \log [f_\omega(X_1)/f_\theta(X_1)]$ if $\omega \in \Omega_i$.) It can then be proved that $\{\delta_c\}$ is asymptotically Bayes, i.e., the conclusion of Theorem 1 holds for $\{\delta_c\}$. The key is to regard δ_c as a test of Ω_i vs $\bigcup_{j \neq i} \Omega_j$ when concerned with $E_\omega N(c)$ for $\omega \in \Omega_i$. In fact all arguments (except Lemma 1, which is trivial) can be reduced to those of Section 2. The Corollary and Theorem 2 follow similarly. The problem can also be treated through simultaneous tests, as exemplified in Figure 1.

When a single indifference region I is present (so that, for each i, if the true state is in Ω_i (resp., I), the loss is positive (resp., 0) if decision d_j is made, where $j \neq i$), the assumptions and methods of Section 3 carry over directly. There are now k sets $I_j = \{\alpha \in I \mid \lambda_{Ij}(\alpha) = \min_i \lambda_{Ii}(\alpha)\}$, etc.

For the more general situation when there may exist several "semi-indifference" regions, i.e., regions where there may be several possible correct decisions and also several possible incorrect decisions, we only need slightly more care to see that our methods carry over. Let Ω_0 denote the entire parameter space. For convenience we again take the loss function to be 0-1. For $1 \leqq i_1 < i_2 < \cdots < i_m \leqq k$, $1 \leqq m \leqq k$ define $\Omega_{i_1 \cdots i_m} = \{\omega \varepsilon \Omega_0 \mid L(\omega, i_1) = \cdots = L(\omega, i_m) = 0;$ $L(\omega, j) = 1$ if $j \not\varepsilon \{i_1, \cdots, i_m\}\}$. Thus, $\Omega_{i_1 \cdots i_m}$ is the set of ω's where the decisions $i_1, \cdots, i_m$ are correct and all others incorrect; it may or may not be empty. When $m = k$, $\Omega_{i_1 \cdots i_k} = \Omega_{1 \cdots k}$ is a "true" indifference region while, if $2 \leqq m < k$, $\Omega_{i_1 \cdots i_m}$ is a "semi-indifference" region. Let $B_i = \{\omega \mid L(\omega, i) = 0\}$, $i = 1, \cdots, k$. Let F be an a priori distribution on Ω_0. Let $F_{n,x}$ be the a posteriori distribution after n observations when $x = (x_1, \cdots, x_n)$ is observed. Let us define δ_c to be the procedure which stops as soon as $F_{n,x}(B_i) > 1 - c$ for some i and makes that decision i for which $F_{n,x}(B_i) > 1 - c$ or randomizes (in any way) among all decisions i for which $F_{n,x}(B_i) > 1 - c$. Let $N_i(c)$ be the first n such that $F_{n,x}(B_i) > 1 - c$ and $N_i(c) = \infty$ if no such n exists, i.e., $N_i(c)$ is the first n such that

$$\int_{B_i} f_\omega(x_1, \cdots, x_n) F(d\omega) \Big/ \int_{\Omega_0} f_\omega(x_1, \cdots, x_n) F(d\omega) > 1 - c. \tag{4.1}$$

Then, if $N(c)$ is the number of observations required by δ_c to terminate, we have

$$N(c) = \min (N_1(c), \cdots, N_k(c)). \tag{4.2}$$

For this procedure δ_c it is easy to check that the "loss" part of the risk must be $O(c)$, i.e., the analogue of Lemma 1 holds here. In fact, it follows, just as in Lemma 1, that

$$\int_{\Omega_0 - B_i} P_\omega\{\delta_c \text{ makes decision } i\} F(d\omega) \leqq c.$$

To proceed further we need to make assumptions and definitions that parallel those of Section 2.

ASSUMPTIONS.

1. $\lambda(\omega, \theta) = E_\omega [\log f_\omega(X) - \log f_\theta(X)]$ is continuous in both variables simultaneously. (f_ω (f_θ) denotes the density when the true state of nature is ω (θ).)
2. $\lambda(\omega, \theta) = 0$ if, and only if, $\omega = \theta$.
3. The obvious generalizations of Assumptions 3 and 5 of Section 2.
4. Ω_0 is compact, $\Omega_0 - B_i$ is compact for each i, and $F(\Omega_0 - B_i) > 0$ for each i.

REMARK 11. 1. and 2. are stronger than necessary; comparison with Assumptions 1 and 2 of Section 2 will provide weaker assumptions.

We define, for $\omega \varepsilon \Omega_{i_1 \cdots i_m}$,

$$\lambda(\omega) = \max_{1 \leqq j \leqq m} \min_{\theta \varepsilon \Omega_0 - B_{i_j}} \lambda(\omega, \theta) \tag{4.3}$$

Since $\{\Omega_{i_1\cdots i_m}\}$ are disjoint sets there is no vagueness in this definition. Note that, when there are two possible decisions and a single indifference region as in Section 3, this definition is in agreement with the definition there of $\lambda(\alpha) = \max(\lambda_{I1}(\alpha), \lambda_{I2}(\alpha))$. The compactness of Ω_0 and $\Omega_0 - B_i$ and continuity of $\lambda(\omega, \theta)$ guarantees that $\min \lambda(\omega) > 0$.

From (4.2) we have, for $\omega \varepsilon \Omega_{i_1\cdots i_m}$, that

$$E_\omega N(c) \leqq \min_{1 \leqq j \leqq m} E_\omega N_{i_j}(c). \tag{4.4}$$

From (4.1) it follows that $N_i(c)$ is no greater than the first n for which

$$\int_{B_i} f_\omega(x_1, \cdots x_n) F(d\omega) \Big/ \int_{\Omega_0 - B_i} f_\omega(x_1, \cdots x_n) F(d\omega) > \frac{1}{c}.$$

The methods of Lemma 2 enable us to conclude, therefore, that, for $\omega \varepsilon B_i$,

$$E_\omega N_i(c) \leqq [1 + o(1)] \, |\log c| / \min_{\theta \varepsilon \Omega_0 - B_i} \lambda(\omega, \theta). \tag{4.5}$$

Since $\omega \varepsilon \Omega_{i_1\cdots i_m}$ implies that $\omega \varepsilon B_{i_1} \cap \cdots \cap B_{i_m}$, we can conclude from (4.4) and (4.5) that

$$E_\omega N(c) \leqq [1 + o(1)] \, |\log c| / \lambda(\omega) \tag{4.6}$$

as $c \to 0$, for each $\omega \varepsilon \Omega_{i_1\cdots i_m}$. Thus the analogue of Lemma 2 is established. To obtain the analogues of Lemmas 3 and 3′ we can proceed as follows: Define

$$C_i = \{\omega \mid \lambda(\omega) = \min_{\theta \varepsilon \Omega_0 - B_i} \lambda(\omega, \theta)\}, \qquad i = 1, \cdots, k.$$

Some, but not all, C_i may be empty; the C_i's may not be disjoint; the union of the C_i's is Ω_0. Each C_i is closed (and therefore compact) because of the continuity of $\lambda(\omega, \theta)$. Since $\lambda(\omega) > 0$ for all ω we have $C_i \cap \Omega_0 - B_i$ empty for each i. Now argue, as in Lemma 3, by suitably covering C_i and $\Omega_0 - B_i$ as Ω and Θ were covered in proof of Lemma 3 to obtain $E_\omega N(c) \leqq M_i |\log c|$ for all $\omega \varepsilon C_i$. Taking $M = \max(M_1, \cdots, M_k)$ we obtain $E_\omega N(c) \leqq M |\log c|$ for all $\omega \varepsilon \Omega_0$.

To prove the result of Theorem 1, we note that it follows, as in Theorem 1, that if $\{\delta_c\}$ is not asymptotically Bayes then there is a sequence $\{c_j\}$ and a point $\omega_0 \varepsilon \Omega_{1\cdots m}$ (or some $\Omega_{i_1\cdots i_m}$ which for convenience of notation we take to be $\Omega_{1\cdots m}$) such that, for the sequence of Bayes solutions $\{\delta^*_{c_j}\}$

$$E_{\omega_0} N^*(c_j) \leqq (1\text{–}2\epsilon) \, |\log c_j| / \lambda(\omega_0) \tag{4.7}$$

for some $\epsilon > 0$ and all j, and

$$\sum_{t=m+1}^{k} P_{\omega_0} \{\delta^*_{c_j} \text{ makes decision } t\} \leqq A c_j \, |\log c_j| \tag{4.8}$$

for all j. (If $\omega_0 \varepsilon \Omega_{1\cdots k}$ then (4.8) is vacuous). Suppose for convenience that $\omega_0 \varepsilon C_1$ (ω_0 must be in $C_1 \cup \cdots \cup C_m$). Then, again as in the argument of Theorem 1, we can find $\theta_1 \varepsilon \Omega_0 - B_1$ such that $\lambda(\omega_0, \theta_1) \leqq [(1 - \epsilon)/(1 - 2\epsilon)] \lambda(\omega_0)$

and $P_{\theta_1}\{\delta^*_{c_j}$ makes decision 1$\} < Ac_j\,|\log c_j|$ for all j. (Here and in (4.10) a subsequence of $\{c_j\}$ is rewritten, for convenience, as $\{c_j\}$, just as in the proof of Theorem 1.) We can continue and find, in addition to θ_1, (not necessarily distinct) values $\theta_2, \cdots, \theta_m$ with $\theta_i \varepsilon \Omega_0 - B_i$ such that

$$\lambda(\omega_0, \theta_i) \leqq [(1-\epsilon)/(1-2\epsilon)]\lambda(\omega_0) \tag{4.9}$$

$$P_{\theta_i}\{\delta^*_{c_j} \text{ makes decision } i\} \leqq Ac_j\,|\log c_j| \tag{4.10}$$

for all j and $1 \leqq i \leqq m$. We now apply Lemma 5 to $\delta^*_{c_j}$ with ω_0 playing the role of ω in the lemma (the conditions of the lemma are satisfied because of (4.7), (4.8), (4.10) and our assumptions) to obtain, with the help of (4.9),

$$E_{\omega_0}N^*(c_j) \geqq \frac{[1 + o(1)]|\log c_j|}{\max_{1 \leqq i \leqq m} \lambda(\omega_0, \theta_i)} \geqq \frac{1 - 2\epsilon}{1 - \epsilon}[1 + o(1)]\frac{|\log c_j|}{\lambda(\omega_0)}.$$

This contradicts (4.7) and thereby establishes the result of Theorem 1 for $\{\delta_c\}$ as defined above (4.1).

The results of Theorems 2 and 4 can be obtained by a modification of $\{\delta_c\}$ like the one made in Section 3. In fact, for all pairs i, j ($i \leqq k$, $j \leqq k$) such that $C_i \cap B_j$ is not empty, let δ_{ijc} be the *test* of Section 2 of the closure $(\overline{C_i \cap B_j})$ of $C_i \cap B_j$ against $\Omega_0 - B_i$ (recall that C_i is defined after (4.6)). We now proceed by stopping the first time one of the δ_{ijc}'s terminates *and* accepts $\overline{C_i \cap B_j}$ and we make decision j (or randomize among those j's for which there are several pairs i, j).

THEOREM 5. *With* $\lambda(\omega)$ *as defined in* (4.3) *the results of Theorems* 1, 2, 3, *and* 4 *and their corollaries remain valid in k-decision problems with or without indifference and semi-indifference regions.*

The remarks in Section 3 concerning the dependence of the asymptotically optimum procedures on the support of the a priori distribution pertain here with the role of the indifference region played by $\Omega_{1\cdots k}$. The semi-indifference regions cannot be incorporated because (roughly) they affect the asymptotic value of $E_\omega N$ whereas the indifference region, as pointed out in Section 3, does not; and, in addition, the semi-indifference regions appear in the "loss" part of the risk whereas the indifference region does not.

We conclude the discussion in this section by considering the three-decision problem treated by Sobel and Wald (1949). The density functions are normal with (unknown) mean θ and variance 1. Numbers $\theta_1 < \theta_2 \leqq \theta_3 < \theta_4$ are specified and the problem is to decide whether $\theta \leqq \theta_1$, $\theta_2 \leqq \theta \leqq \theta_3$, or $\theta \geqq \theta_4$, with indifference between the first two decisions if the "true" $\theta \varepsilon (\theta_1, \theta_2)$, and with indifference between the second two decisions if $\theta \varepsilon (\theta_3, \theta_4)$. Thus, in the notation of this section, we have $\Omega_0 = (-\infty, +\infty)$, $\Omega_1 = (-\infty, \theta_1]$, $\Omega_2 = [\theta_2, \theta_3]$, $\Omega_3 = [\theta_4, \infty)$, $\Omega_{12} = (\theta_1, \theta_2)$, $\Omega_{23} = (\theta_3, \theta_4)$, and Ω_{13} and Ω_{123} are empty. According to (4.3) $\lambda(\theta) = (\theta - \theta_2)^2/2$ for $\theta \varepsilon \Omega_1$, $\lambda(\theta) = (\theta - \theta_3)^2/2$ for $\theta \varepsilon \Omega_3$, $\lambda(\theta) = \min\{(\theta - \theta_4)^2/2, (\theta - \theta_1)^2/2\}$ for $\theta \varepsilon \Omega_2$, etc. The procedure of Sobel and Wald is to test simultaneously (by sequential probability ratio tests) θ_1 against θ_2 and θ_3

against θ_4 and to decide that $\theta \leqq \theta_1$ if θ_1 and θ_3 are accepted, to decide $\theta_2 \leqq \theta \leqq \theta_3$ if θ_2 and θ_3 are accepted, and to decide $\theta \geqq \theta_4$ if θ_2 and θ_4 are accepted (it is impossible to accept θ_1 and θ_4 simultaneously under their restrictions). By comparing our asymptotic value of $E_\theta N(c)$ with that of their procedure (e.g. (6.8) or (6.9) of Sobel and Wald) it becomes clear that their procedure is asymptotically optimum if and only if $\theta_2 = \theta_3$ and the a priori distribution is concentrated on $\{\theta_1, \theta_2, \theta_4\}$. Moreover, when $\theta = (\theta_1 + \theta_2)/2$ their procedure gives $E_\theta N'(c)/|\log c| \to \infty$ where $N'(c)$ is the number of observations required by the Sobel-Wald procedure to terminate and give error probabilities of order $O(c)$. This last remark is analogous to the behavior of the SPRT for the problem of Figure 4.

5. The design problem. In this section we consider the problems of Sections 2, 3, and 4 when there are design questions developing from the additional feature that, at each stage of observation, the decision to take another observation is accompanied by a choice of an experiment (i.e., a design) to perform in order to make the additional observation. We shall always assume that the same design choices are available at each stage of observation. As mentioned in the summary and introduction, our aim here is to show that there are sequential procedures (each of which, of course, includes the choice of design at each stage) which can be described briefly, can be proved asymptotically optimum easily, and can be used in applications with a minimum of calculations and switching of designs. These procedures make use of an extension of the idea, first implemented by Wald (1951) in certain simpler estimation problems, of taking a preliminary sample which (when c is small) is large but is small relative to the total expected sample size, using this preliminary sample to estimate the "true" state of nature, and then deciding once and for all on the future course of experimentation. Before introducing detailed design considerations, let us make this idea more transparent. Here and throughout Section 5 we will give details only for the design problem associated with Section 2; the arguments in the context of Sections 3 and 4 will be quite similar.

Suppose first, to make the statement simple, that there are a finite number k of possible states of nature. For $1 \leqq i \leqq k$ suppose $\{\delta_c^i\}$ is a family of procedures (including choices of designs) such that, for all i and j,

$$\begin{aligned} E_j N(\delta_c^i) &= O(|\log c|), \\ E_i N(\delta_c^i) &= \mu_i |\log c| \, [1 + o(1)], \qquad (5.1) \\ P_j\{\text{wrong decision using } \delta_c^i\} &= o(c \log c). \end{aligned}$$

Let d_i be the "decision" that i is the true state (this is not necessarily the original decision space). Let $\{\delta_c^0\}$ be any family of procedures for which

$$\begin{aligned} E_j N(\delta_c^0) &= o(|\log c|), \qquad (5.2) \\ P_i\{\delta_c^0 \text{ reaches decision } d_i\} &= 1 - o(1), \end{aligned}$$

for $1 \leqq i, j \leqq k$. Consider the procedure δ_c^* defined as follows: First use δ_c^0. If "decision" d_i is made, then use δ_c^i (as though starting anew; we assume the same design choices are available at each stage). The final decision is reached using δ_c^i. Then, supposing for the moment that each observation costs the same amount c, (5.1) and (5.2) trivially yield, for the risk when i is true,

$$r_i(\delta_c^*) = cE_i N(\delta_c^*) \, [1 + o(1)] = \mu_i \, c|\log c| \, [1 + o(1)].$$

If μ_i can be shown to be the minimum possible value of $E_i N(\delta)$ as $c \to 0$ for any procedure which is Bayes relative to an F with support $\{1, 2, \cdots, k\}$, we will thus have, in $\{\delta_c^*\}$, an asymptotic Bayes procedure relative to any such F.

[Added in revision: L. R. Abramson, in his Columbia thesis, has, independently, employed this same technique in dealing with the case of two states of nature, two decisions, and k experiments, and his $\{\delta_c^0\}$ is a SPRT as is $\{\delta_c^i\}$, $i = 1, 2$. Under Abramson's restrictions, which imply that $\lambda^{e_j}(\theta) > 0$ for $\theta = \theta_1, \theta_2$ and $j = 1, \cdots k$ (see (5.9) and previous for the definition of λ^{e_j}), our construction of $\{\delta_c^i\}$ by any of the Methods I, II, III below shows that, in this special case, $\{\delta_c^i\}$ will be a SPRT for $i = 1, 2$ (this results from the fact that when there are two states of nature the maximin $\bar{e}$ (see (5.10) et seq.) is non-randomized), and one of the possibilities for our construction of $\{\delta_c^0\}$ (see the paragraph following (5.10)) is to take $\{\delta_c^0\}$ as a SPRT (the key here is that $\bar{e}^0$ can be taken, under Abramson's assumptions, as non-randomized).]

In the case where $\Omega \cup \Theta = \Phi$ consists of infinitely many states of nature we have to proceed with somewhat extra care. As in Section 2 let F be an a priori distribution on Φ and we suppose that the support of F is all of Φ. Suppose we can select designs and make observations $\{X_1, \cdots\}$ (not necessarily independent nor identically distributed) and, on the basis of these observations, suppose there is a sequence $\{t_n ; n \geqq 1\}$ of consistent estimators of the true parameter, i.e., for any $\epsilon > 0$

$$\lim_{n\to\infty} P_\phi \{ |t_n(X_1, \cdots, X_n) - \phi| < \epsilon\} = 1 \tag{5.3}$$

for each $\phi \in \Phi$. How to make the observations $X_1, \cdots$ in terms of the available designs is immaterial for the moment; we are merely supposing that it is possible to do so and to define $\{t_n\}$ which will satisfy (5.3). Let us further suppose that, for each $\gamma \in \Phi$ and each $\epsilon > 0$, there is a family $\{\delta_c^{\gamma,\epsilon}\}$ of decision procedures (which incorporates the choice of design at each stage) such that

$$\begin{aligned} &\text{(a)} \quad \textstyle\int_\Phi P_\phi\{\text{wrong decision using } \delta_c^{\gamma,\epsilon}\} F(d\phi) = O(c); \\ (5.4) \quad &\text{(b)} \quad \sup_{\phi\in\Phi} E_\phi N(\delta_c^{\gamma,\epsilon}) = O(|\log c|) \\ &\text{(c)} \quad E_{\phi'} N(\delta_c^{\gamma,\epsilon}) \leqq [1 + \epsilon + o(1)] \, |\log c| \, \mu(\phi'), \text{ for } |\phi' - \gamma| \leqq \epsilon'. \end{aligned}$$

for some $\epsilon' > 0$ and where $\mu(\phi')$ is some positive number. Now define a family of procedures $\{\delta_c^{*\epsilon}\}$ as follows: Let $n(c)$ be a sequence of integers with $n(c) = o(|\log c|)$. Take $n(c)$ observations $x_1, \cdots, x_{n(c)}$ and compute $t_{n(c)}(x_1, \cdots, x_{n(c)})$. If $t_{n(c)} = \gamma$ use $\delta_c^{\gamma,\epsilon}$ and continue observation *independently of the* $n(c)$ *observations used to form* $t_{n(c)}$ until $\delta_c^{\gamma,\epsilon}$ terminates and then make whatever decision $\delta_c^{\gamma,\epsilon}$ makes.

Thus, when c is the cost of observation,

$$
\begin{aligned}
r_c(\phi, \delta_c^{*\epsilon}) \leqq\ & E_\phi P_\phi\{\text{wrong decision using } \delta_c^{\gamma,\epsilon} \mid t_{n(c)} = \gamma\} + cn(c) \\
& + cO(|\log c|)P_\phi\{|t_{n(c)} - \phi| > \epsilon'\} \\
& + c\,|\log c|\,[1 + \epsilon + o(1)]\mu(\phi)P_\phi\{|t_{n(c)} - \phi| < \epsilon'\}.
\end{aligned}
\tag{5.5}
$$

Now, integrating both sides with respect to F and using (b) of (5.4) we obtain

$$
\begin{aligned}
r_c(F, \delta_c^{*\epsilon}) &\leqq o(c\,|\log c|) + (1 + \epsilon)c\,|\log c| \int_\Phi \mu(\phi)F\,(d\phi) \\
&= [1 + \epsilon + o(1)]c\,|\log c| \int_\Phi \mu(\phi)F\,(d\phi)
\end{aligned}
\tag{5.6}
$$

where the $o(1)$ term may depend on ϵ.

REMARK 12. If we assume that $\{t_n\}$ is uniformly consistent i.e., that the limit in (5.3) holds, for each $\epsilon > 0$, uniformly for $\phi \,\varepsilon\, \Phi$; if we assume that the $o(1)$ term in (c) of (5.4) is uniform in ϕ; and, finally, if we replace (a) of (5.4) by (a$'$)P_ϕ {wrong decision using $\delta_c^{\gamma,\epsilon}$} $= o(c\,|\log c|)$ uniformly in ϕ, we can obtain from (5.5)

$$
r_c(\phi, \delta_c^{*\epsilon}) \leqq c\,|\log c|\,[1 + \epsilon + o(1)]\mu(\phi)
\tag{5.7}
$$

with the $o(1)$ term in (5.7) independent of ϕ but perhaps depending on ϵ. The difference between the uniform and non-uniform statements is related, as we shall see, to the difference between Theorem 1 and Theorem 2.

The $o(1)$ term in (5.6) depends on ϵ, but is $<\epsilon$ for $c \leqq C_\epsilon$ (say). Thus, for each ϵ, the bracketed expression in (5.6) is $<1 + 2\epsilon$ for $c \leqq C_\epsilon$, and if we define δ_c^* to be $\delta_c^{*\epsilon}$ for the smallest ϵ satisfying $C_\epsilon \geqq c$ (with the obvious modification if the infimum of such ϵ is not attained), we have

$$
r_c(F, \delta_c^*) \leqq [1 + o(1)]c\,|\log c| \int_\Phi \mu(\phi)F\,(d\phi).
\tag{5.8}
$$

Our problem will be to show how to construct $\{t_n\}$ and $\{\delta_c^{\gamma,\epsilon}\}$ (resp. $\{\delta_c^0\}$ and $\{\delta_c^i\}$) so that (5.3) and (5.4) (resp. (5.2) and (5.1)) are satisfied with $\mu(\phi)$ (resp. μ_i) in (5.4) (resp. (5.1)) the minimum possible value; it will then follow that $\{\delta_c^*\}$, as constructed in the above, is asymptotically Bayes with respect to F. This will yield the result of Theorem 1. To obtain the result of Theorem 2 we must construct $\{t_n\}$ and $\{\delta_c^{\gamma,\epsilon}\}$ so that the uniform versions (see Remark 12) of (5.3) and (5.4) hold. (When Φ is finite, Theorem 2 is the same as Theorem 1).

We are now ready for our design considerations. Let $\mathcal{E} = \{e\}$ be the set of one-observation experiments available at each stage. Each such experiment will be assumed to cost the same amount c, since the more general case can be treated without difficulty in the same way merely by dividing "information numbers" by experimental costs. For each k, given the outcomes of the first k experiments, the conditional distribution of the $(k + 1)$st observation depends only on the experiment chosen for this $(k + 1)$st stage. We alter the notation of Section 2 slightly and take f_θ^e to be the density of an observation when e is the experiment and $\theta \varepsilon \Phi$ is true, while $\lambda^e(\theta, \phi) = E_\theta \log [f_\theta^e(X_1)/f_\phi^e(X_1)]$ is the corresponding "information" concerning ϕ when θ is true; θ and ϕ are arbitrary elements in Φ. (In Section 2 we limited the notation so that $\theta \varepsilon \Theta$; this is no longer so.) For simplicity we can assume one measure μ applies for all e, although this is not really necessary. As in Chernoff (1959), Albert (1961), and Bessler (1960), we must also consider mixtures of experiments, i.e., probability measures $\bar{e}$ on $\mathcal{E}$. (We shall see that it always suffices, under the assumptions considered in the next paragraph, to limit consideration to measures $\bar{e}$ with finite support. It may nevertheless be convenient in applications to consider $\bar{e}$'s with infinite support. If $\mathcal{E}$ is nondenumerable, such $\bar{e}$'s are measures relative to a separable Borel field generated by open sets in a topology relative to which $\lambda^e(\theta, \phi)$ is continuous, etc.) Then any information number $\lambda^{\bar{e}}$ is defined to be the probabilistic mixture under $\bar{e}$ of λ^e's, etc.

The assumptions we require are similar to those of previous sections and will usually be easy to verify if f_θ^e is sufficiently regular in θ and e as the latter vary over domains which are assumed compact, although such compactness is not necessary. For brevity and simplicity, we assume the following:

ASSUMPTIONS.

(1) Ω, Θ, $\mathcal{E}$ are compact with Ω, Θ subsets of some Euclidean space.

(2) Assumption 1 of Section 2.

(3) Assumption 3 with a supremum over e inserted immediately to the left of each expectation sign.

(4) Assumption 2 is altered by requiring boundedness and continuity of $\lambda^e(\theta, \phi)$ in all three variables and

$$\sup_{\bar{e}} \inf_{\theta \varepsilon \Omega, \phi \varepsilon \Theta} \lambda^{\bar{e}}(\theta, \phi) > 0; \qquad \sup_{\bar{e}} \inf_{\theta \varepsilon \Theta, \phi \varepsilon \Omega} \lambda^{\bar{e}}(\theta, \phi) > 0. \tag{5.9}$$

(5) Assumption 5 with the continuity there uniform in e.

For the proof of the analogue of Theorem 2 we will also need

(6) Assumption 6 is altered with the insertion of a supremum over e to the left of the expectation sign and the continuity there is uniform in e.

We define

$$\begin{aligned} \lambda(\theta) &= \sup_{\bar{e}} \inf_{\phi \varepsilon \Theta} \lambda^{\bar{e}}(\theta, \phi) & \text{if } \theta \varepsilon \Omega \\ &= \sup_{\bar{e}} \inf_{\phi \varepsilon \Omega} \lambda^{\bar{e}}(\theta, \phi) & \text{if } \theta \varepsilon \Theta \end{aligned} \tag{5.10}$$

Under the compactness and continuity Conditions (1) and (4) just mentioned, the game with kernel $\lambda^e(\theta, \phi)$ for each fixed $\theta \varepsilon \Omega$ is determined and has positive value $\lambda(\theta)$ which by (5.9) is bounded away from zero. The same holds for $\theta \varepsilon \Theta$. Moreover, for each $\epsilon > 0$ we can find a finite covering of Ω (or of Θ) such that, for each set S of the covering, there is an $\bar{e}_s$ with finite support and which is ϵ-maximin for all θ in S. Also, there is an $\bar{e}^0$ (say) with finite support and rational probabilities for which $\inf_{\theta \varepsilon \Omega, \phi \varepsilon \Theta} \lambda^{\bar{e}^0}(\theta, \phi) > 0$, and $\inf_{\theta \varepsilon \Theta, \phi \varepsilon \Omega} \lambda^{\bar{e}^0}(\theta, \phi) > 0$. These conclusions can be derived more generally from the work of Wald (1950) or

LeCam (1955) under assumptions analogous to Assumption 4, but our object here is brevity rather than greatest generality. $\lambda(\theta)$ turns out, as we shall see, to be the reciprocal of the minimum value of $\mu(\theta)$ for which (5.4) (or (5.1)) holds.

We now describe several simple ways of constructing asymptotically optimum designs. First, if Φ is finite, δ_c^0 can be constructed very easily: letting b be the denominator of the rational probabilities $\pi_1, \pi_2, \cdots, \pi_m$ (say) associated with $\bar{e}^0$, we need only take "blocks" of b experiments, $b\pi_j$ in each block being of the jth type making up $\bar{e}^0$; we can then view the vector of the outcomes from each of these blocks as a single "observation" and with such observations we take δ_c^0 to be the procedure $\delta_{q(c)}$ of Section 4 (without indifferences) with $q(c) = o(1)$ and $|\log q(c)| = o(|\log c|)$, e.g., $q(c) = |\log c|^{-1}$. Alternatively we could use the randomized experiment $\bar{e}^0$ at each stage.

When Φ is infinite there will often be a simple choice for $\{t_n\}$ using either the "block" method above or the randomized experiment $\bar{e}^0$ at each stage to obtain observations. In fact, under our assumptions, the sequence of maximum likelihood estimators, based on observations obtained as just described, is consistent (see Kiefer and Wolfowitz (1956)).

An alternative program, when Φ is infinite, is to construct $\{\delta_c^{*\epsilon}\}$ to satisfy (5.5) in a way analogous to the construction of $\{\delta_c^*\}$ from $\{\delta_c^0\}$ and $\{\delta_c^i\}$ following (5.2). We first construct, for each $\epsilon' > 0$, a family $\{\delta_c^{0,\epsilon'}\}$ (not to be confused with $\{\delta_c^{\gamma,\epsilon}\}$) as follows: Suppose that observations are taken either by the "block" method or by $\bar{e}^0$ at each stage. Let $U_1, \cdots, U_{k(\epsilon')}$ be a collection of open spheres of radius $\epsilon'/2$ which covers Φ and let $\gamma_1, \cdots, \gamma_{k(\epsilon')}$ be the centers of the spheres, so that we have $\gamma_i \in \Phi$ for $i = 1, \cdots, k(\epsilon')$. Consider the $k(\epsilon')$-decision problem with Φ the space of states of nature, decisions $d_1, \cdots, d_{k(\epsilon')}$, and loss function L satisfying $L(\phi, d_j) = 0$ if $\phi \in U_j$, $L(\phi, d_j) = 1$ if $\phi \notin U_j$. We are now in the context of Section 4 (with indifferences and semi-indifferences) and we let $\delta_c^{0,\epsilon'}$ be the procedure there with the c in Section 4 replaced by $q(c)$ with $|\log q(c)| = o(|\log c|)$ and $q(c) = o(1)$, e.g., $q(c) = |\log c|^{-1}$. We now define $\{\delta_c^{*\epsilon}\}$ by taking ϵ' as in (5.4c), using $\delta_c^{0,\epsilon'}$ until it terminates, and if it leads to decision d_j we "estimate" the true parameter by γ_j and then use $\delta_c^{\gamma_j,\epsilon}$ until the latter terminates and we make the decision that $\delta_c^{\gamma_j,\epsilon}$ leads to. That (5.5) is satisfied for this definition of $\delta_c^{*\epsilon}$ follows because $E_\phi N(\delta_c^{0,\epsilon'}) = O\,(|\log q(c)|) = o(|\log c|)$ uniformly for $\phi \in \Phi$ and, if $\phi \in U_i$, $\sum' P_\phi\{\delta_c^{0,\epsilon'}$ makes decision $d_j\} = o(1)$ where $\sum'$ means summing over those j such that $U_i \cap U_j$ is empty.

We now turn to the construction of $\{\delta_c^{\gamma,\epsilon}\}$. We will give two methods, both of which can be used to obtain $\{\delta_c^i\}$. A third method which we present is appropriate when Φ is finite, i.e., for the construction of $\{\delta_c^i\}$. To describe the first two methods suppose that $\gamma \in \Omega$ and $\epsilon > 0$ are fixed. Let $\bar{e}_\gamma$, $\epsilon' > 0$, $\epsilon'' > 0$ be such that

$$\text{(5.11)}\qquad \begin{aligned} &\text{(a)}\quad 0 < \epsilon'' < 2\epsilon \inf_{\theta\in\Omega} \lambda(\theta) \\ &\text{(b)}\quad \lambda(\theta) \leqq \inf_{\phi\in\Theta} \lambda^{\bar{e}_\gamma}(\theta, \phi) + \epsilon'' \quad \text{for} \quad |\theta - \gamma| < \epsilon' \\ &\text{(c)}\quad \inf_{\theta\in\Omega, \phi\in\Theta} \lambda^{\bar{e}_\gamma}(\theta, \phi) > 0. \end{aligned}$$

Here $\bar{e}_\gamma$ may be chosen to have finite support and with rational probabilities because of the comments following (5.10). Condition (c) can be satisfied because we may always combine $\bar{e}_\gamma$ with a small multiple of $\bar{e}^0$. The fact that the remainder of (5.11) can be satisfied follows from our assumptions.

Method I. Define $\{\delta_c^{\gamma,\epsilon}\}$ by using the randomized experiment $\bar{e}_\gamma$ at each stage and the stopping and terminal decision rule of $\{\delta_c\}$ in Section 2 with the likelihood function there being replaced after n observations by $\prod_{j=1}^n f_\phi^{e_j}(x_j)$ where e_j is the actual (nonrandomized) experiment chosen at stage j (after randomization as prescribed by $\bar{e}_\gamma$). This is in the spirit of Chernoff's treatment, except that, once the "preliminary experiment" (associated with $\{t_n\}$) has been performed, we select the same $\bar{e}_\gamma$ to be used at every stage, rather than having to allow the randomized experiment to be changed continually with sequential calculations. Note that the role of the preliminary experiment is to obtain the "right" design. The relevant random walk considerations of Section 2 proceed as before with $\lambda(\theta, \phi)$ replaced by $\lambda^{\bar{e}_\gamma}(\theta, \phi)$ and without any additional design considerations, thus yielding (5.4) with (c) of (5.4) following from Lemma 2 and (a) of (5.11).

Method II. Instead of utilizing the randomized experiment $\bar{e}_\gamma$ we can use the "block" method described earlier for δ_c^0, replacing $\bar{e}^0$ in that description by $\bar{e}_\gamma$ and observing that the methods of Section 2 will yield

$$E_\theta\,\{\text{number of blocks}\} \sim [1 + \epsilon + o(1)]\,|\log c|/b\lambda(\theta)$$

since the appropriate information number when θ is true and $\phi \in \Phi$ is $b\lambda^{\bar{e}_\gamma}(\theta, \phi)$. This method thus yields a procedure with non-randomized choices of designs and which satisfies (5.4). (The possibility of using nonrandomized choices with the right asymptotic frequency was mentioned explicitly by Chernoff (1959) in the heuristic discussion of his "prototype" example.)

Method III. Here we assume that Φ is finite. The method will yield a direct construction of $\{\delta_c^i\}$ to satisfy (5.1). For each possible state i we specify a fixed sequence of experiments $e_{i1}, e_{i2}, \cdots$ such that the frequency of occurrences of each different e among $e_{i1}, \cdots, e_{in}$ tends as $n \to \infty$ to the probability assigned by a maximin $\bar{e}^i$ to e. (Since the set of possible states is finite, a maximin $\bar{e}^i$ with finite support is known to exist. Of course, if $\bar{e}^i$ has rational probabilities, we can consider blocks.) We assume here, for simplicity, that $\min_{\phi\neq\theta} \lambda^{\bar{e}^i}(\theta, \phi) > 0$ for each i, the modification needed otherwise being simple. Using a standard argument usually applied to identically distributed random variables, we again bound $E_\theta N$ by showing the smallness of

$$P_\theta\left\{\sum_{j=1}^{(1+\epsilon)|\log c|/\lambda(\theta)} \log\,[f_\theta^{e_{ij}}(X_j)/f_\phi^{e_{ij}}(X_j)] < |\log c|\right\}$$

by applying exponentiation and Chebyshev's inequality. It is important that the sequences $\{e_{ij}\}$ do not change with c.

To summarize, then, our methods reduce the construction of procedures and proofs in the design setting to those of Sections 2, 3, and 4 where there is no design problem, and we obtain (under Assumptions (1)–(5)) for a given a priori distribution F (whose support is Φ for convenience), a family $\{\delta_c^*\}$ such that (5.8) holds with $\mu(\phi) = 1/\lambda(\phi)$ where $\lambda(\phi)$ is defined in (5.10). By appropriate use of Remark 12, adding Assumption (6), and using the procedures $\{\delta_c'\}$ of Theorem 2, we can also obtain a family $\{\delta_c^{**}\}$ such that (5.8) holds with $\{\delta_c^*\}$ replaced by $\{\delta_c^{**}\}$ and F replaced by any G having the same support as F (without changing δ_c^{**}) with the $o(1)$ term in (5.8) uniform for all such G.

As the last step in our discussion we will show that, if $\{\delta_c^B\}$ is Bayes with respect to F, then

$$r_c(F, \delta_c^*)/r_c(F, \delta_c^B) \to 1 \tag{5.12}$$

i.e., $\{\delta_c^*\}$ is asymptotically Bayes as $c \to 0$.

To establish (5.12) we proceed as in Theorem 1 and conclude that, if (5.12) fails, there is a sequence $\{c_i\}$ and a $\theta \in \Omega$ (say) such that

$$\begin{aligned} E_\theta N(\delta_{c_i}^B) &\leqq (1 - 2\epsilon)\, |\log c_i|/\lambda(\theta) \\ P_\theta\{\text{wrong decision using } \delta_{c_i}^B\} &\leqq Kc_i\, |\log c_i|. \end{aligned} \tag{5.13}$$

By Assumption (4), we can find $\{\phi_1, \cdots, \phi_k\} \subset \Theta$ such that

$$0 < \frac{1 - 2\epsilon}{\lambda(\theta)} \leqq \frac{1 - \epsilon}{\sup_{\bar{e}} \inf_{1 \leqq j \leqq k} \lambda^{\bar{e}}(\theta, \phi_j)} < \infty \tag{5.14}$$

and $\{\phi_j\}$ can be selected so that, for a subsequence of $\{c_i\}$ (which we take to be $\{c_i\}$ for notational convenience),

$$P_{\phi_j}\{\text{wrong decision using } \delta_{c_i}^B\} \leqq Kc_i\, |\log c_i|$$

for all i and $1 \leqq j \leqq k$. Considering $\delta_{c_i}^B$ as a test of θ vs. $\phi_1, \cdots, \phi_k$ with available experiments $\mathcal{E}$ we see, by reference to Lemmas 4 and 5 of Chernoff (although Chernoff assumes $\mathcal{E}$ is finite his arguments remain valid under our assumptions on $\mathcal{E}$ and (5.14)), that

$$E_\theta N(\delta_{c_i}^B) \geqq \frac{[1 + o(1)]\, |\log c_i|}{\sup_{\bar{e}} \inf_{1 \leqq j \leqq k} \lambda^{\bar{e}}(\theta, \phi_j)}$$

which, together with (5.14), contradicts (5.13). Thus (5.12) is proved.

For the design problems in the context of Sections 3 and 4 the assumptions, definitions, and constructions are analogous to those described here for Section 2, and the verification of (5.12) proceeds by using the extension of Lemmas 4 and 5 of Chernoff to the multi-decision problem in the same way we used our Lemma 5 in Sections 3 and 4.

Interpreting the assumptions, definitions, and constructions in the appropriate manner to suit the context (Sections 2, 3, or 4) we can now state

THEOREM 6. *Under Assumptions* (1)–(5) *and with F the a priori distribution, the family $\{\delta_c^*\}$, constructed by any of the described methods, is asymptotically Bayes with respect to F for the problems of Sections* 2, 3, *or* 4 *when $\mathcal{E}$ is the set of available one-observation experiments. If Assumption* (6) *is added the family $\{\delta_c^{**}\}$ is asymptotically Bayes with respect to any G having the same support as F.*

The corollaries to Theorem 1 and Theorem 3 are also valid here; the corollary to Theorem 1 in the context of this section is, in fact, the optimality result of Chernoff, Albert, and Bessler. Less restrictive hypotheses, under which Theorem 6 remains valid can be obtained in analogy with Remark 7 in Section 2.

The dependence of the procedures on the support of the a priori distribution is subject to the same remarks as made in Section 3 and Section 4. Thus, for the design problem associated with Section 3, if the support F is $\Omega \cup \Theta \cup I$ then $\{\delta_c^{**}\}$ is asymptotically Bayes with respect to any G whose support is $\Omega \cup \Theta \cup I'$ for any subset I' of I.

The following is an example of these considerations (an example related to Bessler's Example 5): Suppose there are three possible states of nature and three corresponding decisions. The three possible states of nature specify (f, g, g), (g, f, g), and (g, g, f) as the possible vector of densities of three populations where f and g are specified. e_i is the experiment which takes an observation from population i. Bessler's solution to the design problem is to observe, at each stage, that population which, on the basis of previous observations, is the maximum likelihood estimator of the population with density f. A solution using Method II above is first to take $n(c) = o(|\log c|)$ (but with $n(c) \to \infty$) observations (e.g., $[|\log c|^{\frac{1}{2}}]$) from each population and, on the basis of these observations, find the maximum likelihood estimator of the population with density f. If this is population i we conduct the remaining experimentation from population i except that out of every $n(c)$ observations we reserve two observations one from each of the other two populations. (This modification is needed since λ^{e_i} can be zero.) Since the $o(1)$ term of the second line of (5.2) is then of order $e^{-n(c)}$ while the first line of (5.1) which it will multiply in the final expression for EN is of order $|\log c|\, n(c)$, it is easy to verify that this design is asymptotically optimum. (We have not satisfied (5.1), but have proceeded in a modified manner which is more expeditious.)

If f and g are normal with unit variance and means 1 and 0 respectively, Bessler's solution is to observe the population i for which $\sum_{j=1}^{n_i} x_{ij} - n_i/2$ is largest, where $\{x_{ij}, j = 1, \cdots, n_i\}$ are the previous observations from population i. However, for less simple f and g the calculations are more formidable, and for slightly more complicated problems, e.g., involving three different normal densities instead of two, Bessler shows how much more complicated his solution can become, with the necessity of prescribing randomization probabilities which vary from stage to stage.

APPENDIX

Let $\{\xi_i\}$ be a sequence of independent and identically distributed random variables with distribution G. Let $\mu(G) = E_G\xi_1$, $\sigma^2(G) = \operatorname{Var}_G \xi_1$, and for $\mu_0 > 0$, $0 < \sigma_0 < \infty$ put $\mathcal{G}(\mu_0, \sigma_0) = \{G \mid \mu(G) \leqq -\mu_0, \sigma^2(G) \leqq \sigma_0^2\}$, and for $\mu_0 > 0$ and $\rho_0 > 0$ let $\mathcal{G}^*(\mu_0, \rho_0) = \{G \mid \mu(G) \leqq -\mu_0, \mu(G)/\sigma(G) \leqq -\rho_0\}$. Let $S_k = \sum_1^k \xi_i$, $k = 1, 2, \cdots$; $S_0 = 0$.

THEOREM A.

$$\sup_{G \varepsilon \mathcal{G}(\mu_0, \sigma_0)} \sum_{k=1}^{\infty} P_G\{S_k > 0\} < \infty. \tag{A.1a}$$

(A.1b) $$\sup_{G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)} \sum_{k=1}^{\infty} P_G\{S_k > 0\} < \infty.$$

PROOF. Since

$$P_G\{S_k > 0\} = P_G\{S_k - \mu(G)k > \mu(G)k\} \leqq P_G\{S_k - \mu(G)k > \mu_0 k\}$$

for all G we can dominate (A.1a) by

(A.2) $$\sup_{G \varepsilon \mathcal{G}'} \sum_{k=1}^{\infty} P_G\{S_k > k\}$$

where $\mathcal{G}' = \{G \mid \mu(G) = 0, \sigma^2(G) \leqq \sigma_0^2/\mu_0^2\}$. For fixed $G \varepsilon \mathcal{G}'$ the sum in (A.2) converges; see Erdös (1949). An inspection of Erdös' proof reveals that the only way the distribution of ξ_1 enters in the argument (outside of its mean being 0) is through its variance, and the use of Chebyshev's inequality in Erdös' argument shows that variance 1 (which is what he assumes) can be replaced by an upper bound on the variance. Thus (A.2) is finite and (A.1a) is thereby established.

(A.1b) follows from (A.1a) by observing that

$$\sup_{G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)} \sum P_G\{S_k > 0\} = \sup_{G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)} \sum P_G\{S_k/\sigma_0 > 0\} = \sup_{G \varepsilon \mathcal{G}(\rho_0, 1)} \sum P_G\{S_k > 0\}.$$

THEOREM B.

(A.3) $$\sup_{G \varepsilon \mathcal{G}(\mu_0, \sigma_0)} E_G[\max_{k \geqq 0} S_k] < \infty.$$

PROOF. It is well known that the expectation considered here is finite when $\mu(G) < 0$, $\sigma^2(G) < \infty$. To obtain the uniformity expressed in (A.3) we use Theorem 4.1 in Spitzer (1956) which states

$$E_G \max_{k \geqq 0} S_k = \sum_{k=1}^{\infty} (1/k) E_G S_k^+ .$$

Suppressing the dependence on G we can now write

$$\begin{aligned} E \max_{k \geqq 0} S_k &= \sum_{k=1}^{\infty} \frac{1}{k} E S_k^+ = \sum_{k=1}^{\infty} \frac{1}{k} \int_{S_k > 0} (\xi_1 + \cdots + \xi_k)\, dP \\ &= \sum_{k=1}^{\infty} \int_{S_k > 0} \xi_1\, dP \leqq \sum_{k=1}^{\infty} \int_0^{\infty} r\, d_r P\{\xi_1 < r, \xi_1 + \cdots + \xi_k > 0\}. \end{aligned}$$

Since

$$\begin{aligned} &P\{r' < \xi_1 < r, \xi_2 + \cdots + \xi_k > -\xi_1\} \\ &\qquad \leqq P\{r' < \xi_1 < r, \xi_2 + \cdots + \xi_k > -r\} \leqq P\{r' < \xi_1 < r\}, \end{aligned}$$

we have (writing $\xi_k' = \xi_k - \mu/2$, $S_k' = S_k - k\mu/2$, $h = k - 1$)

$$\begin{aligned} E \max_{k \geqq 0} S_k \leqq \sum_{k=1}^{\infty} \int_0^{(-h\mu/2)} & rP\left\{\xi_2 + \cdots + \xi_k > \frac{h\mu}{2}\right\} d_r P\{\xi_1 < r\} \\ &+ \int_{(-h\mu/2)}^{\infty} r\, d_r P\{\xi_1 < r\} \end{aligned}$$

$$\leqq \sum_{k=1}^{\infty} E\xi_1^+ P\{S'_{k-1} > 0\} + \int_0^{\infty} \sum_{h=0}^{[-2r/\mu+1]} r\, d_r P\{\xi_1 < r\}$$

$$\leqq E\xi_1^+ \sum_{k=1}^{\infty} P\{S'_{k-1} > 0\} + \int_0^{\infty} r(2 - 2r/\mu)\, d_r P\{\xi_1 < r\},$$

and this is finite by (A.1a) and the finiteness of $E\xi_1^+$ and $E\xi_1^2$.

THEOREM C. *Let N_t be the first k such that $S_k < -t$ $(t > 0)$. Then, for $G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)$,*

$$E_G N_t \leqq -t/\mu(G) + \chi(t, G)$$

where

$$\sup_{G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)} \chi(t, G) = o(t)$$

as $t \to +\infty$.

PROOF. Let $\{\epsilon_t\}$ be a sequence of positive numbers which will go to zero in a way to be chosen below. Then

$$EN_t = \sum_{k=0}^{\infty} P\{N_t > k\} = \sum_{k=0}^{\infty} P\{\min_{j \leqq k} S_j > -t\} \leqq \sum_{k=0}^{\infty} P\{S_k > -t\}$$

$$= \sum_{k \leqq (1+\epsilon_t)(-t/\mu)} P\{S_k > -t\} + \sum_{k > (1+\epsilon_t)(-t/\mu)} P\{S_k > -t\}$$

$$\leqq (1 + \epsilon_t) \frac{t}{-\mu} + \sum_{k > (1+\epsilon_t)(-t/\mu)} P\left\{S_k - \frac{k\mu}{1 + \epsilon_t} > \frac{-k\mu}{1 + \epsilon_t} - t\right\}$$

$$\leqq \frac{t}{-\mu} + \frac{\epsilon_t t}{-\mu} + \sum_{k=0}^{\infty} P\{S'_k(t) > 0\}$$

where $S'_k(t)$ is the kth partial sum of independent and identically distributed random variables with mean $\epsilon_t \mu(G)/(1 + \epsilon_t)$ and variance $\sigma^2(G)$ when G is the distribution of S_1. By (A.1b),

$$\sup_{G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)} \sum_{k=0}^{\infty} P\{S'_k(t) > 0\} = B(t) \text{ (say) } < \infty.$$

Choose ϵ_t to go to zero so slowly that $B(t) = o(t)$. Since $\mu(G) \leqq -\mu_0$ for $G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)$, we have

$$-\epsilon_t t/\mu + \sum_{k=0}^{\infty} P\{S'_k(t) > 0\} = o(t)$$

uniformly for $G \varepsilon \mathcal{G}^*(\mu_0, \rho_0)$, and thus Theorem C is proved.

THEOREM D. *Let ν be the last k such that $S_k > 0$. Then, for any $G \varepsilon \mathcal{G}(\mu_0, \sigma_0)$,*

$$E_G \nu \leqq (2/\mu_0) E_G[\max_{k \geqq 0} S_k] + \sum_{k=1}^{\infty} P_G\{S_k + k\mu_0/2 > 0\}.$$

PROOF. Suppressing the dependence on G we have

$$\begin{aligned} E\nu &= \sum_{k=1}^{\infty} P\{\nu \leqq k\} = \sum_{k=1}^{\infty} P\{\max_{j \geqq k} S_j > 0\} \\ &= \sum_{k=1}^{\infty} P\{\max_{j \geqq k} (S_j - S_k) + S_k > 0\} = \sum_{k=1}^{\infty} P\{M + S_k > 0\} \end{aligned}$$

where M has the distribution of $\max_{j \geqq 0} S_j$ and is independent of $\{S_k\}$. Consequently, denoting by $[r]$ the greatest integer $\leqq r$,

$$\begin{aligned} \sum_{k=1}^{\infty} P\{M + S_k > 0\} &= \sum_{k=1}^{\infty} \int P\{S_k > -m\} \, d_m P\{M \leqq m\} \\ &= \sum_{k=1}^{[2m/\mu_0]} \int P\{S_k > -m\} \, d_m P\{M \leqq m\} \\ &\quad + \sum_{k=[1+2m/\mu_0]} \int P\left\{S_k + \frac{\mu_0 k}{2} > \frac{\mu_0 k}{2} - m\right\} d_m P\{M \leqq m\} \\ &\leqq \frac{2}{\mu_0} EM + \sum_{k=1}^{\infty} P\left\{S_k + \frac{\mu_0 k}{2} > 0\right\} \end{aligned}$$

and Theorem D is proved.

REFERENCES

ALBERT, A. E. (1961). The sequential design of experiments for infinitely many states of nature. *Ann. Math. Statist.* **32** 774–799.

ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

BESSLER, S. A. (1960). Theory and applications of the sequential design of experiments, k-actions, and infinitely many experiments. Technical Report no. 55, Department of Statistics, Stanford Univ.

CHERNOFF, H. (1959). Sequential design of experiments. *Ann. Math. Statist.* **30** 755–770.

DONNELLY, T. G. (1957). A family of truncated sequential tests. Doctoral dissertation, Univ. of North Carolina.

ERDÖS, P. (1949). On a theorem of Hsu and Robbins. *Ann. Math. Statist.* **20** 286–291.

HOEFFDING, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368.

KIEFER, J. and WEISS, L. (1957). Some properties of generalized sequential probability ratio tests. *Ann. Math. Statist.* **28** 57–75.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

LECAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.

LEHMANN, E. (1957a, 1957b). A theory of some multiple decision problems. I and II. *Ann. Math. Statist.* **28** 1–25 and 547–572.

SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

SOBEL, M. and WALD, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Statist.* **20** 502–522.

SPITZER, F. (1956). A combinatorial lemma and its application to probability theory. *Trans. Amer. Math. Soc.* **82** 332–339.

WALD, A. (1947). *Sequential Analysis*. Wiley, New York.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York

WALD, A. (1951). Asymptotic minimax solutions of sequential point estimation problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 1–11. Univ. of California Press.