ASYMPTOTIC SHAPES OF BAYES SEQUENTIAL TESTING REGIONS[1]

By Gideon Schwarz

Columbia University and Stanford University[2]

Summary. The large-sample limiting shapes of the Bayes sequential testing regions of composite hypotheses are found explicitly. The result obtained is related to the Sequential Probability Ratio Test in the same way that the likelihood ratio statistic for composite hypotheses is related to the Neyman-Pearson test for simple hypotheses.

1. Introduction. For many sequential decision problems the optimal solutions have been fully characterized, but the characterization is far from being explicit and often cannot be utilized in practical applications even with the aid of high speed computers. On the other hand, heuristic approaches and simplifying assumptions have led to procedures that are easy to apply and sometimes easy to evaluate, but there is no reason to believe that those procedures are optimal in any sense of the word. It is possible to bridge the gap between the extremes by developing an asymptotic theory.

In Wald's work [16] there are some approximations which are asymptotically valid for large samples; a more systematic study of asymptotic questions was done by Chernoff [5], [6] and Anscombe [2]. As sample size is not one of the given parameters of a sequential problem, large sample theory has to be defined in terms of a different parameter, and Chernoff defined large sample theory as the study of sequential problems when the cost of an observation approaches zero. The results obtained by Chernoff are mainly concerned with certain heuristically plausible procedures and include a study of the asymptotic performance of those procedures. In this paper we also use the cost parameter to define large sample theory, but we deal with the exactly optimal procedures; and, though we cannot describe them explicitly, their optimality properties enable us to obtain explicit asymptotic formulae for them.

The nature of these asymptotic formulae is best explained in geometric terms. The main difficulty in the problems we shall consider is that of obtaining an optimal sampling plan. The choice of terminal decisions is relatively easy. Geometrically, a sampling plan can be represented by a region in the space of all (n, S_n), where n is the ordinal number of the observation, and S_n is a statistic which sums up the relevant information obtained through the n first observations. Our results are based on the fact that, for a fixed problem with cost of an observation equal to c, the optimal region grows with diminishing c in such a

Received November 18, 1960; revised June 26, 1961.

[1] This research was supported in part by the Office of Naval Research under Contract Number Nonr-266(59), Project Number 042-205. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Present address: Hebrew University.

manner that its shape approaches a limiting shape while its size goes to infinity in all directions. It is that limiting shape for which we obtain explicit formulae.

The problems we treat can also be described from a point of view that is more oriented towards applications. In fixed sample size theory, the practical problem of testing whether a parameter of a distribution is larger than some preassigned number is most easily solved when it is formalized as a problem of a simple hypothesis versus a simple alternative. This is justified for two reasons. First, in practice it becomes unimportant which action is finally chosen when the parameter lies in some intermediate "indifference region." Second, certain monotonicity properties of the procedures assure a performance at least as good as achieved at the two theoretically assumed parameter values, when the true value of the parameter lies outside those two values.

The problem of testing a simple hypothesis versus a simple alternative was also the first one to be treated in sequential theory, and the Bayes solution of the problem is given by Wald's Sequential Probability Ratio Test. However, the justifications for formalizing the practical problem in this manner do not apply to the sequential case. It still holds true that an operating characteristic of any plausible procedure which yields acceptable error probabilities at the two chosen parameter values controls the error probabilities adequately in the practical problem, but the performance of a sequential procedure cannot be judged by the operating characteristic alone: the expected sample size must be taken into account as well. It has been pointed out (Kiefer and Weiss [11]) that the performance of the Sequential Probability Ratio Test becomes quite unsatisfactory when the true parameter value lies between the hypotheses. Combining this fact with Chernoff's result [6], [7] that the relative contribution of sampling costs to the total risk of an optimal sequential procedure tends to 100 per cent as c approaches zero, we must come to the conclusion that for a large-sample theory the restriction to simple hypotheses is not permissible.

Thus one is led to consider problems with more than two possible parameter values. A possible approach has been suggested by Kiefer and Weiss [11] who obtained some properties of optimal solutions. The *ad hoc* "straight line procedures" of Anderson [1], Armitage [3], and Donnelly [8] are also aimed at this problem. Basically, Kiefer and Weiss suggest incorporating an indifference region by having a region where there is no loss for wrong decisions, but its parameter points are possible values, and the sample size for those values must therefore be considered. The existence of an indifference region can be incorporated in the formal framework through an appropriate loss function. The existence of an indifference region turns out to be essential for the success of our method, and we do not know whether the results obtained remain valid when no indifference region exists.

Another requirement for applicability of our approach concerns the existence of a sufficient statistic whose dimension does not increase with the number of observations. In the present paper we treat only the case where S_n, the sum of the observation, is a real sufficient statistic. The parameter space is in that case a real line or interval, and we restrict our considerations to testing one-sided hypotheses about this parameter.

In a forthcoming paper the results will be generalized to other types of statistics and to different kinds of hypotheses.

2. Bounds for Bayes regions. The observations are assumed to be independent and identically distributed according to a probability law that belongs to a one-parameter family of the Koopman-Darmois type [12]. The defining property of the Koopman-Darmois type is the existence of a cumulative sufficient statistic, that is, a sufficient statistic for a single observation which, when summed over n observations yields a sufficient statistic for n observations. We denote the statistic for the ith observation by X_i and, as there is no danger of confusion, we shall refer to X_i itself as the ith observation.

According to Koopman [12] there exists a measure F dominating the family, and a real function $b(\theta)$ of the parameter θ, such that the probability density functions of X_i relative to the dominating measure can be written

$$f(x, \theta) = \exp [\theta x - b(\theta)]. \tag{1}$$

Similarly, the sufficient statistic $S_n = \sum_{i=1}^{n} X_i$ has a density function

$$f_n(s, \theta) = \exp [\theta s - nb(\theta)]. \tag{2}$$

To see the significance of $b(\theta)$, we consider the moment generating function

$$E_\theta(e^{tX}) = \int e^{(t+\theta)x-b(\theta)}\, dF = e^{-b(\theta)} \int e^{(t+\theta)x}\, dF; \tag{3}$$

on the other hand we have

$$E_{t+\theta}(1) = 1 = \int e^{(t+\theta)x-b(t+\theta)}\, dF = e^{-b(t+\theta)} \int e^{(t+\theta)x}\, dF. \tag{4}$$

Dividing (3) by (4) we see that the moment generating function exists whenever t is such that $t + \theta$ is in the domain of definition of $b(\theta)$, and is given by

$$E_\theta(e^{tX}) = \exp [b(\theta + t) - b(\theta)]. \tag{5}$$

Existence of the moment generating function implies existence of all moments, as well as all semiinvariants. We may therefore differentiate the logarithm of (5) k times at $t = 0$, and obtain $d^k b/d\theta^k$ for the kth semiinvariant; in particular, $\mu(\theta) = E_\theta(X) = b'(\theta)$ and $\text{Var}_\theta(X) = b''(\theta)$. Hence, $b''(\theta)$ is non-negative. If for some θ, $b''(\theta)$ equals zero, the distribution of X for that θ is degenerate. The logarithm of its moment generating function is linear, and by (5), $b(\theta)$ is linear as well, and $b''(\theta)$ is equal to zero identically.

Ruling out this degenerate case, we have $b''(\theta)$ positive, $b(\theta)$ strictly convex, and $E_\theta(X) = b'(\theta)$ is a monotone increasing function, whose inverse function we denote by $\theta(\mu)$.

The parameter space Ω is the domain of existence of $b(\theta)$. As $b(\theta)$ can fail to exist only by becoming infinite, its convexity implies that the domain consists of an interval, possibly unbounded on either side. We now define a null hypothesis $H_0 = \{\theta \mid \theta \leqq m_0\}$ and an alternative hypothesis $H_1 = \{\theta \mid \theta \geqq m_1\}$, where m_0 and m_1 lie in the parameter space, and $m_0 < m_1$. The loss structure is defined by a non-negative bounded measurable function $l(\theta)$, which is the penalty for making a wrong decision when θ is the true parameter value. We assume the utility unit chosen so that $l(\theta)$ is bounded above by one. The open interval (m_0, m_1) will be called the "indifference region", and throughout that interval $l(\theta)$ is equal to zero. Outside (m_0, m_1), $l(\theta)$ is assumed to be positive. The supremum of $l(\theta)$ over H_i we denote by L_i. The cost of each observation is a constant c, expressed in the same utility unit as $l(\theta)$.

As a "measure of distinguishability" of the hypotheses, we introduce the second difference of $b(\theta)$ for the indifference region,

$$\Delta = b(m_0) + b(m_1) - 2b[\tfrac{1}{2}(m_0 + m_1)]. \tag{6}$$

The number Δ is positive (because $b(\theta)$ is a convex function) and plays a role similar to the Kullback-Leibler [13] information number.

A terminal decision may be made at any stage of sampling.

Finally, we assume a given *a priori* probability distribution W on the parameter space. For any decision procedure $\mathfrak{F}$ the risk $\rho(\mathfrak{F}, W)$ is now a well defined number, and we are interested in the procedure $\mathfrak{B}$, the Bayes procedure, for which this number is minimal. As we may restrict ourselves to procedures based on the sufficient statistic (n, S_n), the procedures can be defined by partitioning the (n, S_n)-plane into three sets: a sampling region, and two stopping regions. Given such a partition, the corresponding procedure consists of sampling until the point (n, S_n) lies for the first time in one of the stopping regions, and deciding "H_0" or "H_1" according to which stopping region it is. Once a Bayes sampling region is known, the division of its complement to obtain the Bayes stopping regions is relatively simple; we therefore formulate our problem as that of finding the Bayes sampling regions in the (n, S_n) plane. Accordingly, the same symbol will be used to denote a procedure and its sampling region.

An implicit characterization of the Bayes sampling region was given by Wald and Wolfowitz in Theorem 3.7 of [18]. An important role in that characterization is played by the "stopping risk" $R(n, S_n)$, which in our case can be defined by

$$R(n, S_n) = \min_{i=0,1} \int_{H_i} e^{\theta S_n - nb(\theta)} l(\theta)\, dW(\theta) \Big/ \int_{\Omega} e^{\theta S_n - nb(\theta)}\, dW(\theta). \tag{7}$$

For every n and S_n the number $R(n, S_n)$ measures the expected loss due to wrong decisions, given that sampling was stopped when the point (n, S_n) was reached, and the terminal decision for which the integral in (7) is smaller was taken. In terms of the function $R(n, S_n)$ a family of regions can be defined by putting for positive r, $C(r) = \{(n, S_n) \mid R(n, S_n) \geqq r\}$.

These regions will be used with appropriate choices of r as bounds for the Bayes region $\mathfrak{B}$.

THEOREM I. *For sufficiently small cost of sampling c,*

$$C(c) \supset \mathcal{B} \supset C(3c\Delta^{-1} \log c^{-1}). \tag{8}$$

PROOF. To prove the left-hand side, we observe that if $(n, S_n) \varepsilon \mathcal{B}$, the Bayes procedure leads to taking at least one more observation, at a cost of c; this could not be the case if stopping at that stage and taking one of the terminal actions would lead to an expected loss smaller than c. Therefore $R(n, S_n)$ is at least c, and $(n, S_n) \varepsilon C(c)$ follows. The other inclusion will be proved if we show that whenever $(n, S_n) \varepsilon C(3\Delta^{-1}c \log c^{-1})$, there exists a procedure that leads to taking at least one more observation, and whose expected loss due to error and due to sampling add up to a number less than $3c\Delta^{-1} \log c^{-1}$. This can be shown to hold for an appropriately defined fixed sample size procedure.

Let us now consider the fixed sample size procedure $\mathfrak{F}_N$ which consists of taking N observations where N is the smallest integer larger than $(2/\Delta) \log c^{-1}$ and deciding as follows: Let $f_{i,N}$ denote the density $\exp(m_i S_N - Nb(m_i))$ and let $\delta = 0$ or 1 according as $L_0 f_{0,N} - L_1 f_{1,N} > 0$ or $\leqq 0$. Then in procedure $\mathfrak{F}_N$ we decide "H_0" if $\delta = 0$ and "H_1" if $\delta = 1$. If $\theta = m_0$ or m_1, the probability of an error is $E(\delta \mid \theta = m_0)$ or $E(1 - \delta \mid \theta = m_1)$ respectively.

$$\delta L_0 f_{0,N} + (1 - \delta) L_1 f_{1,N} = \min(L_0 f_{0,N}, L_1 f_{1,N}) \leqq (L_0 f_{0,N} L_1 f_{1,N})^{\frac{1}{2}}. \tag{9}$$

Hence

$$\begin{aligned} L_0 \int \delta f_{0,N}\, dF^N + L_1 \int (1 - \delta) f_{1,N}\, dF^N &\leqq (L_0 L_1)^{\frac{1}{2}} \\ &\cdot \left[\int (f_{0,1} f_{1,1})^{\frac{1}{2}}\, dF\right]^N = (L_0 L_1)^{\frac{1}{2}} e^{-N\Delta/2} \leqq c. \end{aligned} \tag{10}$$

This bound for the expected loss due to error when the true parameter value is one of the end points of the indifference region holds for all parameter values, as can easily be seen from the monotonicity of Koopman-Darmois distributions: the derivative of the cumulative distribution function of a Koopman-Darmois variable with respect to the parameter is

$$\begin{aligned} \frac{\partial}{\partial\theta} \int_{-\infty}^{a} \exp(\theta x - b(\theta))\, dF &= \int_{-\infty}^{a} (x - \mu(\theta)) \exp(\theta x - b(\theta))\, dF \\ &\text{Prob}\{X < a\} [E(X \mid X < a) - E(X)] \leqq 0. \end{aligned} \tag{11}$$

With X, the sum S_n has a distribution of the Koopman-Darmois type, and accordingly the error probabilities can only decrease when θ moves away from the indifference region.

As the cost of taking N observations is Nc the total expected loss incurred by $\mathfrak{F}_N$ is bounded from above by $c + Nc$, which in turn is bounded by $c(2 + 2\Delta^{-1} \log c^{-1})$. Now, for sufficiently small c, $\log c^{-1}$ will exceed 2Δ, and we have

$$\Delta^{-1} \log c^{-1} > 2 \tag{12}$$

and therefore

$$3\Delta^{-1} \log c^{-1} > 2 + 2\Delta^{-1} \log c^{-1}. \tag{13}$$

Thus continuing according to $\mathfrak{F}_N$ leads to a smaller expected loss than stopping when the risk of stopping is greater than $3c\Delta^{-1} \log c^{-1}$, and the theorem is proved.

3. Asymptotic shapes. In view of Theorem I, the region $C(c)$ can be used as an approximation to the Bayes sampling region when c is small compared to unity. In fact, using $C(c)$ would yield an operating characteristic uniformly better or at least as good as that of the Bayes region, provided that after stopping, the best action is always taken. On the other hand, using $C(c)$ leads to oversampling. For practical applications, the use of $C(c)$ has a more serious disadvantage. In order to plot the boundary of $C(c)$ in the (n, S_n) plane, the *a priori* distribution W has to be given, and moreover, there is no explicit formula for the boundary, and the computational labour involved is immense. The use of $C(c)$ can therefore not be suggested as a solution to the problem of finding an explicit, easily applicable procedure.

It is here that the concept of "asymptotic shape" comes to our aid. As the region $C(c)$ approximates the Bayes sampling region for c approaching zero, the problem would be solved if $C(c)$ could in turn be approximated for c approaching zero by some explicitly given region. If, for instance, $C(c)$ would tend to a finite limiting region, the limiting region could be used as an approximation. However, considering the unboundedly increasing sample sizes that one expects heuristically when the cost of sampling approaches zero, one cannot hope for $C(c)$ to have such a finite limit. On the other hand, we shall see that $C(c)$ grows with diminishing c in a particular way, namely, its shape approaches a definite limit. By "shape" we mean all that is invariant under homothetic transformations, and as a formal definition of the shape of a bounded region we can define that homothetic transform of the region whose boundary intersects the horizontal axis at its unit point, and at no point to the right of the unit point. By "convergence" of shapes we mean pointwise convergence of their boundaries. If the shapes of a parametrized family of regions converge to a limiting shape when the parameter approaches a limit, that limiting shape is called "the asymptotic shape of the family."

Before we state and prove a theorem about the existence of an explicitly describable asymptotic shape for the family $C(c)$, we define two functions in the (n, S_n)-plane, in terms of which the asymptotic shapes will be given. Those functions are the likelihood ratio statistics $\lambda_0(n, S_n)$ and $\lambda_1(n, S_n)$ whose occurrence in large sample sequential theory dates back to Wald [19].

In the fixed sample size theory of testing hypotheses there is usually only one likelihood ratio statistic defined. The fact that in our case the parameter space includes, in addition to H_0 and H_1, also the indifference region, enables us to define two such statistics, as follows

$$\lambda_i(n, S_n) = \frac{\sup_{\theta \varepsilon H_i}(\text{mod } W)(e^{\theta S_n - nb(\theta)})}{\sup_{\theta \varepsilon \Omega}(\text{mod } W)(e^{\theta S_n - nb(\theta)})}, \qquad i = 0, 1, \tag{14}$$

where sup (mod W) denotes the essential supremum relative to W.

Unlike $R(n, S_n)$, the λ_i do not depend on the *a priori* distribution W except through its null sets. In analogy to the definition of $C(r)$ by $R(n, S_n)$ we now define a family of regions $\Lambda(r)$, as follows:

(15) $$\Lambda(r) = \{(n, S_n) \mid \lambda_i(n, S_n) \geqq r, \quad i = 1, 2\}.$$

From the definition of the λ_i it is easily seen that their logarithms are homogeneous of order one, that is, for arbitrary $\alpha > 0$

(16) $$\log \lambda_i(\alpha n, \alpha S_n) = \alpha \log \lambda_i(n, S_n).$$

In view of this equation, $\Lambda(r)$ for arbitrary r is a homothetic transform of $\Lambda(r_0)$ for any fixed r_0. Choosing $r_0 = 1/e$ and denoting the operation of transforming homothetically simply by multiplication by the factor involved, we have

(17) $$\Lambda(r) = \mathcal{B}_0 \log r^{-1},$$

where

(18) $$\mathcal{B}_0 = \Lambda(1/e) = \{(n, S_n) \mid \log \lambda_i(n, S_n) \geqq -1, \quad i = 0, 1\}.$$

Clearly all the regions $\Lambda(r)$ have the same shape, which is therefore also their asymptotic shape for r approaching zero.

We are going to show that the same asymptotic shape is also that of the family $C(r)$, and finally, that of the Bayes regions $\mathcal{B}(c)$.

First, to study $\mathcal{B}_0$ geometrically, we observe that it is the intersection of two sets Λ_0 and Λ_1 defined by

(19) $$\Lambda_i = \{(n, S_n) \mid \log \lambda_i \geqq -1\}, \qquad i = 0, 1.$$

Substituting the definition of λ_i for, say, $i = 0$, we obtain

(20) $$\Lambda_0 = \{(n, S_n) \mid \sup_{\theta\varepsilon\Omega}(\text{mod } W)(\theta S_n - nb(\theta)) \leqq 1 + \sup_{\theta\varepsilon H_0}(\text{mod } W)(\theta S_n - nb(\theta))\}$$

This can be simplified further, if we take into account the special form of the function under the supremum signs. First we divide the inequality in (20) by n, and introducing the sample mean $\bar{X} = S_n/n$, we obtain

(21) $$\Lambda_0 = \{(n, S_n) \mid \sup_{\theta\varepsilon\Omega}(\text{mod } W)(\theta\bar{X} - b(\theta)) \leqq n^{-1} + \sup_{\theta\varepsilon H_0}(\text{mod } W)(\theta\bar{X} - b(\theta))\}.$$

Let us now fix $\bar{X}$. Geometrically, this amounts to regarding the intersection of Λ_0 with a line of slope $\bar{X}$ through the origin. Now $\theta\bar{X} - b(\theta)$ is a downward-convex function of θ, and for a value $\bar{X}$ in the range of $\mu(\theta) = b'(\theta)$ it attains a single maximum at $\theta(\bar{X})$, the solution of $\bar{X} - b'(\theta) = 0$. As $\theta(\mu)$ is monotone increasing, the location of the maximum moves to the left or to the right with $\bar{X}$. For values of $\bar{X}$ such that $\theta(\bar{X}) \varepsilon H_0$, the maximum is obtained in H_0, and consequently the essential suprema on both sides of (21) agree. For such an $\bar{X}$, any positive n will make the inequality hold, and the entire ray of slope $\bar{X}$ will be contained in Λ_0. If we now move $\bar{X}$ to the right, the essential supremum in H_0

can only decrease, and the essential supremum in the complement $\Omega - H_0$ can only increase. For some $k_0 \geqq m_0$ the supremum in Ω will exceed the supremum in H_0 whenever $\bar{X} > k_0$.

For those values of $\bar{X}$, the intersection of Λ_0 with the line of slope $\bar{X}$ through the origin will consist of those points on that line whose n-coordinates fulfill the inequality in (21).

The value k_0 can be obtained when H_0 and W are given. Consider the points $\theta_0 = \sup(\text{mod } W)H_0$ and $\theta^0 = \inf(\text{mod } W)(\Omega - H_0)$. When $\bar{X} = k_0$, the maximum of $\theta\bar{X} - b(\theta)$ is obtained between θ_0 and θ^0, and the essential suprema in H_0 and $\Omega - H_0$ are obtained at θ_0 and θ^0. They are equal to each other; therefore $k_0 = [b(\theta^0) - b(\theta_0)]/[\theta^0 - \theta_0]$.

We can now simplify the description of the boundary of Λ_0 by restricting ourselves to values of $\bar{X} > k_0$. For those values the essential supremum in H_0 is attained at $\theta = \theta_0$, and therefore

$$(22)\quad \text{Bdry}\Lambda_0 = \{(n, S_n) \mid \sup_{\theta\epsilon\Omega}(\text{mod } W)(\theta S_n - nb(\theta)) = 1 + \theta_0 S_n - nb(\theta_0)\}.$$

Replacing the index 0 by 1, similar results for Λ_1 are readily obtained. It is easily seen that $k_0 \leqq k_1$, and the equality sign holds only if $W(H_0 \cup H_1) = 1$, that is, no weight is given to the indifference region. Consequently, the region $\mathcal{B}_0 = \Lambda_0 \cap \Lambda_1$ intersects every line through the origin whose slope lies in the range of $b'(\theta)$ at a finite interval, except possibly the line $S = k_0 n = k_1 n$. A further consequence of (23) is that Λ_0, and hence also Λ_1 and $\mathcal{B}_0$, are convex regions.

In the statement of Theorem II we shall use a notation that expresses simultaneously the asymptotic shape and the rate of growth of a family of regions. If $Q(r)$ is a parametrized family of regions such that for a real function $q(r)$ the homothetic transform $[1/q(r)]Q(r)$ approaches a region Q_0, we write $Q(r) = Q_0 q(r) + o(q(r))$.

A further remark is necessary if the regions $Q(r)$ consist only of points whose n-coordinate is an integer. As homothetic transformations do not preserve this property, such a family of regions must be represented as an intersection at the set of all points with integral n with a family of regions of the form $Q_0 q(r) + o(q(r))$.

THEOREM II. *Denote by* $\mathcal{I}$ *the set of all* (n, S_n) *such that n is a positive integer, and, by* $\mathcal{B}_0$, *the set*

$$(24)\quad \{(n, S_n) \mid 1 + \min_{i=0,1}(\theta_i S_n - nb(\theta_i)) \geqq \sup_\theta(\text{mod } W)(\theta S_n - nb(\theta))\}.$$

Then

$$C(r) = (\mathcal{B}_0 \log r^{-1} + o(\log r^{-1})) \cap \mathcal{I},$$

when r approaches zero.

PROOF. The *a posteriori* risk of deciding H_1 without any further observations is given by

$$(25)\qquad R = \int_{H_0} \exp(\theta S_n - nb(\theta))l(\theta)\, dW \Big/ \int_\Omega \exp(\theta S_n - nb(\theta))\, dW.$$

For fixed $R = r$, (25) represents a curve in the (n, S_n) plane, and we can find the intersection of the curve with a line of slope k through the origin by substituting $S_n = kn$. We now obtain for fixed r and k, an equation for the n-coordinate of the intersection

$$(26) \qquad r = \int_{H_0} \exp\left[(\theta k - b(\theta))n l(\theta)\; dW \right/ \int_{\Omega} \exp[(\theta k - b(\theta))n]\; dW.$$

The integral in the denominator is simply the nth power of the L_n-norm of $\exp(\theta k - b(\theta))$ relative to the measure W, and the numerator is the nth power of the L_n-norm of the same function restricted to H_0 relative to the measure $\int l(\theta) dW$. Denoting the two norms by $\|g\|_n$ and $\|h\|_n$, we have

$$(27) \qquad r^{1/n} = \frac{\| g \|_n}{\| h \|_n}.$$

Now this fraction is bounded away from zero for all $n \geqq 1$. This is true because the denominator is bounded from above by $\max \exp(\theta k - b(\theta))$, which is finite because $b(\theta) - \theta k$ is convex, and the numerator is bounded away from zero by $\min \exp(\theta k - b(\theta))$ over any finite interval, times the $\int l(\theta) dW$ measure of that interval. However, if the right hand side of (27) is bounded away from zero, so is the left hand side, and $r^{1/n} >$ constant implies $\lim_{r\to 0} n = \infty$. We now put $n = t \log r^{-1}$, and obtain an equation for t

$$(28) \qquad e^{-1/t} = \frac{\| g \|_n}{\| h \|_n}.$$

Now we pass to the limit as $r \to 0$, and make use of the limiting property of L_n norms ([14], p. 160)

$$(29) \qquad \lim_{n\to\infty} \|g\|_n = \sup(\mathrm{mod}\, W)\; |g|$$

to obtain for $\tau = \lim_{r\to 0} t$

$$(30) \qquad e^{-\tau^{-1}} = \frac{\sup_{H_0}\left(\mathrm{mod} \int l\, dW\right) \exp\,(\theta k - b(\theta))}{\sup_{\Omega}\,(\mathrm{mod}\, W) \exp\,(\theta k - b(\theta))};$$

here we can replace $\int l dW$ by just W, which has the same null sets, and solve for τ:

$$(31) \quad \tau = [\sup_{\Omega}\,(\mathrm{mod}\, W)(\theta k - b(\theta)) - \sup_{H_0}\,(\mathrm{mod}\, W)(\theta k - b(\theta))]^{-1}.$$

Clearly, this curve is identical to (22), and by interchanging H_0 and H_1 a similar result is obtained for the region where the expected loss of stopping and deciding H_0 exceeds r. The intersection of the two regions is by definition $C(r)$, and Theorem II is proved.

The main theorem of the paper can now be proved by applying both Theorems I and II.

THEOREM III. *Let* $\mathcal{B}(c, W)$ *be the Bayes sampling region for* W *as the a priori distribution and* c *be the cost of one observation. Then, with* $\mathcal{B}_0$ *defined as in Theorem II,*

$$\mathcal{B}(c, W) = [\mathcal{B}_0 \log c^{-1} + o(\log c^{-1})] \cap \mathcal{J}.$$

PROOF. Applying Theorem II to $r = c$ yields

$$C(c) = [\mathcal{B}_0 \log c^{-1} + o(\log c^{-1})] \cap \mathcal{J}. \tag{32}$$

Applying it to $r = (3c/\Delta) \log c^{-1}$ yields

$$\begin{aligned} C((3c/\Delta) &\log c^{-1} \\ &= [\mathcal{B}_0 \log ((3c/\Delta) \log c^{-1})^{-1} + o (\log ((3c/\Delta) \log c^{-1})^{-1})] \cap \mathcal{J} \\ &= [\mathcal{B}_0 \log c^{-1} - \mathcal{B}_0 \log \log c^{-1} + \mathcal{B}_0 \log (\Delta/3) \\ &\qquad + o(\log ((3c/\Delta) \log c^{-1})^{-1})] \cap \mathcal{J}, \end{aligned} \tag{33}$$

but as all the terms except the first one are $o(\log c^{-1})$, we can collect them into one term, and write

$$C((3c/\Delta) \log c^{-1}) = [\mathcal{B}_0 \log c^{-1} + o(\log c^{-1})] \cap \mathcal{J} \tag{34}$$

Thus $C(c)$ and $C((3c/\Delta) \log c^{-1})$ have the same asymptotic representation. By Theorem I, this representation is also shared by the Bayes region $\mathcal{B}(c, W)$.

In view of this result, the family of regions $\mathcal{B}_0 \log c^{-1}$ is a large sample approximation to the Bayes regions, and there is an explicit formula for $\mathcal{B}_0 \log c^{-1}$. Through the essential suprema, $\mathcal{B}_0 \log c^{-1}$ still depends on W, but this dependence is of a special nature, and it can be removed by treating various classes of distributions W. This will be done in the next section.

4. Completely mixed a priori distributions. As long as we hold up the level of generality for which Theorems I, II, and III were proved, there is no way of simplifying the definition of the region $\mathcal{B}_0$ as given in the statement of Theorem II. In this chapter, we treat special cases of the general theory, and study the geometric nature of the regions arising.

The first possibility of specialization stems from the dependence of the regions on the *a priori* distribution W. The definition of the essential supremum involves only the sets of measure zero. Hence, the region $\mathcal{B}_0$ depends only on the "kernel" of W, that is, on the collection of its sets of measure zero, and we can study the implications of various assumptions on the kernel.

An important class of *a priori* measures is the class of "completely mixed measures". Those are the measures that dominate the Lebesgue measure, that is, they give positive weight to every set of positive Lebesgue measure. This property can be expressed in terms of the kernel; it is equivalent to the condition that the kernel of the Lebesgue measure include the kernels of all the measures in the class.

The importance of this class of measures lies in the fact that any procedure that is Bayes against a measure in that class is essentially admissible, in the sense that no procedure can have a risk function that is as good for almost all θ, and better for a set of θ of positive Lebesgue measure.

As we are concerned here mostly with geometrical properties of the region $\mathcal{B}_0$, we rename the coordinates in the (n, S_n)-plane and denote them by x and y instead of n and S_n respectively.

THEOREM IV. *When W is a measure that dominates the Lebesgue measure on the θ-line, the boundary of $\mathcal{B}_0$ consists of two curves, whose equations are*

$$\begin{aligned} xb(\theta(y/x)) - xb(\theta_0) - y\theta(y/x) + y\theta_0 + 1 &= 0, \\ xb(\theta(y/x)) - xb(\theta_1) - y\theta(y/x) + y\theta_1 + 1 &= 0. \end{aligned} \tag{35}$$

PROOF. The essential supremum of a bounded continuous function *modulo* the Lebesgue measure is simply the maximum of the functions.

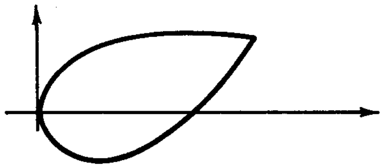
The function $y\theta - xb(\theta)$ obtains its maximum at the solution of

$$y - xb'(\theta) = 0, \tag{36}$$

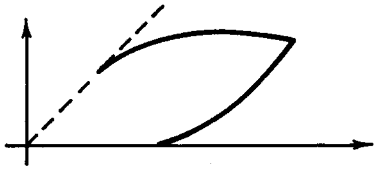
and that solution is $\theta = \theta(y/x)$, where again $\theta(k)$ is the inverse function of $b'(\theta)$. Now, when $\sup\,(\text{mod } W)(y\theta - xb(\theta))$ is replaced by $y\theta(y/x) - xb(\theta(y/x))$, (43) is readily obtained.

Acknowledgments. I wish to thank Professor T. W. Anderson, who was my advisor when this thesis was written, for the help and encouragement I received from him. Also, I would like to thank Professor Herman Chernoff, who suggested to me the area of large-sample sequential testing as a research topic during my stay at Stanford University in the summer of 1959, and was very helpful in more than one way while this work was being done.

Examples of Asymptotic Shapes, when W is completely mixed

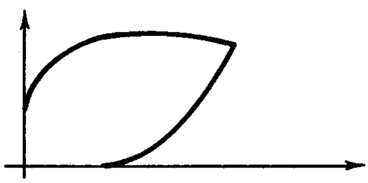


Normal Distribution

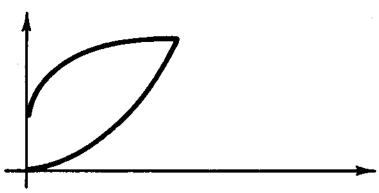


Bernoulli Distribution

Poisson Distribution



Geometrical Distribution



Exponential Distribution

Example	Parameter	$b(\theta)$	Boundary of $\mathcal{B}_0$
Normal $(\mu, 1)$	$\theta = \mu$	$\frac{1}{2}\theta^2$	$y = \theta_i x + (2x)^{\frac{1}{2}}(-1)^i$
Bernoulli (p)	$\theta = \log \frac{p}{1-p}$	$\log(1 + e^\theta)$	$(x - y)\log(x - y) = 1 + x\log[x/(1 + e^{\theta_i})] - y\log(y/e^{\theta_i})$
Poisson (λ)	$\theta = \log \lambda$	e^θ	$y\log(y/x) = 1 - xe^{\theta_i} + y(1 + \theta_i)$
Geometrical (μ)	$\theta = \log \frac{\mu}{1+\mu}$	$-\log(1 - e^\theta)$	$(x + y)\log(x + y) = -1 + x\log[x/(1 - e^{\theta_i})] + y\log(y/e^{\theta_i})$
Exponential (μ)	$\theta = -1/\mu$	$-\log(-\theta)$	$x\log(-x/e^{\theta_i}) = 1 + x\log y + \theta_i y$

REFERENCES

[1] T. W. ANDERSON, "A modification of the sequential probability ratio test to reduce the sample size," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 165–197.

[2] F. J. ANSCOMBE, "Large sample theory of sequential estimation," *Biometrika*, Vol. 36 (1949), pp. 455–458.

[3] P. ARMITAGE, "Restricted sequential procedures," *Biometrika*, Vol. 44 (1957), pp. 9–26.

[4] K. J. ARROW, D. BLACKWELL, AND M. A. GIRSHICK, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213–244.

[5] H. CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

[6] H. CHERNOFF, "Large-sample theory; parametric case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1–22.

[7] H. CHERNOFF, "Sequential design of experiments," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 755–770.

[8] T. G. DONNELLY, "A family of truncated sequential tests," doctoral dissertation, University of North Carolina, 1957.

[9] A. DVORETZKY, J. KIEFER, AND J. WOLFOWITZ, "Sequential decision problems for processes with continuous time parameter — testing hypotheses," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 254–264.

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

[11] J. KIEFER AND LIONEL WEISS, "Some properties of generalized sequential probability ratio tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 57–75.

[12] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, Vol. 39 (1939), pp. 399–415.

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

[14] M. LOÈVE, *Probability Theory*, D. Van Nostrand, New York, 1955.

[15] A. WALD, *Sequential Analysis*, John Wiley and Sons, New York, 1947.

[16] A. WALD, "Sequential tests of statistical hypotheses," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 117–186.

[17] A. WALD AND J. WOLFOWITZ, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–339.

[18] A. WALD AND J. WOLFOWITZ, "Bayes solutions of sequential decision problems," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 82–99.

[19] A. WALD, "Asymptotic minimax solutions of sequential point estimation problems," pp. 1–12, *Proceedings of the Second Berkeley Symposium on Probability and Statistics*, University of California Press, 1951.