Asymptotically Pointwise Optimal and Asymptotically Optimal Stopping Rules for Sequential Bayes Confidence Interval Estimation*

by

Sudhakar Kunte
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #328

August 1973

*This research was supported by the Air Force Office of Scientific Research under the Grant No. 73-2432. The United States Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

In Chapter 2, we further consider the sequential version of this decision problem by taking the loss function

$$L(\theta,I,n) = L(\theta,I) + cn,$$

where $c > 0$ represents the cost per observation. In this case we show that for any stopping rule the sequence of Bayes terminal decision rules is given by $\{I_n^*\}$.

Let $\{Y_n\}$ represent the sequence of posterior Bayes risks for $\{I_n^*\}$. In the last part of Chapter 2, we consider the example of the normal distribution with unknown mean μ and known variance. Using the conjugate prior for μ, we show that the sequence of $\{Y_n\}$ behaves in such a way, that we can use the theory developed in Chapter 1 and obtain a family $\{t(c):c > 0\}$ of A.P.O. and A.O. stopping rules for this problem. As expected in this case, for each $c, t(c)$ yields a fixed sample size.

In Chapter 3, we consider the example of finding a confidence interval for the mean μ of a normal distribution where the variance σ^2 is also unknown. Using the conjugate prior for μ and σ^2, we again show that the sequence of posterior risks $\{Y_n\}$ behaves in such a way that we can use the results of Chapter 1. In the example of this chapter and in Chapter 2, the f that works to give the A.P.O. and A.O. stopping rules $\{t(c):c > 0\}$ defined by (2), is given by $f(x) = \{x/\log x\}^{1/2}$ for $x \geqslant 3$.

Motivated by the two normal distribution examples, in Chapter 4, we consider the general Bayes confidence interval estimation problem, where $f(x|\theta)$ and ψ are assumed to have any general functional form. Under certain regularity assumptions on f and ψ, we show that the sequence $\{I_n^*\}$ defined in (4) is the sequence of terminal Bayes confidence intervals. Further we show that the sequence of posterior Bayes risks $\{Y_n\}$ of $\{I_n^*\}$ behaves in such a way that for this problem the class of stopping rules $\{t(c):c > 0\}$ defined in (2) is A.P.O. Here again the function f which defines $t(c)$ is given by $f(x)=(x/\log x)^{1/2}$, for $x \geq 3$.

for the process $X(n,c)$:

$t(c) = n$, if n is the first integer such that

(2) $(1 - \frac{f(n)}{f(n+1)})Y_n < c.$

If we further assume that $\sup_n E(f(n)Y_n) < \infty$, then we show that the class of stopping rules $\{t(c):c > 0\}$ defined in (2) is also A.O. for the process $X(n,c)$. Further we give an example in which the condition $\sup_n E(f(n)Y_n) < \infty$ is not satisfied and the class of stopping rules $\{t(c):c > 0\}$, even though A.P.O., is not A.O. for the process $X(n,c)$.

Suppose $\{X_n\}$ is a sequence of independent identically distributed random variables on some probability space (Ω,F,P_0), $\theta \varepsilon \Theta$, where Θ is an open sub-interval of the real line. Suppose that P_θ has the density $f(x|\theta)$ with respect to some σ-finite measure μ. Further, suppose ψ is the density of a prior distribution for θ on Θ. In Chapter 1, we give a decision theoretic formulation of the confidence interval estimation problem as follows: Let $\mathcal{I}$ be the class of all sub-intervals of Θ. This is our action space. For $\theta \varepsilon \Theta$ and $I \varepsilon \mathcal{I}$ the loss due to taking the action I when θ is the parameter, is given by

(3) $L(\theta,I) = a\ell(I) + b\{1 - \delta_I(\theta)\}$

where ℓ is the length of I and $\delta_I(\theta)$ is 1 or 0, depending on whether $\theta \varepsilon I$ or not. Under this loss function, if we assume that for every n almost surely the posterior density $\psi(\theta|X_1,\ldots,X_n)$ of θ is strictly unimodal and continuous, then we show that the Bayes rule under the prior ψ is given by $I_n^*=[\alpha_{1n}^*,\alpha_{2n}^*]$ if α_{1n}^* and $\alpha_{2n}^*(> \alpha_{1n}^*)$ are the two solutions of the equation

(4) $\psi(\theta|X_1,\ldots,X_n) = \frac{a}{b}$.

Otherwise I_n^* is a single point set if (4) has only one solution, and I_n^* is the null set if (4) has no solutions.

ABSTRACT

Kunte, Sudhakar. Ph.D. Purdue University. August 1973. Asymptotically Pointwise Optimal and Asymptotically Optimal Stopping Rules for Sequential Bayes Confidence Interval Estimation. Major Professor: Leon J. Gleser.

Let $\{Y_n\}$ be a sequence of positive random variables defined on some probability space (Ω, F, P) and let Y_n be F_n measurable, where $F_1 \subset F_2 \subset \ldots \subset F$. For every $c > 0$ define

(1) $X(n,c) = Y_n + nc.$

$X(n,c)$ is a random process defined on the sequence $\{Y_n\}$ and $c \in (0,\infty)$. Suppose T is the class of all stopping rules defined for this process, in the sense that for every $t \in T$, the event $[t=n]$ is F_n measurable.

Suppose $\{t(c): c > 0\}$ is a class of stopping rules contained in T.

<u>Definition 1</u>: $\{t(c): c > 0\}$ is an asymptotically pointwise optimal (A.P.O.) class of stopping rules for the process $X(n,c)$, if $X(t(c),c)\{\inf_{s \in T} X(s,c)\}^{-1} \to 1$ a.s. P, as $c \to 0$.

<u>Definition 2</u>: $\{t(c): c > 0\}$ is an asymptotically optimal (A.O.) class of stopping rules for the process $X(n,c)$ if $E[X(t(c),c)]\{\inf_{s \in T} E[X(s,c)]\}^{-1} \to 1$, as $c \to 0$.

In Chapter 2 we assume that there exists a positive strictly increasing function $f:[0,\infty) \to [0,\infty)$, such that $f(n)Y_n \to V$, a.s., as $n \to \infty$, where V is a positive constant. Under some other not too restrictive assumptions [see A1.1) to A1.5) of Chapter 1] we show that the class of stopping rules $\{t(c): c > 0\}$ defined below is A.P.O.

INTRODUCTION

The theory of confidence intervals has always been of interest to statisticians, particularly since it is realized that in most typical statistical estimation problems, no matter how good a point estimate may be, it can almost surely never be equal to the true value of the parameter. If possible, one tries to obtain a random interval or, more generally, a random subset in the parameter space, such that with a specified amount of probability one can correctly say that this interval (or the subset) covers the true value of the parameter.

In most cases, every test statistic for testing the hypothesis that the parameter has a particular value yields one such interval. There are also other well-known methods for obtaining the confidence intervals directly. However, a confidence interval constructed in this way may lose much of its practical value on two counts: (1) Its length may be too large; (2) its probability of coverage may be too small. Since a confidence interval which is good on one count may fail to be so on the other count, some kind of a optimality criterion is needed for choosing one interval among the class of all possible intervals.

There are two classical approaches to define such a criterion. In the first approach, it is required that the probability of coverage of the interval be not less than a preassigned value (rather arbitrarily chosen), and among the intervals which satisfy this requirement, we are to choose the one which has smallest expected length. In the second approach, it is required that the length of the interval be not more than a preassigned value (again, rather arbitrarily chosen), and among the intervals which satisfy this requirement, we are to choose the one which has the largest probability of covering the true parameter value. In most cases, it is possible to obtain intervals satisfying the criterion of either approach. The theory of obtaining such intervals is thoroughly discussed by many textbooks, for example, in Lehmann [1959], and more recently in Zacks [1970].

One natural generalization of the criteria named above that comes to mind is simply to demand that simultaneously the length of the interval be not more than a preassigned value and the probability of coverage be not less than a preassigned value. As long as a given confidence interval satisfies both of these restrictions, it seems unnecessary to worry any further about the optimality or uniqueness of the interval. Such an interval after all can not be very bad. However, if the sample size is fixed in advance, no such confidence interval may exist. This fact was probably realized for the first time by Dantzig [1940]. One possible solution to this problem was suggested by Stein [1945], when he suggested the use of a two stage sampling plan. This method achieves the desired goal of obtaining a confidence interval which satisfies the restrictions on length and probability of coverage of the interval at the same time. However, it does so by using a random number of observations, thus raising the question of the cost of sampling. A natural generalization of Stein's two stage procedure is to consider sequential sampling plans for obtaining the confidence intervals with desired properties. Work on such procedures, among others, is done by Stein & Wald [1947], Chow & Robbins [1965], Paulson [1969], and Farrell[1959]. Chow and Robbins also have defined optimality criteria for choosing one among various sampling plans. The aim of all such sampling plans is to achieve the desired goals for the confidence interval while using a minimum of observations.

All these approaches, being essentially classical in nature, still contain the arbitrariness for the choice of numbers that are placed as restrictions on the length and/or the probability of coverage of the interval. This being the case, they are open to all the criticism that can be and has been levied against similar arbitrary choices in the classical approaches to point estimation and hypothesis testing. One way to get around the criticism mentioned above is to consider the confidence interval problem from the decision-theoretic point of view, by defining a suitable loss function.

Such an approach has also been discussed previously by Lehmann [1959], Joshi [1966], Winkler [1972], and Cohen & Strawderman [1973], among others.

Having posed the problem in decision theoretic terms, one can also try to find Bayes solutions to the problem under suitable prior distributions. Hence, to some extent, the nature of the interpretation of the solution changes from that of classical solutions. Using the observed values of the observable random variables, we are trying to set up the limits for the value that the unobservable random variable (the parameter) has already taken. The relation between the value of the unobservable random variable and the values of the observable random variables that is exploited is that the observable random variables have a distribution which has the value of the unobservable random variable as its parameter(s). Thus, knowing the values the observable random variables have taken, we should presumably know more about the distribution from which the value of the parameter has come [posterior analysis]. Hence, under this approach, the confidence interval, instead of being a random interval possibly set up to cover the constant unknown value of the parameter, is an interval expected to cover the random value of the parameter [i.e., a prediction interval].

Such a Bayesian problem can again be considered sequentically. The first attempt at considering sequential Bayesian decision problems was probably made by Arrow, Blackwell &

Girshick [1949]. In their paper, they show that, in most cases of sequential Bayesian problems, the terminal decision rule is the Bayes rule for the actual sample size attained by the stopping rule. Since there are many sampling plans available, in order to choose a proper stopping rule, the need of some optimality criterion comes up. One such optimality criterion is suggested in Kiefer and Sacks [1963], for the case when the cost of sampling is proportional to the number of observations. If c is the cost per observation, Kiefer and Sacks define a class of stopping rules $\{t(c):c>0\}$ to be asymptotically optimal if the expected risk under this class of stopping rules, in the limit as $c \to 0$, is not larger than the expected risk of any other class of stopping rules.

In this thesis we consider the sequential Bayesian confidence interval problem, and try to obtain asymptotically optimal stopping rules for this problem in the sense of Kiefer and Sacks. Similar work for the problems of point estimation and of hypothesis testing has previously been done by Bickel & Yahav [1967a, 1968]. In fact, the present work essentially is motivated by the final remark that Bickel & Yahav make in their paper [1968], and also by a personal communication with Professor Bickel.

In Chapter 1, we develop some theory of asymptotically optimal stopping rules. We prove that if $\{Y_n\}$ is a sequence of random variables defined on some probability space $(\Omega,F.P)$, if $\{Y_n\}$ satisfies the condition $f(n)Y_n \to$ constant

> 0 a.s., where $f(n)$ is an increasing function of n which satisfies certain other regularity conditions (see Chapter 1, Theorem 1.2), and if $\sup_n E(f(n)Y_n) < \infty$, then the class of stopping rules $\{f(c):c > 0\}$ given below is asymptotically optimal:

For $c > 0$, define $t(c) = n$ if n is the first positive integer such that

$$(0.1) \qquad \{1 - \frac{f(n)}{f(n+1)}\}Y_n \leq c \ .$$

A similar theorem in the case where $f(n)$ is a positive power of n was proved earlier in Bickel & Yahav [1968]. Our result generalizes this theorem of Bickel & Yahav in the sense of generalizing the function f permitted. However, in contrast to Bickel & Yahav's results, we demand the convergence of $f(n)Y_n$ to a constant instead of a random variable. There are some technical difficulties involved in this generalization, which make the present proofs somewhat interesting. The condition $\sup_n E(f(n)Y_n) < \infty$ is in general rather difficult to verify. However, some condition like this seems to be necessary for the asymptotic optimality of the class $\{t(c):c > 0\}$. An example where only this condition is not satisfied and the class $\{t(c):c > 0\}$ is not asymptotically optimal is given at the end of Chapter 1.

In Chapter 2, we develop the theory of fixed sample size Bayes confidence intervals. We obtain a representation for such intervals and also for their posterior Bayes risks.

The loss function considered here is given as follows:

(0.2) $L(\theta, I) = a\ell(I) + b[1 - \delta_I(\theta)]$

where a and b are positive constants, $\ell(I)$ is the length of I, and

(0.3) $$\delta_I(\theta) = \begin{cases} 1, & \text{if } \theta \varepsilon I, \\ 0, & \text{otherwise.} \end{cases}$$

We also consider the sequential version of this problem, and show that for any stopping rule, the Bayes terminal decision rule is to take the Bayes confidence interval for the actual sample size achieved by the stopping rule. Thus, once we have obtained the Bayes confidence intervals for every fixed sample size (as is done in the earlier part of this chapter), the problem reduces to that of obtaining the proper stopping rule.

In Chapter 2, we discuss the example of the normal distribution with unknown mean and known variance to illustrate the theory. Taking a normal prior for the mean, if we define Y_n to be the posterior Bayes risk of the Bayes rule in this case, we show that the sequence $\{Y_n\}$ satisfies all the assumptions on $\{Y_n\}$ of Chapter 1. Now, using the theory of Chapter 1, we show that the asymptotically optimal class of stopping rules $\{t(c):c > 0\}$ as defined in (0.1), in this case for each c takes only a fixed number of observations. However, the calculations for obtaining this fixed number involves inverting the normal distribution function.

In Chapter 3, we consider the more interesting example of the normal distribution with both the mean and variance unknown. We consider the "normal-inverted gamma" prior on μ and σ^2 in this case. This is the conjugate class of prior distributions for the normal distribution. We continue to take the same loss function as defined in (0.2), so that the loss function is independent of the unknown value of σ^2. Now, using the Remark 2.2 given in Chapter 2, we obtain the fixed sample size Bayes confidence intervals $\{[\alpha^*_{1n}, \alpha^*_{2n}]\}$ and also the posterior Bayes risk of these intervals given by $\{Y_n\}$. Again, using the theory of Chapter 1, we show that the class of stopping rules $\{t(c):c > 0\}$ defined in (0.1) is asymptotically optimal in this case. In Appendix 1 at the end of this chapter, we prove a theorem which gives an upper bound on the probability in the tail of the student's t-distribution. Even though the proof of this theorem is easy, we have not previously seen such a bound used in the literature.

Finally, in Chapter 4, we try to attack the general problem. Suppose $\{X_n\}$ are independently, identically distributed random variables with distribution function $F(x|\theta)$ and θ has some prior distribution with density given by $\psi(\theta)$ on Θ. We assume that the functions F and ψ are such that for each n, the posterior density of θ given $X_1,\ldots,X_n$ exists and is almost surely unimodal and continuous. We put some more regularity conditions on F and ψ

which essentially ensure that the posterior distribution converges to normality (for details see the conditions given in Chapter 4), and under these conditions we show that the posterior Bayes risk Y_n of the Bayes rule $[\alpha^*_{1n}, \alpha^*_{2n}]$ behaves asymptotically in such a way that we can use the theory of Chapter 1 to obtain a family of stopping rules $\{t(c): c > 0\}$ which is A.P.O. for this problem. Our claim that this family of stopping rules is also A.O. for the problem is inconclusive to the extent that the results obtained in this chapter do not conform to the assumptions of Theorem 1.2 of Chapter 1. We hope to be able to overcome this difficulty in our future research.

CHAPTER I

ASYMPTOTICALLY OPTIMAL STOPPING RULES

Introduction

In this chapter, we prove some general theorems in the theory of asymptotically optimal stopping rules. These theorems are applied later, in Chapters 3 and 4, to the problem of obtaining asymptotically optimal Bayesian confidence intervals. Theorems of the type proved in this chapter were earlier proved by Bickel & Yahav [1968] for special cases. Hence, we generalize those theorems to make them applicable to a wider class of problems, including those discussed in Chapter 3 and 4. The mathematical structure involved in this chapter is conceptually independent of the problems arising from the confidence interval problems. Hence, we give the theory of this chapter in a rather general setting in the hope that these theorems may find use also in other problems. ||

Let $\{Y_n\}$ be a sequence of random variables defined on a probability space (Ω,F,P), where Y_n is F_n measurable and $F_1 \subset F_2 \subset \ldots\ldots \subset F$. For $c > 0$ and n a positive integer, define $X(n,c) = Y_n + nc$. $X(n,c)$ is a random process depending on the sequence $\{Y_n\}$ and the constant $c > 0$. We consider the problem of obtaining stopping rules for such a process which are asymptotically pointwise optimal (A.P.O.) or asymptotically optimal (A.O.) in the following sense.

These definitions have been used earlier by Bickel & Yahav [1968].

<u>Definition 1</u>: A class of stopping rules $\{t(c): c > 0\}$ is called asymptotically pointwise optimal (A.P.O.) if

$$(1.1) \quad X(t(c),c)[\inf_{s \varepsilon T} X(s,c)]^{-1} \to 1, \quad \text{a.s.,} \quad \text{as} \quad c \to 0.$$

<u>Definition 2</u>: A class of stopping rules $\{t(c): c > 0\}$ is called asymptotically optimal (A.O.) if

$$(1.2) \quad \lim_{c \to 0} E[X(t(c),c)] \{\inf_{s \varepsilon T} E[X(s,c)]\}^{-1} = 1.$$

In both of the above definitions T is the class of all stopping rules defined on the sequence $\{Y_n\}$. As usual here a stopping rule t is considered to be a non-negative integer-valued random variable such that the event $\{t = n\}$ is measurable with respect to the σ-field F_n.

In order to obtain such A.P.O. and A.O. stopping rules, we find it necessary to make certain assumptions. We list these assumptions here and then prove two theorems, Theorem 1.1 and Theorem 1.2, which give us A.P.O. and A.O. stopping rules, respectively.

<u>Assumptions</u>:

A 1.1) $P[Y_n > 0] = 1$ for each n.

A 1.2) $P[Y_n \to 0] = 1$.

A 1.3) $f(n)Y_n \to V > 0$ a.s. as $n \to \infty$, where V is a constant and $f(x)$ is a <u>strictly</u> <u>increasing</u> function of x on $[0,\infty)$. Further if we define

$$(1.3)\qquad F(x) = \frac{x\{f(x+1) - f(x)\}}{f(x+1)},$$

then we assume that $F(x)$ is a bounded function on $[0,\infty)$ such that $F(x) \to M$ as $x \to \infty$, where M is a positive constant.

A 1.4) For each $x > 0$ and $c > 0$ there exists an integer $N(x,c)$ which minimizes the function

$$(1.4)\qquad h(x,c,n) = (f(n))^{-1} x + nc.$$

Further, this $N(x,c)$ may be taken as the first integer n such that $\Delta(h(x,c,n)) \geq 0$, where Δ is the one step difference operator (on n). In particular, this last condition is always satisfied if the difference $\{\frac{1}{f(x)} - \frac{1}{f(x+1)}\}$ is non-increasing in x.

A 1.5) There exists an increasing function g and a number X_o such that for $x > X_o$.

$$\lim_{y\to\infty} \frac{f(xy)}{f(x)g(y)} > 1,$$

and

$$(1.5)\qquad \sum_{n=1}^{\infty} \frac{1}{ng(n)} < \infty .$$

Now, we state the two main theorems of this chapter, which we will subsequently prove.

Let us define a class of stopping rules $\{t(c): c > 0\}$ as follows:

<u>Definition 3</u>: $t(c) = n$, if n is the first integer such that

$$(1.6)\qquad \left(1 - \frac{f(n)}{f(n+1)}\right) Y_n \leq c .$$

<u>Theorem 1.1</u>: Under the assumptions A 1.1) to A 1.4), the class of stopping rules $\{t(c): c > 0\}$ defined in (1.6) is A.P.O.

<u>Theorem 1.2</u>: Under the assumptions A 1.1) to A 1.5), if

$$\sup_n E\{f(n)\, Y_n\} < \infty, \tag{1.7}$$

then the class of stopping rules $\{t(c): c > 0\}$ defined in (1.6) is A.O.

In order to prove these theorems, we will need some preliminary results, which are the consequences of our assumptions from A 1.1) to A 1.4).

In the following lemmas, it is assumed that A1.1) to A1.4) hold.

We introduce some more variables.

<u>Definition 4</u>: Define a random variable $n_o(c)$ such that

$$X(n_o(c),c) = \min_n (X(n,c)). \tag{1.8}$$

<u>Definition 5</u>: Define a r.v. $n^*(c)$ such that

$$h\{(f(t(c)))^{-1}\, Y_{t(c)}, c, n^*(c)\} = \min_n\{h\{(f(t(c)))^{-1}\, Y_{t(c)}, c, n\}\}. \tag{1.9}$$

<u>Definition 6</u>: Define $m_o(c)$ such that

$$h(V,c,m_o(c)) = \min_n \{h(V,c,n)\}. \tag{1.10}$$

All these variables, $t(c), m_o(c), n^*(c)$ and $m_o(c)$, obviously depend on c. However, for notational simplicity, from now on we supress their dependence on c and just write them as t, n_o, n^* and m_o. Since for fixed $c, V > 0$, $X(n,c) \to \infty$, a.s., and $h(V,c,n) \to \infty$ as $n \to \infty$, it is clear that n_o, n^*, m_o are well-defined and almost surely finite, all $c > 0$.

We note that the function $F(x)$ defined in (1.3), can be written as

$$F(x) = x\{1 - \frac{f(x)}{f(x+1)}\}$$

$$(1.11) \qquad = xf(x)\{\frac{1}{f(x)} - \frac{1}{f(x+1)}\} ,$$

and that $xf(x) \to \infty$. From these facts and assumption A 1.3), we obtain:

<u>Lemma 1.1</u>: As $x \to \infty$,

(a) $1 - \frac{f(x)}{f(x+1)} = 0(x^{-1})$,

(b) $\frac{f(x)}{f(x+1)} \to 1$,

(c) $\{\frac{1}{f(x)} - \frac{1}{f(x+1)}\} = 0(xf(x))^{-1}\} = o(1)$,

(d) $\frac{x \; f(x)}{(x-1) \; f(x-1)} \to 1$.

<u>Lemma 1.2</u>: As $c \to 0$,

(a) $t \to \infty$ and $n^* \to \infty$, a.s.,

(b) $m_o \to \infty$.

<u>Proof</u>: From the definition of t, t is first $n \geq 1$, s.t.

$$(1.12) \qquad \{1 - \frac{f(n)}{f(n+1)}\} Y_n \leq c.$$

By A 1.1) and A 1.3) the left-hand side of the inequality in (1.12) is positive for each n almost surely. Also, from A 1.3) and Lemma 1.1(c), the left-hand side of the inequality in (1.12) goes to 0, a.s., as $n \to \infty$. From these facts, it directly follows that $t \to \infty$, a.s., as $c \to 0$.

From A 1.4), we see that n^* and m_o are both special cases of the variable

(1.13) $\tilde{n}$ = first $n \geq 1$ s.t. $h(V(c),c,n+1) - h(V(c),c,n) \geq 0$,

where $V(c) \to V$, a.s., as $c \to 0$. Now,

$$h(V(c),c,n+1) - h(V(c),c,n) = V(c) \left\{\frac{1}{f(n+1)} - \frac{1}{f(n)}\right\} + c,$$

so that

(1.14) $\tilde{n}$ = first $n \geq 1$ s.t. $\left\{\frac{1}{f(n)} - \frac{1}{f(n+1)}\right\} V(c) \leq c$.

Now the same argument used for t shows that $\tilde{n} \to \infty$, a.s., as $c \to 0$, which proves that $n^* \to \infty$, a.s., and $m_o \to \infty$ as $c \to 0$. (Q.E.D.)

<u>Lemma 1.3</u>: As $c \to 0$, $n_o \to \infty$, a.s.

<u>Proof</u>: We recall that $n_o = n_o(c)$ is defined to be such that $X(n_o,c) = \min_n X(n,c)$.

We first show that $X(n_o,c) \to 0$ a.s. as $c \to 0$. Let $\{c_k\} \to 0$ be a sequence of positive numbers and $\{n_k\} \to \infty$ be defined such that $c_k n_k \to 0$. Note that for every sequence $\{c_k\} \to 0$ such a sequence $\{n_k\}$ can be chosen. Now, for this sequence $\{c_k\}$ we know

$$X(n_k,c_k) = Y_{n_k} + c_k n_k \to 0 \text{ a.s., as } k \to \infty.$$

Thus, since

$$X(n_o(c_k),c_k) \leq X(n_k,c_k),$$

it follows that

(1.15) $X(n_o(c_k),c_k) \to 0$ a.s., as $k \to \infty$.

This being true for every sequence $\{c_k\} \to 0$, we have

$$X(n_o,c) \to 0, \text{ a.s., as } c \to 0.$$

This in turn implies that

$$Y_{n_o(c)} \to 0, \quad \text{a.s.,} \quad \text{as} \quad c \to 0.$$

Also, by A 1.1), we have $P[Y_i > 0] = 1$ for all i, and thus given any N we have

$$P[\inf_{n \le N} Y_n > 0] = 1.$$

Therefore, almost surely for every sample point, we can find a $\delta > 0$ s.t.

$$\sup_{c<\delta} Y_{n_o(c)} \le \inf_{n \le N} Y_n .$$

This inequality implies that for all N we can find $\delta > 0$ s.t.

$$\inf_{c<\delta} n_o(c) \ge N,$$

This proves the lemma. (Q.E.D.)

<u>Lemma 1.4</u>: $f(t)ct \to MV$ a.s. as $c \to 0$.

<u>Proof</u>: By definition of t

$$(1.16) \quad \left\{\frac{1}{f(t)} - \frac{1}{f(t+1)}\right\} f(t)\, Y_t \le c < \left\{\frac{1}{f(t-1)} - \frac{1}{f(t)}\right\} f(t-1) Y_{t-1},$$

or

(1.17)

$$\frac{t\{f(t+1) - f(t)\}}{f(t+1)} f(t) Y_t \le f(t)ct \le$$

$$\frac{(t+1)\{f(t) - f(t-1)\}}{f(t)} \quad \frac{t\, f(t)}{(t-1) f(t-1)} f(t-1) Y_{t-1}.$$

Since as $c \to 0$, $t(c) \to \infty$ a.s., the results of Lemma 1.1 imply that when we take the limit as $c \to 0$, we get

$$\lim_{c \to 0} f(t) \cdot c \cdot t \to MV, \quad \text{a.s.} \qquad \text{(Q.E.D.)}$$

Lemma 1.5: $n^*f(n^*)c \to MV$, a.s., as $c \to 0$.

Proof: We have already shown (in Lemma 1.2) that

$$(1.18)\quad \left\{\frac{1}{f(n^*)} - \frac{1}{f(n^*+1)}\right\} f(t)Y_t \le c < \left\{\frac{1}{f(n^*-1)} - \frac{1}{f(n^*)}\right\} f(t)Y_t ,$$

or

(1.19)

$$\frac{n^*\{f(n^*+1) - f(n^*)\}}{f(n^*+1)} f(t)Y_t \le n^*f(n^*)c \le$$

$$\frac{(n^*-1)\{f(n^*) - f(n^*-1)\}}{f(n^*)} \; \frac{n^*f(n^*)}{(n^*-1)f(n^*-1)} f(t)Y_t .$$

Recall from Lemma 1.2 that $t \to \infty$, $n^* \to \infty$, a.s., as $c \to 0$. Thus, from A 1.2), $f(t)Y_t \to V$, a.s., as $c \to 0$. Making use of Lemma 1.1) and taking the limit in (1.19) as $c \to 0$, we find that $f(n^*)cn^* \to MV$, a.s., as $c \to 0$. (Q.E.D.)

Lemma 1.6: $\limsup_{c\to 0} \dfrac{X(t,c)}{\inf_n h(f(t)Y_t,c,n^*)} < 1$, a.s.

Proof: By definition,

$$\frac{X(t,c)}{\inf_n h(f(t)Y_t,c,n)} = \frac{X(t,c)}{h(f(t)Y_t c,n^*)}$$

$$(1.20)\quad = \frac{Y_t+ct}{[f(n^*)]^{-1}f(t)Y_t+cn^*}$$

$$= \frac{f(t)Y_t+ctf(t)}{f(t)Y_t+cn^*f(n^*)} \cdot \frac{f(n^*)}{f(t)} .$$

However, by the definitions of n^* and t and by A 1.4), we can show that $n^* \le t$ (a.s.). Thus, since f is a strictly increasing function of its argument, we have.

$$(1.21)\quad \frac{X(t,c)}{\inf_n h(f(t)Y_t,c,n)} \le \frac{f(t)Y_t+ctf(t)}{f(t)Y_t+cn^*f(n^*)} .$$

However, as $c \to 0$ the right-hand side of (1.21) converges to 1, a.s., by the fact that $t(c) \to \infty$ as $c \to 0$, and by Lemma 1.4 and 1.5.

Therefore, we have

$$\limsup_{c \to 0} \frac{X(t,c)}{\inf_n h(f(t)Y_t, c, n)} \leq 1, \quad \text{a.s.} \qquad \text{(Q.E.D)}$$

<u>Lemma 1.7</u>: $\lim_{c \to 0} \dfrac{h(f(t)Y_t, c, n_0)}{X(n_o, c)} = 1$, a.s.

<u>Proof</u>:
$$\frac{h(f(t)Y_t, c, n_o)}{X(n_o, c)} = \frac{(f(n_o))^{-1} f(t)Y_t + n_o c}{Y_{n_o} + n_o c}$$

$$= \frac{f(t)Y_t + n_o c f(n_o)}{f(n_o)Y_{n_o} + n_o c f(n_o)} \; .$$

Now, since $t \to \infty$ and $n_o \to \infty$ as $c \to 0$ (Lemma 1.2), it follows from A 1.3) that $f(t)Y_t \to V$ and $f(n_o)Y_{n_o} \to V$, a.s. Hence since $n_o c f(n_o) \geq 0$, and thus is bounded away from $-V$, it follows that

$$\lim_{c \to o} \frac{h(f(t)Y_t, c, n_o)}{X(n_o, c)} = 1, \quad \text{a.s.}$$

We are now ready to prove Theorem 1.1. (Q.E.D.)

<u>Proof of Theorem 1.1</u>: We first show that for each $c > 0$, $t(c)$ is a proper stopping rule, in the sense that

$$P[t < \infty] = 1.$$

From the definition of t, we know that $t = n$, if n is the first integer such that

$$\left(1 - \frac{f(n)}{f(n+1)}\right) Y_n \leq c \; ,$$

or

$$(1.22) \qquad \left(\frac{1}{f(n)} - \frac{1}{f(n+1)}\right) f(n)\, Y_n \le c.$$

From Lemma 1.1 we know $\left(\frac{1}{f(n)} - \frac{1}{f(n+1)}\right) \to 0$ and from A 1.3) we know $f(n)\, Y_n \to V > 0$, a.s. Hence, the left hand side of (1.22) converges to zero almost surely. It follows that the inequality is satisfied by some n, almost surely, and thus $P[t < \infty] = 1$.

Next, we have to prove that

$$\lim_{c \to 0} \frac{X(t,c)}{\{\inf_{s \varepsilon T} X(s,c)\}} = 1, \quad \text{a.s.}$$

However, since we have shown above that $t \varepsilon T$ for every $c > 0$, we have

$$\frac{X(t,c)}{\inf_{s \varepsilon T} X(s,c)} \ge 1, \quad \text{a.s.}$$

Consequently, in order to prove the theorem, it is enough to show that

$$(1.23) \qquad \limsup_{c \to 0} \frac{X(t,c)}{\inf_{s \varepsilon T} X(s,c)} \le 1, \quad \text{a.s.,}$$

or, equivalently,

$$(1.24) \qquad \limsup_{c \to 0} \frac{X(t,c)}{X(n_o,c)} \le 1, \quad \text{a.s.}$$

Now,

$$(.25) \qquad \frac{X(t,c)}{X(n_o,c)} \le \frac{X(t,c)}{\inf_n h(f(t)Y_t,c,n)} \cdot \frac{h(f(t)Y_t,c,n_o)}{X(n_o,c)} .$$

Hence, using Lemma 1.6 and Lemma 1.7, we see that

$$(1.26) \qquad \limsup_{c \to 0} \frac{X(t,c)}{X(n_o,c)} \le 1, \quad \text{a.s.,}$$

and this proves the theorem. (Q.E.D.)

Note: Theorem 1.1 is a generalization of Theorem 2.1 of Bickel & Yahav [1968] in the case when their $k(n) = n$. However, as can be noticed, the steps in our proof take a different approach than that of Bickel & Yahav. In fact, the proof of Bickel & Yahav [1968] is in error because the following statement which they make in their paper is not, in general, true:

$$x(\tilde{t}(c),c) = \min_n h(\tilde{t}^\beta(c)Y_{\tilde{t}(c)},c,n).$$

This statement is not true because $\min_n h(\tilde{t}^\beta(c)Y_{\tilde{t}(c)},c,n)$ is not necessarily achieved by their stopping rule $\tilde{t}(c)$, but is achieved at some $n^*(c)$, where $n^*(c) \leq \tilde{t}(c)$. Hence, the steps following this statement in their proof also need not hold true. However, the proof given here now provides rigorous support for their Theorem 2.1 in the case $K(n) = n$. We note that in the case $f(n) = n^\beta$, our stopping rule $t(c)$ is identical to the stopping rule $\tilde{t}(c)$ in Bickel & Yahav [1968].

Cor 1.1: Under assumptions A.1.1 - A.1.4,

$$\lim_{c\to 0} \frac{X(t,c)}{\inf_n h(V,c,n)} = 1 \quad \text{a.s.}$$

Proof: Since $f(n)Y_n \to V$ a.s., given any $\varepsilon > 0$, there exists N_ε, possibly depending on the sample sequence, s.t. for $n > N_\varepsilon$,

$$(1 - \varepsilon)[f(n)]^{-1} V \leq Y_n \leq (1 + \varepsilon)[f(n)]^{-1} V ,$$

or

$$(1-\varepsilon)[[f(n)]^{-1} V + nc] \le Y_n + nc \le (1+\varepsilon)[[f(n)]^{-1} V + nc],$$

or

$$(1-\varepsilon)\inf_{n>N_\varepsilon}[h(V,c,n)] \le \inf_{n>N_\varepsilon} X(n,c) \le (1+\varepsilon)\inf_{n>N_\varepsilon}[h(V,n,c)].$$

However, from Lemma 1.2 we know that $m_o \to \infty$ as $c \to 0$, and so we can find a δ_1 small enough s.t. for $c < \delta_1$,

$$m_o(c) > N_\varepsilon ,$$

and thus

$$\inf_{n>N_\varepsilon} h(V,c,n) = \inf_n h(V,c,n).$$

Also since by Lemma 1.3 $n_o(c) \to \infty$, a.s., as $c \to 0$, we can find a δ_2 small enough (which may again depend on the sample sequence) such that for $c < \delta_2$,

$$n_o(c) > N_\varepsilon.$$

Thus, choosing $\delta^* = \min[\delta_1,\delta_2]$, we have for $c < \delta^*$, that

$$(1-\varepsilon)\inf_n h(V,c,n) \le X(n_o,c) \le (1+\varepsilon)\inf_n h(V,c,n).$$

Now $\varepsilon > 0$ being arbitrary, we get

$$\lim_{c\to 0} \frac{X(n_o,c)}{\inf_n h(V,c,n)} = 1, \quad \text{a.s.}$$

However, from Theorem 1.1 we have

$$\lim_{c\to 0} \frac{X(t,c)}{X(n_o,c)} \to 1 \quad \text{a.s.},$$

and so

$$\lim_{c\to 0} \frac{X(t,c)}{\inf_n h(V,c,n)} \to 1, \quad \text{a.s.} \qquad \text{(Q.E.D.)}$$

<u>Lemma 1.8</u>: i) $f(m_o)cm_o \to MV$, as $c \to 0$,

ii) $f(m_o)X(t,c) \to (1+M)V$, a.s., as $c \to 0$.

<u>Proof</u>: In the proof of Lemma 1.2 we showed that

$$m_o = \text{first } n \geq 1 \text{ s.t. } \left\{\frac{1}{f(n)} - \frac{1}{f(n+1)}\right\} V \leq c.$$

Thus,

$$(1.27)\quad \left\{\frac{1}{f(m_o)} - \frac{1}{f(m_o+1)}\right\} V \leq c < \left\{\frac{1}{f(m_o-1)} - \frac{1}{f(m_o)}\right\} V.$$

Now using exactly the same argument as in the proof of Lemma 1.5, we get $f(m_o)cm_o \to MV$, as $c \to 0$. Thus, since

$$(1.28)\quad f(m_o)h(V,c,m_o) = V + f(m_o)\, cm_o ,$$

using i) we have

$$f(m_o)\ h(V,c,m_o) \to (1+M)V. \quad \text{as} \quad c \to 0,$$

and thus, using Cor.1.1, we find that

$$f(m_o)\ X(t,c) \to (1+M)V, \quad \text{a.s.,} \quad \text{as} \quad c \to 0. \qquad \text{(Q.E.D.)}$$

<u>Lemma 1.9</u>: $f(m_o)ct \to MV$, a.s., as $c \to 0$.

<u>Proof</u>: Since from Lemma 1.4, we know that

$$f(t)ct \to MV, \quad \text{a.s.,} \quad \text{as} \quad c \to 0 ,$$

in order to prove this lemma it is enough to prove that

$$\frac{f(m_o)}{f(t)} \to 1, \quad \text{a.s.,} \quad \text{as} \quad c \to 0.$$

Now,

$$f(t)X(t,c) = f(t)Y_t + f(t)ct,$$

from which, using A 1.3), Lemma 1.2, and Lemma 1.4, it follows that

$$(1.29)\quad f(t)X(t,c) \to (1+M)V \quad \text{a.s.,} \quad \text{as} \quad c \to 0.$$

Also,

$$\frac{f(m_o)}{f(t)} = \frac{f(m_o)X(t,c)}{f(t)X(t,c)} ,$$

so using Lemma 1.8(ii), we see that

$$\frac{f(m_o)}{f(t)} \to \frac{(1+M)V}{(1+M)V} = 1, \quad \text{a.s.,} \quad \text{as} \quad c \to 0.$$

This proves the lemma. (Q.E.D.)

We need one more lemma before we get to the proof of Theorem 1.2.

<u>Lemma 1.10</u>: Under Assumptions A1.1) to A 1.5),

$$E\{f(m_o)ct\} \to MV.$$

<u>Proof</u>: In order to prove this lemma we use the following theorem from Bickel & Yahav [1968]:

<u>Theorem</u>: Let $\{R_n\}$ be a sequence of r.v.s. on some probability space such that $R_n \to R$ in law, Let

$$a_{m,n} = P[|R_m| > n]$$

and

$$a_n = \sup_m a_{m,n}.$$

Then $\sum_n a_n < \infty \Rightarrow E(R_n) \to E(R)$.

We now make use of this theorem to prove Lemma 1.10. Note that in Lemma 1.9 we have proved that

$$f(m_o)ct \to MV, \quad \text{a.s.,} \quad \text{as} \quad c \to 0,$$

which is in fact stronger than convergence in law.

Also, in our case, let $N = \left[\frac{n}{f(m_o)c}\right]$, and

(1.30) $a_{cn} = P[f(m_o)ct > n]$.

Then

$$a_{cn} = P[t > N]$$

$$\leq P[f(N)Y_N > c\{\frac{1}{f(N)} - \frac{1}{f(N+1)}\}^{-1}] ,$$

The last inequality follows from the definition of t.

Using Markov's inequality, we obtain

$$a_{cn} \leq \frac{\alpha}{c}\{\frac{1}{f(N)} - \frac{1}{f(N+1)}\}$$

$$(1.31) \qquad \leq \frac{\alpha}{cNf(N)} \frac{N\{f(N+1)-f(N)\}}{f(N+1)}$$

$$\leq \frac{\alpha\beta}{cNf(N)}$$

where $\alpha = \sup_n E\{f(n)Y_n\}$ and $\beta = \sup_n \frac{n\{f(n+1)-f(n)\}}{f(n+1)} < \infty$.

Now,

$$(1.32) \qquad N=[\frac{n}{cf(m_o)}] = [\frac{n\ m_o}{cm_o f(m_o)}] \geq n[\frac{m_o}{cm_o f(m_o)}] = nM_o,$$

where $M_o=[\frac{m_o}{cm_o f(m_o)}]$. Again, we supress the dependence of N and M_o on c for simplicity of notation.

Because f is strictly increasing, we have from (1.31) and (1.32) that

$$a_{cn} \leq \frac{\alpha\beta}{cnM_o f(nM_o)}$$

$$(1.33) \qquad \leq \frac{\alpha\beta}{cM_o f(m_o)} \cdot \frac{1}{n} \frac{f(m_o)}{f(nM_o)} .$$

Since $M_o = [\frac{m_o}{cm_o f(m_o)}]$,

$$\frac{m_o}{cm_o f(m_o)} - 1 \leq M_o \leq \frac{m_o}{cm_o f(m_o)} ,$$

or

$$(1.34)\quad \frac{1}{cm_o f(m_o)} - \frac{1}{m_o} \le \frac{M_o}{m_o} \le \frac{1}{cm_o f(m_o)} .$$

Taking the limit in (1.34) as $c \to 0$, we see that

$$(1.35)\quad \frac{M_o}{m_o} \to \frac{1}{MV} \quad \text{as} \quad c \to 0.$$

Also,

$$c\, M_o f(m_o) = \frac{M_o}{m_o} \cdot c\, m_o f(m_o),$$

and so

$$(1.36)\quad c\, M_o f(m_o) \to \frac{1}{MV} \cdot MV = 1, \quad \text{as} \quad c \to 0.$$

Now,

$$f(nM_o) = f\left(m_o \cdot \frac{nM_o}{m_o}\right)$$

and since $m_o \to \infty$ as $c \to 0$, we can find a c_1^* s.t. for $0 < c < c_1^*$ and large enough n, using A 1.5), we have

$$\frac{f\left(m_o \frac{nM_o}{m_o}\right)}{f(m_o) g\left(\frac{nM_o}{m_o}\right)} \ge 1 ,$$

or

$$(1.37)\quad \frac{1}{g\left(\frac{nM_o}{m_o}\right)} \ge \frac{f(m_o)}{f(nM_o)} .$$

Therefore, for $0 < c < c_1^*$ and large enough n, we have from (1.33) and (1.37) that

$$a_{cn} \le \frac{\alpha\beta}{cM_o f(m_o)} \cdot \frac{1}{ng\left(\frac{nM_o}{m_o}\right)} ,$$

and thus

(1.38) $a_{cn} \leq \frac{\alpha\beta}{cM_o f(m_o)} \qquad \frac{1}{ng(\hat{n}(c))}$,

where $\hat{n}(c) = \left[\frac{nM_o}{m_o}\right]$.

Now,

$$\frac{nM_o}{m_o} - 1 \leq \hat{n}(c) \leq \frac{nM_o}{m_o} ,$$

and so for each n,

(1.39) $\lim_{c\to 0} \frac{\hat{n}(c)}{n} \to \frac{1}{MV}$ as $c \to 0$.

In consequence, given an $\varepsilon>0$, s.t., $\frac{1}{MV} - \varepsilon > 0$, we can find a c_2^* s.t. for $0 < c < c_2^*$

$$\frac{\hat{n}(c)}{n} \geq \left(\frac{1}{MV} - \varepsilon\right) ,$$

or for $0 < c < c_2^*$

$$\hat{n}(c) \geq \left(\frac{1}{MV} - \varepsilon\right) n > \bar{n} ,$$

where $\bar{n} = [(\frac{1}{MV} - \varepsilon)n]$.

Hence, for $0 < c < c_2^*$,

(1.40) $\frac{1}{g(\hat{n}(c))} \leq \frac{1}{g(\bar{n})}$.

Also, since $cM_o f(m_o) \to 1$ as $c \to 0$, given $1 > \delta > 0$ we can find a c_3^* s.t. for $0 < c < c_3^*$,

(1.41) $\frac{1}{cM_o f(m_o)} < \frac{1}{1-\delta}$.

Thus, if we take $c^* = \min(c_1^*, c_2^*, c_3^*)$, we have

for $0 < c < c^*$ and large enough n,

(1.42) $a_n = \sup_{0<c<c^*} a_{cn} \leq \frac{\alpha\beta}{1-\delta} \frac{1}{ng(\bar{n})}$,

where $\frac{\bar{n}}{n} \to (\frac{1}{MV} - \varepsilon) > 0$ as $n \to \infty$.

Finally, since $\sum_{n=1}^{\infty} \frac{1}{n\, g(n)} < \infty$, we get

$$(1.43) \quad \sum_{n=1}^{\infty} a_n \le \frac{\alpha\beta}{1-\delta} \sum_{n=1}^{\infty} \frac{1}{ng(\bar{n})} < \infty .$$

Thus from the theorem of Bickel & Yahav quoted earlier, it follows that

$$E\{f(m_o)ct\} \to MV, \quad \text{as} \quad c \to 0. \qquad \text{(Q.E.D.)}$$

Proof of Theorem 1.2.

From Lemma 1.8, we know

$$f(m_o)X(t,c) \to (1+M)V \quad \text{a.s.,} \quad \text{as} \quad c \to 0.$$

Also from Theorem 1.1, we have for any $s \varepsilon T$

$$\limsup_{c \to 0} \frac{X(t,c)}{X(s,c)} \le 1, \quad \text{a.s.}$$

Thus, we see that, for any $s \varepsilon T$

$$\liminf_{c \to 0} f(m_o)X(s,c) \ge (1+M)V, \quad \text{a.s.}$$

Using Fatou's lemma,

$$\liminf_{c \to 0} E\{f(m_o)X(s,c)\} \ge (1+M)V.$$

Because $t \varepsilon T$, it follows that in order to prove Theorem 1.2, it is enough to prove that

$$\limsup_{c \to 0} \frac{E(X(t,c))}{\inf_{s \varepsilon T} E(X(s,c))} \le 1.$$

Equivalently, for any $s \varepsilon T$, we need to show that

$$\limsup_{c \to 0} \frac{E(X(t,c))}{E(X(s,c))} \le 1.$$

Now,

$$\limsup_{c\to 0} \frac{E(X(t,c))}{E(X(s,c))} = \limsup_{c\to 0} \frac{f(m_o)E(X(t,c))}{f(m_o)E(X(s,c))}$$

$$\leq \limsup_{c\to 0} \frac{f(m_o)E(X(t,c))}{\liminf_{c\to 0} f(m_o)E(X(s,c))}$$

$$\leq \limsup_{c\to 0} \frac{f(m_o)E(X(t,c))}{(1+M)V} .$$

Therefore, in order to prove the theorem, it is enough to prove

$$f(m_o)E(X(t,c)) \to (1+M)V, \quad \text{as} \quad c \to 0.$$

From the definition of t we have

$$\frac{f(t+1) - f(t)}{f(t+1)} Y_t \leq c,$$

or

$$Y_t \leq ct \frac{f(t+1)}{t\{f(t+1)-f(t)\}}$$

$$\leq \frac{1}{\gamma} ct$$

where $\gamma = \inf_n \frac{n\{f(n+1)-f(n)\}}{f(n+1)}$. Thus,

$$X(t,c) = Y_t + ct \leq (1+\frac{1}{\gamma}) ct$$

and

$$f(m_o)X(t,c) \leq (1+\frac{1}{\gamma}) f(m_o) ct.$$

From Lemma 1.9 and Lemma 1.10, we know that

$$f(m_o) ct \to MV, \quad \text{a.s.},$$

and

$$E\{f(m_o) ct\} \to MV.$$

From Lemma 1.8, we also have,

$$f(m_o)X(t,c) \to (1+M)V.$$

Using a well-known generalization of the dominated convergence theorem (Royden [1968: page 89]), we see that

$$E\{f(m_o)X(t,c)\} \to (1+M)V,$$

and this proves the theorem. (Q.E.D.)

A condition like $\sup_n E\{f(n)Y_n\} < \infty$ is, in general, rather difficult to verify in practical situations. However, some condition like this seems to be necessary in order to prove the asymptotic optimally of the class of rules $\{t(c):c > 0\}$. In case this condition is not satisfied, the family $\{t(c);c > 0\}$ may not be asymptotically optimal, as can be seen from the following example.

Example: Let the probability space be the interval [0,1] with Lebesgue measure. Let f be some function satisfying the conditions on f given in A 1.3 to A 1.5. For each $\omega\varepsilon[0,1]$ and each positive integer n, define a random variable Y_n such that

$$f(n)Y_n(\omega) = \begin{cases} \{\frac{1}{f(n)} - \frac{1}{f(n+1)}\}^{-1}V, & \text{if } \omega\varepsilon[0,\frac{1}{n}), \\ V & \text{, otherwise,} \end{cases}$$

where $V > 0$ is a positive constant.

Then, we have

$$P[Y_n > 0] = 1.$$

Also,

$$P[Y_n \to 0] = 1$$

and

$$f(n)Y_n \to V, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty,$$

so from this definition of Y_n, Assumptions A 1.1) to A 1.5) are satisfied.

Further

$$E\{f(n)Y_n\} = \frac{f(n+1)}{n\{f(n+1)-f(n)\}} \cdot f(n)V + (1-\frac{1}{n})V$$

$$\to \infty \quad \text{as} \quad n \to \infty,$$

which shows that

$$\sup_n E\{f(n)Y_n\} = \infty.$$

Also

$$P[t>n] = P[\{\frac{1}{f(i)} - \frac{1}{f(i+1)}\} f(i)Y_i > c, \quad \text{for} \quad i \leq n].$$

However,

$$\{\frac{1}{f(i)} - \frac{1}{f(i+1)}\} f(i)Y_i(\omega) = \begin{cases} V & , \text{ if } \omega\varepsilon[0,\frac{1}{i}), \\ \{\frac{1}{f(i)} - \frac{1}{f(i+1)}\}V, & \text{if } \omega\varepsilon[\frac{1}{i}, 1], \end{cases}$$

so that if $c < V$, we have

$$P[t > n] \geq \frac{1}{n},$$

and thus,

$$E(t) = \sum_{n=1}^{\infty} P[t > n] = \infty .$$

This implies that

$$EX(t,c) = E\{f(t)Y_t\} + cE(t) = \infty, \text{ for each } 0 < c < V.$$

However, if we define a rule s such that s always stops after the first observation, then

$$X(s,c) = Y_1 + c,$$

and thus

$$E(X(s,c)) = E(Y_1) + c < \infty, \quad \text{for all } c.$$

Hence, the rule $\{t(c):c > 0\}$ cannot be asymptotically optimal.

CHAPTER II

SEQUENTIAL BAYES CONFIDENCE INTERVALS

2.0. Introduction

In this chapter, we obtain representation for the fixed-sample size Bayes confidence intervals, and for the posterior Bayes risk of this Bayes rule. We also consider the sequential version of the Bayesian confidence interval problem, and show that the terminal decision rule when we stop after taking n observations is the same as the fixed-sample Bayes rule for n observations, so that the problem of finding the Bayes rule reduces to that of finding the appropriate stopping rule. At this point, we can use the theory developed in Chapter 1 and obtain A.P.O. and A.O. Bayes sequential confidence interval estimates. As an example of the theory developed in this chapter, we consider the example of obtaining the confidence interval for the mean μ of a normal distribution when the variance is known and μ is assumed to have a normal prior distribution. As might be expected, the asymptotically optimal Bayes confidence interval rule in this case comes out to be a fixed-sample procedure.

2.1. Fixed-Sample Bayesian Confidence Interval Estimation.

Suppose we observe a sequence $\{X_i\}_{i=1}^{n}$ of random variables, independent and identically distributed according to a probability law given by a density $f(x|\theta)$, where $\theta \in \Theta$

CHAPTER II

SEQUENTIAL BAYES CONFIDENCE INTERVALS

2.0. Introduction

In this chapter, we obtain representation for the fixed-sample size Bayes confidence intervals, and for the posterior Bayes risk of this Bayes rule. We also consider the sequential version of the Bayesian confidence interval problem, and show that the terminal decision rule when we stop after taking n observations is the same as the fixed-sample Bayes rule for n observations, so that the problem of finding the Bayes rule reduces to that of finding the appropriate stopping rule. At this point, we can use the theory developed in Chapter 1 and obtain A.P.O. and A.O. Bayes sequential confidence interval estimates. As an example of the theory developed in this chapter, we consider the example of obtaining the confidence interval for the mean μ of a normal distribution when the variance is known and μ is assumed to have a normal prior distribution. As might be expected, the asymptotically optimal Bayes confidence interval rule in this case comes out to be a fixed-sample procedure.

2.1. Fixed-Sample Bayesian Confidence Interval Estimation.

Suppose we observe a sequence $\{X_i\}_{i=1}^{n}$ of random variables, independent and identically distributed according to a probability law given by a density $f(x|\theta)$, where $\theta \in \Theta$

and Θ is an open subinterval of the real line. Suppose nature chooses a value θ_o of θ according to a probability law with density $\psi(\theta)$ on Θ. Let $\mathscr{I}$ represent the class of all subintervals of Θ. This class includes in particular single point sets and also the null set. $\mathscr{I}$ is our action space. Further, for each $\theta \in \Theta$ and $I \in \mathscr{I}$, define the loss function $L(\theta,I)$ as follows:

(2.1) $L(\theta,I) = a\ell(I) + b[1 - \delta_I(\theta)]$,

where a,b are positive constants,

$\ell(I)$ = length of I

and

$$\delta_I(\theta) = \begin{cases} 1, & \text{if } \theta \in I, \\ 0, & \text{otherwise.} \end{cases}$$

Under this loss function, the Bayes risk of a decision rule $I(X_1,X_2,\ldots,X_n) = I_{\underset{\sim}{X}_n}$ is given by

(2.2) $$R(\psi,I) = \int_{\Theta} \left\{ \int_{R^n} L(\theta,I_{\underset{\sim}{X}_n}) \prod_{i=1}^{n} f(x_i|\theta) d\underset{\sim}{X}_n \right\} \psi(\theta) d\theta .$$

Since $L(\theta,I_{\underset{\sim}{X}_n})$ is a non-negative function, we can use Theorem 20 of Royden [1968: page 270] to change the order of integration in equation (2.2). We obtain:

(2.3) $$R(\psi,I) = \int_{R^n} \left\{ \int_{\Theta} L(\theta,I_{\underset{\sim}{X}_n}) \psi(\theta|x_1,\ldots,x_n) d\theta \right\} g(\underset{\sim}{x}_n) d\underset{\sim}{x}_n$$

where $\psi(\theta|x_1,\ldots,x_n)$ is the posterior density of θ given $x_1,\ldots,x_n$, and $g(\underset{\sim}{x}_n)$ is the joint marginal density of $(X_1,\ldots,X_n)$.

By definition, a Bayes decision rule is given by $I^*(x_1,\ldots,x_n) = I^*_{\underset{\sim}{x}_n}$, the decision rule which minimizes (2.3). The Bayes rule equals I^* if for almost all $(x_1,\ldots,x_n)$, $I^*_{\underset{\sim}{x}_n}$ minimizes,

(2.4) $$\rho(\psi, I_{\underset{\sim}{x}_n}) = \int_{\Theta} L(\theta, I_{\underset{\sim}{x}_n})\, \psi(\theta|x_1,\ldots,x_n)d\theta.$$

This is the posterior Bayes risk of the interval $I_{\underset{\sim}{x}_n}$.

Our first task is to obtain a representation for $I^*_{\underset{\sim}{x}_n}$.

Now,

$$\rho(\psi, I_{\underset{\sim}{x}_n}) = \int_{\Theta} L(\theta, I_{\underset{\sim}{x}_n})\, \psi(\theta|x_1,\ldots,x_n)d\theta$$

$$= a\ell(I_{\underset{\sim}{x}_n}) + b\{1 - \int_{I_{\underset{\sim}{x}_n}} \psi(\theta|x_1,\ldots,x_n)\, d\theta\}.$$

Hence, if we define $I_{\underset{\sim}{x}_n} = [\alpha_1(\underset{\sim}{x}_n), \alpha_2(\underset{\sim}{x}_n)] = [\alpha_1, \alpha_2]$, then

(2.5) $$\rho(\psi, I_{\underset{\sim}{x}_n}) = a(\alpha_2 - \alpha_1) + b\{1 - \Psi(\alpha_2|x_1,\ldots,x_n) + \Psi(\alpha_1|x_1,\ldots,x_n)\}$$

where $\Psi(\theta|x_1,\ldots,x_n)$ is the posterior distribution function evaluated at θ. The Bayes rule $I^*_{\underset{\sim}{x}_n} = [\alpha^*_{1n}, \alpha^*_{2n}]$ for the prior distribution ψ is the interval which minimizes (2.5).

Let us make the change of variables:

$$\alpha_1 = \alpha_1,$$

$$d = \alpha_2 - \alpha_1.$$

Then

(2.6) $\rho(\psi, I_{\underset{\sim}{x}_n}) = ad + b\{1-\Psi(\alpha_1+d|x_1,\ldots,x_n)+\Psi(\alpha_1|x_1,\ldots,x_n)\}$.

Differentiating this partially with respect to d and α_1, we obtain

$$\frac{\partial\rho}{\partial d} = a - b\psi(\alpha_1 + d|x_1,\ldots,x_n),$$

and

$$\frac{\partial\rho}{\partial\alpha} = b\{-\psi(\alpha_1+d|x_1,\ldots,x_n) + \psi(\alpha_1|x_1,\ldots,x_n)\}.$$

Setting these partial derivatives equal to zero, we find that it is necessary for the values α_{1n}^* and d_n^* minimizing (2.6) with respect to α_1 and d to satisfy the following equations:

$$\begin{aligned} &\psi(\alpha_{1n}^* + d_n^*|x_1,\ldots,x_n) = \psi(\alpha_{1n}^*|x_1,\ldots,x_n), \\ (2.7)\quad &\psi(\alpha_{1n}^*|x_1,\ldots,x_n) = \frac{a}{b}. \end{aligned}$$

Now, let us assume that for each $(x_1,\ldots,x_n)$, $\psi(\theta|x_1,\ldots,x_n)$ is a strictly unimodal, continuous density. It then follows that, depending on the values of a and b, the equation

(2.8) $\psi(\theta|x_1,\ldots,x_n) = \frac{a}{b}$

has either zero, one, or two solutions, depending on whether the density ever goes above $\frac{a}{b}$. In consequence, depending on the values of a and b, for each n, the Bayes confidence interval is given by:

$$(2.9)\quad I^*_{\underset{\sim}{x}_n} = \begin{cases} \text{null set, if } \psi(\theta|x_1,\ldots,x_n) < \frac{a}{b} \text{ for all } \theta\varepsilon\Theta \\ \{\alpha^*\}, \text{ if equation 2.8 has just one solution } \alpha^*, \\ [\alpha^*_{1n},\alpha^*_{2n}], \text{ if } \alpha^*_{1n} \text{ and } \alpha^*_{2n} \text{ are the two} \\ \qquad \text{solutions of (2.8).} \end{cases}$$

Also, the posterior Bayes risk of this rule is given by

$$(2.10a)\quad Y_n = \rho(\psi, I^*_{\underset{\sim}{x}_n}) = a(\alpha^*_{2n}-\alpha^*_{1n}) + b\{1-\psi(\alpha^*_{2n}(x_1,\ldots,x_n) + \Psi(\alpha^*_{1n}|x_1,\ldots,x_n)\},$$

where $I^*_{\underset{\sim}{x}_n} = [\alpha^*_{1n},\alpha^*_{2n}]$. Otherwise, when $I^*_{\underset{\sim}{x}_n}$ is a null set or a single point set, we have

$$(2.10b)\quad Y_n = \rho(\psi, I^*_{\underset{\sim}{x}_n}) = b.$$

2.2. Sequential Problem.

Let us now consider the sequential version of the Bayes confidence interval decision problem considered in Section 2.1 Here, we can go on sampling as long as we wish. However, if we stop sampling after observing $(x_1,\ldots,x_n) = \underset{\sim}{x}_n$ and take the action $I_{\underset{\sim}{x}_n}$, when θ is the true parameter value, the loss is given by

$$(2.11)\quad L(\theta, I_{\underset{\sim}{x}_n}, n) = a\ell(I_{\underset{\sim}{x}_n}) + b[1-\delta_{I_{\underset{\sim}{x}_n}}(\theta)] + cn$$
$$= L(\theta, I_{\underset{\sim}{x}_n}) + cn,$$

where c is a positive constant. The constant c can be interpreted as the cost per observation of sampling.

In this sequential problem, our decision is a pair (τ, I_τ), where τ is a stopping rule and I_τ is a terminal decision rule which tells us what decision to make after the stopping rule τ tells us to stop. Both the stopping rule τ and the terminal decision rule I_τ generally depend on the observations already made. The Bayes risk of the rule (τ, I_τ) for the prior density $\psi(\theta)$ is given by

$$(2.12)\quad R(\psi,\tau,I_\tau) = \int_\Theta E_\theta(L(\theta,I_\tau) + c\tau)\psi(\theta)d\theta$$

$$= \sum_{n=0}^{\infty} \int_\Theta \left\{ \int_{[\tau=n]} [L(\theta,I_n)+cn] \prod_{i=1}^{n} f(x_i|\theta)d\underset{\sim}{x}_n \right\} \psi(\theta)d\theta .$$

Again changing the order of integration as done earlier, we obtain

$$(2.13)\quad R(\psi,\tau,I_\tau) = \sum_{n=0}^{\infty} \int_{[\tau=n]} \left\{ \int_\Theta [L(\theta,I_n)+cn]\psi(\theta|x_1,\ldots,x_n) \right\} g(\underset{\sim}{x}_n)d\underset{\sim}{x}_n .$$

Because (2.13) is a sum of non-negative terms, in order to minimize this sum, it is obvious that we have to try to minimize each term. Thus, as soon as a stopping rule tells us to stop after n observations, the best we can do is to choose the interval I_n which minimizes the inside integral in (2.13). However, we have already solved this problem in Section 2.1; the interval $I^*_{\underset{\sim}{x}_n}$ given by (2.9) is the interval which minimizes the inside integral. In other words, for every stopping rule τ, the terminal decision rule which minimizes the Bayes risk when $(\tau = n)$ is the fixed sample size Bayes rule for n observations. This fact is a property which is true of most Bayesean sequential problems and was perhaps realized for the first time by Arrow, Blackwell and

Girshick [1949].

Using our earlier notation, if we denote the posterior Bayes risk of the Bayes rule $I^*_{\underset{\sim}{x}_n}$ by Y_n, i.e.,

$$Y_n = \int_{\Theta} L(\theta, I^*_{\underset{\sim}{x}_n}) \psi(\theta | x_1, \ldots, x_n) d\theta,$$

then from (2.13) we see that the Bayes risk of the rule $(\tau, I^*_{\underset{\sim}{x}_\tau})$ is given by

(2.14) $R(\psi, \tau, I^*_{\underset{\sim}{x}_\tau}) = E\{Y_{\tau(c)} + c\tau(c)\}$.

Hence, in order to find the sequential Bayes rule for our problem, we have to find a stopping rule τ which minimizes (2.14). Such a stopping rule would naturally depend on c, the cost of observation. Consequently, if we want to find an asymptotically optimal Bayes rule for our problem, we can in fact use the theory developed in Chapter 1, with Y_n now equal to the posterior Bayes risk of the Bayes rule for the sample of size n.

Before we go on to give an example of the theory developed here, we stop to make some remarks.

Remark 2.1: In order that our representation of the Bayes confidence interval $I^*_{\underset{\sim}{x}_n}$ and the posterior Bayes risk of this Bayes rule Y_n hold, we have made one important assumption in this chapter, namely that for each n the posterior density of θ given by $\psi(\theta | x_1, \ldots, x_n)$ is strictly unimodal and continuous. It would be nice to have a set of necessary and sufficient conditions on $\psi(\theta)$, the prior density, and

$f(x|\theta)$, the common conditional density given θ of the observations X_i, which assure that this assumption holds. At present, we cannot demonstrate the existence of a set of such conditions. However, for the unimodality assumption to hold, it is sufficient that $\psi(\theta)$ and $f(x|\theta)$ both be log concave in θ. In particular, if both $\psi(\theta)$ and $f(x|\theta)$ belong to the exponential family, then this latter condition is satisfied and $\psi(\theta|X_1,X_2,\ldots,X_n)$ is unimodal in θ for almost all $X_1,X_2,\ldots,X_n$.

Remark 2.2: Suppose that there is a p-dimensional nuisance parameter φ in the problem. That is, the common density of the random variables $X_1,X_2,\ldots,$ is given by $f(x|\theta,\varphi)$ where $\theta\varepsilon\Theta$ as before and $\varphi\varepsilon\Phi$, an open subset of p-dimensional Euclidean space. If we do not let our loss function depend on the value of φ, i.e., we continue to have

$$L(\theta,\varphi,I) = a\ell(I) + b[1-\delta_\theta(I)],$$

then the above problem essentially remains the same as the one discussed in this chapter with the understanding that $\psi(\theta|x_1,\ldots,x_n)$ now represents the marginal posterior density of θ given $x_1,\ldots,x_n$, with respect to a prior density $\psi(\theta,\varphi)$ for the parameter (θ,φ).

2.3. Example.

Suppose that $\{X_n\}$ is a sequence of independent, identically distributed $N(\theta,1)$ random variables, and that the prior distribution of θ is $N(\mu,\sigma^2)$ where μ and σ

are known constants. Then it is well known that the posterior distribution of θ given $x_1,\ldots,x_n$ is given by

(2.15) $\psi(\theta | x_1,\ldots,x_n) = N(\mu_n,\sigma_n^2)$,

where

(2.16) $$\mu_n = \frac{n\bar{x} + \mu/\sigma^2}{n + 1/\sigma^2} ,$$

and

(2.17) $$\sigma_n^2 = (n + \frac{1}{\sigma^2})^{-1}$$

Also, we know that

$$\underset{\theta}{\text{Max}}\ \psi(\theta | x_1,\ldots,x_n) = \frac{1}{\sqrt{2\pi}\ \sigma_n} = \frac{1}{\sqrt{2\pi}} (n + \frac{1}{\sigma^2})^{\frac{1}{2}} ,$$

so that for n chosen such that $\frac{1}{\sqrt{2\pi}} (n + \frac{1}{\sigma^2})^{\frac{1}{2}} > \frac{a}{b}$, we have $I^*_{\underset{\sim}{x}_n} = [\alpha^*_{1n}, \alpha^*_{2n}]$, where α^*_{in} are the solutions of the equation

$$\frac{1}{\sqrt{2\pi}\ \sigma_n} e^{-\frac{1}{2\sigma_n^2} (\alpha^*_{in}-\mu_n)^2} = \frac{a}{b} .$$

That is,

$$-\frac{1}{2\sigma_n^2}(\alpha^*_{in} - \mu_2)^2 = \log \frac{a\sqrt{2\pi}\ \sigma_n}{b} ,$$

or equivalently,

$$(\alpha^*_{in} - \mu_n)^2 = \sigma_n^2 \log \frac{b^2}{2\pi\ \sigma_n^2\ a^2} .$$

Solving these equations, we obtain

(2.18) $$\alpha^*_{1n} = \mu_n - \sigma_n \left\{ \log \frac{b^2}{2\pi\ \sigma_n^2\ a^2} \right\}^{\frac{1}{2}}$$

and

$$(2.19) \quad \alpha^*_{2n} = \mu_n + \sigma_n \left\{ \log \frac{b^2}{2\pi\, \sigma_n^2\, a^2} \right\}^{\frac{1}{2}} .$$

Also,

$$\alpha^*_{2n} - \alpha^*_{1n} = 2\sigma_n \left\{ \log \frac{b^2}{2\pi\, \sigma_n^2\, a^2} \right\}^{\frac{1}{2}} ,$$

and thus,

$$1 - \Psi(\alpha^*_{2n} | x_1, \ldots, x_n) = \frac{1}{\sqrt{2\pi}\, \sigma_n} \int_{\alpha^*_{2n}}^{\infty} e^{-\frac{1}{2\sigma_n^2}(\theta - \mu_n)^2} d\theta$$

$$(2.20) \quad = \frac{1}{\sqrt{2\pi}} \int_{\{\log \frac{b^2}{2\pi\, a^2\, \sigma_n^2}\}^{\frac{1}{2}}}^{\infty} e^{-\frac{1}{2}t^2}\, dt$$

and

$$(2.21) \quad \Psi(\alpha^*_{1n} | x_1, \ldots, x_n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\{\log \frac{b^2}{2\pi\, a^2\, \sigma_n^2}\}^{\frac{1}{2}}} e^{-\frac{1}{2}t^2}$$

$$= 1 - \bar{\Psi}(\alpha^*_{2n} | x_1, \ldots, x_n) .$$

We conclude that the Bayes risk of the Bayes confidence interval $[\alpha^*_{1n}, \alpha^*_{2n}]$ is

$$(2.22) \quad Y_n = 2 \left\{ a\, \sigma_n \{\log \frac{b^2}{2\pi\, a^2\, \sigma_n^2}\}^{\frac{1}{2}} + \frac{b}{\sqrt{2\pi}} \int_{\{\log \frac{b^2}{2\pi\, a^2\, \sigma_n^2}\}^{\frac{1}{2}}}^{\infty} e^{-\frac{1}{2}t^2}\, dt \right\} .$$

Since $\sigma_n^2 = \dfrac{1}{n + \frac{1}{\sigma^2}}$, we have

$$\left\{\log \frac{b^2}{2\pi\, a^2\, \sigma_n^2}\right\}^{\frac{1}{2}} = \left\{\log(n + \frac{1}{\sigma^2}) + \log \frac{b^2}{2\pi\, a^2}\right\}^{\frac{1}{2}} .$$

Using the well-known bound on the probability of the normal tail, we find that

$$\frac{b}{\sqrt{2\pi}} \int_{\{\log \frac{b^2}{2\pi a^2 \sigma_n^2}\}^{\frac{1}{2}}}^{\infty} e^{-\frac{1}{2}t^2}\, dt \le \frac{b}{\sqrt{2\pi}\left\{\log \frac{b^2}{2\pi\, a^2\, \sigma_n^2}\right\}^{\frac{1}{2}}}\; e^{-\frac{1}{2}\left\{\log \frac{b^2}{2\pi a^2 \sigma_n^2}\right\}}$$

$$= \frac{a\, \sigma_n}{\left\{\log \frac{b^2}{2\pi\, a^2\, \sigma_n^2}\right\}^{\frac{1}{2}}}$$

$$= \frac{a}{(n + \frac{1}{\sigma^2})^{\frac{1}{2}} \left\{\log(n + \frac{1}{\sigma^2}) + \log \frac{b^2}{2\pi a^2}\right\}^{\frac{1}{2}}} = 0\left(\frac{1}{\sqrt{n \log n}}\right) .$$

Hence,

$$(2.23) \quad Y_n = 2a\, \{ (n + \frac{1}{\sigma^2})^{-\frac{1}{2}} \{\log \frac{b^2}{2\pi\, a^2} + \log(n + \frac{1}{\sigma^2})\}^{\frac{1}{2}} + 0(\frac{1}{\sqrt{n \log n}})\} ,$$

and thus

$$(2.24) \quad \sqrt{\frac{n}{\log n}}\; Y_n \to 2a .$$

Now, if we restrict x to be ≥ 3 and define

$$(2.25) \quad f(x) = \sqrt{\frac{x}{\log x}} ,$$

then this f satisfies all the requirements for f stated in assumptions A 1.3) to A 1.5) of Chapter 1. Further, $f(n)Y_n \to 2a > 0$. Also notice that since σ_n does not depend on $x_1,\ldots,x_n$, Y_n is not a random variable, but instead takes a fixed value for each n, no matter what the observed values of the X's are. Thus, obviously

$$E(f(n)Y_n) = f(n)Y_n$$

and so

$$\sup_n E\{f(n)Y_n\} < \infty .$$

Theorem 1.1 and 1.2 of Chapter 1 now apply in this case to show that class of stopping rules $\{t(c): c > 0\}$ is A.P.O. and also is A.O., where, for each $c > 0$, $t(c)$ is defined as follows: $t(c) = n$, if n is the first integer, $n \geq 3$, such that

$$(2.26) \quad \left(1 - \sqrt{\frac{n \log(n+1)}{(n+1)\log n}}\right) Y_n \leq c.$$

Further, the corresponding sequence of Bayes confidence intervals is given by $\{I^*_{\underset{\sim}{x}_n}\} = \{[\alpha^*_{1n}, \alpha^*_{2n}]\}$, where α^*_{1n} and α^*_{2n} are given by (2.18) and (2.19) respectively. We reemphasize that since Y_n is independent of the data, the stopping rule $t(c)$ is a fixed integer satisfying (2.26). Although $t(c)$ does not define the optimal sample size $\tau^*(c)$ for the Bayes procedure (that sample size is found by minimizing $Y_n + nc$ over all $n \geq 0$), the theory of Chapter 1 tells us that

$$\lim_{c \to 0} \frac{E(Y_{t(c)} + ct(c))}{E(Y_{\tau^*(c)} + c\tau^*(c))} = 1 .$$

Solution for the optimal sample size $\tau^*(c)$, since it involves inverting the normal c.d.f., is considerably more complicated than finding $t(c)$, so that $t(c)$ may offer a more practical alternative in cases where the cost of sampling is small. Our major motivation, however, for giving this example is to illustrate the theory of this chapter and to motivate the results of the next chapter.

CHAPTER III

CONFIDENCE INTERVAL ESTIMATION OF THE MEAN OF THE NORMAL DISTRIBUTION WITH UNKNOWN VARIANCE

3.0 Introduction.

In this chapter, we consider the problem of obtaining the asymptotically optimal sequential Bayes confidence interval procedure for the mean of the normal distribution where the variance is unknown. For this problem, if we stop sampling after n observations and choose the interval I, our loss function is given by

$$L(\mu,\sigma^2,n,I) = a\ell(I) + b[1 - \delta_I(\mu)] + nc.$$

$$(3.1) \qquad = L(\mu,I) + nc,$$

where μ, σ^2 are the true values of the parameters. Thus the loss function is independent of the true value of σ^2, and so we can use Remark 2.1 of Chapter 2 to get the solution in this case. In order to do this, we first have to find the fixed sample size Bayes confidence intervals $[\alpha^*_{1n},\alpha^*_{2n}]$ and also the posterior Bayes risk $\{Y_n\}$ of these intervals. Further, we have to show that $\{Y_n\}$ satisfies all the assumptions of Theorem 1.1 and Theoren 1.2 of Chapter 1, so that we can use these theorems to obtain A.O. and A.P.O. stopping rules.

The class of prior distributions that we consider here is given by

$$\psi(\mu,\sigma^2|n_o,\nu_o,m_o,v_o) =$$

$$(3.2) \qquad \sqrt{\frac{n_o}{2\pi\sigma^2}}\; e^{-\frac{1}{2}\frac{n_o}{\sigma^2}(\mu-m_o)^2} \cdot \frac{1}{\Gamma(\frac{1}{2}\nu_o)\left(\frac{2}{\nu_o v_o}\right)^{\frac{1}{2}\nu_o}}$$

$$\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}\nu_o+1} \; e^{-\frac{1}{2}\frac{\nu_o v_o}{\sigma^2}} ,$$

where n_o, ν_o, m_o, v_o are known constants s.t., $\nu_o > 2$, $v_o > 0$, $n_o > 0$. Under this prior distribution, the distribution of σ^2 is inverted gamma type, and the distribution of μ given σ^2 is normal. As is well known (Raiffa and Schlaifer [1961]), the prior distribution (3.2) belongs to the family of conjugate priors for the normal distribution with mean and variance unknown.

3.1 <u>Fixed Sample Size Bayes Confidence Intervals</u>.

Using the prior distribution (3.2), it is well known (Raiffa and Schlaifer [1961, page 303]) that the posterior distribution of μ and σ^2, given the first n observations $x_1, x_2, \ldots, x_n$, is of the same functional form as (3.2) with the new parameters now given by n_1, ν_1, m_1, v_1, where

$$n_1 = n_o + n,$$

$$\nu_1 = \nu_o + n,$$

(3.3)
$$m_1 = \frac{1}{n_1}\left\{n_o m_o + \sum_{i=1}^{n} x_i\right\},$$

$$v_1 = \frac{1}{n_1}\left\{(\nu_o v_o + n_o m_o^2) + \sum_{i=1}^{n} x_i^2 - n_1 m_1^2\right\}.$$

Note that for simplicity of notation we have suppressed the dependence of these new parameters on n. The density of the posterior distribution of μ and σ^2 given $x_1,\ldots,x_n$, is now given by

(3.4) $$\psi(\mu,\sigma^2 | n_1,\nu_1,m_1,v_1).$$

In order to obtain the marginal posterior density of μ, we integrate out σ^2 from (3.4). It can be easily varified that after doing this integration, we get the following marginal density of μ given $x_1,\ldots,x_n$:

(3.5) $$\psi(\mu | n_1,\nu_1,m_1,v_1) = \sqrt{\frac{n_1}{v_1}}\;\frac{1}{\sqrt{\nu_1}\,B(\frac{1}{2},\frac{\nu_1}{2})}\;\frac{1}{\left[\frac{n_1(\mu-m_1)^2}{\nu_1 v_1}+1\right]^{\frac{\nu_1+1}{2}}}.$$

From (3.5) we can see that if we define $t = \sqrt{\frac{n_1}{v_1}}(\mu - m_1)$, then t has student's t-distribution with ν_1 degrees of freedom.

Now, since $X \sim N(\mu,\sigma^2)$ and (μ,σ^2) has the prior distribution given by ψ, we have,

$$E(X) = E\{E(X| \mu,\sigma^2)\},$$

or

(3.6) $E(X) = E(\mu) = m_o$.

Also,

$$E(X^2) = E\{E(X^2|\mu,\sigma^2)\}$$
$$= E\{\mu^2 + \sigma^2\}$$
$$= E(\sigma^2) + E\{E(\mu^2|\sigma^2)\}$$
$$= E(\sigma^2) + E\{m_o^2 + \frac{\sigma^2}{n_o}\}$$
$$= (1 + \frac{1}{n_o})E(\sigma^2) + m_o^2 ,$$

or

(3.7) $E(X^2) = (1 + \frac{1}{n_o}) \frac{\nu_o}{\nu_o-2} v_o + m_o^2$.

Since from (3.6) and (3.7), we see that $E(X)$ and $E(X^2)$ are both finite, using the strong law of large numbers for exchangeable random variables (Loève [1963,p.365,400]).

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to m_o, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty,$$

and also

$$\frac{1}{n} \sum_{i=1}^{n} X_i^{\,2} \to (1 + \frac{1}{n_o}) \frac{\nu_o}{\nu_o-2} v_o + m_o^{\,2}, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty.$$

[Here, and in the rest of this chapter almost sure convergence is in terms of the marginal distribution of the X_i's]. Further, using these results and the definition of v_1, from (3.3) we see that

(3.8) $v_1 \to (1 + \frac{1}{n_o}) \frac{\nu_o}{\nu_o-2} v_o$, a.s., as $n \to \infty$.

From (3.5) and (3.8) it can be shown (see (3.14) for example) that the value of the posterior density of μ at m_1 goes to ∞ as $n \to \infty$. Also, note from (3.5) that $\psi(\mu|n_1,\nu_1,m_1,v_1) \to 0$ as $\mu \to \pm\infty$. Thus, for every value of a and b such that $a > 0$ and $b > 0$, the equation

$$(3.9) \quad \psi(\mu|n_1,\nu_1,m_1,v_1) = \frac{a}{b}$$

will, for large enough n, always have two distinct solutions α^*_{1n} and $\alpha^*_{2n} (> \alpha^*_{1n})$. Now, the theory of Chapter 2 shows that $I^*_n = [\alpha^*_{1n}, \alpha^*_{2n}]$ is in fact the Bayes confidence interval for the fixed sample size n. Further the posterior Bayes risk of this procedure I^*_n is given by,

$$(3.10) \quad Y_n = a(\alpha^*_{2n} - \alpha^*_{1n}) + b[1 - \int_{\alpha^*_{in}}^{\alpha^*_{2n}} \psi(\mu|n_1,\nu_1,m_1,v_1)d\mu].$$

3.2 Asymptotic Behavior of $\{Y_n\}$.

From the definition of α^*_{1n} and α^*_{2n}, we know that

$$(3.11) \quad \psi(\alpha^*_{in}|n_1,\nu_1,m_1,v_1) = \frac{a}{b}, \quad i = 1,2.$$

Also $\alpha^*_{1n} < \alpha^*_{2n}$. If we define

$$(3.12) \quad \beta_{in} = \sqrt{\frac{n_1}{v_1}}\,(\alpha^*_{in} - m_1),$$

then using (3.5) we see that the β_{in} satisfy the equation

$$(3.13) \quad \sqrt{\frac{n_1}{v_1}}\;\frac{c_{\nu_1}}{\left[\frac{\beta_{in}^2}{\nu_1} + 1\right]^{\frac{\nu_1+1}{2}}} = \frac{a}{b},$$

where

$$c_{\nu_1} = \frac{1}{\sqrt{\nu_1} B(\tfrac{1}{2}, \frac{\nu_1}{2})} = \frac{\Gamma\left(\frac{\nu_1+1}{2}\right)}{\sqrt{\nu_1}\sqrt{\pi}\ \Gamma(\frac{\nu_1}{2})} .$$

Since $\nu_1 = \nu_o + n$, from known facts about gamma functions it is straightforward to show that as $n \to \infty$

$$(3.14)\quad c_{\nu_1} \to \frac{1}{\sqrt{2\pi}} .$$

Lemma 3.1. As $n \to \infty$,

$$\text{(i)}\ \frac{\beta_{in}^2}{\nu_1} \to 0, \quad \text{(ii)}\ \frac{\beta_{in}^2}{\log n_1} \to 1, \quad \text{a.s.}$$

Proof: From (3.13) we have

$$\frac{\sqrt{n_1}}{\left[\frac{\beta_{in}^2}{\nu_1} + 1\right]^{\frac{\nu_1+1}{2}}} = \left(\frac{a}{b} \cdot \frac{\sqrt{\nu_1}}{c_{\nu_1}}\right),$$

or

$$(3.15)\quad \left[\frac{e^{\frac{\log n_1}{\nu_1+1}}}{\frac{\beta_{in}^2}{\nu_1} + 1}\right] = \left(\frac{a}{b}\ \frac{\sqrt{\nu_1}}{c_{\nu_1}}\right)^{\frac{2}{\nu_1+1}} .$$

From (3.8) and (3.14), we see that the right hand side of (3.15) goes to 1, a.s., as $n \to \infty$. Also $(\nu_1 + 1)^{-1} \log n_1 \to 0$ as $n \to \infty$, and so we get that

$$\frac{\beta_{in}^2}{\nu_1} \to 0, \text{ a.s., as } n \to \infty.$$

Now, taking logarithms on both sides of (3.13) and dividing on both sides by $\log n_1$, we get

(3.16)

$$\frac{1}{2}\,\frac{\log \nu_1}{\log n} + \frac{\log c_{\nu_1}}{\log n_1} + \frac{1}{2} - \frac{1}{2}\,\frac{\nu_1+1}{\log n_1}\log\left(\frac{\beta_{in}^2}{\nu_1} + 1\right) = \frac{\log(\frac{a}{b})}{\log n_1} \;.$$

Taking the limit as $n \to \infty$ in (3.16), we see that the right hand side of (3.16) goes to zero as $n \to \infty$ and so the left hand side must also go to zero, a.s.

It thus follows from (3.8) and (3.14) that

$$\frac{\nu_1+1}{\log n_1}\log\left(\frac{\beta_{in}^2}{\nu_1} + 1\right) \to 1, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty,$$

or

$$\frac{\nu_1+1}{\nu_1} \cdot \frac{\beta_{in}^2}{\log n_1} \cdot \frac{\log\left(\frac{\beta_{in}^2}{\nu_1} + 1\right)}{\frac{\beta_{in}^2}{\nu_1}} \to 1, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty.$$

However, since we have already shown that $\frac{\beta_{in}^2}{\nu_1} \to 0$ as $n \to \infty$, it follows that $\{\log(\nu_1^{-1}\beta_{in}^2+1)\}/(\nu_1^{-1}\beta_{in}) \to 1$, a.s., as $n \to \infty$, and so from the above statement we conclude that as $n \to \infty$,

$$\frac{\beta_{in}^2}{\log n_1} \to 1. \qquad \text{(Q.E.D.)}$$

Using the definitions of n_1 and β_{in}, and Lemma 3.1, we can show that

$$\frac{1}{\nu_1}\,\frac{n}{\log n}\,(\alpha_{in}^* - m_1)^2 \to 1, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty.$$

However, since from (3.8) we have

$$v_1 \to \left(1 + \frac{1}{n_o}\right) \frac{\nu_o}{\nu_o - 2} v_o, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty,$$

we see that,

$$\frac{n}{\log n} (\alpha^*_{in} - m_1)^2 \to \left(1 + \frac{1}{n_o}\right) \frac{\nu_o}{\nu_o - 2} v_o, \quad \text{a.s., as} \quad n \to \infty,$$

or

$$(3.17) \quad \sqrt{\frac{n}{\log n}} |\alpha^*_{in} - m_1| \to \left\{\left(1 + \frac{1}{n_o}\right)\frac{\nu_o v_o}{\nu_o - 2}\right\}^{\frac{1}{2}}, \quad \text{a.s., as} \quad n \to \infty.$$

Now since α^*_{1n} and α^*_{2n} are the two solutions of Equation (3.9), since $\alpha^*_{1n} < \alpha^*_{2n}$, and since m_1 is the mode of $\psi(\mu|n_1,\nu_1,m_1,v_1)$, we know that $\alpha^*_{1n} < m_1$ and $\alpha^*_{2n} > m_1$

It follows that

$$(3.18) \quad \sqrt{\frac{n}{\log n}} (\alpha^*_{2n} - \alpha^*_{1n}) \to 2\left\{\left(1 + \frac{1}{n_o}\right) \frac{\nu_o v_o}{\nu_o - 2}\right\}^{\frac{1}{2}}, \text{ a.s., as } n \to \infty.$$

<u>Lemma 3.2</u>. As $n \to \infty$,

$$\sqrt{\frac{n_1}{\log n_1}} \int_{\alpha^*_{2n}}^{\infty} \psi(\mu|n_1,\nu_1,m_1,v_1)\,d\mu \to 0, \quad \text{a.s.} \ .$$

<u>Proof</u>: Making the substitution $t = \sqrt{\frac{n_1}{v_1}}(\mu - m_1)$, we get

$$\int_{\alpha^*_{2n}}^{\infty} \psi(\mu|n_1,\nu_1,m_1,v_1)\,d\mu = \frac{1}{\sqrt{\nu_1}B(\frac{1}{2}, \frac{\nu_1}{2})} \int_{\beta_{2n}}^{\infty} \frac{1}{\left[\frac{t^2}{\nu_1} + 1\right]^{\frac{\nu_1 + 1}{2}}} dt.$$

Now, using the bound on the probability in the tail of students' t-distribution from Appendix 1, we get

$$\int_{\alpha_{2n}^*}^{\infty} \psi(\mu|n_1,\nu_1,m_1,v_1)d\mu \le \sqrt{\frac{\nu_1}{\nu_1-2}}\ \frac{\beta_{2n}^{-1}}{\sqrt{\nu_1-2}B(\frac{1}{2},\frac{\nu_1-2}{2})}\frac{1}{\left[\frac{\beta_{2n}^2}{\nu_1}+1\right]^{\frac{\nu_1-1}{2}}}$$

$$\le \frac{\nu_1}{\nu_1-2}\ \frac{1}{\beta_{2n}}\cdot\frac{B(\frac{1}{2},\frac{\nu_1}{2})}{B(\frac{1}{2},\frac{\nu_1-2}{2})}\ \frac{\left[\frac{\beta_{2n}^2}{\nu_1}+1\right]}{\sqrt{\nu_1}B(\frac{1}{2},\frac{\nu_1}{2})}\ \frac{1}{\left[\frac{\beta_{2n}^2}{\nu_1}+1\right]^{\frac{\nu_1+1}{2}}}$$

(3.19)

$$\le (\frac{\nu_1}{\nu_1-1})\ \frac{1}{\beta_{2n}}\ \left[\frac{\beta_{2n}^2}{\nu_1}+1\right]\cdot\ \frac{a}{b}\sqrt{\frac{v_1}{n_1}}\ .$$

The last inequality in (3.19) follows from (3.13) and the fact that

$$\frac{B(\frac{1}{2},\ \frac{\nu_1}{2})}{B(\frac{1}{2},\frac{\nu_1-2}{2})} = \frac{\nu_1-2}{\nu_1-1}\ .$$

Thus,

$$(3.20)\ \sqrt{\frac{n_1}{\log n_1}}\ \int_{\alpha_{2n}^*}^{\infty}\psi(\mu|n_1,\nu_1,m_1,v_1)d\mu$$

$$\le (\frac{\nu_1}{\nu_1-2})\ \frac{a}{b}\sqrt{\frac{1}{\beta_{2n}\log n_1}}\ \left[\frac{\beta_{2n}^2}{\nu_1}+1\right]\sqrt{v_1}\ .$$

From Lemma 3.1 and (3.8), it follows that the right hand side of (3.20) goes to zero almost surely, and so the lemma is proved. (Q.E.D.)

Cor. 3.2. As $n \to \infty$,

$$\sqrt{\frac{n_1}{\log n_1}} \int_{-\infty}^{\alpha_{1n}^*} \psi(\mu | n_1, \nu_1, m_1, v_1) d\mu \to 0, \quad \text{a.s.}$$

Proof: Since the posterior density of μ is symmetric about m_1 (see(3.5)), we have

$$(3.21) \quad \int_{-\infty}^{\alpha_{1n}^*} \psi(\mu | n_1, \nu_1, m_1, v_1) d\mu = \int_{\alpha_{2n}^*}^{\infty} \psi(\mu | n_1, \nu_1, m_1, v_1) d\mu \ .$$

Hence, from Lemma 3.2, the corollary follows. (Q.E.D.)

Theorem 3.1. If Y_n is defined as in Equation (3.10), then

$$\sqrt{\frac{n}{\log n}} \; Y_n \to 2a\{(1 + \frac{1}{n_o}) \frac{\nu_o}{\nu_o - 2} \; v_o\}^{\frac{1}{2}}, \quad \text{a.s.}, \quad \text{as} \quad n \to \infty.$$

Proof: Using (3.18), Lemma 3.2 and Cor. 3.2, it follows from (3.10) that

$$\sqrt{\frac{n_1}{\log n_1}} \; Y_n \to 2a \; \{(1 + \frac{1}{n_o}) \frac{\nu_o}{\nu_o - 2} \; v_o\}^{\frac{1}{2}}, \quad \text{a.s.}, \quad \text{as} \quad n \to \infty.$$

However, since $n_1 = n + n_o$, we have

$$\sqrt{\frac{n_1 \log n}{n \log n_1}} \to 1 \ , \quad \text{as} \quad n \to \infty,$$

and so,

$$\sqrt{\frac{n}{\log n}} \, Y_n \to 2a\{(1 + \frac{1}{n_o}) \frac{\nu_o}{\nu_o - 2} \, v_o\}^{\frac{1}{2}} \ , \quad \text{a.s.}, \quad \text{as} \quad n \to \infty.$$

(Q.E.D.)

For $x \geq 0$ define the function,

$$f(x) = \begin{cases} \sqrt{x} & , \text{ if } 0 \leq x \leq e, \\ \sqrt{\frac{x}{\log x}} & , \text{ if } x > e. \end{cases}$$

This function satisfies all of the requirements in Assumptions A1.3) to A 1.5) of Chapter 1. Further, since we have just now proved that

$$f(n)\ Y_n \to 2a\{(1+\frac{1}{n_o})\ \frac{\nu_o}{\nu_o-2}\ v_o\}^{\frac{1}{2}} > 0, \quad \text{a.s., as} \quad n \to \infty,$$

we can use Theorem 1.1 of Chapter 1 and show that the class of stopping rules $\{t(c):c > 0\}$ is A.P.O. for our problem, where for each $c > 0$, $t(c)$ is defined as follows:

$t(c) = n$, if n is the first integer n, $n \geq 1$, such that

$$(3.22) \quad (1 - \frac{f(n)}{f(n+1)})\ Y_n \leq c.$$

In order that we can also use Theorem 1.2 of Chapter 1 in this example, we have to show further that

$$\sup_n E\{\sqrt{\frac{n}{\log n}}\ \ Y_n\} < \infty.$$

A direct proof of this result is rather difficult; consequently, we use a rather indirect argument to get this result.

Notice that Y_n is the posterior Bayes risk of the Bayes rule $I_n^* = [\alpha_{1n}^*, \alpha_{2n}^*]$, so that if we consider any other rule $I_n' = [\alpha_{1n}', \alpha_{2n}']$ and if Y_n' is the posterior Bayes risk of I_n', then Y_n' almost surely exceeds Y_n.

Thus, in order to prove that

$$\sup_n E\{f(n)Y_n\} < \infty ,$$

it is enough to show that there exists some rule I_n' with

$$\sup_n E\{f(n)Y_n'\} < \infty .$$

Let us define,

$$(3.23) \quad I_n' = [\alpha_{1n}', \alpha_{2n}'] = [m_1 - \sqrt{v_1}\sqrt{\frac{\log \nu_1}{n_1}},\ m_1 + \sqrt{v_1}\sqrt{\frac{\log \nu_1}{n_1}}] .$$

Then we know that

$$(3.24) \quad Y_n' = 2a\sqrt{v_1} \quad \frac{\log \nu_1}{n_1}$$

$$+ 2b\ P[\sqrt{\frac{n_1}{v_1}}(\mu - m_1) > (\log \nu_1)^{\frac{1}{2}} | n_1, \nu_1, m_1, v_1].$$

However, since $\sqrt{\frac{n_1}{v_1}}(\mu - m_1)$ given n_1, ν_1, m_1, v_1 has the students' t-distribution with ν_1 degrees of freedom, using the result from Appendix 1 we get

$$P\left[\sqrt{\frac{n_1}{v_1}}(\mu - m_1) > (\log \nu_1)^{\frac{1}{2}} | n_1, \nu_1, m_1, v_1\right]$$

$$\leq \sqrt{\frac{\nu_1}{\nu_1 - 2}} \quad \frac{1}{(\log \nu_1)^{\frac{1}{2}}} \quad f_{\nu_1 - 2}\left(\sqrt{\frac{(\nu_1 - 2)}{\nu_1}}(\log \nu_1)^{\frac{1}{2}}\right),$$

where $f_{\nu_1 - 2}(t)$ is the density of students' t-distribution with $(\nu_1 - 2)$ degrees of freedom.

Thus

$$(3.25)\ \sqrt{\frac{n}{\log n}}\ P\left[\sqrt{\frac{n_1}{v_1}}\ (\mu-m_1) > (\log \nu_1)^{\frac{1}{2}} \middle| n_1,\nu_1,m_1,v_1\right] \leq$$

$$\sqrt{\frac{\nu_1}{\nu_1-2}}\ \sqrt{\frac{n}{\nu_1}}\ \sqrt{\nu_1}\ \frac{1}{\sqrt{(\log \nu_1)(\log n)}}\ f_{\nu_1-2}\left(\sqrt{\frac{(\nu_1-2)}{\nu_1}}(\log \nu_1)^{\frac{1}{2}}\right).$$

From (3.24), (3.25), and the fact that $\nu_1 = n + \nu_o$, in order to prove that $\sup_n E\{f_{(n)}Y_n'\} < \infty$, it is enough to show that

(i) $\sup_n E\{\sqrt{v_1}\} < \infty$

and

(ii) $\sup_n \sqrt{\nu_1}\ f_{\nu_1-2}\left(\sqrt{\frac{\nu_1-2}{\nu_1}}\ (\log \nu_1)^{\frac{1}{2}}\right) < \infty$.

The next two lemmas will precisely prove these results. Having shown by these lemmas that

$$\sup_n \{f(n)Y_n'\} < \infty,$$

we will have then shown that

$$\sup_n \{f(n)Y_n\} < \infty .$$

<u>Lemma 3.3</u>: If $f_{\nu_1-2}(t)$ is the density of student's t-distribution with (ν_1-2) degrees of freedom, where $\nu_1 = n + n_o$, then

$$\sup_n \sqrt{\nu_1}\; f_{\nu_1-2}\left(\sqrt{\frac{(\nu_1-2)}{\nu_1}}\,(\log \nu_1)^{\frac{1}{2}}\right) < \infty .$$

<u>Proof</u>: We have,

$$(3.26)\quad \sqrt{\nu_1}\; f_{\nu_1-2}\left(\sqrt{\frac{(\nu_1-2)}{\nu_1}}(\log \nu_1)^{\frac{1}{2}}\right)$$

$$= \frac{1}{\sqrt{(\nu_1-2)}\,B(\frac{1}{2},\frac{\nu_1-2}{2})} \; \frac{\sqrt{\nu_1}}{\left[\frac{\log \nu_1}{\nu_1} + 1\right]^{\frac{\nu_1-1}{2}}}$$

$$= \frac{1}{\sqrt{(\nu_1-2)}\,B(\frac{1}{2},\frac{\nu_1-2}{2})} \left[\frac{\log \nu_1}{\nu_1} + 1\right]^{\frac{1}{2}} \frac{\sqrt{\nu_1}}{\left[\frac{\log \nu_1}{\nu_1} + 1\right]^{\frac{\nu_1}{2}}} .$$

Now, we know that

$$\frac{1}{\sqrt{(\nu_1-1)}\;B(\frac{1}{2},\frac{\nu_1-2}{2})} \rightarrow \frac{1}{\sqrt{2\pi}} \quad \text{as} \quad n \rightarrow \infty,$$

and so,

$$\sup_n \frac{1}{\sqrt{(\nu_1-1)}\;B(\frac{1}{2},\frac{\nu_1-2}{2})} < \infty .$$

Also,

$$\sup_n \left[\frac{\log \nu_1}{\nu_1} + 1\right]^{\frac{1}{2}} < \infty,$$

and from Appendix 2 we know that

$$\sup_n \frac{\sqrt{\nu_1}}{\left[\frac{\log \nu_1}{\nu_1} + 1\right]^{\frac{\nu_1}{2}}} < \infty .$$

Using these results, from (3.26) we see that

$$\sup_n \sqrt{\nu_1} \; f_{\nu_1 - 2}\left(\sqrt{\frac{(\nu_1-2)}{\nu_1}} \; (\log \nu_1)^{\frac{1}{2}}\right) < \infty. \qquad \text{(Q.E.D.)}$$

Lemma 3.4. If v_1 is defined as in (3.3), then

$$\sup_n E(\sqrt{v_1}) < \infty.$$

Proof: According to Raiffa and Schlaifer [1961, page 307], we know that the marginal distribution of v_1 is inverted beta 1 type with density given by

$$(3.27) \quad f_{i\beta 1}(v_1 | \tfrac{1}{2}\nu_o, \tfrac{1}{2}\nu_1, \frac{\nu_o v_o}{\nu_1}),$$

where the functional form of (3.27) is given by

$$f_{i\beta 1}(y|r,n,b) = \frac{1}{B(r,n-r)} \; \frac{(y-b)^{n-r-1} \cdot b^r}{y^n} ,$$

where, $0 \leq b \leq y < \infty$ and $n > r > 0$.

Further, if y is a random variable with this inverted beta 1 type density with parameters r,n and b, then

$$(3.28) \quad E(\sqrt{Y}) = \frac{b^{\frac{1}{2}} B(r - \frac{1}{2}, n - r)}{B(r,n-r)} .$$

Using (3.27) and (3.28) we get

$$E(\sqrt{v_1}) = \left(\frac{\nu_o v_o}{\nu_1}\right)^{\frac{1}{2}} \frac{B(\frac{\nu_o - 1}{2}, \frac{n}{2})}{B(\frac{\nu_o}{2}, \frac{n}{2})}$$

$$(3.29) \qquad = (\nu_o v_o)^{\frac{1}{2}} \sqrt{\frac{n}{\nu_1}} \cdot \frac{\Gamma(\frac{\nu_o - 1}{2}) \quad \Gamma(\frac{n+\nu_o}{2})}{\sqrt{n}\,\Gamma(\frac{\nu_o}{2}) \quad \Gamma(\frac{n+\nu_o - 1}{2})} .$$

Using Sterling's approximations for gamma functions, it can be easily seen that the right hand side of (3.29), in fact, converges to a finite constant. In consequence,

$$\sup_n E(\sqrt{v_1}) < \infty . \qquad \text{(Q.E.D.)}$$

To summarize, our arguments in this chapter have proved the following theorem.

<u>Theorem 3.2</u>: If $\{X_i\}$ is a sequence of independent identically distributed $N(\mu, \sigma^2)$ random variables and the prior distribution of μ and σ^2 is given by (3.2), then the sequence of terminal decision rules $\{I_n^*\}=\{[\alpha_{1n}^*, \alpha_{2n}^*]\}$ is the sequence of terminal Bayes decision rules against any stopping rule. Further, the class of stopping rules $\{t(c): c > 0\}$ defined in (3.22) is A.P.O. and also A.O. for this problem.

APPENDIX 1 TO CHAPTER III

<u>Theorem</u>: If T is a random variable having student's t-distribution with ν degrees of freedom, then for any $a > 0$,

$$P[T > a] \leq \frac{1}{a\sqrt{(\frac{\nu-2}{\nu})}} \; f_{\nu-2}\left(a\sqrt{\frac{(\nu-2)}{\nu}}\right),$$

where $f_\nu(t)$ is the density of students' t-distribution with ν degrees of freedom.

<u>Proof:</u> We know that

$$f_\nu(t) = \frac{1}{\sqrt{\nu}\, B(\frac{1}{2},\frac{\nu}{2})} \qquad \frac{1}{\left[\frac{t^2}{\nu} + 1\right]^{\frac{\nu+1}{2}}} .$$

Thus,

$$P[T > a] = \int_{(t>a)} f_\nu(t)\,dt$$

$$\leq \int_{(t>a)} \frac{t}{a} f_\nu(t)\,dt$$

$$= \frac{1}{\sqrt{\nu}\, B(\frac{1}{2},\frac{\nu}{2})\, a} \int_{(t>a)} \frac{t}{\left[\frac{t^2}{\nu} + 1\right]^{\frac{\nu+1}{2}}} \; dt.$$

Making the substitution $\frac{t^2}{\nu} = \ell$, we obtain

$$P[T>a] \le \frac{\sqrt{\nu}}{2a\ B(\frac{1}{2},\frac{\nu}{2})} \int_{(t > \frac{a^2}{\nu})} \frac{1}{(\ell+1)^{\frac{\nu+1}{2}}}\, d\ell$$

$$= \sqrt{\frac{\nu}{\nu-1}} \quad \frac{1}{a\ B(\frac{1}{2},\frac{\nu}{2})} \quad \frac{1}{[\frac{a^2}{\nu}+1]^{\frac{\nu-1}{2}}}$$

$$= \sqrt{\frac{\nu}{\nu-1}} \quad \frac{1}{a} \cdot \frac{B(\frac{1}{2},\frac{\nu-2}{2})}{B(\frac{1}{2},\frac{\nu}{2})} \cdot \frac{1}{B(\frac{1}{2},\frac{\nu-2}{2})} \quad \frac{1}{[\frac{a^2}{\nu}+1]^{\frac{\nu-2+1}{2}}} ,$$

and so,

$$P[T>a] \le \frac{1}{\sqrt{\frac{(\nu-2)}{\nu}}\, a} \cdot f_{\nu-2}\left(\sqrt{\frac{\nu-2}{\nu}}\, a\right). \qquad \text{(Q.E.D.)}$$

APPENDIX 2 TO CHAPTER III

Theorem:

$$\sup_{0 < n < \infty} \frac{\sqrt{n}}{\left[\frac{\log n}{n} + 1\right]^{\frac{n}{2}}} < e^{\frac{1}{2}} .$$

Proof: We know

$$(1) \qquad \frac{\sqrt{n}\; e^{-\frac{1}{2}}}{\left[\frac{\log n}{n} + 1\right]^{\frac{n}{2}}} = \left[\frac{e^{\frac{1}{n}\log n - \frac{1}{n}}}{\frac{\log n}{n} + 1}\right]^{\frac{n}{2}} .$$

From (1) it is clear that in order to prove the theorem, it is enough to prove that for positive integers n the following inequality holds true:

$$e^{\frac{1}{n}\log n - \frac{1}{n}} < \frac{\log n}{n} + 1,$$

or, equivalently, it is enough to prove that for all positive integers n,

$$(2) \qquad \frac{1}{n}\log n - \frac{1}{n} < \log\left(\frac{\log n}{n} + 1\right).$$

Now for $x > e$, define

$$(3) \quad g(x) = \log\left(\frac{\log x}{x} + 1\right) - \frac{1}{x}\log x + \frac{1}{x} .$$

Then,

$$(4) \quad \frac{dg(x)}{dx} = \frac{1}{x^2[\frac{\log x}{x} + 1]} \left[- 1 - \frac{2 \log x}{x} + \frac{(\log x)^2}{x} \right].$$

For $x > e$, the right hand side of (4) is always negative, and so for $x > e$, the function $g(x)$ is a decreasing function. Further as $x \to \infty$, $g(x) \to \log (1)$ and so we see that for $x > e$

$$g(x) > \log 1,$$

or

$$\left(\frac{\log x}{x} + 1\right) > e^{\frac{\log x}{x} - \frac{1}{x}} .$$

It thus follows from (2) that

$$(5) \quad \sup_{3 \leq n < \infty} \frac{\sqrt{n}}{\left[\frac{\log n}{n} + 1\right]^{\frac{n}{2}}} < e^{\frac{1}{2}} .$$

For $n = 1$ or 2, the inequality

$$\frac{\sqrt{n}}{\left[\frac{\log n}{n} + 1\right]^{\frac{n}{2}}} < e^{\frac{1}{2}}$$

can be numerically verified. This, together with (5), proves the theorem. (Q.E.D.)

CHAPTER IV

ASYMPTOTICS OF POSTERIOR BAYES RISK FOR THE GENERAL CONFIDENCE INTERVAL PROBLEM

4.0. Introduction.

In the general Bayes confidence interval problem we are allowed to observe a sequence $\{X_n\}$ of independent random variables identically distributed according to a common probability law P_θ, which is one of a class $\{P_\theta: \theta \in \Theta\}$ of possible distributions. Nature chooses a fixed θ_o from the parameter space Θ, and we are allowed to observe the X_i's sequentially - these X_i's now have the probability law P_{θ_o}. Our permissible strategy is to keep observing the X_i's according to a stopping rule t which tells us when to stop based on the observations already seen, and when we stop at $t = n$, say, to estimate θ_o with a confidence interval $I_n = [\alpha_{1n}(X_1,\ldots,X_n), \alpha_{2n}(X_1,\ldots,X_n)]$. If we stop after $t = n$ observations and choose the interval I_n, then our loss is

$L(\theta_o, n, I_n) = a\,\ell(I_n) + b\{1 - \delta_{I_n})\} + nc$, where $\ell(I_n)$ is the length of I_n, $a > 0$, $b > 0$, $c > 0$, and $\delta_{I_n}(\theta_o) = 1$ if $\theta_o \varepsilon I_n$, and $= 0$ otherwise. Our goal is to choose a stopping rule t^* and a sequence of intervals $\{I_n^*\}$, where I_n^* is the interval used if we stop at $t^* = n$, so as to minimize the Bayes risk under the prior distribution ψ and the loss function L.

In Chapter 2, we showed for any stopping rule t how to find optimal choices for the intervals I_n. The Bayes stopping rule t^* in general would depend on c (the cost per observation). In Chapter 2, we gave a formula for the posterior Bayes risk Y_n for using the Bayes interval I_n^* when observation ceased after taking n observations. In Chapter 1, we provided conditions under which the class of stopping rules $\{t(c): c > 0\}$ defined by:

$$t(c) = \text{first } n \geq 1 \text{ such that } \left(1 - \frac{f(n)}{f(n+1)}\right) Y_n \leq c,$$

where the function $f(n)$ is specified in the conditions, is asymptotically pointwise optimal as $c \to 0$. The results of Chapter 3 suggest that in fact $f(x)$ should be $\sqrt{x/\log x}$ for $x > e$. In the present chapter we give a set of assumptions on the prior $\psi(\theta)$ and the class of probability laws $\{P_\theta: \theta \varepsilon \Theta\}$ under which this choice of $f(x)$ provides asymptotically pointwise optimal stopping rules for minimizing the posterior Bayes risk.

In order to fulfill our stated aims, we begin in

Section 4.1 by obtaining some basic asymptotic results on the posterior distribution of θ given $X_1, X_2, \ldots, X_n$ and on the posterior Bayes risk of the Bayes fixed-sample confidence intervals I_n^*. This is followed in Section 4.2 by a proof of the asymptotically pointwise optimality of the family of stopping rules $\{t(c): c > 0\}$ defined by $f(x) = \sqrt{x/\log x}$ for $x \geq e$.

4.1 Asymptotic Theory.

Let $\{X_n\}$ be a sequence of independent, identically distributed random variables with common probability law P_θ belonging to the family $\{P_\theta: \theta \in \Theta\}$. We assume that each P_θ has density $f(x|\theta)$ with respect to a σ-finite measure μ. The parameter space, Θ, is assumed to be an open subinterval of the real line. In addition, we adopt the following regularity conditions.

A4.1) There is a prior measure Ψ on Θ which has a density $\psi(\theta)$ with respect to Lebesgue measure. The density $\psi(\theta)$ is assumed to be continuous, positive, and bounded on Θ.

A4.2) For all $\theta_o \in \Theta$,

$$(4.1) \quad 0 < \int \prod_{i=1}^{n} f(X_i|\theta)\, \psi(\theta)\, d\theta < \infty, \quad \text{a.s. } P_{\theta_o}.$$

A4.3) Let $\varphi(\theta, X) = \log f(x|\theta)$. Then it is assumed that the partial derivatives $\partial\varphi(\theta,X)/\partial\theta$ and $\partial^2\varphi(\theta,X)/\partial\theta^2$ exist and are continuous in θ a.s., P_{θ_o} in X for all $\theta_o \varepsilon \Theta$.

A4.4) For each $\theta \varepsilon \Theta$, there exists an $\varepsilon(\theta) > 0$, such that

(4.2) $E_\theta[\sup\{|\frac{\partial^2\varphi(s,X)}{\partial\theta^2}| : |s-\theta| < \varepsilon(\theta),\ s \varepsilon \Theta\}] < \infty.$

[Note: As usual, E_θ denotes computation of the expectation assuming that the true value of the parameter is θ.]

A4.5) The function

(4.3) $A(\theta) = - E_\theta\ (\frac{\partial^2\varphi(\theta,X)}{\partial\theta^2}) = E_\theta\ (\frac{\partial\varphi(\theta,X)}{\partial\theta})^2$

is a positive function of $\theta \varepsilon \Theta$.

A4.6) For all $\theta, \theta_1 \varepsilon \Theta$, $\theta_1 \neq \theta$,

(4.4) $E_\theta[\varphi(\theta_1,X) - \varphi(\theta,X)] < 0.$

If θ_o is a value of θ chosen according to the prior density $\psi(\theta)$, and if $X_1, X_2, \ldots, X_n$ have common density $f(X|\theta_o)$ given θ_o, then under Assumptions A4.1) to A4.6) Bickel and Yahav [1967] have shown that a.s. P_{θ_o} for all large enough n the likelihood equation

$$\sum_{i=1}^{n} \frac{\partial\varphi(\theta,X_i)}{\partial\theta} = 0$$

has a solution $\hat{\theta}_n$,

and

(4.5) $\hat{\theta}_n \to \theta_o$, a.s. P_{θ_o} .

Note: All of the convergence statements for random sequences made in the rest of this chapter will hold true almost surely under P_{θ_o}, unless otherwise noted. Thus, we will not mention the fact of almost sure P_{θ_o} convergence in the rest of this chapter unless a statement of the fact is needed for the sake of clarity.

Let us denote the posterior density of θ given $X_1, X_2, \ldots, X_n$ by $\psi(\theta|\underset{\sim}{X})$. Making a change of variable

(4.6) $t = n^{\frac{1}{2}}(\theta - \hat{\theta}_n)$, and denoting the posterior density of t given $X_1, X_2, \ldots, X_n$ by $\psi^*(t|\underset{\sim}{X})$, we have

(4.7) $\psi^*(t|\underset{\sim}{X}) = n^{-\frac{1}{2}}\, \psi(n^{-\frac{1}{2}} t + \hat{\theta}_n|\underset{\sim}{X})$.

Under our assumptions [A4.1) to A4.6)], Bickel and Yahav [1967; Theorem 2.2] have proved that

(4.8) $\int_{-\infty}^{\infty} |\psi^*(t|\underset{\sim}{X}) - N(t, A^{-1}(\theta_o))|\,dt \to 0,$

where

$$N(t, A^{-1}(\theta_o)) = \sqrt{\frac{A(\theta_o)}{2\pi}}\; e^{-\frac{1}{2}A(\theta_o)t^2}, \; -\infty < t < \infty.$$

Let us define

(4.9) $\nu_n(t) = \exp\{\sum_{i=1}^{n} \varphi(n^{-\frac{1}{2}} t + \hat{\theta}_n, X_i) - \sum_{i=1}^{n} \varphi(\hat{\theta}_n, X_i)\}$.

It is easily verified that

$$(4.10)\quad \psi^*(t|\underset{\sim}{X}) = \frac{\nu_n(t)\ \psi(n^{-\frac{1}{2}}\, t + \hat{\theta}_n)}{c_n},$$

where

$$c_n = \int_{-\infty}^{\infty} \nu_n(t)\ \psi(n^{-\frac{1}{2}}\, t + \hat{\theta}_n|\underset{\sim}{X})\ dt.$$

Bickel and Yahav [1967; proof of Theorem 2.2] demonstrate that

$$(4.11)\quad \int_{-\infty}^{\infty} \psi(n^{-\frac{1}{2}}t + \hat{\theta}_n)|\nu_n(t) - (\frac{2\pi}{A(\theta_o)})^{\frac{1}{2}}\ N(t,A^{-1}(\theta_o))|dt$$

$$\to 0 \quad \text{as} \quad n \to \infty,$$

which, using (4.10) and the fact that the absolute value of an integral is less than or equal to the integral of the absolute value, implies

$$(4.12)\quad |c_n - (\frac{2\pi}{A(\theta_o)})^{\frac{1}{2}} \int_{-\infty}^{\infty} \psi(n^{-\frac{1}{2}}t + \hat{\theta}_n)N(t,A^{-1}(\theta_o))dt|$$

$$\to 0, \quad \text{as} \quad n \to \infty.$$

Now since $\hat{\theta}_n \to \theta_o$ and ψ is bounded and continuous, we can apply the dominated convergence theorem to show that

$$\int_{-\infty}^{\infty} \psi(n^{-\frac{1}{2}}t + \hat{\theta}_n)\ N(t,A^{-1}(\theta_o))dt \to \psi(\theta_o).$$

Hence, from (4.12) we conclude that

$$(4.13)\quad c_n \to (\frac{2\pi}{A(\theta_o)})^{\frac{1}{2}}\ \psi(0).$$

From (4.7) and (4.12),

$$(4.14)\quad \psi(\theta|\underset{\sim}{X}) = n^{-\frac{1}{2}}\ \psi^*(n^{\frac{1}{2}}(\theta - \hat{\theta}_n)|\underset{\sim}{X}) = \frac{n^{\frac{1}{2}}\ \psi(\theta)\nu_n(n^{\frac{1}{2}}(\theta - \hat{\theta}_n))}{c_n}.$$

Thus,

$$\psi(\hat{\theta}_n|\underset{\sim}{X}) = \frac{n^{\frac{1}{2}}\nu_n(0)\psi(\hat{\theta}_n)}{c_n} = \frac{n^{\frac{1}{2}}\psi(\hat{\theta}_n)}{c_n},$$

since $\nu_n(0)\equiv 1$ by (4.9). Since $\psi(\theta)$ is continuous and since $\hat{\theta}_n \to \theta_o$, it follows that $\psi(\hat{\theta}_n) \to \psi(\theta_o) > 0$. Using (4.13) and Assumption A4.1), it follows that

(4.15) $\psi(\hat{\theta}_n|\underset{\sim}{X}) \to \infty$, as $n \to \infty$.

Let $\theta_1 \neq \theta_o$ be a fixed element of Θ. From (4.14) we have

(4.16) $$\psi(\theta_1|\underset{\sim}{X}) = \frac{n^{\frac{1}{2}}\psi(\theta_1)\ \nu_n(n^{\frac{1}{2}}(\theta_1 - \hat{\theta}_n))}{c_n}.$$

However, from (4.9)

(4.17) $$\log \nu_n(n^{\frac{1}{2}}(\theta_1-\hat{\theta}_n)) = \sum_{i=1}^{n}[\varphi(\theta_1,X_i)-\varphi(\hat{\theta}_n,X_i)]$$

$$\leq \sum_{i=1}^{n}[\varphi(\theta_1,X_i) - \varphi(\theta_o,X_i)],$$

where the inequality in (4.17) follows from the definition of $\hat{\theta}_n$ as the maximum likelihood estimator. Now, from A4.6)

$$E_{\theta_o}[\varphi(\theta_1,X_i) - \varphi(\theta_o,X_i)] = \delta(\theta_1) < 0,$$

for $i = 1,2,\ldots,n$. Thus, for any $\delta > 0$ such that $\delta(\theta_1) + \delta < 0$, using the S.L.L.N., we know that for all large enough n

$$\frac{1}{n}\sum_{i=1}^{n}[\varphi(\theta_1,X_i) - \varphi(\theta_o,X_i)] < \delta(\theta_1) + \delta.$$

Thus, from (4.17), we have that for all large enough n,

$$\log \nu_n(n^{\frac{1}{2}}(\theta_1 - \hat{\theta}_n)) \leq e^{n(\delta(\theta_1) + \delta)} .$$

Hence, from (4.16) we find that

$$\psi(\theta_1|\underset{\sim}{X}) \leq \frac{n^{\frac{1}{2}} e^{n(\delta(\theta_1) + \delta)} \psi(\theta_1)}{c_n} .$$

Since $\delta(\theta_1) + \delta < 0$ and since c_n converges to a positive constant, we conclude that for $\theta_1 \neq \theta_o$

(4.18) $\psi(\theta_1|\underset{\sim}{X}) \to 0$ as $n \to \infty$.

Suppose now that a and b are given positive constants.

<u>Lemma 4.1.</u> Given $a,b > 0$, for large enough n there exist (a.s. P_{θ_o}) at least two distinct solutions α_{1n} and α_{2n} of the equation

(4.19) $\psi(\alpha|\underset{\sim}{X}) = \frac{a}{b}$.

<u>Proof</u> From A4.1) and A4.3) it is easily demonstrated that $\psi(\theta|\underset{\sim}{X})$ is (a.s. P_{θ_o}) continuous in θ. Using (4.15) and (4.18), we see that almost surely for large enough n the posterior density $\psi(\theta|\underset{\sim}{X})$ will exceed a/b near $\hat{\theta}_n$ and will be near 0 for θ in a neighborhood of $\theta_o - \epsilon$ and in a neighborhood of $\theta_o + \epsilon$, where $\epsilon > 0$ is a fixed positive constant. The lemma now follows using the continuity of $\psi(\theta|\underset{\sim}{X})$ in θ.

To ensure that <u>exactly</u> two distinct solutions α^*_{1n} and α^*_{2n}, $\alpha^*_{1n} < \alpha^*_{2n}$, of (4.19) exist (a.s. P_{θ_o}) for large enough n, we make the following assumption.

<u>A4.7</u>) For every n, the posterior density of θ given $X_1, X_2, \ldots, X_n$, denoted by $\psi(\theta|\underset{\sim}{X})$, is almost surely strictly unimodal in θ.

We now proceed to study the asymptotic behavior as $n \to \infty$ of the two solutions α^*_{1n} and α^*_{2n} of (4.19).

Define $B^*_{in} = n^{\frac{1}{2}}(\alpha^*_{in} - \hat{\theta}_n)$. Then the B^*_{in} satisfy

$$(4.20) \quad \psi^*(B^*_{in}|\underset{\sim}{X}) = \frac{a}{b\sqrt{n}} .$$

Notice that since $\psi^*(\theta|\underset{\sim}{X})$ is obtained from $\psi(\theta|\underset{\sim}{X})$ by a non-trivial change of location and scale, and since $\psi(\theta|\underset{\sim}{X})$ is strictly unimodal, $\psi^*(\theta|\underset{\sim}{X})$ is also strictly unimodal. Thus, (4.20) almost surely for large enough n has exactly two solutions B^*_{1n} and B^*_{2n}, $B^*_{1n} < B^*_{2n}$.

<u>Lemma 4.2</u>. Under Assumptions A4.1) to A4.7),

$$(4.21) \quad B^*_{1n} \to -\infty, \ \beta^*_{2n} \to \infty \ \text{ as } n \to \infty.$$

<u>Proof</u>. We prove the result for B^*_{2n}; the result for B^*_{1n} follows by a similar proof.

Recall from (4.8) that

$$(4.22) \quad \int_{-\infty}^{\infty} |\psi^*(t|\underset{\sim}{X}) - N(t, A^{-1}(\theta_o)|dt \to 0 \ \text{ as } n \to \infty.$$

Suppose $\{x_n\}$ is a sequence of observations for which (4.22) holds. Then for this sequence $\{x_n\}$, we show that (4.21) must also hold. Our proof is by contraposition.

Thus, suppose there exists a sequence $\{x_n\}$ for which (4.21) does not hold $(i = 2)$. Now $B^*_{2n} \not\to \infty$ implies that there exists $0 < M < \infty$ and a subsequence $\{k_n\}$, $k \to \infty$ as $n \to \infty$, for which

(4.23) $B^*_{2,k_n} < M$, all n.

From (4.20) and the fact that ψ^* is strictly unimodal, we know that for all $t > B^*_{2n}$,

$$\psi(t|x_1,x_2,\ldots,x_{k_n}) < \frac{a}{b\sqrt{k_n}} .$$

In particular, for $M \leq t \leq M + 1$,

(4.24) $\psi^*(t|x_1,\ldots,x_{k_n}) < \dfrac{a}{b\sqrt{k_n}}$.

Thus, for $M \leq t \leq M + 1$, we have by (4.23) and (4.24) that $|\psi^*(t|x_1,\ldots,x_{k_n}) - N(t,A^{-1}(\theta_o))| \to N(t,A^{-1}(\theta_o))$, as $n \to \infty$.

Applying the dominated convergence theorem, we get

(4.25) $\displaystyle\int_M^{M+1} |\psi^*(t|x_1,\ldots,x_n)-N(t,A^{-1}(\theta_o))|\,dt$

$$\to \int_M^{M+1} N(t,A^{-1}(\theta_o))\,dt > 0 .$$

This in turn implies that

(4.26) $\int |\psi^*(t|x_1,\ldots,x_n) - N(t,A^{-1}(\theta_o))|dt \not\to 0.$

Thus, on this sequence $\{x_n\}$, (4.22) does not hold, which by contraposition proves the lemma. (Q.E.D.)

Lemma 4.3. Under Assumptions A4.1) to A4.7),

(4.27) $(\alpha^*_{2n} - \hat{\theta}_n) \to 0, \quad (\hat{\theta}_n - \alpha^*_{1n}) \to 0, \quad \text{as} \quad n \to \infty.$

Proof. We prove only that $(\alpha^*_{2n} - \hat{\theta}_n) \to 0$.

The proof that $(\hat{\theta}_n - \alpha^*_{1n}) \to 0$ follows similar steps and is omitted.

Let q_n be the larger solution of the equation

(4.28) $N(q,A^{-1}(\theta_o)) = \dfrac{a}{b\sqrt{n}}$.

[As long as $n > (2\pi)^{-1} a^2 b^{-2} A(\theta_o)$, exactly two solutions of (4.28) exist since $N(t,A^{-1}(\theta_o))$ is strictly unimodal and $N(t,A^{-1}(\theta_o)) \le N(0,A^{-1}(\theta_o)) = (A(\theta_o)/2\pi)^{1/2}$.]

Thus,

(4.29) $q_n = A^{-1/2}(\theta_o)\{\log n + \log \dfrac{A(\theta_o)\ b^2}{2\pi\ a^2}\}^{1/2}$.

From Lemma 4.2, $B^*_{2n} \to \infty$ as $n \to \infty$, and thus $B^*_{2n} > 0$ for large enough n. Assume n is large enough so that $B^*_{2n} > 0$.

Let J_n denote the interval $[\min(B^*_{2n},q_n),\ \max(B^*_{2n},q_n)]$.

Notice that on J_n, one of $\psi^*(t|\underset{\sim}{X})$ and $N(t,A^{-1}(\theta_o))$

always exceeds a/b, while the other is always exceeded by a/b. Thus, it follows that for $t \in J_n$,

(4.30) $$|\psi^*(t|\underset{\sim}{X}) - N(t,A^{-1}(\theta_o))| \geq |\frac{a}{b\sqrt{n}} - N(t,A^{-1}(\theta_o))|.$$

Using (4.22) and (4.30), we find that

(4.31) $$\int_{J_n} |\frac{a}{b\sqrt{n}} - N(t,A^{-1}(\theta_o))| \, dt \to 0 .$$

Now, for every n, the definition of q_n implies that for each $t \in J_n$,

(4.32) $$\text{sign}(\frac{a}{b\sqrt{n}} - N(t,A^{-1}(\theta_o))) = \text{sign}\,(B^*_{2n} - q_n),$$

and so

(4.33) $$\int_{J_n} |\frac{a}{b\sqrt{n}} - N(t,A^{-1}(\theta_o))|dt$$

$$= |\frac{a}{b\sqrt{n}} |B^*_{2n} - q_n| - \int_{J_n} N(t,A^{-1}(\theta_o))dt| .$$

However, we know that $B^*_{2n} \to \infty$ and $q_n \to \infty$ as $n \to \infty$. Hence, it follows that $\min(B^*_{2n},q_n) \to \infty$ and thus

(4.34) $$\int_{J_n} N(t,A^{-1}(\theta_o))dt \to 0, \quad \text{as} \quad n \to \infty.$$

Also, from (4.29) we have

(4.35) $$\frac{q_n}{\sqrt{n}} \to 0, \quad \text{as} \quad n \to \infty.$$

Thus, using (4.31) and (4.33), we see that

$$\frac{B_{2n}^*}{\sqrt{n}} \cong (\alpha_{2n}^* - \hat{\theta}_n) \to 0, \quad \text{as} \quad n \to \infty. \qquad \text{(Q.E.D.)}$$

<u>Cor. 4.1</u>. Under Assumptions A4.1) to A4.7),

(4.36) $\alpha_{in}^* \to \theta_o$, as $n \to \infty$, $i = 1,2$.

<u>Proof</u>. From Lemma 4.3, we obtain

$$|\alpha_{in}^* - \hat{\theta}_n| \to 0, \quad \text{as} \quad n \to \infty.$$

From (4.5), $|\hat{\theta}_n - \theta_o| \to 0$ as $n \to \infty$. From these two results, (4.36) easily follows. (Q.E.D.)

We now state and prove the two main theorems of this section.

<u>Theorem 4.1</u>. Under Assumptions A4.1) to A4.7),

(4.37) $\sqrt{\frac{n}{\log n}} \; (\alpha_{2n}^* - \alpha_{1n}^*) \to 2\, A^{-\frac{1}{2}}(\theta_o)$.

<u>Proof</u>. We first prove that

(4.38) $\sqrt{\frac{n}{\log n}} \; (\alpha_{2n}^* - \hat{\theta}_n) \to A^{-\frac{1}{2}}(\theta_o)$.

This, of course, is equivalent to proving that

(4.39) $(\log n)^{-\frac{1}{2}} B_{2n}^* \to A^{-\frac{1}{2}}(\theta_o)$.

Now, from the definition of B_{2n}^*, we know

$$\psi^*(B_{2n}^*|\underset{\sim}{X}) = \frac{a}{b\sqrt{n}}.$$

Using (4.10), we obtain

$$\frac{\nu_n(B^*_{2n})\ \psi(n^{-\frac{1}{2}} B^*_{2n} + \hat{\theta}_n)}{c_n} = \frac{a}{b\sqrt{n}} ,$$

or

$$\frac{\nu_n(B^*_{2n})\ \psi(\alpha^*_{2n})}{c_n} = \frac{a}{b\sqrt{n}} .$$

From (4.9),

$$\log \nu_n(B^*_{2n}) = \sum_{i=1}^{n} \varphi(n^{-\frac{1}{2}} B^*_{2n} + \hat{\theta}_n, X_i) - \sum_{i=1}^{n} \varphi(\hat{\theta}_n, X_i).$$

Expanding $\varphi(n^{-\frac{1}{2}} B^*_{2n} + \hat{\theta}_n, X_i)$ in a Taylor series around $\hat{\theta}_n$,

$$\log \nu_n(B^*_{2n}) = \sum_{i=1}^{n} \{ n^{-\frac{1}{2}} B^*_{2n} \frac{\partial\varphi(\hat{\theta}_n, X_i)}{\partial\theta} + \frac{1}{2} \frac{(B^*_{2n})^2}{n} \frac{\partial^2\varphi(\theta^*_n, X_i)}{\partial\theta^2} \} ,$$

where $\theta^*_n \in [\hat{\theta}_n, \hat{\theta}_n + n^{-\frac{1}{2}} B^*_{2n}] = [\hat{\theta}_n, \alpha^*_{2n}]$. It follows that

$$(4.41)\quad \log \nu_n(B^*_{2n}) = \frac{1}{2} \frac{(B^*_{2n})^2}{n} \sum_{i=1}^{n} \frac{\partial^2\varphi(\theta^*_n, X_i)}{\partial\theta^2} ,$$

because from the definition of $\hat{\theta}_n$, we have

$$\sum_{i=1}^{n} \frac{\partial\varphi(\hat{\theta}_n, X_i)}{\partial\theta} = 0.$$

From Corollary 4.1 and (4.5), $\alpha^*_{2n} \to \theta_o$ and $\hat{\theta}_n \to \theta_o$ as $n \to \infty$. Thus, $\theta^*_n \to \theta_o$ as $n \to \infty$.

Further, from A4.3) and A4.4), using the monotone convergence theorem, it can be verified that $\frac{\partial^2 \varphi(\theta,x)}{\partial\theta^2}$ is a super-continuous function of θ. [Super-continuity of a random function is defined in DeGroot (1970, Section 10.6) and the proof of the asserted super-continuity of $\frac{\partial^2 \varphi(\theta,X)}{\partial\theta^2}$ follows as in the proof of Theorem 10.6.1, p.206, of DeGroot's book (1970).] Now, using Theorem 10.8.1 of DeGroot [1970], we can establish that

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \varphi(\theta_n^*, X_i)}{\partial\theta^2} \to -A(\theta_o), \quad \text{a.s. } P_{\theta_o}.$$

Consequently, for large n,

$$(4.42) \quad \log \nu_n(B_{2n}^*) \approx -\tfrac{1}{2}(B_{2n}^*)^2 A(\theta_o).$$

Now, taking natural logarithms on both sides of (4.40), we get

$$(4.43) \quad \log \nu_n(B_{2n}^*) + \log \psi(\alpha_{2n}^*) - \log c_n = \log \frac{a}{b} - \tfrac{1}{2} \log n,$$

and dividing (4.43) through by $\log n$, and then taking the limit as $n \to \infty$, we find from (4.13) and (4.42) that

$$\frac{(B_{2n}^*)^2}{\log n} A(\theta_o) \to 1,$$

since ψ is a bounded function. This result in turn implies that

$$\frac{B_{2n}^*}{\sqrt{\log n}} \to [A(\theta_o)]^{-\frac{1}{2}},$$

since B_{2n}^* is positive for large enough n (Lemma 4.1).

Using a similar argument, it can be shown that

$$\frac{-B_{1n}^*}{\sqrt{\log n}} \to [A(\theta_o)]^{-\frac{1}{2}} .$$

Putting the two results together gives

$$\sqrt{\frac{n}{\log n}}\,(\alpha_{2n}^* - \alpha_{1n}^*) = \frac{1}{\sqrt{\log n}}\,(B_{2n}^* - B_{1n}^*) \to 2[A(\theta_o)]^{-\frac{1}{2}}.$$

(Q.E.D.)

<u>Theorem 4.2</u>. Under Assumptions A4.1) to A4.7),

$$(4.44)\quad \sqrt{\frac{n}{\log n}} \int_{\alpha_{2n}^*}^{\infty} \psi(\theta|\underset{\sim}{X})\,d\theta \to 0,$$

and

$$(4.45)\quad \sqrt{\frac{n}{\log n}} \int_{-\infty}^{\alpha_{1n}^*} \psi(\theta|X)\,d\theta \to 0.$$

<u>Proof</u>. We will prove (4.44). A similar argument shows that (4.45) holds.

From (4.7) and the definition of B_{2n}^*,

$$(4.46)\quad \sqrt{\frac{n}{\log n}} \int_{\alpha_{2n}^*}^{\infty} \psi(\theta|\underset{\sim}{X})\,d\theta = \sqrt{\frac{n}{\log n}} \int_{B_{2n}^*}^{\infty} \psi^*(t|\underset{\sim}{X})\,dt .$$

From Equation (2.35) of Bickel and Yahav [1968], we know that for every $\delta > 0$ there exists $\varepsilon(\delta) > 0$ such that

$$\int_{\delta\sqrt{n}}^{\infty} \psi^*(t|\underset{\sim}{X})dt = \frac{1}{c_n} \int_{\delta\sqrt{n}}^{\infty} \psi(n^{-\frac{1}{2}} t + \hat{\theta}_n) V_n(t)dt$$

$$\leq \frac{n^{\frac{1}{2}}}{2\, c_n} e^{-n\, \varepsilon(\delta)} (1 + |\hat{\theta}_n|) .$$

Thus

$$(4.47) \quad \sqrt{\frac{n}{\log n}} \int_{\delta\sqrt{n}}^{\infty} \psi^*(t|\underset{\sim}{X})dt \leq \frac{1}{2\, c_n} \frac{n e^{-n\varepsilon(\delta)}(1+|\hat{\theta}_n|)}{\sqrt{\log n}} .$$

Now, from (4.13) and the fact that $\hat{\theta}_n \to \theta_o$, we see that the right hand side of (4.47) goes to 0 as $n \to \infty$. In consequence, for every $\delta > 0$, we have

$$(4.48) \quad \sqrt{\frac{n}{\log n}} \int_{\delta\sqrt{n}}^{\infty} \psi^*(t|\underset{\sim}{X}) \to 0, \quad \text{as} \quad n \to \infty .$$

Bickel and Yahav [1967; Equation (2.40)] further show that for every $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ and $N \equiv N(\delta, x_1, x_2, \ldots)$, possibly depending on the sample sequence $\{x_n\}$, such that $n > N$ and $|n^{-\frac{1}{2}}t| < \delta^* = \delta(\varepsilon)$ implies

$$\log \nu_n(t) \leq t^2(-A(\theta_o) - 2\varepsilon) = -t^2(A(\theta_o) + 2\varepsilon).$$

Thus, for $n > N$ and $|n^{-\frac{1}{2}}t| < \delta^*$, $\nu_n(t) \leq e^{-t^2(A(\theta_o) + 2\varepsilon)}$, or, for $N > n$ and $|n^{-\frac{1}{2}}t| < \delta^*$,

$$\psi^*(t|\underset{\sim}{X}) = \frac{\nu_n(t)\ \psi(n^{-\frac{1}{2}}\,t + \hat{\theta}_n)}{c_n}$$

$$\leq \frac{\psi(n^{-\frac{1}{2}}t + \hat{\theta}_n)}{c_n}\ e^{-t^2(A(\theta_o) + 2\varepsilon)} .$$

Thus,

$$(4.49)\quad \int_{B_{2n}^*}^{\delta\sqrt{n}} \psi^*(t|\underset{\sim}{X})dt \leq \int_{B_{2n}^*}^{\delta\sqrt{n}} \frac{\psi(n^{-\frac{1}{2}}t + \hat{\theta}_o)\ e^{-t^2(A(\theta_o)+2\varepsilon)}dt}{c_n} .$$

By A4.1) ψ is a bounded function of θ. Suppose the bound of ψ is M, then from (4.49) we get

$$(4.50)\quad \int_{B_{2n}^*}^{\delta\sqrt{n}} \psi^*(t|\underset{\sim}{X})dt \leq \frac{M}{c_n} \int_{B_{2n}^*}^{\infty} e^{-t^2(A(\theta_o)+2\varepsilon)}\ dt .$$

Now from Theorem 4.1, we know that

$$B_{2n}^* = 0(\sqrt{\log n}) .$$

Thus, using the well known bound on the tail of the normal distribution function, and the fact that $\{c_n\}$ converges to a constant as $n \to \infty$, we get

$$\int_{B_{2n}^*}^{\delta\sqrt{n}} \psi^*(t|X)dt = 0\left(\frac{1}{\sqrt{n \log n}}\right) .$$

Thus,

$$(4.51)\quad \sqrt{\frac{n}{\log n}} \int_{B_{2n}^*}^{\delta\sqrt{n}} \psi^*(t|\underset{\sim}{X})dt \to 0, \quad \text{as} \quad n \to \infty .$$

Results (4.48) and (4.51) together now imply that

$$\sqrt{\frac{n}{\log n}} \int_{B_{2n}^*}^{\infty} \psi^*(t|\underset{\sim}{X})dt \to 0, \quad \text{as} \quad n \to \infty. \qquad \text{(Q.E.D.)}$$

4.2. Solution for the General Problem.

Let us assume that the assumptions A4.1) to A4.7) are satisfied. Thus among other things, we have assumed that for each n, the posterior density of θ, given by $\psi(\theta|X_1,\ldots,X_n)$, is unimodal and continuous. We now can use the results of Chapter 2, which state that for every stopping rule, the sequence of terminal Bayes confidence intervals is given by $\{I_n^*\}$ defined as follows: Let

(4.52) $$\psi(\theta|X_1,\ldots,X_n) = \frac{a}{b}.$$

Then, define,

(4.53) $$I_n^* = \begin{cases} \text{Empty set, if (4.52) has no solutions.} \\ \{\alpha^*\}, \text{ if } \alpha^* \text{ is the only solution of (4.52).} \\ [\alpha_{1n}^*, \alpha_{2n}^*], \text{ if } \alpha_{1n}^* \text{ and } \alpha_{2n}^* (>\alpha_{1n}^*) \text{ are two} \\ \text{solutions of (4.52).} \end{cases}$$

[Recall that since $\psi(\theta|X_1,\ldots,X_n)$ is strictly unimodal, there can be no more than two solutions to Equation (4.52).]

If the posterior Bayes risk of this sequence of rules $\{I_n^*\}$ is given by $\{Y_n\}$, we know that:

If I_n^* is a null set or a single point set, then

$Y_n = b,$

otherwise if $I_n^* = [\alpha_{1n}^*, \alpha_{2n}^*]$, then

(4.54) $$Y_n = a(\alpha_{2n}^* - \alpha_{1n}^*) + b\{\int_{-\infty}^{\alpha_{1n}^*} \psi(\theta|X_1,\ldots,X_n)d\theta$$

$$+ \int_{\alpha_{2n}^*}^{\infty} \psi(\theta|X_1,\ldots,X_n)d\theta\}.$$

From our earlier discussion in this chapter we already know that under our assumptions, almost surely P_{θ_o} for every sequence of observations, Equation (4.52) will eventually have exactly two solutions and so for large enough n, Y_n will in fact be given by (4.54)

Now for $x \geq 0$, let us define a function f as follows

(4.55) $$f(x) = \begin{cases} \sqrt{x} & \text{if } x \leq e \\ \sqrt{\frac{x}{\log x}} & \text{if } x \geq e. \end{cases}$$

Using Theorem 4.1 and Theorem 4.2, we see that

(4.56) $$f(n)Y_n \to 2a\{A(\theta_o)\}^{-\frac{1}{2}} > 0.$$

Also it is easy to verify that this sequence $\{Y_n\}$ and the function f satisfy all of Assumptions A1.1) to A1.5), of Chapter 1 except Assumption A1.3). Assumption A1.3) is not satisfied only in the sense that $f(n)Y_n$ in this case converges to $2a\{A(\theta_o)\}^{-\frac{1}{2}}$, which is a positive random variable, and not a positive constant, under the prior $\psi(\theta)$.

If we go through the proof of Theorem 1.1 of Chapter 1, the only place where we have used the fact that V is a positive constant is to claim that $t(c)$ is a proper stopping rule and further that $t(c) \to \infty$, a.s., as $c \to 0$. These claims hold even if we assume that V is a random variable such that $P[0 < V < \infty] = 1$. Hence, as far as Theorem 1.1 of Chapter 1 is concerned, we can modify our Assumption A1.3) by

A1.3') $f(n)Y_n \to V$ a.s., as $n \to \infty$, where $P[0 < V < \infty] = 1$ and $f(x)$ is strictly increasing function of x on $[0\ \infty]$. Further if we define

$$F(x) = \frac{x\{f(x+1) - f(x)\}}{f(x+1)}$$

then we assume that $F(x)$ is a bounded function of $[0,\infty]$ such that $F(x) \to M$ as $x \to \infty$, where M is a positive constant.

Now since our $\{Y_n\}$ and f satisfy this modified A1.3), Theorem 1.1 applies in our case, and we can claim that a class of A.P.O. stopping rules for this Bayesian sequential confidence interval is given by $\{t(c):c > 0\}$ where for each $c > 0$, $t(c)$ is defined as follows:

$t(c) = n$, if n is the first integer

such that

$$(4.56)\quad (1 - \frac{f(n)}{f(n+1)})Y_n \le c.$$

We would also like to be able to use Theorem 1.2 in this case, and claim that the family $\{t(c):c > 0\}$ is also A.O. for our problem. However, we can not do this right away, because the proof that we have given for Theorem 1.2 does not go through under the modified Assumption A1.3'). There are two possible ways out of this problem: (i) To modify the proof of Theorem 1.2 so as to make it go through under the Assumption A1.3'); (ii) instead of proving the conditional convergence of $f(n)Y_n$ to $2a(A(\theta_o))^{-1}$ under the condition that θ_o is the true parameter value, we should try to prove the marginal convergence [marginal on the unconditional distribution of the X_n's] to possibly $2a(E(A(\theta))^{-\frac{1}{2}}$, expectation being carried out with respect to the prior distribution of θ. In future, we plan to continue the study of both of these approaches to see if we can in fact prove that the family of stopping rules $\{t(c):c > 0\}$ defined in (4.56) is also A.O. for this general problem of Bayesian sequential confidence intervals.

BIBLIOGRAPHY

(1) Arrow, K., Blackwell, D. and Girshick (1949). Bayes and Minimax Solution of Sequential Decision Problems. Econometrica. 17, 213-244.

(2) Bickel, P.J. and Yahav, J.A. (1967). Some Contributions to the Asymptotic Theory of Bayes Solutions. Technical Report. Stanford University.

(3) Bickel, P.J. and Yahav, J.A. (1967a). Asymptotically Pointwise Optimal Procedures in Sequential Analysis. Proc. Fifth Berk Symp Math. Stat. Prob. 401-413.

(4) Bickel, P.J. and Yahav, J.A. (1968). Asymptotically Optimal Bayes and Minimax Procedures in Sequential Estimation. Ann. Math. Stat. 39. 442-456.

(5) Chow, Y.S. and Robbins, H. (1965). On the Asymptotic Theory of Fixed Width Sequential Intervals for the Mean. Ann. Math. Stat. 36. 36-50.

(6) Chung, K.L. (1968). A Course in Probability Theory. Harcourt, Brace and World, Inc., New York.

(7) DeGroot, M.H. (1970). Optimal Statistical Decisions. McGraw-Hill Book Company.

(8) Dantzig, G.B. (1940). On the Non-existance of Test of Students Hypothesis having Power Function Independent of σ^2. Ann. Math. Stat. 11. 186-192.

(9) Farrell, R.H. (1959). Sequentially Determined Bounded Length Confidence Intervals. Ph.D. thesis, Univ. of Illinois.

(10) Joshi, V.M. (1966). Admissability of Confidence Intervals. Ann. Math. Stat. 37. 629-638.

(11) Kiefer, J. and Sacks, J. (1963). Asymptotically Optimal Sequential Inference and Design. Ann. Math. Stat. 34. 705-750.

(12) Lehmann, E.L. (1959). Testing Statistical Hypothesis. John Wiley & Sons, New York.

(13) Loève, M. (1963). Probability Theory. D. Van Nostramd Company, Inc., New York.

(14) Paulson, E. (1969). Sequential Interval Estimation for the Mean of Normal Populations. Ann. Math. Stat. 40. 509-516.

(15) Raiffa, H. and Schlaifer (1961). Applied Statistical Decision Theory. Division of Research Harvard Business School, Boston.

(16) Royden, H.L. (1968). Real Analysis. Second Edition. The MacMillan Company, Collier-MacMillan Limited, London.

(17) Strawderman, W.E. and Cohen, A. (1973). Admissible Confidence Intervals and Point Estimation for Translation or Scale Parameters. Ann. Stat. 1. 545-550.

(18) Stein, C. (1945). A Two-sample Test for a Linear Hypothesis whose Power is Independent of the Variance. Ann. Math. Stat. 16. 243-258.

(19) Stein, C. and Wald, A. (1947). Sequential Confidence Intervals for the Mean of a Normal Distribution with Known Variance. Ann. Math. Stat. 18. 427-433.

(20) Winkler, R.L. (1972). Decision Theoretic Approach to Interval Estimation. J. Amer. Stat. Asso. 67. 187-191.

(21) Zacks, S. (1970). The Theory of Statistical Inference. John Wiley & Sons, New York.