ASYMPTOTICALLY OPTIMAL MULTIPARAMETER SEQUENTIAL BAYES REGIONAL ESTIMATION PROCEDURES

by

Richard A. Sundheim
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-15

August, 1979

INTRODUCTION

The need for sequential estimation procedures became evident when Dantzig (1940) showed that a fixed-sample confidence interval for the mean μ of a normal distribution with unknown variance σ^2 cannot achieve both a fixed-width 2d and a specified coverage probability $1-\alpha$ for all values of μ and σ^2. A two-stage sequential estimator for μ was proposed by Stein (1945), and Stein and Wald (1947), Stein (1949), and Anscombe (1952, 1953), among others, suggested some general sequential procedures.

For general univariate distributions, Chow and Robbins (1965) give a fixed-width confidence interval procedure for the mean whose stopping rule is required to have certain properties asymptotically as $d \to 0$ with α fixed. Serfling and Wackerly (1976) attack this same problem with a different sequential approach whose asymptotic analysis lets $\alpha \to 0$ with d fixed.

To achieve a rational balance in choosing d and α when estimating the parameter of a one-parameter family of distributions, Gleser and Kunte (1976) adopt a sequential Bayes approach where the loss function is a linear combination of the length of the interval, the indicator function for noncoverage, and the sample size. They develop stopping rules which are asymptotically pointwise optimal (A.P.O.) and asymptotically optimal (A.O.) in the sense used by Kiefer and Sacks (1963) and Bickel and Yahav (1967, 1968, 1969a) for sequential Bayes hypothesis testing and point estimation problems, respectively. Here, the asymptotic analysis is carried out by letting the cost, c, per observation tend to 0.

Extending the sequential fixed-width confidence intervals to higher dimensional parameters requires defining the shape (sphere, ellipsoid, rectangle, cone, etc.) and size (maximum diameter, volume, circumference, etc.) of the region to be formed. Chatterjee (1959) extends Stein's two-stage procedure to the multivariate normal case. Fully sequential fixed-size confidence regions for a vector of parameters have been proposed by Sidak (1967) and Callahan (1969) using rectangles, Gleser (1965, 1966) and Srivastava (1967) using spheres, and Albert (1966) preferring ellipsoids. Their asymptotic analysis lets the size of the region tend to 0 while the confidence remains fixed.

Chapter 1 of the present thesis extends Gleser and Kunte's (1976) approach to the multiparameter case by developing A.P.O. and A.O. sequential Bayes regional estimation procedures for an r-dimensional vector of parameters in the presence of nuisance parameters. The "size" of the region is defined to be its volume. A balance between size, noncoverage probability and cost of sampling is achieved by taking as the loss function a linear combination of the volume, the indicator function for noncoverage, and the sample size.

In Section 1, the fixed-sample Bayes regional estimation problem is solved. That is, given n observations, $\underset{\sim}{X}_n$, representations are found for the optimal region, $C^*(\underset{\sim}{X}_n)$, and for the posterior Bayes risk Y_n. Section 2 looks at the sequential case. Since an explicit form of the Bayes optimal stopping rule appears intractible, a search for A.P.O. and A.O. stopping rules begins.

Among the set of sufficient conditions given by Gleser and Kunte (1976) for stopping rules to be A.P.O. and A.O. is that

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

where f(n) is a strictly increasing, positive function and V an almost surely positive random variable. They then show that the proper f(n) for their interval estimation problem is $(n/\log n)^{1/2}$. The example of Section 3 finds A.P.O. and A.O. sequential Bayes regional estimation procedures for estimating the first r components of the mean vector of a p-variate normal distribution when using the conjugate prior. This example suggests that the correct f(n) in the multiparameter case is $(n/\log n)^{r/2}$. That this is indeed the case is shown in Sections 4 and 5 where A.P.O. and A.O. stopping rules for the regional estimation problem are developed. Comments are made in Section 6 toward generalizing these results to the case of estimating vector-valued functions of the parameters.

The special case of estimating the mean of a univariate normal distribution, with the unknown variance acting as a nuisance parameter, is treated in Kunte (1973). Unfortunately, Kunte's stopping rule is complicated, and it appears difficult even to find a recursive formula for utilizing this procedure in practice. Indeed, in general, to apply the procedures of Chapter 1, the form of the stopping rule needs to be simplified. This simplication is the goal of Sections 1 and 2 of Chapter 2 where a theory is developed for approximating the stopping rules of Gleser and Kunte (1976, Theorems 3.1 and 4.1) in such a way that the new approximate stopping rules retain A.P.O. and A.O. characteristics. Much of the motivation for this theory came from the A.P.O. proof in Chapter 1 which indicates that asymptotically, as $n \to \infty$, the posterior Bayes risk of the Bayes optimal terminal decision rule $C^*(\underline{X}_n)$ is dominated by the volume, $V(C^*(\underline{X}_n))$ of $C^*(\underline{X}_n)$, in the sense that the volume $V(C^*(\underline{X}_n)) = O((n/\log n)^{-r/2})$, while the posterior probability of noncoverage is $o((n/\log n)^{-r/2})$. Returning to the estimation problem, it is shown that the approximate stopping rule which is based only on the volume, completely ignoring the posterior probability of noncoverage, is A.P.O., as well as being much less complicated. Comments toward a solution of the A.O. character of this stopping rule are presented.

Finally, in Section 2.3, the results on the approximate stopping rules are used to show that the stopping rules of Chapter 1 are robust with respect to the prior distribution, in the sense that if the prior information is not too badly misspecified, the stopping rules are still A.P.O. and A.O.

CHAPTER 1

MULTIPARAMETER SEQUENTIAL BAYES REGIONAL ESTIMATION

For general multivariate probability distributions asymptotically pointwise optimal (A.P.O.) and asymptotically optimal (A.O.) sequential Bayes regional estimation procedures are developed for an r-dimensional vector of parameters in the presence of nuisance parameters. In Section 1, the fixed-sample Bayes regional estimation problem is solved. Section 2 looks at the sequential case and indicates a need to take an asymptotic approach. The example of Section 3 finds A.P.O. and A.O. sequential Bayes regional estimation procedures for estimating the first r components of the mean vector of a p-variate normal distribution when using the conjugate prior. A.P.O. and A.O. stopping rules for the general regional estimation problem are developed in Sections 4 and 5. Section 6 makes comments toward generalizing these results to the case of estimating an r-dimensional vector-valued function of $p(p \geq r)$ parameters.

1.1. The Fixed-sample Problem

Suppose that we observe independent identically distributed q-dimensional random vectors $X_1, X_2, \ldots$ whose common probability measure P_θ belongs to a family $\{P_\theta: \theta \in \Theta\}$ of measures defined on a measure space $(\mathcal{X}, \mathcal{B})$, indexed by an open subset Θ of p-dimensional Euclidean space and dominated by a σ-finite measure μ. Let $f(x|\theta) = dP_\theta/d_\mu$ be the density function of P_θ with respect to μ and let $\psi(\theta)$ be a prior density (with respect to Lebesgue measure) on Θ.

A regional estimator is desired for θ_r, the r-dimensional $(r \leq p)$ vector consisting of the first r components of θ. The terminal action space is taken to be the class $\mathcal{C}$ of all Lebesgue measurable subsets of $\Theta_r = \{\theta_r: (\theta_r', \theta_{p-r}')' \in \Theta\}$, the parameter space of θ_r. A regional estimation procedure, $C(\underline{X}_n)$, is a measurable assignment of regions $C \in \mathcal{C}$ to samples $\underline{X}_n = (X_1, X_2, \ldots, X_n)$. The loss function chosen is

$$(1.1.1) \quad L(\theta_r, C(\underline{X}_n)) = aV(C(\underline{X}_n)) + b(1-\delta_{C(\underline{X}_n)}(\theta_r)) + cn,$$

where a, b, c are finite positive constants, $V(C(\underline{X}_n))$ is the Lebesgue measure (volume) of $C(\underline{X}_n)$ and $\delta_{C(\underline{X}_n)}(\theta_r) = 1$ if $\theta_r \in C(\underline{X}_n)$ and 0 otherwise.

To determine the fixed-sample Bayes procedure, $C^*(\underline{X}_n)$, against the prior ψ, only procedures with finite Bayes risk need be considered. [There is at least one procedure with finite Bayes risk - namely, $C(\underline{X}_n) \equiv$ empty set, all $\underline{X}_n$.] Using the Fubini-Tonelli theorem, it is straightforward to show that the Bayes risk of a regional estimator $C(\underline{X}_n)$ is given by

$$R(\psi,C(\underline{X}_n)) = \int_{\Theta}[\int_{\mathcal{X}^n} L(\theta_r,C(\underline{X}_n)) \prod_{i=1}^{n} f(x_i|\theta)d\underline{X}_n]\psi(\theta)d\theta$$

$$(1.1.2) \qquad = \int_{\mathcal{X}^n} [\int_{\Theta} L(\theta_r,C(\underline{X}_n))\psi(\theta|\underline{X}_n)d\theta]dG_n(\underline{X}_n)$$

$$= \int_{\mathcal{X}}[\rho(\psi,C(\underline{X}_n))+cn]dG_n(\underline{X}_n),$$

where

$$(1.1.3) \qquad \rho(\psi,C(\underline{X}_n)) = aV(C(\underline{X}_n))+b[1-\int_{C(\underline{X}_n)} \psi_r(\theta_r|\underline{X}_n)d\theta_r]$$

$$= b[1+\int_{C(\underline{X}_n)} (ab^{-1}-\psi_r(\theta_r|\underline{X}_n))d\theta_r]$$

is the posterior Bayes risk of $C(\underline{X}_n)$, $\psi_r(\theta_r|\underline{X}_n)$ is the marginal posterior density of θ_r given $\underline{X}_n$, and $G_n(\underline{X}_n)$ is the marginal distribution function of $\underline{X}_n$. From (1.1.2) and (1.1.3) a fixed-sample Bayes regional estimation procedure is seen to be:

$$(1.1.4) \qquad C^*(\underline{X}_n) = \text{closure (in } \Theta_r\text{) of } \{\theta_r\colon \psi_r(\theta_r|\underline{X}_n) \geq ab^{-1}\},$$

since (1.1.4) clearly minimizes (1.1.3) and can be shown to be a measurable assignment of subsets of Θ_r to samples $\underline{X}_n$.

1.2. The Sequential Problem

Now consider the sequential Bayes regional estimation problem with decisions as pairs $(C(\underset{\sim}{X}_t),t)$, where t is a stopping rule, and $C(\underset{\sim}{X}_t)$ is a terminal decision rule (a subregion of Θ_r). It follows from Arrow, Blackwell and Girshick (1949) that the "conditional" Bayes optimal decision, when t = n, is $(C^*(\underset{\sim}{X}_n),n)$. The "unconditional" Bayes optimal decision is $(C^*(\underset{\sim}{X}_{t^*}),t^*)$ where t* is the stopping rule which minimizes the Bayes risk

$$(1.2.1)\quad R(\psi,C^*(\underset{\sim}{X}_t),t) = \sum_{n=0}^{\infty} \int_{\{t=n\}} [\rho(\psi,C^*(\underset{\sim}{X}_n))+cn]dG_n(\underset{\sim}{X}_n)$$
$$= E(Y_t+ct),$$

where Y_n is the posterior Bayes risk of $C^*(\underset{\sim}{X}_n)$ and the expectation of Y_t+ct is taken over the joint distribution of θ_r and $X_1,X_2,\ldots$. This minimization problem appears to be intractible in general. Thus, following Gleser and Kunte (1976), it is assumed that the cost c of sampling is very small ($c \to 0$). A search is then made for asymptotically pointwise optimal (A.P.O.) and asymptotically optimal (A.O.) stopping rules in the sense defined in Bickel and Yahav (1967, 1968). That is, a class $\{t(c): c > 0\}$ of stopping variables is said to be A.P.O. if

$$(1.2.2)\qquad P\{\lim_{c\to 0} \frac{X(t(c),c)}{\inf_{s\in T} X(s,c)} = 1\} = 1,$$

and A.O. if

(1.2.3) $$\lim_{c\to 0} \left\{\frac{E(X(t(c),c))}{\inf_{s\in T} E(X(s,c))}\right\} = 1,$$

where $X(n,c) = Y_n + cn$, and T is the class of all permissible stopping rules.

In order to make our exposition as self-contained as possible, Gleser and Kunte's (1976) Theorem 3.1 and Theorem 4.1, along with their sufficient conditions, are restated here. Recall that in their paper $\{Y_i,\ i = 1,2,\ldots\}$ is a sequence of observable random variables which come to us one at a time. We wish to choose a stopping rule t based on $Y_1, Y_2, \ldots$ under the loss function $Y_t + cK(t)$ and under assumptions B.0 - B.3' given below.

Assumptions (Gleser and Kunte (1976)):

B.0. $P\{Y_n > 0\} = 1$, all n, and $P\{\lim_{n\to\infty} Y_n = 0\} = 1$.

B.1. There exists a strictly increasing, positive function f(x) defined on $[0,\infty)$ and an almost surely positive random variable V such that

$$P\{\lim_{n\to\infty} f(n)Y_n = V\} = 1.$$

B.2. For each $x > 0$ and $c > 0$, there exists an integer $N(x,c)$ which minimizes the function

$$h(x,c,n) = (f(n))^{-1}x + cK(n).$$

Further, N(x,c) may be taken as the first integer n such that

$$\Delta h(x,c,n) = h(x,c,n+1) - h(x,c,n) \geq 0.$$

B.3. The function

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

is bounded, and

$$\lim_{x\to\infty} G(x) = M, \qquad 0 \le M < \infty.$$

B.4. Either $f(x)/f(x+1) \to 1$ or

$$K(x+1)/K(x) \to 1 \quad \text{as} \quad x \to \infty.$$

B.3'. When $K(x) = x$, the function $G(x)$ is bounded and

$$\lim_{x\to\infty} G(x) = M, \qquad 0 < M < \infty.$$

Theorem 1.2.1. (Gleser and Kunte (1976) Theorem 3.1): Under assumptions B.0 - B.4, for each $c > 0$, let

(1.2.4) $t(c)$ = first $n \geq 1$ such that

$$(1 - \frac{f(n)}{f(n+1)})Y_n \leq c\,\Delta K(n),$$

where $\Delta K(n) = K(n+1) - K(n)$. Then the class of stopping rules $\{t(c): c > 0\}$ is A.P.O.

Theorem 1.2.2. (Gleser and Kunte (1976) Theorem 4.1): When $K(x) = x$ and assumptions B.0 - B.2 and B.3' hold, the class $\{t(c): c > 0\}$ of stopping rules defined by

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

is A.P.O. Further, if

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

this class of stopping rules is A.O.

The example of the next section serves to illustrate and motivate a general theory of A.P.O. and A.O. stopping rules for our regional estimation problem. This general theory is then developed in Sections 4 and 5.

1.3. Example

If $X_1, X_2, \ldots, X_n$ are i.i.d. $N_p(\theta,\Sigma)$, Σ is assumed known and positive definite, and θ has a multivariate normal prior $N_p(\mu,\Delta)$, then the posterior density $\psi(\theta|\underset{\sim}{X}_n)$ [cf. DeGroot, p. 175] is $N_p(\hat{\theta}(\underset{\sim}{X}_n), W_n)$, where

$$\hat{\theta}(\underset{\sim}{X}_n) = W_n(\Delta^{-1}\mu + \Sigma^{-1} \sum_{i=1}^{n} X_i),$$

and

$$W_n = (\Delta^{-1} + n\Sigma^{-1})^{-1}.$$

A regional estimator for θ_r will be obtained. The corresponding point estimator is $\hat{\theta}_r(\underset{\sim}{X}_n)$, where $\hat{\theta}_r(\underset{\sim}{X}_n)$ is the vector consisting of the first r coordinates of $\hat{\theta}(\underset{\sim}{X}_n)$. The covariance matrices and their inverses are assumed to be partitioned such that the leading principal submatrix is rxr. Subscripts are used on the covariance submatrices, and superscripts on submatrices of the corresponding inverses. Thus, for example,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} .$$

Integrating out the nuisance parameter vector θ_{p-r} from $\psi(\theta|\underset{\sim}{X}_n)$ yields the marginal posterior density $\psi_r(\theta_r|\underset{\sim}{X}_n)$:

$$(1.3.1) \quad \psi_r(\theta_r|\underset{\sim}{X}_n) = (2\pi)^{-r/2}|W_{n11}|^{-1/2}\exp\{-\tfrac{1}{2}(\theta_r-\hat{\theta}_r(\underset{\sim}{X}_n))'W_{n11}^{-1}(\theta_r-\hat{\theta}_r(\underset{\sim}{X}_n))\}.$$

It is thus clear that

$$\max_{\theta_r} \psi_r(\theta_r|\underset{\sim}{X}_n) = (2\pi)^{-r/2}|W_{n11}|^{-1/2}. \tag{1.3.2}$$

From the definition of W_n and results about patterned matrices [cf. Rao, Problems 2.7 and 2.9, p. 33], it can be shown that

$$\begin{aligned}
|W_{n11}^{-1}| &= |(\Delta^{11}+n\Sigma^{11})-(\Delta^{12}+n\Sigma^{12})(\Delta^{22}+n\Sigma^{22})^{-1}(\Delta^{21}+n\Sigma^{21})| \\
&= n^r|(n^{-1}\Delta^{11}+\Sigma^{11})-(n^{-1}\Delta^{12}+\Sigma^{12})(n^{-1}\Delta^{22}+\Sigma^{22})^{-1}(n^{-1}\Delta^{21}+\Sigma^{21})| \qquad (1.3.3)\\
&= n^r|\Sigma^{11}-\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}|(1+o(1)) \\
&\to \infty \quad \text{as} \quad n \to \infty.
\end{aligned}$$

From (1.1.4), the fixed-sample Bayes procedure is

$$C^*(\underset{\sim}{X}_n) = \{\theta_r: \ (\theta_r-\hat{\theta}_r(X_n))'W_{n11}^{-1}(\theta_r-\hat{\theta}_r(X_n)) \le k\},$$

where

$$k = -2\log(ab^{-1}(2\pi)^{r/2}|W_{n11}|^{1/2}). \tag{1.3.4}$$

Note from (1.3.2) and (1.3.3) that $C^*(\underset{\sim}{X}_n)$, an ellipsoid, is nonempty for large enough n. The volume of $C^*(\underset{\sim}{X}_n)$ is

$$V(C^*(\underset{\sim}{X}_n)) = (k\pi)^{r/2}|W_{n11}|^{1/2}/\Gamma(\tfrac{r}{2}+1). \tag{1.3.5}$$

Using (1.3.3) - (1.3.5), it can be shown that

$$\begin{aligned}
\lim_{n\to\infty}\left(\frac{n}{\log n}\right)^{r/2}V(C^*(\underset{\sim}{X}_n)) &= \frac{(r\pi)^{r/2}|\Sigma^{11}-\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}|^{-1/2}}{\Gamma(\frac{r}{2}+1)} \qquad (1.3.6)\\
&= (r\pi)^{r/2}|\Sigma_{11}|^{1/2}/\Gamma(\tfrac{r}{2}+1), \quad \text{a.s.}
\end{aligned}$$

The probability of noncoverage is now investigated by considering the transformation

(1.3.7) $$t = n^{1/2}(\theta_r - \hat{\theta}_r(\underset{\sim}{X}_n)).$$

This transforms the ellipsoid $C^*(\underset{\sim}{X}_n)$ into the ellipsoid

$$T(\underset{\sim}{X}_n) = \{t\colon t'(nW_{n11})^{-1}t \leq k\}.$$

The posterior density of t, $\psi_r^*(t|\underset{\sim}{X}_n)$, is multivariate normal with mean 0 and covariance matrix nW_{n11}. Applying the inequality in Appendix 1 bounds the probability of noncoverage as follows:

(1.3.8) $$\int_{C^{*c}(\underset{\sim}{X}_n)} \psi_r(\theta_r|\underset{\sim}{X}_n)d\theta_r = \int_{T^c(\underset{\sim}{X}_n)} \psi_r^*(t|\underset{\sim}{X}_n)dt$$

$$\leq \frac{2^{2-r/2}k^{r/2-1}e^{-k/2}}{\Gamma(\frac{r}{2})},$$

where A^c represents the complement of A. From (1.3.3), (1.3.4) and (1.3.8), it follows that

(1.3.9) $$\lim_{n\to\infty} \left(\frac{n}{\log n}\right)^{r/2} \int_{C^{*c}(\underset{\sim}{X}_n)} \psi_r(\theta_r|\underset{\sim}{X}_n)d\theta_r = 0, \quad \text{a.s.}$$

Finally, from (1.3.6) and (1.3.9) the posterior Bayes risk, Y_n, of the Bayes region $C^*(\underset{\sim}{X}_n)$ satisfies

(1.3.10) $$\lim_{n\to\infty} \left(\frac{n}{\log n}\right)^{r/2} Y_n = a(r\pi)^{r/2}|\Sigma_{11}|^{1/2}/\Gamma(\frac{r}{2}+1), \quad \text{a.s.}$$

It should be noted that $f(x) = (x/\log x)^{r/2}$ is well-defined for $x > 1$, positive and strictly increasing for $x > e$ and $(f(x))^{-1}$ is convex for $\log x > (r+2)^{-1}(r+1+\sqrt{5+2r})$. Thus, it is clearly possible to define at least one function $f^*(x)$ which equals $f(x)$ for $\log x > (r+2)^{-1}(r+1+\sqrt{5+2r})$ and which is positive, strictly increasing and is such that $(f^*(x))^{-1}$ is convex on $[0,\infty)$. It is easily seen that $f^*(x)$ and $Y_1, Y_2, \ldots$ satisfy assumptions B.0 - B.2 and B.3' with $M = r/2$. For asymptotic purposes, the beginning few values of $f^*(x)$ are unimportant. Now, since W_n does not depend on X_n, Y_n is not a random variable. Thus (1.3.10) implies that

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

Theorem 1.2.2 now applies to show that the class $\{t(c): c > 0\}$ of stopping rules defined by

$$(1.3.11) \quad t(c) = \text{first } n \geq 2 \text{ such that } \left[1-\left(\frac{n \log(n+1)}{(n+1)\log n}\right)^{r/2}\right]Y_n \leq c$$

is A.P.O. and A.O.

1.4. A.P.O. Sequential Bayes Multiparameter Regional Estimation

The example of Section 3 suggests that establishing sufficient conditions for the posterior density $\psi_r(\theta_r|\underset{\sim}{X}_n)$ to have approximately the form of a multivariate normal density for large n should produce

$$(1.4.1)\quad \lim_{n\to\infty} (n/\log n)^{r/2} Y_n = a(r\pi)^{r/2} |K_r(\theta_0)|^{1/2}/\Gamma(\tfrac{r}{2}+1), \quad \text{a.s. } (P_{\theta_0}),$$

where $K_r(\theta_0)$ is a certain matrix function of the parameter θ, and a.s. (P_θ) refers to the conditional distribution of $X_1, X_2, \ldots,$ given θ. It then will follow from Theorem 1.2.1 that the class of stopping rules defined by (1.3.11) is A.P.O. [To verify assumption B.1 it is necessary to show almost sure convergence with respect to the joint probability distribution of $X_1, X_2, \ldots,$ and θ_0. However, since probabilities are bounded, use of (1.4.1) for all $\theta_0 \in \Theta$ and the dominated convergence theorem will establish the desired results.]

One set of sufficient conditions to make $\psi(\theta|\underset{\sim}{X}_n)$ asymptotically of multivariate normal form is Bickel and Yahav's (1969b) assumptions A2.2, A2.6 - A2.9 which are restated here.

<u>A.1</u>. The prior density $\psi(\theta)$ is continuous, positive and bounded on Θ.

<u>A.2</u>. Let $\varphi(\theta, X) = \log f(x|\theta)$. The partial derivatives

$$\varphi_i^{(1)}(\theta, X) = \frac{\partial\varphi(\theta, X)}{\partial\theta_i}, \quad i = 1,2,\ldots,p,$$

$$\varphi_{jk}^{(2)}(\theta, X) = \frac{\partial^2\varphi(\theta, X)}{\partial\theta_j\partial\theta_k}, \quad j,k = 1,2,\ldots,p,$$

exist and are continuous in θ, a.s. (P_{θ_0}), for all $\theta_0 \in \Theta$.

A.3. For each $\theta \in \Theta$, there exist $\epsilon(\theta) > 0$ such that

$$E_\theta(\sup\{|\varphi_{ij}^{(2)}(\theta,X)|: ||s-\theta|| < \epsilon(\theta),\ s \in \Theta\}) < \infty$$

for all i,j, where the expectation E_θ, is taken over X with respect to the density $f(x|\theta)$.

A.4. $A(\theta) = \langle A_{ij}(\theta)\rangle$ is negative definite for all $\theta \in \Theta$, where

$$A_{ij}(\theta) = E_\theta(\varphi_{ij}^{(2)}(\theta,X)) = -E_\theta(\varphi_i^{(1)}(\theta,X)\varphi_j^{(1)}(\theta,X)).$$

A.5. For all $\theta \in \Theta$ and all $\epsilon > 0$,

$$E_\theta(\sup\{[\varphi(s,X) - \varphi(\theta,X)]: ||s-\theta|| \geq \epsilon,\ s \in \Theta\}) < 0.$$

Note that since $\psi(\theta)$ is a density, and from assumptions A.1 and A.2, it follows that for all $\theta_0 \in \Theta$,

$$0 < \int_\Theta \psi(\theta) \prod_{i=1}^n f(x_i|\theta)d\theta < \infty, \quad \text{a.s. } (P_{\theta_0}).$$

This result corresponds to assumption A2.5 of Bickel and Yahav (1969b).

Theorem 1.4.1. Under assumptions A.1 - A.5, the class $\{t(c): c > 0\}$ of stopping rules defined by (1.3.11) is A.P.O.

Proof. The theorem will follow from Theorem 1.2.1 if (1.4.1) can be established. Fix arbitrary $\theta_0 \in \Theta$ and note that a strongly consistent estimator $\hat{\theta}_n \equiv \hat{\theta}_n(\underset{\sim}{X}_n,\theta_0)$ of θ_0 is constructed in Bickel and Yahav (1969b, Lemma 2.1).

It follows from Theorem 2.2 of Bickel and Yahav (1969b)[1] that the transformation

$$w = (w_r', w_{p-r}')' = n^{1/2}(\theta - \hat{\theta}_n)$$

has a posterior density

(1.4.2) $$\psi^*(w|\underset{\sim}{X}_n) = n^{-p/2}\psi(n^{-1/2}w + \hat{\theta}_n|\underset{\sim}{X}_n)$$

which satisfies

(1.4.3) $$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\psi^*(w|\underset{\sim}{X}_n) - N(w;0,K(\theta_0)|dw = 0, \quad \text{a.s. } (P_{\theta_0}),$$

where $K(\theta_0) = (-A(\theta_0))^{-1}$ and $N(w;\mu,\Sigma)$ is the density function of the $N_p(\mu,\Sigma)$ distribution.

Lemma 1.4.1. The posterior density of w_r,

(1.4.4) $$\psi_r^*(w_r|\underset{\sim}{X}_n) = \int_{-\infty}^{\infty} \psi^*(w|\underset{\sim}{X}_n)dw_{p-r}$$

$$= n^{-r/2}\psi_r(n^{-1/2}w_r + \hat{\theta}_{rn}|\underset{\sim}{X}_n),$$

satisfies

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\psi_r^*(w_r|\underset{\sim}{X}_n) - N(w_r;0,K_r(\theta_0)|dw_r = 0, \quad \text{a.s. } (P_{\theta_0}),$$

where $\hat{\theta}_{rn}$ is the rx1 vector of the first r components of $\hat{\theta}_n$ and $K_r(\theta_0)$ is the leading rxr submatrix of $K(\theta_0)$.

---

[1] The "p/2" in the exponent of (1.4.2) is mistakenly a "1/2" in equation (2.21) of Bickel and Yahav (1969b).

Proof. Equation (1.4.4) follows from (1.4.2). Now, from (1.4.3) and Tonelli's theorem

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\psi_r^*(w_r|\underline{X}_n) - N(w_r;0,K_r(\theta_0))|dw_r$$

$$= \lim_{n\to\infty} \int_{-\infty}^{\infty} |\int_{-\infty}^{\infty} \psi^*(w|\underline{X}_n) - N(w;0,K(\theta_0))dw_{p-r}|dw_r$$

$$\leq \lim_{n\to\infty} \int_{-\infty}^{\infty} |\psi^*(w|\underline{X}_n) - N(w;0,K(\theta_0))|dw = 0, \quad \text{a.s. } (P_{\theta_0}). \quad \square$$

While this lemma is not needed for the proof of Theorem 1.4.1, it does provide motivation for the proof by showing in what sense $\psi_r^*(w_r|\underline{X}_n)$ approaches $N(w_r;0,K_r(\theta_0))$. Thus for sufficiently large n, we expect a neglible difference between the region $C^*(\underline{X}_n)$ and the region formed from the appropriate normal posterior density. However, the case where the posterior density is normal has been handled in the example of Section 3. To make this argument rigorous we establish bounds on $\psi_r^*(w)$ which are then used to trap $(n/\log n)^{r/2}Y_n$ between two quantities, both of which can be made arbitrarily close to the right-hand side of (1.4.1). To this end, let

$$(1.4.5) \qquad \nu_n(w) = \exp\{\sum_{i=1}^{n} (\varphi(n^{-1/2}w+\hat{\theta}_n,X_i) - \varphi(\hat{\theta}_n,X_i))\}$$

and

$$(1.4.6) \qquad c_n = \int_{-\infty}^{\infty} \nu_n(w)\psi(n^{-1/2}w+\hat{\theta}_n)dw.$$

It is easily verified that

(1.4.7) $\psi^*(w|\underset{\sim}{X}_n) = c_n^{-1}\psi(n^{-1/2}w+\hat{\theta}_n)\nu_n(w)$.

From equation (2.29) of Bickel and Yahav (1969b), assumption A.1, the strong consistency of $\hat{\theta}_n$, and the dominated convergence theorem, it can be shown that

(1.4.8) $\lim_{n\to\infty} c_n = \psi(\theta_0)(2\pi)^{p/2}|K(\theta_0)|^{1/2}$, a.s. (P_{θ_0}).

Bickel and Yahav (1969b, equation (2.34)) give the following bound on $\nu_n(w)$: For every $\delta > 0$,

(1.4.9) $\sup\{\nu_n(w):\ ||w|| > \delta n^{1/2}\} \sim \exp\{-n\ \epsilon(\delta)\}$.

Their equation (2.40)[2] gives an upper bound on $\nu_n(w)$ for all w inside a sphere; however, their proof can be straightforwardly extended (see Appendix 2) to provide the following upper and lower bounds on $\nu_n(w)$: For every $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ and $N_1 = N(X_1,X_2,\ldots;\ \theta_0,\ \delta(\epsilon))$ such that for every $n \geq N_1$, and all w in the sphere

$$W_n \equiv W_n(\delta(\epsilon)) = \{w:\ ||w|| < n^{1/2}\delta(\epsilon)\},$$

we have

(1.4.10) $\nu_n(w) \leq \exp\{-\frac{1}{2}w'K^{-1}(\theta_0,+\epsilon)w\}$, a.s. (P_{θ_0}),

and

[2]The right-hand side of equation (2.40) of Bickel and Yahav (1969b) should be divided by 2 and the "-" should be a "+".

(1.4.11) $\nu_n(w) \geq \exp\{-\frac{1}{2} w'K^{-1}(\theta_0,-\epsilon)w\}$, a.s. (P_{θ_0}),

where $K^{-1}(\theta_0,+\epsilon) = K^{-1}(\theta_0) - \epsilon I$ and

$K^{-1}(\theta_0,-\epsilon) = K^{-1}(\theta_0) + \epsilon I$. For ϵ sufficiently small, $K(\theta_0,+\epsilon)$ is positive definite.

Note, from assumption A.1 and the strong consistency of $\hat{\theta}_n$, that there exist two sequences, $L_{in} = L_{in}(X_1,X_2,\ldots;\ \theta_0)$, $i = 1,2$, such that,

$$L_{1n} \leq \psi(n^{-1/2}w + \hat{\theta}_n) \leq L_{2n},$$

(1.4.12) $$\lim_{n\to\infty} L_{1n} = \lim_{n\to\infty} L_{2n} = \psi(\theta_0).$$

A lower bound on $\psi_r^*(w|\underset{\sim}{X}_n)$ will now be obtained.

Lemma 1.4.2. For $n \geq N_1$,

$$\psi_r^*(w_r|\underset{\sim}{X}_n) \geq g_n^-(w_r)$$

where

(1.4.13) $$g_n^-(w_r) = c_n^{-1}L_{1n}(2\pi)^{p/2}[|K(\theta_0,-\epsilon)|^{1/2}N(w_r;\ 0,K_r(\theta_0,-\epsilon)) - (2\pi)^{-r/2}|K(\theta_0)|^{1/2}|K_r(\theta_0)|^{-1/2}e^{-\frac{n}{2}\delta^2(\epsilon)}]$$

and $K_r(\theta_0,-\epsilon)$ is the leading rxr submatrix of $K(\theta_0,-\epsilon)$.

Proof. Letting

$$W_{n,p-r} \equiv W_{n,p-r}(\delta(\epsilon),||w_r||) = \{w_{p-r}: ||w_{p-r}|| \leq \delta n^{1/2} - ||w_r||\},$$

and applying (1.4.7), (1.4.11) and (1.4.12), we have

$$(1.4.14) \quad \psi_r^*(w_r|\underline{X}_n) = \int_{W_{n,p-r}} \psi^*(w|\underline{X}_n)dw_{p-r} + \int_{W^c_{n,p-r}} \psi^*(w|\underline{X}_n)dw_{p-r}$$

$$\geq \int_{W_{n,p-r}} \psi^*(w|\underline{X}_n)dw_{p-r}$$

$$\geq \int_{W_{n,p-r}} c_n^{-1}L_{1n}\exp\{-\tfrac{1}{2}w'K^{-1}(\theta_0,-\epsilon)w\}dw_{p-r}$$

$$= \int_{-\infty}^{\infty} c_n^{-1}L_{1n}\exp\{-\tfrac{1}{2}w'K^{-1}(\theta_0,-\epsilon)w\}dw_{p-r}$$

$$- \int_{W^c_{n,p-r}} c_n^{-1}L_{1n}\exp\{-\tfrac{1}{2}w'K^{-1}(\theta_0,-\epsilon)w\}dw_{p-r}$$

$$= c_n^{-1}L_{1n}|K(\theta_0,-\epsilon)|^{1/2}(2\pi)^{p/2}N(w_r;\ 0,K_r(\theta_0,-\epsilon))$$

$$- \int_{W^c_{n,p-r}} c_n^{-1}L_{1n}\exp\{-\tfrac{1}{2}w'K^{-1}(\theta_0)w\}\exp\{-\tfrac{\epsilon}{2}w'w\}dw_{p-r}.$$

To find an upper bound for this last integral, we bound the integrand by noting that $w_{p-r} \in W^c_{n,p-r}$ if and only if $w \in W^c_n$ and thus $\exp\{-\frac{\epsilon}{2}w'w\} \leq \exp\{-\frac{n\epsilon}{2}\delta^2(\epsilon)\}$. Hence, it follows that

$$\int_{W^c_{n,p-r}} c_n^{-1}L_{1n}\exp\{-\tfrac{1}{2}w'K^{-1}(\theta_0)w\}\exp\{-\tfrac{\epsilon}{3}w'w\}dw_{p-r}$$

$$\leq c_n^{-1}L_{1n}\exp\{-\tfrac{n\epsilon}{2}\delta^2(\epsilon)\}|K(\theta_0)|^{1/2}(2\pi)^{p/2}N(w_r;0,K_r(\theta_0))$$

$$\leq c_n^{-1}L_{1n}\exp\{-\tfrac{n\epsilon}{2}\delta^2(\epsilon)\}|K(\theta_0)|^{1/2}(2\pi)^{\frac{p-r}{2}}|K_r(\theta_0)|^{-1/2}.$$

Putting this last inequality into (1.4.13) completes the proof of the lemma. □

From (1.1.4) and (1.4.4), it is apparent that the region

(1.4.15) $B^*(\underline{X}_n) = \{w_r: \psi_r^*(w_r|\underline{X}_n) \geq ab^{-1}n^{-r/2}, n^{-1/2}w_r+\hat{\theta}_{rn} \in \Theta_r\}$

is the image of $C^*(\underline{X}_n)$ under the transformation from θ to w.

From (1.4.15) and the discussion in Section 1.1, the region $B^*(\underline{X}_n)$ can be seen to be the fixed-sample Bayes procedure against the prior $\psi^*(w)$ and loss function

$$L(w_r,B(\underline{X}_n)) = aV(B(\underline{X}_n))+bn^{r/2}(1-\delta_{B(\underline{X}_n)}(w_r))+ cn,$$

where $B(\underline{X}_n)$ is a Lebesgue measurable subset of the parameter space of w_r. Thus, the posterior Bayes risk of $B^*(\underline{X}_n)$ is

(1.4.16) $\rho(\psi^*,B^*(\underline{X}_n)) = aV(B^*(\underline{X}_n))+bn^{r/2}\int_{B^{*c}(\underline{X}_n)}\psi_r^*(w_r|\underline{X}_n)dw_r.$

It follows from (1.1.4), (1.4.4), (1.4.15) and (1.4.16) that

(1.4.17) $\lim_{n\to\infty}(n/\log n)^{r/2}\rho(\psi,C^*(\underline{X}_n)) = \lim_{n\to\infty}(\log n)^{-r/2}\rho(\psi^*,B^*(\underline{X}_n)).$

We now attempt to establish upper and lower bounds on $\rho(\psi^*, B^*(\underline{X}_n))$ such that they force (1.4.17) to the limit in (1.4.1). Consider the ellipsoids

(1.4.18) $Q_n(+\epsilon) = \{w_r: w_r'K_r^{-1}(\theta_0,+\epsilon)w_r \leq q_{2n}^2\},$

and

$$(1.4.19) \qquad Q_n(-\epsilon) = \{w_r: w_r' K_r^{-1}(\theta_0,-\epsilon)w_r \le q_{1n}^2\},$$

where

$$(1.4.20) \quad q_{2n}^2 = r \log n - \log[ab^{-1}L_{2n}^{-1}c_n(2\pi)^{-(p-r)/2}|K(\theta_0,+\epsilon)|^{-1/2} |K_r(\theta_0,+\epsilon)|^{1/2}]^2,$$

and

$$(1.4.21) \quad q_{1n}^2 = r \log n - \log\{[ab^{-1}L_{1n}^{-1}c_n(2\pi)^{-p/2}+n^{r/2}e^{-\frac{n\epsilon}{2}\delta^2(\epsilon)} (2\pi)^{-r/2}|K(\theta_0)|^{1/2}|K_r(\theta_0)|^{-1/2}\}^2 + \log|K(\theta_0,-\epsilon)|.$$

Now, from (1.4.13), (1.4.19) and (1.4.21), $w_r \in Q_n(-\epsilon)$ if and only if $g_n^-(w_r) \ge ab^{-1}n^{-r/2}$. Thus, for $n \ge N_1$, Lemma 1.4.2 implies that $Q_n(-\epsilon) \subset B^*(\underset{\sim}{X}_n)$ and hence

$$(1.4.22) \qquad aV(Q_n(-\epsilon)) \le \rho(\psi^*,B^*(\underset{\sim}{X}_n)).$$

The property of $B^*(\underset{\sim}{X}_n)$ being Bayes yields the inequality

$$(1.4.23) \qquad \rho(\psi^*,B^*(\underset{\sim}{X}_n)) \le \rho(\psi^*,Q_n(+\epsilon)),$$

and hence

$$(1.4.24) \quad aV(Q_n(-\epsilon)) \le \rho(\psi^*,B^*(\underset{\sim}{X}_n)) \le aV(Q_n(+\epsilon))+bn^{r/2} \int_{Q_n^c(+\epsilon)} \psi_r^*(w_r|\underset{\sim}{X}_n)dw_r.$$

From (1.4.8), (1.4.12), (1.4.18) - (1.4.21), the volumes

$$V(Q_n(-\epsilon)) = (q_{1n}^2\pi)^{r/2}|K_r(\theta_0,-\epsilon)|^{1/2}/\Gamma(\tfrac{r}{2}+1),$$

and

$$V(Q_n(+\epsilon)) = (q_{2n}^2\pi)^{r/2}|K_r(\theta_0,+\epsilon)|^{1/2}/\Gamma(\frac{r}{2}+1),$$

satisfy

$$(1.4.25)\quad \lim_{n\to\infty} (\log n)^{-r/2} V(Q_n(-\epsilon)) = (r\pi)^{r/2}|K_r(\theta_0,-\epsilon)|^{1/2}/\Gamma(\frac{r}{2}+1),$$

and

$$(1.4.26)\quad \lim_{n\to\infty} (\log n)^{-r/2} V(Q_n(+\epsilon)) = (r\pi)^{r/2}|K_r(\theta_0,+\epsilon)|^{1/2}/\Gamma(\frac{r}{2}+1).$$

To handle the probability of noncoverage part of the risk in (1.4.24), define the cylinder set

$$H_n(\epsilon) = \{w\colon w_r \in Q_n(+\epsilon)\}$$

and observe that

$$(1.4.27)\quad \int_{Q_n^c(+\epsilon)} \psi_r^*(w_r|\underline{X}_n)dw_r = \int_{H_n^c(\epsilon)} \psi^*(w|\underline{X}_n)dw$$

$$= \int_{H_n^c(\epsilon)\cap W_n} \psi^*(w|\underline{X}_n)dw + \int_{H_n^c(\epsilon)\cap W_n^c} \psi^*(w|\underline{X}_n)dw.$$

For $n \geq N_1$, it follows from (1.4.7), (1.4.10), (1.4.12) and the inequality in Appendix 1 that

$$\int_{H_n^c(\epsilon)\cap W_n} \psi^*(w|\underline{X}_n)dw \leq c_n^{-1}L_{2n} \int_{H_n^c(\epsilon)} \exp\{-\frac{1}{2}w'K^{-1}(\theta_0,+\epsilon)w\}dw$$

$$= c_n^{-1}L_{2n}|K(\theta_0,+\epsilon)|^{1/2}(2\pi)^{p/2} \int_{Q_n^c(+\epsilon)} N(w_r;0,K_r(\theta_0,+\epsilon))dw_r$$

$$\leq c_n^{-1}L_{2n}|K(\theta_0,+\epsilon)|^{1/2}(2\pi)^{p/2}\, 2^{2-r/2}(q_{2n}^2)^{r/2-1}e^{-1/2q_{2n}^2}/\Gamma(\frac{r}{2}+1).$$

From (1.4.8), (1.4.12), (1.4.20) and this last inequality, it can be seen that

(1.4.28) $$\lim_{n\to\infty} (n/\log n)^{r/2} \int_{H_n^c(\epsilon)\cap W_n} \psi_r^*(w_r|\underline{X}_n)dw_r = 0, \quad \text{a.s. } (P_{\theta_0}).$$

For n large enough, (1.4.2), (1.4.7) and (1.4.9) yield

$$\int_{H_n^c(\epsilon)\cap W_n^c} \psi^*(w|\underline{X}_n)dw \leq M_1 c_n^{-1} e^{-n\epsilon(\delta)} \int_{W_n^c} \psi(n^{-1/2}w+\hat{\theta}_n)dw$$

$$\leq M_1 c_n^{-1} e^{-n\epsilon(\delta)} n^{p/2} \int_{-\infty}^{\infty} \psi(\theta)d\theta$$

$$\leq M_1 c_n^{-1} e^{-n\epsilon(\delta)} n^{p/2},$$

where $M_1 > 0$ is a finite constant. Hence, from (1.4.8) it follows that

(1.4.29) $$\lim_{n\to\infty} (n/\log n)^{r/2} \int_{H_n^c(\epsilon)\cap W_n^c} \psi^*(w|\underline{X}_n)dw = 0, \quad \text{a.s. } (P_{\theta_0}).$$

Combining (1.4.27), (1.4.28) and (1.4.29) gives

(1.4.30) $$\lim_{n\to\infty} (n/\log n)^{r/2} \int_{Q_n^c(+\epsilon)} \psi_r^*(w_r|\underline{X}_n)dw_r = 0, \quad \text{a.s. } (P_{\theta_0}).$$

Now, since $\lim_{\epsilon\to 0}|K_r(\theta_0,-\epsilon)| = \lim_{\epsilon\to 0}|K_r(\theta_0,+\epsilon)| = |K_r(\theta_0)|$, and since ϵ can be chosen arbitrarily small, it follows from (1.4.17), (1.4.24) - (1.4.26) and (1.4.30) that

$$\lim_{n\to\infty} (n/\log n)^{r/2} \rho(\psi, C^*(\underline{X}_n)) = a(r\pi)^{r/2}|K_r(\theta_0)|^{1/2}/\Gamma(\tfrac{r}{2}+1), \text{ a.s. } (P_{\theta_0}).$$

Thus we have established (1.4.1) and hence completed the proof of Theorem 1.4.1. □

Corollary 1.4.1. Under assumptions A.1 - A.5, the Bayes optimal terminal decision rule $C^*(\underline{X}_n)$ satisfies

$$V(C^*(\underline{X}_n)) = O((n/\log n)^{-r/2}) \tag{1.4.31}$$

and

$$\int_{C^{*c}(\underline{X}_n)} \psi_r(\theta_r|\underline{X}_n)d\theta_r = o((n/\log n)^{-r/2}). \tag{1.4.32}$$

Proof. Since $Q_n(-\epsilon) \subset B^*(\underline{X}_n)$, we have from (1.4.24) and (1.4.16) that

$$bn^{r/2} \int_{B^{*c}(\underline{X}_n)} \psi_r^*(w_r|\underline{X}_n)dw_r + aV(Q_n(-\epsilon)) \leq \rho(\psi^*, B^*(\underline{X}_n)).$$

Dividing both sides of this inequality by $(\log n)^{r/2}$ and taking the limit as $n \to \infty$, we observe from (1.4.1), (1.4.17) and (1.4.25) that

$$\lim_{n\to\infty} b\left(\frac{n}{\log n}\right)^{r/2} \int_{B^{*c}(\underline{X}_n)} \psi_r^*(w_r|\underline{X}_n)dw_r \leq \frac{a(r\pi)^{r/2}}{\Gamma(\frac{r}{2}+1)} [|K_r(\theta_0)|^{1/2} - |K_r(\theta_0,-\epsilon)|^{1/2}]. \tag{1.4.33}$$

Now, since $\lim_{\epsilon\to 0}|K_r(\theta_0,-\epsilon)| = |K_r(\theta_0)|$, and since ϵ can be chosen arbitrarily small, it follows that the left-hand side of (1.4.33) is 0. Here, (1.4.32) follows from (1.4.17). Now, we can see that (1.4.31) follows from (1.4.1) and (1.4.32). □

1.5. A.O. Sequential Bayes Multiparameter Regional Estimation.

As in Gleser and Kunte (1976), to show that the class $\{t(c): c > 0\}$ of stopping rules defined by (1.3.11) is A.O., it is sufficient to exhibit a sequence $\{C_n(\underset{\sim}{X}_n)\}$ of regional estimation procedures satisfying

$$(1.5.1) \qquad \sup_n\{(\frac{n}{\log n})^{r/2} \int_{\mathcal{X}^n} \rho(\psi, C_n(\underset{\sim}{X}_n))dG_n(\underset{\sim}{X}_n)\} < \infty.$$

The following assumptions generalize assumptions C.1 - C.3 of Gleser and Kunte (1976):

F.1. There exists an r-dimensional vector-valued function $g(\theta)$ on Θ, a positive integer k, and an r-dimensional vector-valued function $\nu(X_1,X_2,\ldots,X_k)$ from $\mathcal{X}^k$ to the range $g(\Theta)$ of $g(\cdot)$ such that

$$(1.5.2) \qquad E_\theta[\nu(X_1,X_2,\ldots,X_k)] = g(\theta), \quad \text{all } \theta \in \Theta,$$

and

$$(1.5.3) \qquad E_\theta||\nu(X_1,X_2,\ldots,X_k)-g(\theta)||^\alpha < \infty, \quad \text{all } \theta \in \Theta,$$

where $\alpha = \max(r+2+\delta, 2r)$ and $\delta > 0$.

F.2. Let

$$\xi_s(\theta) = \sum_{i=1}^{r} E_\theta|\nu_i(X_1,X_2,\ldots,X_k)-g_i(\theta)|^s,$$

$$\sigma(\theta) = \text{tr } E_\theta\{[\nu(X_1,\ldots,X_k)-g(\theta)][\nu(X_1,\ldots,X_k)-g(\theta)]'\}$$

$$\equiv \text{tr } \Sigma(\theta),$$

$$\gamma_{\alpha_r}(\theta) = (\sigma(\theta))^{-\alpha_r} \sum_{i=1}^{r} E_\theta |\nu_i(X_1,\ldots,X_k)-g_i(\theta)|^{2\alpha_r}.$$

Then

(1.5.4) $$\int_\Theta \xi_{r+2+\delta}(\theta)\psi(\theta)d\theta < \infty$$

and

(1.5.5) $$\int_\Theta \gamma_{\alpha_r}(\theta)\psi(\theta)d\theta < \infty,$$

where $\alpha_r = r$ for $r \geq 2$, $\alpha_1 = 3/2$ and $\nu_i(X_1,X_2,\ldots,X_k)$ and $g_i(\theta)$ are the $i^{\underline{th}}$ components of $\nu(X_1,X_2,\ldots,X_k)$ and $g(\theta)$, respectively.

F.3. The function $g(\theta)$ has an inverse function h: $g(\Theta) \to \Theta$ which satisfies a uniform, Lipschitz condition. That is, assume two positive numbers $\epsilon > 0$ and $M > 0$ exist such that for every a and b in $g(\Theta)$,

$$||a-b|| < \epsilon \Rightarrow ||h(a)-h(b)|| < M||a-b||.$$

For clarity, some inequalities have been included in Appendix 3. An extension of Inequality A.3.2 to the case where $\beta \geq 2$ is given in the next lemma.

Lemma 1.5.1. Let $U_1,U_2,\ldots,U_n$ be i.i.d. with mean μ, variance σ^2, and $\tau = E|U_1-\mu|^{2\beta}$, $\beta \geq 2$. Then there exists constants L_1 and L_2, $0 < L_1, L_2 < \infty$, such that

$$E|s_n^2-\sigma^2|^\beta \leq n^{-\beta/2}\tau[L_1+L_2 n^{-\beta/2}], \quad \text{all } n,$$

where $s_n^2 = n^{-1} \sum_{i=1}^{n} (U_i-\bar{U})^2$ is the sample variance.

Proof. Without loss of generality we can assume that $\mu = 0$. Let $Y_i = U_i^2 - \sigma^2$, and note that by Inequality A.3.1,

$$E|Y_i|^\beta \leq 2^{\beta-1}(E|U_i|^{2\beta} + \sigma^{2\beta}) \leq 2^\beta\tau,$$

while

$$E|s_n^2-\sigma^2|^\beta \leq 2^{\beta-1}(E|n^{-1} \sum_{i=1}^{n} Y_i|^\beta + E|\bar{U}|^{2\beta}).$$

By Inequality A.3.3,

$$E|n^{-1} \sum_{i=1}^{n} Y_i|^\beta \leq n^{-\beta/2}M(\beta)E|Y_1|^\beta,$$

and

$$E|\bar{U}|^{2\beta} \leq n^{-\beta}M(2\beta)\tau.$$

Putting all these inequalities together (with $L_1 = 2^{2\beta-1}M(\beta)$ and $L_2 = 2^{\beta-1}M(2\beta)$), yields the desired result. □

A consequence of assumption F.2 is given below.

Lemma 1.5.2. Equation (1.5.4) implies that

$$\int_\Theta \sigma^{r/2}(\theta)\psi(\theta)d\theta < \infty. \tag{1.5.6}$$

Proof. First note that (1.5.4) implies that

$$\int_\Theta \xi_{r+1}(\theta)\psi(\theta)d\theta < \infty \tag{1.5.7}$$

and

$$\int_\Theta \xi_r(\theta)\psi(\theta)d\theta < \infty. \tag{1.5.8}$$

For $r \geq 2$, applying Jensen's inequality and Inequality A.3.1 to $\sigma^{r/2}(\theta)$ yields

$$\sigma^{r/2}(\theta) = (E_\theta \sum_{i=1}^{r} |\nu_i(X_1,\ldots,X_k)-g_i(\theta)|^2)^{r/2}$$

$$\leq E_\theta(\sum_{i=1}^{r} |\nu_i(X_1,\ldots,X_k)-g_i(\theta)|^2)^{r/2}$$

$$\leq r^{r/2-1} \sum_{i=1}^{r} E_\theta |\nu_i(X_1,\ldots,X_k)-g_i(\theta)|^r.$$

This result together with (1.5.8) proves (1.5.6) for the case when $r \geq 2$. Now, when $r = 1$, note that $\sigma(\theta) = \xi_2(\theta)$ and thus (1.5.7) implies (1.5.6). This completes the proof of Lemma 1.5.2. □

Theorem 1.5.1. Under assumptions A.1 - A.5, F.1 - F.3, the class of stopping rules $\{t(c): c > 0\}$ defined by (1.3.11) is A.0.

Proof. Define $m(n) = [nk^{-1}]$ = greatest integer $\leq nk^{-1}$ and let

(1.5.9) $$\bar{\nu}(\underset{\sim}{X}_n) = (m(n))^{-1} \sum_{i=1}^{m(n)} \nu(X_{(i-1)k+1},\ldots,X_{ik}),$$

(1.5.10) $$S(\underset{\sim}{X}_n) = (m(n))^{-1} \sum_{i=1}^{m(n)} [\nu(X_{(i-1)k+1},\ldots,X_{ik})-\bar{\nu}(\underset{\sim}{X}_n)] [\nu(X_{(i-1)k+1},\ldots,X_{ik})- \bar{\nu}(\underset{\sim}{X}_n)]'.$$

Let

(1.5.11) $$d(\underset{\sim}{X}_n) = M\left(\frac{\log m(n)}{m(n)}\right)^{1/2}(1+\epsilon)^{1/2}(\mathrm{tr}\ S(\underset{\sim}{X}_n))^{1/2}(r+\delta)^{1/2}$$

where $\epsilon > 0$ and $M > 0$ are defined by assumption F.3 and $\delta > 0$ is defined by assumption F.1. Finally, let

$$(1.5.12) \qquad u(\underset{\sim}{X}_n) = h(\bar{\nu}(\underset{\sim}{X}_n)),$$

and

$$(1.5.13) \qquad C_n(\underset{\sim}{X}_n) = \{\theta_r : ||u_r(\underset{\sim}{X}_n) - \theta_r|| \leq d(\underset{\sim}{X}_n)\}$$

where $u_r(\underset{\sim}{X}_n)$ consists of the first r components of $u(\underset{\sim}{X}_n)$. By the Fubini-Tonelli theorem

$$(1.5.14) \qquad \int_{\mathcal{X}^n} \rho(\psi, C_n(\underset{\sim}{X}_n))dG_n(\underset{\sim}{X}_n) = \int_{\Theta}\{\frac{a\pi^{r/2}}{\Gamma(\frac{r}{2}+1)} E_\theta(d^r(\underset{\sim}{X}_n)) + bP_\theta(E_n)\}\psi(\theta)d\theta,$$

where

$$E_n = \{\underset{\sim}{X}_n : ||u_r(\underset{\sim}{X}_n) - \theta_r|| > d(\underset{\sim}{X}_n)\}$$

provided either side of (1.5.14) is finite. To show that the right-hand side is finite, note that $P_\theta(E_n) \leq 1$ and that

$$(1.5.15) \quad E_\theta(d^r(\underset{\sim}{X}_n)) = M(\frac{\log m(n)}{m(n)})^{r/2}(1+\epsilon)^{r/2}E_\theta(\text{tr } S(\underset{\sim}{X}_n))^{r/2}(r+\delta)^{r/2}.$$

To bound this expression we first consider the case when $r \geq 2$. By Inequality A.3.1 we have

$$(1.5.16) \quad E_\theta(\text{tr } S(\underset{\sim}{X}_n))^{r/2} = E_\theta(\sum_{i=1}^{r} s_i^2)^{r/2} \leq r^{r/2-1}\sum_{i=1}^{r} E_\theta|s_i^2|^{r/2}$$

where $s_i^2 = s_i^2(\underset{\sim}{X}_n)$ is the $i^{\underline{th}}$ diagonal element of $S(\underset{\sim}{X}_n)$. Letting

$$W_{ij} = \nu_i(X_{(j-1)k+1},\ldots,X_{jk})-g_i(\theta), \text{ and } \bar{W}_i = (m(n))^{-1}\sum_{j=1}^{m(n)} W_{ij},$$

it follows from Inequality A.3.1 that

$$\begin{aligned} E_\theta|s_i^2|^{r/2} &= E_\theta|(m(n))^{-1}\sum_{j=1}^{m(n)}(W_{ij}-\bar{W}_i)^2|^{r/2} \\ &\leq (m(n))^{-1}\sum_{j=1}^{m(n)} E_\theta|W_{ij}-\bar{W}_i|^r \\ &\leq (m(n))^{-1}\sum_{j=1}^{m(n)} 2^{r-1}(E_\theta|W_{ij}|^r+E_\theta|\bar{W}_i|^r) \\ &\leq 2^{r-1}\{E_\theta|W_{i1}|^r + (m(n))^{-r+1}\, E|W_{i1}|^r\}. \end{aligned}$$

This, together with (1.5.16), yields

$$(1.5.17) \qquad E_\theta(\text{tr } S(\underset{\sim}{X}_n))^{r/2} \leq r^{r/2-1}2^{r-1}(1+(m(n))^{-1})\xi_r(\theta).$$

For the case when $r = 1$, Jensen's inequality gives

$$(1.5.18) \qquad E_\theta(\text{tr } S(\underset{\sim}{X}_n))^{1/2} \leq (E_\theta \text{tr } S(\underset{\sim}{X}_n))^{1/2} = \sigma^{1/2}(\theta).$$

Since $P_\theta(E_n) \leq 1$ and $\psi(\theta)$ is a probability density, we have from (1.5.6), (1.5.8), (1.5.15), (1.5.17) and(1.5.18) that the right-hand side of (1.5.14) is finite.

We now attempt to bound $P_\theta(E_n)$. Let

$$(1.5.19) \qquad \delta_n(\theta) = \left(\frac{\log m(n)}{m(n)}\right)^{1/2} (r+\delta)^{1/2}\sigma^{1/2}(\theta)$$

and let

$$F_n = \{\underline{X}_n\colon \ ||u_r(\underline{X}_n)-\theta_r|| > M\,\delta_n(\theta)\},$$

$$G_n = \{\underline{X}_n\colon \ ||\bar{v}(\underline{X}_n)-g(\theta)|| > \delta_n(\theta)\},$$

$$H_n = \{\underline{X}_n\colon \ d(\underline{X}_n) \leq M\,\delta_n(\theta)\},$$

and

$$D_n = \{\theta\colon \ \delta_n(\theta) \geq \epsilon\}.$$

Note that

(1.5.20) $$\int_\Theta P_\theta(E_n)\psi(\theta)d\theta = \int_\Theta (P_\theta\{E_n \cap H_n\} + P_\theta\{E_n \cap H_n^c\})\psi(\theta)d\theta$$

$$\leq \int_\Theta (P_\theta\{H_n\} + P_\theta\{F_n\})\psi(\theta)d\theta$$

$$\leq \int_\Theta P_\theta\{H_n\}\psi(\theta)d\theta + \int_{D_n} \psi(\theta)d\theta + \int_{\Theta\setminus D_n} P_\theta\{F_n\}\psi(\theta)d\theta.$$

By Markov's inequality, for all θ, n,

$$P_\theta\{H_n\} = P_\theta\{[\mathrm{tr}\ S(\underline{X}_n)]^{1/2} \leq (1+\epsilon)^{-1/2}\sigma^{1/2}(\theta)\}$$

(1.5.21) $$\leq P_\theta\{|\mathrm{tr}\ S(\underline{X}_n)-\sigma(\theta)| > (\tfrac{\epsilon}{1+\epsilon})\sigma(\theta)\}$$

$$\leq \frac{E_\theta|\mathrm{tr}\ S(\underline{X}_n)-\sigma(\theta)|^{\alpha_r}}{(\frac{\epsilon}{1+\epsilon})^{\alpha_r}\ \sigma^{\alpha_r}(\theta)}\ .$$

Now, by Inequality A.3.1

(1.5.22) $$E_\theta|\mathrm{tr}\ S(\underline{X}_n)-\sigma(\theta)|^{\alpha_r} \leq r^{\alpha_r-1}\sum_{i=1}^{r} E_\theta|s_i^2-\sigma_i^2(\theta)|^{\alpha_r}$$

where $\sigma_i^2(\theta)$ is the $i^{\underline{th}}$ diagonal element of $\sigma(\theta)$. Using Lemma 1.5.1 and Inequality A.3.2 we see that the right-hand side of (1.5.22) is bounded above, for $r \geq 2$, by

$$(1.5.23) \qquad r^{r-1}(m(n))^{-r/2}[L_1+L_2(m(n))^{-r/2}]\gamma_r(\theta)$$

and, for r = 1, by

$$(1.5.24) \qquad (m(n))^{-1/2}[8+K(m(n))^{-1}]\gamma_{3/2}(\theta).$$

By Markov's inequality,

$$(1.5.25) \quad \int_{D_n} \psi(\theta)d\theta \leq \frac{(r+\delta)^{r/2}}{\epsilon^r}\left(\frac{\log m(n)}{m(n)}\right)^{r/2} \int_{\Theta} \sigma^{r/2}(\theta)\psi(\theta)d\theta.$$

On the other hand, when $\theta \in \Theta\backslash D_n$, so that $\delta_n(\theta) < \epsilon$, the contrapositive of assumption F.3 can be applied to show that

$$\begin{aligned} F_n &= (F_n \cap G_n) \cup (F_n \cap G_n^c) \\ &\subset G_n \cup \{\underline{X}_n: M||\bar{\nu}(\underline{X}_n)-g(\theta)|| \leq ||u_r(\underline{X}_n)-\theta_r||\} \\ &\subset G_n \cup \{\underline{X}_n: M||\bar{\nu}(\underline{X}_n)-g(\theta)|| \leq ||u(\underline{X}_n)-\theta||\} \\ &\subset G_n \cup \{\underline{X}_n: ||\bar{\nu}(\underline{X}_n)-g(\theta)|| \geq \epsilon\} \\ &\subset (G_n \cup G_n) = G_n. \end{aligned}$$

Hence,

$$(1.5.26) \quad \int_{\Theta\backslash D_n} P_\theta\{F_n\}\psi(\theta)d\theta \leq \int_{\Theta\backslash D_n} P_\theta\{G_n\}\psi(\theta)d\theta.$$

Letting

$$Z_{ij} = \frac{W_{ij}}{\sigma_i(\theta)} \quad \text{and } \bar{Z}_i = \frac{\bar{W}}{\sigma_i(\theta)},$$

we obtain

$$P_\theta\{G_n\} = P_\theta\{m(n)\sum_{i=1}^{r}\sigma_i^2(\theta)\bar{Z}_i^2 > \log(m(n))(r+\delta)\sum_{i=1}^{r}\sigma_i^2(\theta)\}$$

$$\leq P_\theta\{m(n)\sigma_i^2(\theta)\bar{Z}_i^2 > \log(m(n))(r+\delta)\sigma_i^2(\theta), \text{ some } i\}$$

$$\leq \sum_{i=1}^{r} P\{m(n)\bar{Z}_i^2 > (\log m(n))(r+\delta)\}$$

$$\leq 2n^{-(r+\delta)}\{1+ \frac{r+\delta}{2}\frac{\log(m(n))}{m(n)} + [\sqrt{r+\delta} + \frac{(r+\delta)^{3/2}}{3!} e^{\sqrt{r+\delta}}]$$

$$(\frac{\log(m(n))}{m(n)})^{3/2}E_\theta|Z|^3\}^n + (\frac{\log(m(n))}{m(n)})^{\frac{r+2+\delta}{2}} m(n)E_\theta|Z|^{r+2+\delta}.$$

The last inequality follows from the following lemma, which may be of independent interest, and was inspired from the moderate deviation results of Rubin and Sethuraman (1965).

Lemma 1.5.3. Let $Z_1, Z_2, \ldots$ be i.i.d. with mean 0, variance 1 and $E|Z|^s < \infty$ for some $s \geq 3$. Then if $F(Z)$ is the common c.d.f. of the Z_i's, and $c > 0$,

$$P\{|n^{-1}\sum_{i=1}^{n} Z_i| > c\sqrt{\frac{\log n}{n}}\} \leq 2n^{-c^2}\{1+ \frac{c^2}{2}\frac{\log n}{n} +$$

$$(c+ \frac{c^3}{3!} e^c)(\frac{\log n}{n})^{3/2}E|Z|^3\}^n + (\frac{\log n}{n})^{s/2} nE|Z|^s.$$

Proof. See Appendix 4. □

We now present another lemma which, together with assumption F.2, implies

$$(1.5.27)\qquad \sup_n\left(\frac{n}{\log n}\right)^{r/2}\int_\Theta P_\theta\{G_n\}\psi(\theta)d\theta < \infty.$$

Lemma 1.5.4. Suppose that θ, Z_1, $Z_2,\ldots$ are defined on a common probability space in such a way that given θ, the variables $Z_1, Z_2,\ldots$ are i.i.d. with mean 0, variance 1 and $E(|Z|^{r+2}|\theta) < \infty$, a.s.($\theta$) for $1 \leq r \leq c^2$. Suppose also that for some $\delta > 0$

$$E|Z|^{r+2+\delta} = E_\psi[E(|Z|^{r+2+\delta}|\theta] < \infty.$$

Then

$$\sup_n n^{r/2}P\{|n^{-1}\sum_{i=1}^n Z_i| > c\sqrt{\frac{\log n}{n}}\} < \infty.$$

Proof. See Appendix 4. □

It now follows from (1.5.15), (1.5.17), (1.5.18), (1.5.20), (1.5.23) - (1.5.27), assumption F.2 and the fact that

$$\left(\frac{n}{\log n}\right)\left(\frac{\log m(n)}{m(n)}\right) \leq 2k,$$

that (1.5.1) holds. This completes the proof of Theorem 1.5.1. □

1.6. Discussion.

In this section we comment on two special cases of the theory developed in Chapter 1 and then discuss some possible future work in this area. This future work includes extending the results to include vector-valued functions of the parameter vector.

We note that the proofs in Sections 4 and 5 apply to the case where $r = p$ and thus the full p-dimensional estimation problem is included in our general theory. Also, the special case where $r = p = 1$ agrees with the Gleser and Kunte (1976) results except we have eliminated their unnecessary assumption A.0., which restricted the Bayes region to be an interval. This restriction was justified on "grounds of convenience." Presumably, a connected interval is more appealing than a union of disjoint intervals. A similar justification would restrict our r-dimensional Bayes regions to be convex so that the standard projection method would yield interval estimators for linear combinations of the parameters. However, this restriction is not needed in the proofs and we find the Bayes property a stronger justification than "grounds of convenience."

We now consider the problem of finding A.P.O. and A.O. procedures for estimating a vector-valued function, $\tau(\theta)$, of θ. Let $\tau(\theta)$ be a p-dimensional, one-to-one and onto function satisfying the assumption:

A.6. Let the row vectors in the pxp matrix

$$T \equiv T(\theta) = \left(\frac{\partial\tau(\theta)}{\partial\theta_1}, \ldots, \frac{\partial\tau(\theta)}{\partial\theta_p}\right)$$

exist, be continuous and bounded in the sense that

$$\sup\left\{\sum_{i=1}^{p} \sum_{j=1}^{p} \left|\frac{\partial\tau_i(\theta)}{\partial\theta_j}\right| : \theta \in \Theta\right\} < \infty$$

where $\tau_i(\theta)$ and θ_j are the i^{th} and j^{th} components of $\tau(\theta)$ and θ, respectively. Moreover, suppose $|T(\theta)| \neq 0$.

Analogous to Section 1, the fixed-sample Bayes procedure, $C^*_\tau(\underline{X}_n)$, against the prior density, ξ, of $\tau(\theta)$ is

$$C^*_\tau(\underline{X}_n) = \text{closure (in } \tau(\Theta)) \text{ of } \{\tau(\theta)\text{: } \xi(\tau|\underline{X}_n) \geq ab^{-1}\},$$

where $\xi(\tau|\underline{X}_n)$ is the posterior density of $\tau(\theta)$.

Conditions A.1-A.5 were sufficient both to make $\psi(\theta|\underline{X}_n)$ of asymptotic multivariate normal form and to prove that the stopping rule, $t(c)$, defined by (1.3.11) is A.P.O. It is straightforward to show that conditions A.1 - A.6 imply that conditions analogous to A.1 - A.5 hold for the reparametrization $\tau(\Theta)$. Thus, $\xi(\tau|\underline{X}_n)$ is of asymptotic multivariate normal form and, by Theorem 1.4.1, the stopping rule

$$t'(c) = \text{first } n \geq 2 \text{ such that } \left[1-\left(\frac{n \log(n+1)}{(n+1)\log n}\right)^{r/2}\right]Y'_n \leq c$$

is A.P.O., where Y'_n is the posterior Bayes risk of $C^*_\tau(\underline{X}_n)$. Conditions A.1 - A.6 and F.1 - F.3 are also sufficient for $t'(c)$ to be A.O.

These same results should hold for more general $\tau(\theta)$ since after a large number of observations have been taken, much of the posterior probability of θ will be concentrated in a small sphere. Under assumption A.6, $\tau(\theta)$ can be approximated over this sphere by a one-to-one, onto function and hopefully we can show that this approximation is adequate and that the contribution to the Bayes region outside this sphere is negligible.

Now suppose we are interested in estimating an r-dimensional vector-valued function, $\tau_r(\theta)$, of θ. Following Bickel and Yahav (1969b), we can usually embed this problem in a p-dimensional space by reparametrizing the parameter space Θ to make $\tau_r(\theta)$ the first r components of our vector parameter. When this reparametrization is not possible we could generalize to the case where $\tau(\theta)$ admits a Taylor expansion to p terms around $\theta = 0$ with the remainder term uniformly of order $||\theta||^p$. Then the problems connected with integrating out the nuisance parameters should be similar to those in Chapter 1.

CHAPTER 2

A.P.O. AND A.O. APPROXIMATE STOPPING RULES

When applying the results of Chapter 1 to the special case of estimating the mean of a univariate normal distribution with unknown variance (see Kunte (1973)), the computations involving the probability of noncoverage part of the posterior Bayes risk are sufficiently complicated as to render the procedure impractical. In the hope of simplifying calculations, this chapter is concerned with general theoretical methods that permit approximation of A.P.O. and A.O. stopping rules in such a manner as to maintain their asymptotic optimality.

Section 1 generalizes Theorem 1.2.1 by giving sufficient conditions on the random variable $\tilde{Y}_n$, used to approximate Y_n in the stopping rule defined by (1.2.4), such that the resulting approximated stopping rule, $\tilde{t}(c)$, is A.P.O. Similarly, Section 2 generalizes Theorem 1.2.2 by showing that its approximated stopping rule is A.O. Both sections rely heavily on the theory developed in Chapters 3 and 4 of Gleser and Kunte (1976).

Section 3 uses these approximation results to prove that the stopping rules in Chapter 1 are robust with respect to the prior. That is, as long as the prior information is not too badly misspecified, the stopping rules will be A.P.O. and A.O.

2.1. A.P.O. Approximate Stopping Rules

Let $\{Y_n\}$ be a sequence of random variables defined on a probability space $(\Omega,\mathfrak{F},P)$, where Y_n is $\mathfrak{F}_n$-measurable and $\mathfrak{F}_1 \subset \mathfrak{F}_2 \subset \ldots \subset \mathfrak{F}$ is an increasing sequence of sub σ-fields and let $K(x)$ be a strictly increasing positive function of $x \geq 0$ satisfying $\lim_{x\to\infty} K(x) = \infty$.

To approximate the class of stopping rules $\{t(c): c > 0\}$ in (1.2.4) replace Y_n with $\tilde{Y}_n$ where $\tilde{Y}_n$ satisfies the following assumptions:

$\tilde{\text{B}}$.0. $P\{\tilde{Y}_n > 0\} = 1$, all n, and $P\{\lim_{n\to\infty} \tilde{Y}_n = 0\} = 1$.

$\tilde{\text{B}}$.1. $P\{\lim_{n\to\infty} f(n)|\tilde{Y}_n - Y_n| = 0\} = 1$.

Note that Assumptions B.1 and $\tilde{\text{B}}$.1 imply that

$$(2.1.1) \qquad P\{\lim_{n\to\infty} f(n)\,\tilde{Y}_n = V\} = 1.$$

Theorem 2.1.1. Under assumptions B.0 - B.4, $\tilde{\text{B}}$.0 and $\tilde{\text{B}}$.1, for each $c > 0$, let

$$(2.1.2) \qquad \tilde{t}(c) = \text{first } n \geq 1 \text{ such that } \left(1 - \frac{f(n)}{f(n+1)}\right)\tilde{Y}_n \leq c\,\Delta K(n).$$

Then the class of approximate stopping rules $\{\tilde{t}(c): c > 0\}$ is A.P.O. That is,

$$(2.1.3) \qquad P\{\lim_{c\to 0} \frac{Y_{\tilde{t}(c)} + cK(\tilde{t}(c))}{\inf_{s\in T} (Y_s + cK(s))} = 1\} = 1.$$

Proof. Under assumptions B.0 - B.4 Gleser and Kunte's (1976) Lemma 3.1 states, for each $c > 0$, that $t(c)$ is a proper stopping rule and

(2.1.4) $$P\{\lim_{c\to 0} t(c) = \infty\} = 1,$$

(2.1.5) $$P\{\lim_{c\to 0} f(t(c))Y_{t(c)} = V\} = 1,$$

and

(2.1.6) $$P\{\lim_{c\to 0} cK(t(c))f(t(c)) = MV\} = 1.$$

Replacing B.0, B.1, t(c) and Y_n with $\tilde{B}.0$, (2.1.1), $\tilde{t}(c)$ and $\tilde{Y}_n$ in the proof of their lemma shows, for each c > 0, that $\tilde{t}(c)$ is a proper stopping rule and

(2.1.7) $$P\{\lim_{c\to 0} \tilde{t}(c) = \infty\} = 1,$$

(2.1.8) $$P\{\lim_{c\to 0} f(\tilde{t}(c))\tilde{Y}_{\tilde{t}(c)} = V\} = 1$$

and

(2.1.9) $$P\{\lim_{c\to 0} cK(\tilde{t}(c))f(\tilde{t}(c)) = MV\} = 1.$$

Assumptions B.0 - B.4, $\tilde{B}.0$, $\tilde{B}.1$ and Theorem 1.2.1 thus yield

(2.1.10) $$P\{\lim_{c\to 0} \frac{\tilde{Y}_{\tilde{t}(c)}+cK(\tilde{t}(c))}{\inf_{s\in T}(\tilde{Y}_s+cK(s))} = 1\} = 1.$$

Now, letting $t \equiv t(c)$ and $\tilde{t} \equiv \tilde{t}(c)$, it follows from (2.1.4), (2.1.7) and assumption $\tilde{B}.1$ that

(2.1.11) $$P\{\lim_{c\to 0} f(t)|\tilde{Y}_t-Y_t|= 0\} = P\{\lim_{c\to 0} f(\tilde{t})|Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}}| = 0\} = 1.$$

Before completing the proof of Theorem 2.1.1, a lemma is presented which indicates that asymptotically $\tilde{t}(c)$ doesn't differ from t(c) by too much.

Lemma 2.1.1.

(2.1.12) $$P\{\lim_{c\to 0} \frac{f(\tilde{t})}{f(t)} = 1\} = 1.$$

Proof. First it is shown that

(2.1.13) $$\underline{\lim}_{c\to 0} \frac{f(\tilde{t})}{f(t)} \geq 1.$$

This follows from (2.1.5), (2.1.6), (2.1.8) - (2.1.11) since

$$1 = \lim_{c\to 0} \{\inf_{s\in T} \frac{\tilde{Y}_s + cK(s)}{Y_{\tilde{t}} + cK(\tilde{t})}\}$$

$$\leq \underline{\lim}_{c\to 0} \{\frac{\tilde{Y}_t + cK(t)}{\tilde{Y}_{\tilde{t}} + cK(\tilde{t})}\}$$

$$= \underline{\lim}_{c\to 0} \{\frac{f(\tilde{t})}{f(t)} [\frac{f(t)(\tilde{Y}_t - Y_t) + f(t)(Y_t + cK(t))}{f(\tilde{t})(\tilde{Y}_{\tilde{t}} + cK(\tilde{t}))}]\}$$

$$\leq (\underline{\lim}_{c\to 0} \frac{f(\tilde{t})}{f(t)}) [\frac{0 + (1+M)V}{(1+M)V}]$$

$$= \underline{\lim}_{c\to 0} \frac{f(\tilde{t})}{f(t)}.$$

In a similar manner it is shown that

(2.1.14) $$\overline{\lim}_{c\to 0} \frac{f(\tilde{t})}{f(t)} \leq 1.$$

Using (2.1.5), (2.1.6), (2.1.8), (2.1.9), (2.1.11) and (1.2. 2) yields

$$1 = \lim_{c\to 0} \{\frac{Y_t + cK(t)}{\inf_{s\in T}(Y_s + cK(s))}\}$$

$$\geq \overline{\lim_{c\to 0}} \left\{\frac{Y_t+cK(t)}{Y_{\tilde{t}}+cK(\tilde{t})}\right\}$$

$$= \overline{\lim_{c\to 0}} \left\{\frac{f(\tilde{t})}{f(t)} \left[\frac{f(t)(Y_t+cK(t))}{f(\tilde{t})(Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}})+f(\tilde{t})(\tilde{Y}_{\tilde{t}}+cK(\tilde{t}))}\right]\right\}$$

$$\geq \left(\overline{\lim_{c\to 0}} \frac{f(\tilde{t})}{f(t)}\right)\left[\frac{(1+M)V}{0+(1+M)V}\right]$$

$$= \overline{\lim_{c\to 0}} \frac{f(\tilde{t})}{f(t)} .$$

Thus, (2.1.13) together with (2.1.14) completes the proof of Lemma 2.1.1. □

We now return to the proof of Theorem 2.1.1 and note from (1.2.2) and (2.1.3) that it is sufficient to show

$$(2.1.15) \quad P\{\lim_{c\to 0} \frac{Y_t+cK(t)}{Y_{\tilde{t}}+cK(\tilde{t})} = 1\} = 1.$$

But this result follows by applying (2.1.5), (2.1.6), (2.1.8), (2.1.9), (2.1.11) and Lemma 2.1.1 to

$$\lim_{c\to 0} \frac{Y_t+cK(t)}{Y_{\tilde{t}}+cK(\tilde{t})} = \lim_{c\to 0}\left\{\frac{f(\tilde{t})}{f(t)} \cdot \frac{f(t)(Y_t+cK(t))}{f(\tilde{t})(Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}})+f(\tilde{t})(\tilde{Y}_{\tilde{t}}+cK(\tilde{t}))}\right\}$$

$$= 1\cdot \frac{(1+M)V}{0+(1+M)V} = 1.$$

Hence (2.1.15) holds and the proof of Theorem 2.1.1 is complete. □

2.2. A.O. Approximate Stopping Rules.

In this section sufficient conditions are found for the class $\{t(c): c > 0\}$ of approximate stopping rules defined in (2.1.2) to be asymptotically optimal for the special case where $K(x) = x$. Use is made of the results in Section 4 of Gleser and Kunte (1976) which concern replacing the random quantity $f(t(c))$ by the non-random quantity $f(\gamma(c))$ in the expectation operations, where

$$\gamma(c) = \text{first } n \geq 1 \text{ such that } nf(n) \geq c^{-1}M$$

with M defined in assumption B.3'.

Theorem 2.2.1. When $K(x) = x$ and assumptions B.0 - B.2, B.3', $\tilde{B}.0$ and $\tilde{B}.1$ hold, then the class $\{\tilde{t}(c): c > 0\}$ of approximate stopping rules defined by

$$\text{(2.2.1)} \qquad \tilde{t}(c) = \text{first } n \geq 1 \text{ such that } \left(1 - \frac{f(n)}{f(n+1)}\right)\tilde{Y}_n \leq c$$

is A.P.O. Further, if (1.2.6) and

$$\text{(2.2.2)} \qquad \sup_n E(f(n)\tilde{Y}_n) < \infty$$

hold and if there exists a finite constant L such that

$$\text{(2.2.3)} \qquad Y_n \leq L\,\tilde{Y}_n$$

then this class of approximate stopping rules is A.O. That is,

$$\text{(2.2.4)} \qquad \lim_{c\to 0} \left\{\frac{E(Y_{\tilde{t}}+c\tilde{t})}{\inf_{s\in T} E(Y_s+cs)}\right\} = 1.$$

Proof. The class $\{\tilde{t}(c): c > 0\}$ defined by (2.2.1) is just the class of rules defined in Theorem 2.1.1. specialized to the case $K(x) = x$. Since $K(x) = x$ and assumptions B.0-B.2, B.3' imply assumptions B.0 - B.4, the A.P.O. character of the stopping rules (2.2.1) follows from Theorem 2.1.1.

Proof. From (1.2.3) and (2.2.4) it is sufficient to show that

$$(2.2.5) \qquad \overline{\lim_{c\to 0}} \frac{E(Y_{\tilde{t}}+c\tilde{t})}{E(Y_t+ct)}\} \leq 1.$$

The nonrandomness of $f(\gamma)$ allows an upper bound for the left-hand side of (2.2.5) to be expressed as

$$(2.2.6) \quad \frac{\overline{\lim_{c\to 0}}\, E[f(\gamma)(Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}})] + \overline{\lim_{c\to 0}}\, E[f(\gamma)(\tilde{Y}_{\tilde{t}}+c\tilde{t})]}{\underline{\lim_{c\to 0}}\, E[f(\gamma)(Y_t+ct)]} .$$

Gleser and Kunte (1976) prove that

$$P\{\lim_{c\to 0} \frac{f(\gamma)}{f(t)} = V^{-M/(M+1)}\} = 1,$$

and

$$(2.2.7) \quad \lim_{c\to 0} E[(f(\gamma)(Y_t+ct)] = (1+M)E(V^{1/(M+1)}).$$

Replacing t, Y_n, (2.1.5), (2.1.6) and (1.2.6) with $\tilde{t}$, $\tilde{Y}_n$, (2.1.8), (2.1.9) and (2.2.2) in their proof, it can be shown that

$$P\{\lim_{c\to 0} \frac{f(\gamma)}{f(\tilde{t})} = V^{-M/(M+1)}\} = 1,$$

and

$$(2.2.9)\qquad \lim_{c\to 0} E[f(\gamma)(\tilde{Y}_{\tilde{t}}+c\tilde{t})] = (1+M)E(V^{1/(M+1)}).$$

From (2.2.6), (2.2.7) and (2.2.9) we see that (2.2.5) will hold if it can be shown that

$$(2.2.10)\qquad \overline{\lim_{c\to 0}}\, E[f(\gamma)(Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}})] = 0.$$

From (2.2.8) and assumption $\tilde{B}.1$, it follows that

$$(2.2.11)\qquad \lim_{c\to 0} f(\gamma)(Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}}) = \lim_{c\to 0} \frac{f(\gamma)}{f(\tilde{t})} f(\tilde{t})(Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}}) = 0, \text{ a.s.}$$

Using (2.2.3) we have the bound

$$(2.2.12)\qquad f(\gamma)(Y_{\tilde{t}}-\tilde{Y}_{\tilde{t}}) \le f(\gamma)(1+L)\tilde{Y}_{\tilde{t}}.$$

Thus application of a well-known generalization of the dominated convergence theorem [see, Royden (1968, page 89)], along with (2.2.12), (2.2.9) and (2.2.11), proves (2.2.10). This completes the proof of Theorem 2.2.1. □

Remark 1. A condition equivalent to (2.2.2) is

$$\sup_n E[f(n)(\tilde{Y}_n-Y_n)^+] < \infty,$$

as can be seen from condition (1.2.6) and the following inequality:

$$\begin{aligned} f(n)\tilde{Y}_n &\le f(n)(\tilde{Y}_n-Y_n)^+ + f(n)Y_n \\ &\le f(n)\tilde{Y}_n + 2f(n)Y_n. \end{aligned}$$

Remark 2. In the absence of any well-developed second-order properties, it would seem preferable to choose, when possible, $\tilde{Y}_n \leq Y_n$, since both $\tilde{t}(c)$ and $t(c)$ are A.P.O. and A.O. but $\tilde{t}(c) \leq t(c)$. In this case, condition (2.2.2) follows from (1.2.6), although (1.2.6) is not always easy to verify. Gleser and Kunte (1976) give an example which demonstrates that some condition like (1.2.6) seems to be needed.

Remark 3. If we can find an easily computable terminal decision rule $\tilde{C}(\underline{X}_n)$ which has Bayes risk of the same order asymptotically as the Bayes rule $C^*(\underline{X}_n)$, then if $\sup_n E(f(n)Y_n) < \infty$, we also have that $\sup_n E(f(n)\tilde{Y}_n) < \infty$, where $\tilde{Y}_n$ is the posterior risk of $\tilde{C}(\underline{X}_n)$. By the Bayes property of Y_n it follows that $Y_n \leq \tilde{Y}_n$. (Hence, in (2.2.3), $L = 1$.) Theorem 2.2.1 then implies that the stopping rule

$$\tilde{t}(c) = \text{first } n \geq 2 \text{ such that } \left(1 - \frac{f(n)}{f(n+1)}\right)\tilde{Y}_n \leq c$$

is A.P.O. and A.O.

Example. Consider the estimation problem of Chapter 1 and let $\tilde{Y}_n = aV(C^*(\underline{X}_n))$. It is clear that $\tilde{B}.0$ is satisfied. To verify $\tilde{B}.1$ note that

$$f(n)|Y_n - \tilde{Y}_n| = (n/\log n)^{r/2}\, a \int_{C^{*c}(\underline{X}_n)} \psi_r(\theta_r|\underline{X}_n)d\theta_r$$

tends to 0 a.s. (P_{θ_0}) by the remark at the end of Section 1.4. Thus, by Theorem 2.1.1, the class $\{\tilde{t}(c): c > 0\}$ of approximate stopping rules defined by

$$\tilde{t}(c) = \text{first } n \geq 2 \text{ such that } \left[1-\left(\frac{n \log (n+1)}{(n+1)\log n}\right)^{r/2}\right]\tilde{Y}_n \leq c$$

is A.P.O.

Now, from $\tilde{Y}_n \leq Y_n$ and (1.2.6) it follows that (2.2.2) is satisfied. By the remark at the end of Section 1.4 we see that (2.2.3) is satisfied almost surely for large enough n depending on the sample point. Unfortunately, our proof of Theorem 2.2.1 requires (2.2.3) to hold uniformly in n over all sample points.

2.3. Robustness.

Consider the estimation problem of Chapter 1 where the procedure $[C^*(t(c)), t(c)]$ was shown to be asymptotically optimal against the prior ψ. In this section it is shown that if ψ is not the true prior, and if the true prior is, say ψ_0, then the procedure $[C^*(t(c)),t(c)]$ remains asymptotically optimal provided that the following condition holds:

D.1. There exist constants k, K, $0 < k \leq K < \infty$, such that

$$k \leq \frac{\psi(\theta)}{\psi_0(\theta)} \leq K \qquad \text{for all } \theta \in \Theta. \tag{2.3.1}$$

We adopt the notation $E_{\Xi}(\cdot)$ for the expectation over the joint probability distribution of $X_1, X_2, \ldots$, and θ when θ has prior density Ξ. Let $C_n^* = C_n^*(\underline{X}_n)$ and $C_n^0 = C^0(\underline{X}_n)$ be the Bayes optimal terminal decision rules against the priors ψ and ψ_0, respectively, and let Y_n and Y_n^0 be their corresponding posterior risks.

Theorem 2.3.1. If assumptions A.2 - A.5 hold and both ψ_0 and ψ satisfy assumption A.1, then the class $\{t(c): c > 0\}$ of stopping rules, assuming that ψ is the prior, is A.P.O. even when ψ_0 is the true prior. If, in addition, assumption D.1 and either

$$\sup_n E_{\psi_0}[f(n)Y_n^0] < \infty, \tag{2.3.2a}$$

or

$$\sup_n E_{\psi}[f(n)Y_n] < \infty \tag{2.3.2b}$$

hold, then t(c) is A.O.

Proof. Since the limiting results of Section 1.4 are independent of the prior, it is easily seen that assumptions $\tilde{B}.0$ and $\tilde{B}.1$ of Chapter 2 are satisfied and hence, by Theorem 2.1.1, t(c) is A.P.O.

To prove t(c) is A.O., by Theorem 2.2.1, we need only show that

$$Y_n^0 \leq LY_n \tag{2.3.3}$$

and

$$\sup_n E_{\psi_0}[f(n)Y_n] < \infty. \tag{2.3.4}$$

Note that from (2.3.1),

$$\psi_{0r}(\theta_r|\underset{\sim}{X}_n) = \frac{\int_{\Theta_{p-r}} \prod_{i=1}^{n} f(x_i|\theta)\psi_0(\theta)d\theta_{p-r}}{\int_{\Theta} \prod_{i=1}^{n} f(x_i|\theta)\psi_0(\theta)d\theta}$$

$$\leq \frac{\int_{\Theta_{p-r}} \frac{1}{k} \prod_{i=1}^{n} f(x_i|\theta)\psi(\theta)d\theta_{p-r}}{\frac{1}{K}\int_{\Theta} \prod_{i=1}^{n} f(x_i|\theta)\psi(\theta)d\theta} \tag{2.3.5}$$

$$\leq \frac{K}{k}\,\psi_r(\theta_r|\underset{\sim}{X}_n).$$

Since C_n^0 is Bayes against ψ_0, applying (2.3.5) yields

$$Y_n^0 = aV(C_n^0) + b\int_{(C_n^0)^c} \psi_{0r}(\theta_r|\underset{\sim}{X}_n)d\theta_r$$

$$\leq aV(C_n^*) + b\int_{(C_n^*)^c} \psi_{0r}(\theta_r|\underset{\sim}{X}_n)d\theta_r$$

$$\leq aV(C_n^*) + \frac{bK}{k} \int_{(C_n^*)^c} \psi_r(\theta_r|\underline{X}_n)d\theta_r$$

$$\leq (\frac{K}{k})Y_n.$$

Hence, we can let $L = K/k$ in (2.3.3).

Now, a similar argument using (2.3.1) yields

$$E_{\psi_0}[f(n)Y_n] = \int_{\mathcal{X}} f(n)Y_n \int_{\Theta} \prod_{i=1}^{n} f(x_i|\theta)\psi_0(\theta)d\theta d\underline{X}_n$$

$$\leq \frac{1}{k} \int_{\mathcal{X}} f(n)Y_n \int_{\Theta} \prod_{i=1}^{n} f(x_i|\theta)\psi(\theta)d\theta d\underline{X}_n$$

$$= \frac{1}{k} E_{\psi}[f(n)Y_n].$$

Thus, (2.3.4) follows from (2.3.2b). However, use of (2.3.1) and (2.3.5) allows us to show that (2.3.2a) and (2.3.2b) are equivalent conditions. This completes the proof of Theorem 2.3.1. □

BIBLIOGRAPHY

[1] Albert, A. (1966). Fixed size confidence ellipsoids for linear regression parameters. Ann. Math. Statist. **37**, 1602-1630.

[2] Anscombe, F. J. (1952). Large-sample theory of sequential estimations. Proc. Camb. Phil. Soc. **48**, 600-607.

[3] Anscombe, F. J. (1953). Sequential estimation. J. Roy. Statist. Soc. Ser. B. **15**, 1-21.

[4] Arrow, K., Blackwell, D. and Girshick, M. (1949). Bayes and minimax solution of sequential decision problems. Econometrica **17**, 213-244.

[5] Bickel, P. J. and Yahav, J. A. (1965). Asymptotically pointwise optimal procedures in sequential analysis. Proc. Fifth Berkeley Symp. Prob. Statist. **1**, Univ. of California Press.

[6] Bickel, P. J. and Yahav, J. A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. Ann. Math. Statist. **39**, 442-456.

[7] Bickel, P. J. and Yahav, J. A. (1969a). On an A.P.O. rule in sequential estimation with quadratic loss. Ann. Math. Statist. **40**, 417-426.

[8] Bickel, P. J. and Yahav, J. A. (1969b). Some contributions to the asymptotic theory of Bayes solutions. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. **11**, 257-276.

[9] Callahan, J. (1969). On some topics in sequential multiparameter estimation. Ph.D. dissertation, The John Hopkins Univ.

[10] Chatterjee, S. K. (1959). On an extension of Stein's two-sample procedure to the multinomial problem. Calcutta Statist. Assn. Bull. **8**, 121-148.

[11] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. Ann. Math. Statist. **36**, 457-462.

[12] Chung, K. L. (1951). The strong law of large numbers. Proc. Second Berk. Symp. Prob. Statist. Univ. California Press.

[13] Dantzig, G. B. (1940). On the non-existence of tests of "student's" hypothesis having power functions independent of σ. Ann. Math. Statist. **11**, 186-192.

[14] DeGroot, M. H. (1970). Optimal Statistical Decisions. McGraw-Hill, New York.

[15] Gleser, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. Ann. Math. Statist. **40**, 935-941.

[16] Gleser, L. J. (1966). Correction to 'On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters'. Ann. Math. Statist. **37**, 1053-1055.

[17] Gleser, L. J. and Kunte, S. (1976). On asymptotically optimal sequential Bayes interval estimation procedures. Ann. Statist. **4**, 685-711.

[18] Kiefer, J. and Sacks, J. (1963). Asymptotically optimal sequential inference and design. Ann. Math. Statist. **34**, 705-750.

[19] Kunte, S. (1973). Asymptotically pointwise optimal and asymptotically optimal stopping rules for sequential Bayes confidence interval estimation. Mimeograph Series No. 328, Department of Statistics, Purdue Univ.

[20] Loève, M. (1955). Probability Theory. D. Van Nostrand Company, Inc., New York.

[21] Rao, C. R. (1973). Linear Statistical Inference and Its Applications. Second Edition. Wiley, New York.

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

[23] Rubin, H. and Sethuraman, J. (1965). Probabilities of Moderate Deviations, Sankhyā Ser. A. **27**, 325-346.

[24] Serfling, R. and Wackerly, D. (1976). Asymptotic theory of sequential fixed-width confidence interval procedures. J. Amer. Statist. Assoc. **71**, 949-955.

[25] Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. J. Amer. Statist. Assoc. **62**, 626-633.

[26] Srivastava, M. S. (1967). On fixed-width confidence bounds for regression parameters and mean vectors. J. Roy Statist. Soc. Ser. B. **29**, 132-140.

[27] Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Statist. **16**, 243-258.

[28] Stein, C. and Wald, A. (1947). Sequential confidence intervals for the mean of a normal distribution with known variance. Ann. Math. Statist. **18**, 427-433.

[29] Stein, C. (1949). Some problems in sequential estimation. Econometrica **17**, 77-78.

[30] von Bahr, B. and Esseen, C. G. (1965). Inequalities for the r^{th} absolute moment of a sum of random variables, $1 \leq r \leq 2$, Ann. Math. Statist. **36**, 299-303.

APPENDICES

APPENDIX 1

A Tail Inequality for Multivariate Normal Distributions:

Let X be an r-dimensional, $(r \geq 1)$, random vector having the density, $N(x;0,\Sigma)$, of a multivariate normal distribution with mean 0 and positive definite covariance matrix Σ. For $k > 2r-4$ the ellipsoid $E = \{x: x'\Sigma^{-1}x \leq k\}$ satisfies the inequality

$$\int_{E^c} N(x;0,\Sigma)dx \leq \frac{2^{2-r/2}}{\Gamma(\frac{r}{2})} k^{r/2-1}e^{-k/2}.$$

Proof. Note that $Y = X'\Sigma^{-1}X$ is a chi-square random variable with r degrees of freedom. Thus, the probability of being outside E can be written as

$$\text{(A.1.1)} \qquad \int_{E^c} N(x;0,\Sigma)dx = \int_k^\infty g(y)e^{-y/4}dy,$$

where $g(y) = \dfrac{2^{-r/2}y^{r/2-1}e^{-y/4}}{\Gamma(\frac{r}{2})}$.

Elementary calculus shows that for $y > 2r-4$, $g(y)$ is a decreasing function of y. Hence, for $k > 2r-4$, equation (A.1.1) is bounded above by

$$g(k)\int_k^{\infty} e^{-y/4}dy = \frac{2^{2-r/2}}{\Gamma(\frac{r}{2})} k^{r/2-1}e^{-k/2}. \qquad \square$$

APPENDIX 2

<u>Theorem</u>. Under assumptions A.1 - A.5, for every $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ and $N_1 = N(X_1, X_2, \ldots; \theta_0, \delta(\epsilon))$ such that for every $n \geq N_1$, and all w in the sphere

$$W_n(\delta(\epsilon)) = \{w: ||w|| < n^{1/2}\delta(\epsilon)\},$$

the following bound holds:

$$(A.2.1) \qquad |\log \nu_n(w) - \frac{w'A(\theta_0)w}{2}| \leq \epsilon\, w'w.$$

Proof. For every $\delta > 0$, the strong consistency of $\hat{\theta}_n$ implies that there exists $n_1 = N(X_1, X_2, \ldots; \theta_0, \delta)$, such that $|\hat{\theta}_n - \theta_0| < \delta$ for all $n \geq n_1$. Bickel and Yahav (1969b, Lemma 2.1) show that there exists $n_2 = N(X_1, X_2, \ldots; \theta_0)$ such that for $n \geq n_2$,

$$\sum_{i=1}^{n} \text{grad } \varphi(\hat{\theta}_n, X_i) = 0$$

where

$$\text{grad } \varphi(\theta, X) = \left(\frac{\partial\varphi(\theta,X)}{\partial\theta_1}, \ldots, \frac{\partial\varphi(\theta,X)}{\partial\theta_p}\right)'.$$

By Taylor's formula, for $n \geq \max(n_1, n_2)$

$$|\log \nu_n(w) - \frac{w'A(\theta_0)w}{2}| = |n^{-1/2}w' \text{ grad } \varphi(\hat\theta_n, X_i)$$

$$+ \frac{n^{-1}}{2} \sum_{i=1}^{n} \int_0^1 w'A(\hat\theta_n + \lambda w n^{-1/2}, X_i) w d\lambda - w'A(\theta_0)w|$$

$$\leq \frac{n^{-1}}{2} \sum_{i=1}^{n} |\int_0^1 w'[A(\hat\theta_n + \lambda w n^{-1/2}, X_i) - A(\theta_0)] w d\lambda|$$

(A.2.2)

$$\leq \frac{n^{-1}}{2} \sum_{i=1}^{n} \sup\{|w'[A(s,X_i) - A(\theta_0)]w| : |s - \hat\theta_n| \leq \delta\}$$

$$\leq \frac{n^{-1}}{2} \sum_{i=1}^{n} \sup\{|w'[A(s,X_i) - A(\theta_0)]w| : |s - \theta_0| \leq 2\delta\}$$

where the matrix A is given by,

$$A(\theta, X) = \frac{\partial^2 \varphi(\theta, X)}{\partial\theta_i \partial\theta_j} .$$

The last inequality follows since $\{s: |s-\hat\theta_n| \leq \delta\} \subset \{s: |s-\theta_0| \leq 2\delta\}$.

Now,

(A.2.3) $\sup\{|w'[A(s,X) - A(\theta_0)]w| : |s-\theta_0| \leq 2\delta\} \leq |w'[A(\theta_0,X) - A(\theta_0)]w|$

$+ ||w||^2 \sup\{|w'[A(s,X) - A(\theta_0,X)]w| : ||s-\theta_0|| \leq 2\delta, ||w|| = 1\}$.

But, the continuity of the elements of A(s,X) assumed in assumption A.2, plus the compactness of the set $\{(s,w): ||s-\theta_0|| \leq 2\delta, ||w|| = 1\}$, implies that

(A.2.4) $\sup\{|w'[A(s,X) - A(\theta_0,X)]w| : ||s-\theta_0|| \leq 2\delta, ||w|| = 1\} \to 0$ as $\delta \to 0$.

On the other hand, the left-hand side of (A.2.4) is bounded in absolute value by

$$2 \sum_{1\le j,k\le p} \sup\{|\frac{\partial\varphi^2(s,X)}{\partial\theta_j\partial\theta_k}|: ||s-\theta_0|| < 2\delta\},$$

which by assumption A.3 has a finite expectation for δ small enough. Hence, by the dominated convergence theorem, the expected value of the left-hand side of (A.2.4) tends to 0 as $\delta \to 0$.

Thus, for every $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that

(A.2.5) $E\{\sup\{|w'[A(s,X)-A(\theta_0,X)]w|: ||s-\theta_0|| \le 2\delta(\epsilon), ||w||=1\} < \epsilon/3.$

By the S.L.L.N., (A.2.5) and assumption A.4, there exists $n_3 = N(X_1,X_2,\ldots;\theta_0)$ such that for $n \ge n_3$,

(A.2.6) $n^{-1} \sum_{i=1}^{n} w'[A(\theta_0,X_i)-A(\theta_0)]w| \le \frac{\epsilon}{3} w'w$

and

(A.2.7) $n^{-1} \sum_{i=1}^{n} ||w||^2 \sup\{w'[A(s,X_i)-A(\theta_0,X_i)]w: ||s-\theta_0|| \le 2\delta(\epsilon), ||w|| = 1\} \le \frac{2\epsilon}{3} w'w.$

Thus, from (A.2.2), (A.2.3), (A.2.6) and (A.2.7) it follows that for every $\epsilon > 0$, there exists a $\delta(\epsilon) > 0$ and $N_1 = \max\{n_1,n_2,n_3\}$ such that (A.2.1) holds. □

APPENDIX 3

Inequalities

Inequality A.3.1. Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables and put $S_n = \sum_{i=1}^{n} X_i$, then it is well-known that

$$E|S_n|^\beta \leq n^{\beta-1} \sum_{i=1}^{n} E|X_i|^\beta, \quad \beta \geq 1,$$

$$E|S_n|^\beta \leq \sum_{i=1}^{n} E|X_i|^\beta, \quad \beta < 1.$$

The first inequality is an application of Jensen's inequality on convex functions while the second follows directly from another Jensen inequality, which states that for $\beta > 0$, $(\sum_{i=1}^{n} X_i^\beta)^{1/\beta}$ is a decreasing function of β.

Inequality A.3.2. (Gleser and Kunte (1976), Lemma 6.3). Let $U_1, U_2, \ldots, U_n$ be i.i.d. with mean μ, variance σ^2, and $\tau = E|U_1 - \mu|^{2\beta}$, $1 \leq \beta \leq 2$. Then there exists a constant K, $0 < K < \infty$, such that

$$E|s_n^2 - \sigma^2|^\beta \leq n^{1-\beta} \tau [2^{2\beta} + K\, n^{-1}], \quad \text{all } n,$$

where $s_n^2 = n^{-1} \sum_{i=1}^{n} (U_i - \bar{U})^2$ is the sample variance.

Inequality A.3.3. (Chung (1951) p. 348). Let $U_1, U_2, \ldots, U_n$ be i.i.d. with mean 0. Then, if $S_n = \sum_{i=1}^{n} U_i$, $\beta \geq 1$,

$$E|S_n|^{2\beta} \leq n^{\beta}M(\beta)E|U_1|^{2\beta},$$

where $M(\beta)$ is a constant depending only on β.

APPENDIX 4

PROOF OF LEMMA 1.5.3.

Let

$$Y_{ni} = \begin{cases} Z_i & \text{if } |Z_i| \leq a_n = (n/\log n)^{1/2} \\ 0 & \text{otherwise.} \end{cases}$$

Then, using Markov's inequality

$$\text{(A.4.1)} \quad P\{|n^{-1}\sum_{i=1}^{n} Z_i| > ca_n\} \leq P\{|n^{-1}\sum_{i=1}^{n} Y_{ni}| > ca_n\} + \sum_{i=1}^{n} P\{|Z_i| > a_n\}$$

$$\leq P\{|n^{-1}\sum_{i=1}^{n} Y_{ni}| > ca_n\} + n\left(\frac{\log n}{n}\right)^{s/2} E|Z|^s$$

Note that the Y_{ni}'s are i.i.d. with

$$|\mu_n| = |EY_{ni}| = \left|\int_{|z|\leq a_n} zF(dz)\right| = \left|-\int_{|z|>a_n} zF(dz)\right|$$

$$\text{(A.4.2)} \qquad \leq \int_{|z|>a_n} |z|F(dz)$$

$$\leq a_n^{-2}\int|z|^3F(dz) = \left(\frac{\log n}{n}\right)E|z|^3$$

and

$$\text{(A.4.3)} \qquad \mu_n^2+\sigma_n^2 = EY_{ni}^2 = \int_{|Z|\leq a_n} z^2F(dz)$$

$$\leq E(Z^2) = 1.$$

Let

$$\varphi_{ni}(\lambda) = \int_{-a_n}^{a_n} e^{\lambda a_n^{-1} y_{ni}} F_{ni}(dy_{ni})$$

where $F_{ni}(y)$ is the c.d.f. of Y_{ni}. Since Y_{ni} is bounded, $\varphi_{ni}(\lambda)$ exists for all λ. Let,

$$G_{ni}(dy_{ni};\lambda) = \frac{e^{\lambda a_n^{-1} y_{ni}} F_{ni}(dy_{ni})}{\varphi_{ni}(\lambda)}$$

and

$$G_n(dy;\lambda) = \mathop{*}_{i=1}^{n} G_{ni}(dy_{ni};\lambda)$$

be the convolution of the G_{ni}'s. Then for any $\lambda > 0$,

$$P\{|n^{-1}\sum_{i=1}^{n} Y_{ni}| > ca_n\} = \prod_{i=1}^{n}\varphi_{ni}(\lambda)\int_{c\sqrt{n \log n}}^{\infty} e^{-\lambda a_n^{-1} y} G_n(dy;\lambda)$$

$$+ \prod_{i=1}^{n}\varphi_{ni}(-\lambda)\int_{-\infty}^{-c\sqrt{n \log n}} e^{|\lambda| a_n^{-1} y} dG_n(dy;-\lambda)$$

$$\le e^{-\lambda c \log n}[\prod_{i=1}^{n}\varphi_{ni}(\lambda)\int_{c\sqrt{n \log n}}^{\infty} G(dy;\lambda) + \prod_{i=1}^{n}\varphi_{ni}(-\lambda)\int_{-\infty}^{-c\sqrt{n \log n}} G_n(dy;-\lambda)]$$

(A.4.4)

$$\le e^{-\lambda c \log n}[\prod_{i=1}^{n}\varphi_{ni}(\lambda) + \prod_{i=1}^{n}\varphi_{ni}(-\lambda)]$$

$$= n^{-\lambda c}[\prod_{i=1}^{n}\varphi_{ni}(\lambda) + \prod_{i=1}^{n}\varphi_{ni}(-\lambda)].$$

Using the inequalities

$$e^x \le 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} e^x, \qquad x > 0,$$

$$e^x \le 1 + x + \frac{x^2}{2} \qquad\qquad x < 0,$$

we have

$$\varphi_{ni}(b) \le 1 + ba_n^{-1}\mu_n + \frac{b^2}{2}a_n^{-2}(\mu_n^2+\sigma_n^2)+\int_{-a_n}^{a_n} R(by_{ni})F_{ni}(dy_{ni})$$

where

$$R(by) = \begin{cases} \frac{y^3b^3}{3!} a_n^{-3} e^{ba_n^{-1}y} & \text{if } by > 0 \\ 0 & \text{if } by < 0. \end{cases}$$

Hence, using (A.4.2) and (A.4.3),

$$\varphi_{ni}(\lambda) \le 1+\lambda a_n^{-1}|\mu_n| + \frac{\lambda^2}{2}a_n^{-2}(\mu_n^2+\sigma_n^2)+\int_0^{a_n} \frac{\lambda^3}{3!}a_n^{-3}y_{ni}^3e^{\lambda a_n^{-1}y_{ni}}F_{ni}(dy_{ni})$$

(A.4.5) $$\le 1+\lambda a_n^{-3}E|Z|^3+ \frac{\lambda^2}{2}a_n^{-2}+ \frac{\lambda^3}{3!}a_n^{-3}e^{\lambda}\int_0^{a_n} y_{ni}^3dF_{ni}(dy_{ni})$$

$$\le 1+\lambda a_n^{-3}E|Z|^3 + \frac{\lambda^2}{2}a_n^{-2}+ \frac{\lambda^3}{3!}a_n^{-3}e^{\lambda}E|Z|^3$$

and $\varphi_{ni}(-\lambda)$ is similarly bounded by the same quantity. From (A.4.1), (A.4.4), (A.4.5) and the definition of a_n, the result of the lemma follows by taking $\lambda = c$. □

PROOF OF LEMMA 1.5.4.

Let $\Psi(\theta)$ be the c.d.f. of θ,

$$\Delta_n(\theta) = P\{|n^{-1}\sum_{i=1}^{n} Z_i| > c\sqrt{\frac{\log n}{n}}\ |\theta\}$$

$$\mu_s(\theta) = E(|Z|^s|\theta), \quad 1 \le s \le r+2,$$

and

$$D = \{\theta: \mu_3(\theta) \le n^{1/2-\epsilon}, \ \epsilon = \delta/2(r+\delta)\}.$$

Then

$$n^{r/2} P\{|n^{-1}\sum_{i=1}^{n} Z_i| > c\sqrt{\frac{\log n}{n}}\} = \int_{\Theta} \Delta_n(\theta)\Psi(d\theta)$$

$$\text{(A.4.6)} \qquad = n^{r/2}\int_D \Delta_n(\theta)\Psi(d\theta) + n^{r/2}\int_{D^c} \Delta_n(\theta)\Psi(d\theta)$$

$$\le n^{r/2}\int_D \Delta_n(\theta)\Psi(d\theta) + n^{r/2}\int_{D^c} \Psi(d\theta).$$

Now,

$$n^{r/2}\int_{D^c} \Psi(d\theta) \le \frac{n^{r/2}}{n^{\frac{1}{2}-\epsilon}}\int_{D^c} \mu_3(\theta)\Psi(d\theta)$$

$$\le \frac{n^{r/2}}{(n^{\frac{1}{2}-\epsilon})^{r+\delta}}\int_{D^c} \mu_3^{r+\delta}(\theta)\Psi(d\theta)$$

$$= \frac{n^{r/2}}{n^{r/2}}\int_{D^c} \mu_3^{r+\delta}(\theta)\Psi(d\theta).$$

Applying Liapounov's inequality (cf. Loeve (1955), p. 172) and noting that $\mu_2(\theta) = 1$, we obtain

$$\text{(A.4.7)} \qquad n^{r/2}\int_{D^c} \Psi(d\theta) \le \int_{D^c} \mu_{r+2+\delta}(\theta)\Psi(d\theta),$$

which is finite by our given assumption. Also, by Lemma 1.5.4,

$$\text{(A.4.8)} \qquad n^{r/2} \int_D \Delta_n(\theta)\Psi(d\theta) \leq 2n^{r/2} n^{-c^2} [1+ \frac{c^2}{2} \frac{\log n}{n} + (c+ \frac{c^3}{3!} e^c)(\frac{\log n}{n})^{3/2} n^{\frac{1}{2}-\epsilon n}] \int_D \Psi(d\theta) +$$

$$+ \frac{n^{r/2}(\log n)^{\frac{r+2+\delta}{2}}}{n^{\frac{r+\delta}{2}}} \int_D \mu_{r+2+\delta}(\theta)\psi(d\theta)$$

and since the right-hand side of this inequality converges to 0 as $n \to \infty$, the supremum of this quantity is finite. From (A.4.6) - (A.4.8), the result follows. □