BOUNDS AND RATES OF CONVERGENCE FOR THE EXTENDED COMPOUND ESTIMATION PROBLEM IN THE SEQUENCE CASE

by

Donald D. Swain

TECHNICAL REPORT NO. 81

June 17, 1965

Supported by the Army, Navy, Air Force and NASA under

Contract Nonr-225(53) (NR-042-002)

with the Office of Naval Research

Gerald J. Lieberman, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

BLANK PAGE

BOUNDS AND RATES OF CONVERGENCE FOR THE EXTENDED COMPOUND ESTIMATION PROBLEM IN THE SEQUENCE CASE

I. Introduction and Summary.

A. The problem

Let $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_n, \ldots)$ be a countably infinite vector whose components θ_i are elements of some finite interval Ω of the real line. Let $\mathfrak{F} = \{p_\theta(\cdot): \theta \in \Omega\}$ be for some measure μ a family of known probability density functions with parameter θ. Let X_i be a real valued random variable with density $p_{\theta_i}(\cdot)$. Suppose the vector $\underline{\theta}$ is unknown and for each i it is desired to estimate θ_i. The estimates are to be made in sequence and the estimate of θ_i may be based on the independent observations X_j $j = 1, \ldots, i$. Thus for each $i = 1, 2, \ldots$, a non randomized estimator $\varphi_i(\underline{X}_i)$ is sought for θ_i, where $\underline{X}_i$ is the vector of observations $(X_1, \ldots, X_i)$. It is assumed that at each stage of this estimation problem one suffers squared error loss, so that if φ_i is the estimate of θ_i a loss of $(\varphi_i - \theta_i)^2$ units is suffered. The risk of an estimator φ_i is defined to be the expected loss, that is $E[(\varphi_i(\underline{X}_i) - \theta_i)^2]$. The average risk for n estimations becomes $\frac{1}{n}\sum_{i=1}^{n} E[(\varphi_i(\underline{X}_i) - \theta_i)^2]$. One would like to find, for specified Ω and $\mathfrak{F}$, a decision procedure $\underline{\varphi} = (\varphi_1, \varphi_2, \ldots)$ which, on the basis of its average risk for the first n estimations, is in some sense optimal for large n.

One way in which such a problem could arise is as follows: suppose the Navy wishes to screen all new recruits and to classify them on the basis of their "natural aptitudes" to be radar technicians. In an attempt to do this, each recruit is given a test whose outcome can be represented as a number. Suppose also that "natural aptitude" can be represented on a numerical scale. On the basis of prolonged testing and evaluation in the past, the Navy has been able to fit a good probability distribution model for the outcome of a person's test score given his "true" aptitude as a parameter. The Navy now wants to estimate each new recruit's aptitude on the basis of his test score. While squared error loss is somewhat artificial, it is clear that the more the Navy errs in estimating a recruit's aptitude the greater the loss it suffers, and squared error loss is a convenient way to represent this. In this example it is also apparent that many decisions will be made, and from the Navy's point of view, the average risk incurred is a reasonable basis upon which to judge the "optimality" of a decision procedure. In this example, then, θ_i would be the i^{th} recruit's true aptitude and X_i would be his test score.

In the preceding example it is not unreasonable to assume each recruit's aptitude is independent of all other recruits' aptitudes. An example will now be presented in which it is not unreasonable to suppose the θ_i's would occur in "patterns." Suppose a Navy anti-submarine group is on patrol duty to guard against submarine penetration. It is necessary, in deciding what type of patrol to carry out, to have an estimate of the average sonar detection range. This range will depend upon many different factors such as sea temperature and salinity, as well as the sonar equipment involved. Suppose a test is conducted every few hours whose results follow reasonably well a known probability distribution with the true average detection range as a parameter. In this example then, θ_i would be the true detection range and X_i the test result. One would not expect θ_i and θ_{i+1} to be unrelated, however, as the conditions fixing their value, while changing, are changing more or less continuously in time and a high value of θ_i would tend to mean a high value of θ_{i+1} as well. In this example, as in the previous one, a decision about the true detection range will be made many times, and the average risk is a reasonable criterion to use in evaluating a decision procedure.

B. Known results.

The problem of finding a good estimator is really twofold. First some standard of optimality must be established, and secondly a procedure must be found which yields good results according to this standard. Samuel [11] has considered the following standard. Fix $\underline{\theta}_n$. Let $G_n(\cdot)$ be the empirical distribution function of $\underline{\theta}_n$. That is

$$G_n(x) = \frac{1}{n} \text{ (the number of } i \text{ such that } \theta_i \leq x) .$$

Let $\{\Theta_i; i = 1, \ldots, n\}$ be mutually independent identically distributed random variables with a priori distribution function G_n. If we now consider X_i to be an observation of a random variable with the conditional density function p_{θ_i} given that $\Theta_i = \theta_i$, then the usual Bayes argument gives $\varphi_i(\underline{X}_i) = E[\Theta_i | X_i]$ as the estimator achieving the minimum Bayes risk $R(G_n)$. Of course this procedure does not apply to the compound estimation problem since G_n is unknown and in any case the θ_i are not observations of random variables. Nevertheless Samuel has shown $R(G_n)$ is an "optimal" standard to use in evaluating a procedure $\underline{\varphi}$ in the following sense: Let $R_n(\underline{\varphi}, \underline{\theta})$ denote the average risk for the first n decisions incurred by a decision procedure $\underline{\varphi}$ against a parameter vector $\underline{\theta}$. Then $R(G_n)$ is an "optimal" standard in that if one considers only the class of "obvious" procedures $\{\underline{\varphi}_n: \varphi_i(\underline{X}_i) = \varphi(X_i)$ $i = 1, \ldots, n\}$ then $\forall n \; R_n(\underline{\varphi}, \underline{\theta}) \geq R(G_n)$. In other words if one bases his decision about θ_i only on the observation having θ_i as a parameter and uses the same rule for each i, one can never achieve a lower average risk than the number $R(G_n)$.

Samuel also gives several sufficient conditions on Ω, $\{p_\theta: \theta \epsilon \Omega\}$, and $\underline{\varphi}$ which ensure that for each fixed $\underline{\theta}$

$$\overline{\lim_{n \to \infty}} \; (R_n(\underline{\varphi}, \underline{\theta}) - R(G_n)) \leq 0$$

and in several cases she exhibits specific procedures which satisfy the above condition.

Robbins [6] [7] [8] and Johns [3] have done work in the related empirical Bayes problem (see Chapter III, Section G) and many of the decision procedures they derive are also "optimal" in the compound decision problem. Extensions of their estimators will be used in later sections.

C. Summary of new results.

As mentioned in Section B it is first necessary to establish a reasonable standard of "optimality" to use in evaluating a particular decision procedure. Many reasons have been advanced in the literature for considering the risk $E[(\varphi_i - \theta_i)^2]$ ~~to be a~~ good indication of how well a particular decision rule does. In the compound decision problem, it seems even more reasonable to consider the average risk $R_n(\underline{\varphi}, \underline{\theta})$ as a reliable index to be used in evaluating a particular decision procedure $\underline{\varphi}$, and this is the index adopted in this paper. A standard $R(\underline{\theta}_n)$ is now needed such that if for all $\underline{\theta}$ and n $R(\underline{\varphi}_n, \underline{\theta}_n)$ is no greater than $R(\underline{\theta}_n)$, one would be willing to say $\underline{\varphi}$ is a good decision procedure. Samuel has given good intuitive reasons for selecting $R(\underline{\theta}_n) = R(G_n)$, and has made the statement [11] that $R(G_n)$ cannot, in the limit, be improved upon. Based on an idea of Johns [5], a sequence of more stringent standards $\{R_k(\underline{\theta}_n)$: $k = 1, 2, \ldots\}$ will, however, be obtained in this paper such that $R_1(\underline{\theta}_n) = R(G_n)$; and for any fixed $k = 1, 2, \ldots$ and for all $\underline{\theta}$, $R_k(\underline{\theta}_n) = R_{k+1}(\underline{\theta}_n) + f(k, n, \underline{\theta}) + h(k, n, \underline{\theta})$ where $f(k, n, \underline{\theta}) \geq 0$ and $h(k, n, \underline{\theta}) = O(\frac{1}{n})$ uniformly in $\underline{\theta}$. In addition for "most" $\underline{\theta}$, $f(k, n, \underline{\theta})$ is in fact strictly positive. $R_k(\underline{\theta}_n)$ will be shown to be the minimum Bayes risk possible if in fact $\underline{\theta}_n$ is a realization of an n dimensional random vector whose last k components are independently distributed from the first $n - k$ components according to the k dimensional empirical distribution function generated by $\underline{\theta}_n$. The extended compound estimation problem is defined to be the problem of finding procedures which asymptotically achieve these standards. The analogous problem in the empirical Bayes case is being considered by Barndorff-Nielsen [1]. To make these statements more explicit several definitions are needed. These definitions will be used throughout the paper.

Def. 1) Let Ω be a bounded interval of the real line. Let $\mathfrak{F} = \{p_\theta: \theta \in \Omega\}$ be a family of probability density functions with respect to some measure μ. Let $\{\theta_j: \theta_j \in \Omega \ j = 1, 2, 3, \ldots\}$ be an arbitrary sequence. Let $\{X_j: j = 1, 2, \ldots\}$ be a sequence of mutually independent real valued random variables with X_j distributed according to p_{θ_j}. Let $\underline{X}_j = (X_1, \ldots, X_j)$. Let $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_n, \ldots)$ where $\theta_i \in \Omega$ $i = 1, 2, \ldots$. Let $\underline{\theta}_n$ be the vector consisting of the first n components of $\underline{\theta}$.

Def. 2) $\forall \underline{\theta}$, n, $\forall k = 1, \ldots, n$ the k^{th} order empirical distribution function of $\underline{\theta}_n$ is:

$$G_n^k(y_1, y_2, \ldots, y_k) = \frac{1}{n - k + 1}$$

$$(\# \text{ of } j \ (k \leq j \leq n) \text{ such that: } \theta_{j-k+\ell} \leq y_\ell \ \ \ell = 1, 2, \ldots, k)$$

When $k = 1$ this definition yields the usual empirical distribution function.

Let k and m be fixed arbitrary positive integers $k \leq m$. Let $\{\Theta_i: i = 1, \ldots, m\}$ be a sequence of random variables with range space Ω. Let $\Theta_{m-k+1}, \ldots, \Theta_m$ have an a priori joint distribution function G and assume the remaining Θ_i are distributed independently of Θ_m. Let $\{X_i: i = 1, \ldots, m\}$ be a sequence of random variables with conditional density functions p_{θ_i} given $\Theta_i = \theta_i$ such that the X_i are mutually conditionally independent given the Θ_i. For estimating the realization θ_m of Θ_m it is well known that the estimate $E[\Theta_m | X_{m-k+1}, \ldots, X_m]$, which depends only on the last k observations, is a Bayes estimate and achieves the Bayes risk $R(G)$.

Def. 3) $\forall\ \underline{\theta},\ n,\ \forall\ k = 1, 2, \ldots, n$

let $R_k(\underline{\theta}_n) = R(G_n^k)$ where G_n^k is the k^{th} order empirical distribution function of $\underline{\theta}_n$. Thus $R_k(\underline{\theta}_n)$ is the Bayes risk for G_n^k.

Using the above definitions it will be shown in theorem 1) that

$$f(k, n, \underline{\theta}) = E\{[E(\Theta_{k+1}|\underline{X}_{k+1}) - E(\Theta_{k+1}|X_2, \ldots, X_{k+1})]^2\}$$

where $\Theta_1, \ldots, \Theta_{k+1}$ have the a priori joint distribution function G_n^{k+1}. It is clear that f_i is always non negative and will equal zero only if $E[\Theta_{k+1}|\underline{X}_{k+1}] = E[\Theta_{k+1}|X_2, \ldots, X_{k+1}]$ with probability one. This condition is clearly satisfied for most $\mathfrak{F}$ only if $\underline{\theta}_n$ generates an empirical distribution function G_n^{k+1} such that Θ_1 and Θ_{k+1} are independently distributed. It is not unreasonable to suppose that "few" arbitrary sequences, occurring in situations leading to the compound decision problem, will satisfy this condition, even as n approaches ∞. Another necessary condition for $E[\Theta_{k+1}|\underline{X}_{k+1}] = E[\Theta_{k+1}|X_2, \ldots, X_{k+1}]$ is that the sample serial correlation coefficient lag $k + 1$ of $\{\theta_i: i = 1, \ldots, n\}$ be zero. Again it seems unlikely that many sequences of θ_i would have this property, especially for small values of k. In particular if $\underline{\theta}$ has repeated "patterns" of length greater than k, neither of these conditions would be expected to hold. Accepting $R_k(\underline{\theta}_n)$ as a standard to be used in evaluating a decision procedure φ, attention is turned to constructing procedures for specific classes $\mathfrak{F}$ and to evaluating these procedures.

Def. 4) Let $\mathcal{E}_n^k(\underline{\varphi}, \underline{\theta}) = R_n(\underline{\varphi}, \underline{\theta}) - R_k(\underline{\theta}_n)$. Thus $\mathcal{E}_n^k(\underline{\varphi}, \underline{\theta})$ represents the difference after n decisions between the average risk attained by a particular decision procedure and the k^{th} standard.

For many important classes $\mathfrak{F}$, including the normal, gamma, a discrete exponential family, and a "non-parametric" class, decision procedures $\underline{\varphi}^k$ will be found and an upper bound $B(k, n)$ will be given such that $\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta}) \leq B(k, n)$ for all $\underline{\theta} \in \Omega^\infty$ and such that $\lim_{n\to\infty} B(k, n) = 0$. For the discrete exponential family, which includes the geometric, negative binomial, and Poisson families, it will be shown that $B(k, n) = O\left(\frac{\log^k n}{n^{1/4}}\right)$. These results represent a considerable improvement over those obtained by Samuel [11] who considered only the case $k = 1$ and showed

$$\overline{\lim_{n\to\infty}} \, [R_n(\underline{\varphi}, \underline{\theta}) - R_1(\underline{\theta}_n)] \leq 0$$

for any fixed $\underline{\theta}$ in a parameter space more restricted than that considered in this paper. If $\mathfrak{F}$ is the class of binomial probability density functions, a decision procedure is obtained which attains a lower average risk than previously known procedures, and $O\left(\frac{\log n}{n^{1/4}}\right)$ is obtained as the rate of convergence of this risk to its "standard."

II. Preliminary Results.

In this chapter we shall first prove that $R_k(\underline{\theta}_n) = R_{k+1}(\underline{\theta}_n) + f(k, n, \underline{\theta}) + h(k, n, \underline{\theta})$ where f and h have the properties stated in Chapter I. We shall then develop a general theorem and corollary which will enable us to obtain specific decision procedures in Chapter III. Finally we shall prove several lemmas which will be useful in Chapter III.

Def. 5) Let $\underline{y}_n = (y_1, y_2, \dots, y_n)$ be an arbitrary vector. For $k \leq n$ we define $\underline{y}_n^k = (y_{n-k+1}, y_{n-k+2}, \dots, y_n)$.

Def. 6) $\forall \mathcal{G}, \underline{\theta}, \forall k, n$ such that $1 \leq k \leq n$ let

$$Q_n(\underline{x}_k) = \frac{1}{n-k+1} \sum_{j=k}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)$$

$$Q_n^*(\underline{x}_k) = \frac{1}{n-k+1} \sum_{j=k}^{n} \theta_j \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)$$

While both Q and Q^* are functions of several variables which are not explicit in the notation, it will always be clear in context what arguments are intended. We note that $Q_n(\underline{x}_k)$ is the unconditional density function of a random vector $\underline{X}_k$ if the parameters $\theta_1, \dots, \theta_k$ are assumed to be random variables $\Theta_1, \dots, \Theta_k$ with a priori distribution function G_n^k.

Def. 7) $\forall\ \mathfrak{G},\ \underline{\theta},\ \forall\ k, n$ such that $1 \leq k \leq n$, let

$$\psi_n^k(\underline{x}_k) = \begin{cases} \dfrac{Q_n^*(\underline{x}_k)}{Q_n(\underline{x}_k)} = \dfrac{\sum_{j=k}^{n} \theta_j \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}{\sum_{j=k}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)} & \text{if } Q_n > 0 \\ 0 & \text{otherwise} . \end{cases}$$

Let m and n be integers such that $1 \leq k \leq m$, and $n < \infty$, then $\psi_n^k(\underline{X}_m^k)$ is one version of $E[\Theta_m | \underline{X}_m^k]$, and $R_k(\underline{\theta}_n) = E[(\Theta_m - \psi_n^k(\underline{X}_m^k))^2] = E[\Theta_m^2 - (\psi_n^k(\underline{X}_m^k))^2]$.

Def. 8) $\forall\ k, j$ such that $1 \leq k \leq j$ let

$$F_j[\varphi, \underline{\theta}_j^k] = E[(\varphi(\underline{X}_j^k) - \theta_j)^2]$$

where $\varphi(\cdot)$ is an arbitrary non randomized estimator with a k-dimensional argument.

$$\forall\ n \geq k \text{ let } R(\varphi, \underline{\theta}_n) = \frac{1}{n - k + 1} \sum_{j=k}^{n} F_j[\varphi, \underline{\theta}_j^k] .$$

$R(\varphi, \underline{\theta}_n)$ is then the Bayes risk in using the rule $\varphi(\underline{X}_k)$ as an estimate of Θ_k when $\Theta_1, \ldots, \Theta_k$ have the a priori distribution G_n^k.

We now compare $R_k(\underline{\theta}_n)$ with $R_{k+1}(\underline{\theta}_n)$ and prove:

Theorem 1) $\forall\ \mathfrak{G},\ \underline{\theta},\ k, n\ \ 1 \leq k < n < \infty\ \ R_k(\underline{\theta}_n) = R_{k+1}(\underline{\theta}_n) + f(k, n, \underline{\theta}) + h(k, n, \underline{\theta})$ where $f(k, n, \underline{\theta}) \geq 0$ and $|h(k, n, \underline{\theta})| = O\left(\frac{1}{n}\right)$ uniformly in $\underline{\theta}$.

Proof:

Let $E_1[\cdot]$ refer to expectation with respect to G_n^k and $E_2[\cdot]$ refer to expectation with respect to G_n^{k+1}. For any estimator $\varphi(\underline{X}_{k+1})$

$$R(\varphi(\underline{X}_{k+1}), \underline{\theta}_n) = E_2[(\varphi(\underline{X}_{k+1}) - \Theta_{k+1})^2]$$

$$= E_2[\Theta_{k+1}^2] - E_2[E_2^2(\Theta_{k+1}|\underline{X}_{k+1})]$$

$$+ E_2\{[E_2(\Theta_{k+1}|\underline{X}_{k+1}) - \varphi(\underline{X}_{k+1})]^2\}$$

which of course is minimized for $\varphi(\underline{X}_{k+1}) = E_2[\Theta_{k+1}|\underline{X}_{k+1}]$. Letting $\varphi(\underline{X}_{k+1}) = E_2[\Theta_{k+1}|\underline{X}_{k+1}^k]$ we obtain

$$R_k(\underline{\theta}_n) - R_{k+1}(\underline{\theta}_n) = R_k(\underline{\theta}_n) - R(E_2[\Theta_{k+1}|\underline{X}_{k+1}^k], \underline{\theta}_n)$$

$$+ E_2\{[E_2(\Theta_{k+1}|\underline{X}_{k+1}) - E_2(\Theta_{k+1}|\underline{X}_{k+1}^k)]^2\} .$$

Let $\quad f(k, n, \underline{\theta}) = E_2\{[E_2(\Theta_{k+1}|\underline{X}_{k+1}) - E_2(\Theta_{k+1}|\underline{X}_{k+1}^k)]^2\}$

$$h(k, n, \underline{\theta}) = R_k(\underline{\theta}_n) - R(E_2[\Theta_{k+1}|\underline{X}_{k+1}^k], \underline{\theta}_n)$$

Then it remains only to show $|h(k, n, \underline{\theta})| = O\left(\frac{1}{n}\right)$ uniformly in $\underline{\theta}$. But since

$$R_k(\underline{\theta}_n) = E_1[\Theta_k^2] - E_1\{E_1^2[\Theta_k|\underline{X}_k]\} \qquad \text{and}$$

$$R(E_2[\Theta_{k+1}|\underline{X}_{k+1}^k], \underline{\theta}_n) = E_2[\Theta_{k+1}^2] - E_2\{E_2^2[\Theta_{k+1}|\underline{X}_{k+1}^k]\}$$

and since Ω is a bounded interval, say $\theta \epsilon \Omega \Rightarrow |\theta| \leq B < \infty$ we have:

$$|h(k, n, \underline{\theta})| \leq |E_1(\Theta_k^2) - E_2(\Theta_{k+1}^2)| + |E_1\{E_1^2[\Theta_k|\underline{X}_k]\} - E_2\{E_2^2[\Theta_{k+1}|\underline{X}_{k+1}^k]\}|$$

and

$$|E_1(\Theta_k^2) - E_2(\Theta_{k+1}^2)| = \left|\frac{1}{n - k + 1} \sum_{j=k}^{n} \theta_j^2 - \frac{1}{n - k} \sum_{j=k+1}^{n} \theta_j^2\right|$$

$$= \frac{1}{(n - k + 1)(n - k)}\left| - \sum_{j=k+1}^{n} \theta_j^2 + (n - k)\theta_k^2\right|$$

$$\leq \frac{B^2}{n - k + 1} .$$

Also

$$|E_1\{E_1^2[\Theta_k|\underline{X}_k]\} - E_2\{E_2^2[\Theta_{k+1}|\underline{X}_{k+1}^k]\}|$$

$$= \left|\int_{R^k} \left\{\left(\frac{\sum_{j=k}^{n} \theta_j \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}{\sum_{j=k}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}\right)^2 \frac{\sum_{j=k}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}{n - k + 1}\right.\right.$$

$$\left.\left. - \left(\frac{\sum_{j=k+1}^{n} \theta_j \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}{\sum_{j=k+1}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}\right)^2 \frac{\sum_{j=k+1}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}{n - k}\right\} \mu(d\underline{x}_k)\right| .$$

For fixed $\underline{x}_k$ the expression inside the braces may be written as

$$\left\{\left(\frac{a_n + a}{b_n + b}\right)^2 \frac{b_n + b}{n - k + 1} - \left(\frac{a_n}{b_n}\right)^2 \frac{b_n}{n - k}\right\}$$

where $a_n = \sum_{j=k+1}^{n} \theta_j \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)$ $\qquad a = \theta_k \prod_{\ell=1}^{k} p_{\theta_\ell}(x_\ell)$

$$b_n = \sum_{j=k+1}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell) \qquad b = \prod_{\ell=1}^{k} p_{\theta_\ell}(x_\ell)$$

and

$$\left| \left(\frac{a_n + a}{b_n + b}\right)^2 \frac{b_n + b}{n - k + 1} - \left(\frac{a_n}{b_n}\right)^2 \frac{b_n}{n - k} \right|$$

$$= \left| \frac{(n - k)b_n(a_n + a)^2 - (n - k + 1)(b_n + b)a_n^2}{(n - k)(n - k + 1)(b_n + b)b_n} \right|$$

$$= \left| \frac{(n - k)(2b_n a\, a_n + b_n a^2 - b\, a_n^2) - b_n a_n^2 - b\, a_n^2}{(n - k)(n - k + 1)(b_n + b)b_n} \right|$$

$$\leq \frac{1}{n - k + 1} \left\{ \left|\frac{2a\, a_n}{b_n + b}\right| + \left|\frac{a^2}{b_n + b}\right| + \left|\frac{b\, a_n^2}{(b_n + b)b_n}\right| \right\}$$

$$+ \frac{1}{(n - k)(n - k + 1)} \left\{ \left|\frac{a_n^2}{b_n + b}\right| + \left|\frac{b\, a_n^2}{(b_n + b)b_n}\right| \right\}.$$

From the definition of a_n, b_n, a, and b it is clear that

$\left|\frac{a_n}{b_n}\right| \leq B \quad \left|\frac{a}{b}\right| \leq B \quad b_n \geq 0 \quad b \geq 0$ so that

$$\left|\frac{2a\, a_n}{b_n + b}\right| \leq \left|\frac{2a\, a_n}{b_n}\right| = 2\left|\frac{a_n}{b_n}\right| \left|\frac{a}{b}\right| \left|b\right| \leq 2B^2 b$$

$$\left|\frac{a^2}{b_n + b}\right| \le \left|\frac{a^2 b}{b^2}\right| \le B^2 b$$

$$\left|\frac{b\, a_n^2}{(b_n + b)(b_n)}\right| \le \left|\frac{b\, a_n^2}{b_n^2}\right| \le B^2 b$$

$$\left|\frac{a_n^2}{b_n + b}\right| \le \left|\frac{a_n^2 b_n}{b_n^2}\right| \le B^2 b_n$$

Thus we have

$$|E_1\{E_1^2[\Theta_k|\underline{X}_k]\} - E_2\{E_2^2[\Theta_{k+1}|\underline{X}_{k+1}^k]\}|$$

$$\le \int_{R^k} \left(\frac{4B^2 b}{n - k + 1} + \frac{B^2(b_n + b)}{(n - k)(n - k + 1)}\right)\mu(d\underline{x}_k)$$

$$= \frac{4B^2}{n - 1 + 1} \int_{R^k} \left(\prod_{\ell=1}^{k} p_{\theta_\ell}(x_\ell)\right)\mu(d\underline{x}_k)$$

$$+ \frac{B^2}{(n - k)(n - k + 1)} \int_{R^k} \left(\prod_{\ell=1}^{k} p_{\theta_\ell}(x_\ell)\right)\mu(d\underline{x}_k)$$

$$+ \frac{B^2}{n - k + 1} \int_{R^k} \frac{\sum_{j=k+1}^{n} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)}{n - k}\, \mu(d\underline{x}_k)$$

$$= \frac{4B^2}{n - k + 1} + \frac{B^2}{(n - k)(n - k + 1)} + \frac{B^2}{n - k + 1}\ .$$

Hence $$|h(k,\ n,\ \underline{\theta})| \leq \frac{7B^2}{n - k + 1}.$$

Q.E.D.

The implications of theorem 1) were discussed in Chapter I. We now state and prove a generalized form of a lemma of Samuel [11].

Lemma 1) $\forall\ \underline{\theta},\ k \geq 1,\ n \geq k$

$$\frac{1}{n - k + 1} \sum_{i=k}^{n} E[(\psi_i^k(\underline{X}_i^k) - \theta_i)^2] \leq R_k(\underline{\theta}_n)$$

Proof: Fix $\underline{\theta}$, k, and n. Using the expressions $F_j(\psi_i^k,\ \underline{\theta}_j^k)$ and $R(\psi_i^k,\ \underline{\theta}_i)$ given in definition 8), and observing that $R(\psi_{i+1}^k,\ \underline{\theta}_i) \geq R_k(\underline{\theta}_i)$ we have:

$$\frac{1}{n - k + 1} \sum_{i=k}^{n} E[(\psi_i^k(\underline{X}_i^k) - \theta_i)^2]$$

$$= \frac{1}{n - k + 1} \sum_{i=k}^{n} \left\{ \sum_{j=k}^{i} F_j(\psi_i^k,\ \underline{\theta}_j^k) - \sum_{j=k}^{i-1} F_j(\psi_i^k,\ \underline{\theta}_j^k) \right\}$$

$$= \frac{1}{n - k + 1} \sum_{i=k}^{n} \left\{ (i - k + 1)R(\psi_i^k,\ \underline{\theta}_i) - (i - k)R(\psi_i^k,\ \underline{\theta}_{i-k}) \right\}$$

$$= \frac{1}{n - k + 1} \sum_{i=k}^{n-1} (i - k + 1)[R(\psi_i^k,\ \underline{\theta}_i) - R(\psi_{i+1}^k,\ \underline{\theta}_i)] + R(\psi_n^k,\ \underline{\theta}_n)$$

$$\leq R(\psi_n^k,\ \underline{\theta}_n) = R_k(\underline{\theta}_n).$$

Q.E.D.

Def. 9) A decision procedure $\underline{\varphi}^k = (\varphi_1^k(X_1), \varphi_2^k(\underline{X}_2), \ldots, \varphi_n^k(\underline{X}_n), \ldots)$ is asymptotically optimal of k^{th} order if

$$\forall \underline{\theta} \quad \overline{\lim_{n\to\infty}} \left[\frac{1}{n} \sum_{i=1}^{n} E[(\varphi_i^k - \theta_i)^2] - R_k(\underline{\theta}_n)\right] \le 0 .$$

If $\overline{\lim_{n\to\infty}} \left\{\sup_{\underline{\theta}} \frac{1}{n} \sum_{i=1}^{n} E[(\varphi_i^k - \theta_i)^2] - R_k(\underline{\theta}_n) \right\} \le 0$ then $\underline{\varphi}^k$ is uniformly asymptotically optimal of k^{th} order.

We shall now state and prove a general theorem, which with its corollary will enable us to obtain the results of Chapter III.

Theorem 2) $\forall$ bounded interval $\Omega = [\alpha, \beta]$, family of densities $\mathfrak{F}$, and integer $k \ge 1$, let $\underline{\varphi}^k$ be a decision procedure such that $\forall\ i \ge k$ $P[\varphi_i(\underline{X}_i) \epsilon\, \Omega] = 1$. Suppose there exist non negative functions $\xi_i(\underline{\theta}, \underline{x}_k)$, $\zeta_i(\underline{\theta}, \underline{x}_k)$, and $a_i(\underline{\theta})$ such that $\forall\ \underline{\theta}, \underline{x}_k,\ i \ge k$

a) $P[|\varphi_i^k(\underline{X}_i) - \psi_i^k(\underline{x}_k)| \ge \xi_i(\underline{\theta}, \underline{x}_k) | \underline{X}_i^k = \underline{x}_k] \le \zeta_i(\underline{\theta}, \underline{x}_k)$

b) $$\lim_{n\to\infty} \left\{\frac{1}{n} \sum_{i=k}^{n} E[\xi_i(\underline{\theta}, \underline{X}_i^k) + \zeta_i(\underline{\theta}, \underline{X}_i^k) | Q_i(\underline{X}_i^k) \ge a_i(\underline{\theta})] + \frac{1}{n} \sum_{i=k}^{n} P[Q_i(\underline{X}_i^k) < a_i(\underline{\theta})]\right\} = 0$$

uniformly in $\underline{\theta}$

where the functions Q_i and ψ_i^k are as given by definitions 6) and 7).

Then the decision procedure $\underline{\varphi}^k$ is uniformly asymptotically optimal of k^{th} order, and moreover

$$\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta}) \leq \frac{(k-1)(\beta-\alpha)^2}{n} + \frac{2(\beta-\alpha)}{n} \sum_{i=k}^{n} E[\xi_i + (\beta-\alpha)\zeta_i | Q_i \geq a_i]$$

$$+ \frac{2(\beta-\alpha)^2}{n} \sum_{i=k}^{n} P[Q_i < a_i]$$

where $\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta})$ is given by definition 4).

Proof: Clearly it is sufficient to prove the upper bound for $\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta})$ is correct.

We first represent $\mathcal{E}_n^k$ as a sum of several functions, and then examine each of these functions. We shall consider k and n fixed.

Let:

$$H_1(\underline{\varphi}^k, \underline{\theta}) = \frac{1}{n} \sum_{i=1}^{k-1} \{E[(\varphi_i^k - \theta_i)^2] - R_k(\underline{\theta})]\}$$

$$H_2(\underline{\varphi}^k, \underline{\theta}) = \frac{1}{n} \sum_{i=k}^{n} \{E[(\varphi_i^k - \theta_i)^2] - E[(\psi_i^k - \theta_i)^2]\}$$

$$H_3(\underline{\varphi}^k, \underline{\theta}) = \frac{1}{n} \sum_{i=k}^{n} \{E[(\psi_i^k - \theta_i)^2] - R_k(\underline{\theta}_n)\} .$$

For the remainder of the proof we delete the superscript k. Clearly

$$\mathcal{E}_n(\underline{\varphi}, \underline{\theta}) = \frac{1}{n} \sum_{i=1}^{n} E[(\varphi_i - \theta_i)^2] - R_k(\underline{\theta}_n)$$

$$= H_1(\underline{\varphi}, \underline{\theta}) + H_2(\underline{\varphi}, \underline{\theta}) + H_3(\underline{\varphi}, \underline{\theta}) .$$

Let $B = (\beta - \alpha)$. Then $E[(\varphi_i - \theta_i)^2] \leq B^2$ and $R_k(\underline{\theta}_n) \leq B^2$; hence $H_1(\underline{\varphi}, \underline{\theta}) \leq \frac{(k-1)B^2}{n}$. It follows immediately from Lemma 1) that $H_3(\underline{\varphi}, \underline{\theta}) \leq 0$. It remains only to examine $H_2(\underline{\varphi}, \underline{\theta})$.

$$H_2(\underline{\varphi}, \underline{\theta}) = \frac{1}{n} \sum_{i=k}^{n} E[(\varphi_i - \psi_i)(\varphi_i + \psi_i - 2\theta_i)]$$

$$\leq \frac{2B}{n} \sum_{i=k}^{n} E[|\varphi_i - \psi_i|]$$

$$\leq \frac{2B}{n} \sum_{i=k}^{n} \min\, [E[|\xi_i + B\zeta_i|], B]$$

$$\leq \frac{2B}{n} \sum_{i=k}^{n} E[\xi_i + B\zeta_i | Q_i \geq a_i] P[Q_i \geq a_i] + \frac{2B^2}{n} \sum_{i=k}^{n} P[Q_i < a_i]$$

The desired result follows immediately.

Q.E.D.

Corollary. If condition b) of theorem 2) is replaced by

b') $\lim\limits_{i \to \infty} E[\xi_i(\underline{\theta}, \underline{X}_i^k) + \zeta_i(\underline{\theta}, \underline{X}_i^k)] = 0$

uniformly in $\underline{\theta}$

then $\underline{\varphi}^k$ is uniformly asymptotically optimal in the k^{th} order.

Proof: From the proof of theorem 2) it is enough to show

$\overline{\lim\limits_{n \to \infty}} \{\sup\limits_{\underline{\theta}} H_2(\underline{\varphi}^k, \underline{\theta})\} \leq 0$, recalling that $H_2(\underline{\varphi}, \underline{\theta})$ is a function of n.

But $H_2(\underline{\varphi}, \underline{\theta}) \leq \frac{2B}{n} \sum_{i=k}^{n} E[|\varphi_i - \psi_i|]$

$$\leq \frac{2B}{n} \sum_{i=k}^{n} E\{\xi_i(\underline{\theta}, \underline{X}_i^k) + B\zeta_i(\underline{X}_i^k)\}$$

$\to 0$ uniformly in $\underline{\theta}$ as $n \to \infty$ since uniform convergence implies uniform convergence in Cesaro mean.

Q.E.D.

We turn now to several lemmas which will be useful in Chapter III. lemmas 2), 3), and 4) will be used to establish condition a) of theorem 2), while lemma 5) will be used in evaluating certain limits.

The first of these is an inequality proved by Hoeffding [2], which we state here without proof.

Lemma 2) If $X_1, X_2, \ldots, X_n$ are independent and $a \leq X_i \leq b$ for $i = 1, \ldots, n$ then $\forall\, t \geq 0$

$$P[|\overline{X} - E[\overline{X}]| \geq t] \leq 2e^{-2n\left(\frac{t}{b-a}\right)^2},$$

where $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Lemma 3) Let $X_1, X_2, \ldots, X_n$ $0 \leq X_i \leq b$ $i = 1, \ldots, n$ be a sequence of random variables such that for some $k > 0$ and $\forall$ $i = 1, 2, \ldots, k$ the random variables $X_i, X_{i+k}, X_{i+2k}, \ldots$ are mutually independent. Then

$$P[|\overline{X} - E[\overline{X}]| \geq \delta] \leq 2ke^{-2\frac{n^2}{k(n+k)}\delta^2}$$

Proof: Let $S_i = \sum_{j=0}^{m} X_{jk+i}$ where m is defined as the integer such that $\frac{n}{k} - 1 < m \leq \frac{n}{k}$ and $X_\ell \equiv 0$ for $\ell > n$. Let $\gamma_i = E[S_i]$. Let A_i = event $|S_i - \gamma_i| \geq \frac{\delta n}{k}$. But S_i is the sum of $m + 1$ independent random variables and from lemma 2) we have:

$$P[A_i] = P\left[\left|\sum_{j=0}^{m} X_{jk+i} - \gamma_i\right| \geq \frac{\delta n}{k}\right]$$

$$= P\left[\frac{1}{m+1}\left|\sum_{j=0}^{m} X_{jk+i} - \gamma_i\right| \geq \frac{\delta n}{k(m+1)}\right]$$

$$\leq 2e^{-2(m+1)\frac{\delta^2 n^2}{k^2(m+1)^2 b^2}}$$

$$\leq 2e^{-2\frac{n^2}{k(n+k)}\frac{\delta^2}{b^2}}$$

So that:

$$P[|\bar{X} - E[\bar{X}]| \geq \delta] = P\left[\left|\sum_{i=1}^{k} (S_i - \gamma_i)\right| \geq \delta n\right]$$

$$\leq P\left[\sum_{i=1}^{k} |S_i - \gamma_i| \geq \delta n\right]$$

$$= 1 - P\left[\sum_{i=1}^{k} |S_i - \gamma_i| < \delta n\right]$$

$$\leq 1 - P\left[\bigcap_{i=1}^{k} A_i^c\right]$$

$$= P\left[\bigcup_{i=1}^{k} A_i\right] \leq \sum_{i=1}^{k} P[A_i]$$

$$\leq 2ke^{-2\frac{n^2}{k(n+k)}\frac{\delta^2}{b^2}}$$

Q.E.D.

Lemma 1.) Suppose for non negative random variables X_1, X_2 $P[|X_i - \mu_i| \geq \delta_i] \leq \epsilon_i$ $i = 1, 2$; $\mu_1 \geq 0$, $\mu_2 > 0$; and $B \neq 0$ is any number such that $\frac{\mu_1}{\mu_2} \leq B < \infty$. Define $W = \min[\frac{X_1}{X_2}, B]$ (if $X_1 = X_2 = 0$ we take $W = 0$). Then

$$P\left[\left|W - \frac{\mu_1}{\mu_2}\right| \geq \frac{1}{\mu_2} (\delta_1 + B\delta_2)\right] \leq \epsilon_1 + \epsilon_2$$

Proof: We first show $|W - \frac{\mu_1}{\mu_2}| \leq \frac{1}{\mu_2}(|X_1 - \mu_1| + B|X_2 - \mu_2|)$

Case i) $W < B$. If $X_2 = 0$ then $X_1 = 0$ and

$$|W - \frac{\mu_1}{\mu_2}| = |\frac{\mu_1}{\mu_2}| \leq \frac{\mu_1}{\mu_2} + B = \frac{1}{\mu_2}(|X_1 - \mu_1| + B|X_2 - \mu_2|)$$

If $X_2 \neq 0$ then

$$|W - \frac{\mu_1}{\mu_2}| = |\frac{X_1}{X_2} - \frac{\mu_1}{\mu_2}| = \frac{1}{X_2\mu_2}|(X_1 - \mu_1)X_2 - (X_2 - \mu_2)X_1|$$

$$\leq \frac{1}{\mu_2}(|X_1 - \mu_1| + B|X_2 - \mu_2|)$$

Case ii) $W = B$. Let $Y = \frac{X_1}{B}$. Then $Y \neq 0$ and $Y \geq X_2$.

$$|W - \frac{\mu_1}{\mu_2}| = B - \frac{\mu_1}{\mu_2} = \frac{X_1}{Y} - \frac{\mu_1}{\mu_2} = \frac{1}{\mu_2}[(X_1 - \mu_1) + B(\mu_2 - Y)]$$

$$\leq \frac{1}{\mu_2}(|X_1 - \mu_1| + B(\mu_2 - X_2) \leq \frac{1}{\mu_2}(|X_1 - \mu_1| + B|X_2 - \mu_2|)$$

From this fact we have

$$|X_1 - \mu_1| < \delta_1 \text{ and } |X_2 - \mu_2| < \delta_2 \Rightarrow |W - \frac{\mu_1}{\mu_2}| < \frac{1}{\mu_2}(\delta_1 + B\delta_2)$$

and

$$P[|W - \frac{\mu_1}{\mu_2}| \geq \frac{1}{\mu_2}(\delta_1 + B\delta_2)] = 1 - P[|W - \frac{\mu_1}{\mu_2}| < \frac{1}{\mu_2}(\delta_1 + B\delta_2)$$

$$\leq 1 - P[|X_1 - \mu_1| < \delta_1 \text{ and } |X_2 - \mu_2| < \delta_2]$$

$$= P[|X_1 - \mu_1| \geq \delta_1 \text{ or } |X_2 - \mu_2| \geq \delta_2]$$

$$\leq \epsilon_1 + \epsilon_2$$

Q.E.D.

Lemma 5) Let F be an absolutely continuous distribution function with corresponding density function f. Let $C_M = \{x: |f''(x)| \leq M \text{ and } f''(x) \text{ is continuous}\}$. Let $D = \{x: |f'(x)| > 0\}$. Let $y \epsilon (-\infty, \infty)$.

i) If there exist $M < \infty$ $\epsilon_o > 0$ such that: $y \epsilon D$ and $\{x: y \leq x \leq y + \epsilon_o\} \subset C_M$ then $\forall \epsilon$ $0 < \epsilon \leq \epsilon_o$

$$\frac{1}{\epsilon}\int_y^{y+\epsilon} f(x)dx = f(y + p\epsilon) \quad \text{with} \quad |p - \tfrac{1}{2}| \leq \frac{2M}{3|f'(y)|}\epsilon$$

ii) If there exist $M < \infty$ $\epsilon_o > 0$ such that: $y \epsilon D$ and $\{x: y - \epsilon_o \leq x \leq y\} \subset C_M$ then $\forall \epsilon$ $0 < \epsilon \leq \epsilon_o$.

$$\frac{1}{\epsilon}\int_{y-\epsilon}^{y} f(x)dx = f(y - p\epsilon) \quad \text{with} \quad |p - \tfrac{1}{2}| \leq \frac{2M}{3|f'(y)|}\epsilon$$

Proof: Using the mean value theorem and the Taylor expansion we have in case i)

$$\frac{1}{\epsilon}\int_{y}^{y+\epsilon} f(x)dx = f(y + p\epsilon) = f(y) + f'(y)p\epsilon + \frac{f''(x^*)}{2}(p\epsilon)^2$$

where $0 \leq p \leq 1$ and $y \leq x^* \leq y + \epsilon$.

But:

$$\frac{1}{\epsilon}\int_{y}^{y+\epsilon} f(x)dx = \frac{1}{\epsilon}(F(y+\epsilon) - F(y)) = \frac{1}{\epsilon}\left(f(y)\epsilon + \frac{f'(y)}{2}\epsilon^2 + \frac{f''(x')\epsilon^3}{6}\right)$$

where $y \leq x' \leq y + \epsilon$.

Thus

$$f'(y)p\epsilon + \frac{f''(x^*)(p\epsilon)^2}{2} = \frac{f'(y)\epsilon}{2} + \frac{f''(x')\epsilon^2}{6}$$

$$\Rightarrow (p - \tfrac{1}{2})f'(y) = \left(\frac{f''(x')}{6} - \frac{f''(x^*)}{2}p^2\right)\epsilon$$

$$\Rightarrow |p - \tfrac{1}{2}||f'(y)| \leq \left(\left|\frac{f''(x')}{6}\right| + \left|\frac{f''(x^*)}{2}\right|p^2\right)\epsilon$$

$$\Rightarrow |p - \tfrac{1}{2}| \leq \frac{1}{|f'(y)|}\left(\frac{M}{6} + \frac{M}{2}\right)\epsilon = \frac{2M\epsilon}{3|f'(y)|}.$$

The proof in case ii) is similar.

Q.E.D.

III. Main results.

We now turn to the task of finding asymptotically optimal procedures, bounds, and rates of convergence for specific classes of distributions. We shall also look at a modification of our problem in a very general class of distributions. Finally we shall consider the "empirical Bayes" problem.

The notation we shall develop and use is inherently cumbersome; to ease its burden somewhat we shall not always indicate all possible dependencies and shall not always indicate one or more of the arguments of a function. Hopefully no misunderstanding will arise because of this practice.

A. A special discrete class of distributions.

We first consider a special discrete class of distributions defined on the non negative integers as follows:

$$P[X = x|\theta] = p_\theta(x) = \theta^x h(\theta) g(x) \qquad x = 0, 1, 2, \ldots$$

where: i) $\theta \in \Omega = [0, \beta] \quad 0 < \beta < \infty \quad h(\beta) > 0$;

ii) There exists $M^* < \infty$ such that: $\forall\, \theta \in \Omega \;\; E_\theta\left[\frac{g(X)}{g(X + 1)}\right] \leq M^*$

iii) There exists $M' < \infty$ such that $\forall$ $i, j = 0, 1, 2, \ldots$ such that $g(i)\, g(j) \neq 0$

$$\frac{g(i + j)}{g(i)g(j)} \leq M'$$

iv) If $\beta < 1$ then there exists a constant b such that $g(x) \leq x^b$ for all but finitely many integers x.
If $\beta \geq 1$ then there exists a constant b such that $g(x) \leq \frac{1}{x-b}$ for all but finitely many integers x.

All of these restrictions are quite mild. The third prevents g from oscillating wildly as its argument progresses through the integers. The second and fourth conditions restrict slightly the rate at which $\theta^x g(x) \to 0$ as $x \to \infty$.

Examples of such a class are:

Type	θ	$h(\theta)$	$g(x)$
Poisson	$[0, \beta], \beta < \infty$	$e^{-\theta}$	$\frac{1}{x!}$
geometric	$[0, \beta], \beta < 1$	$(1 - \theta)$	1
negative binomial	$[0, \beta], \beta < 1$	$(1 - \theta)^a$, $a > 0$ a assumed known	$\binom{a + x - 1}{x}$

The conditions are easily seen to be satisfied.

Recalling that $\underline{x}_j^k = (x_{j-k+1}, \ldots, x_j)$ $j = k, k + 1, \ldots$

we define:

$$Y_j(\underline{x}_k) = \begin{cases} 1 & \text{if } \underline{x}_j^k = \underline{x}_k \\ 0 & \text{otherwise} \end{cases} \qquad j = k, k + 1, \ldots$$

$$P_i(\underline{x}_k) = \frac{1}{i - k + 1} \sum_{j=k}^{i} Y_j(\underline{x}_k) \qquad i = k, k + 1, \ldots$$

$$P_i^*(\underline{x}_k) = \begin{cases} \dfrac{g(x_k)P_i(x_1, x_2, \cdots, x_{k-1}, x_k + 1)}{g(x_k + 1)P_i(\underline{x}_k)} & \text{if } g(x_k + 1)P_i(\underline{x}_k) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$i = k, k + 1, \cdots$$

$$\varphi_i^k(\underline{X}_i) = \begin{cases} \dfrac{\beta}{2} & \text{if } i = 1, \cdots, k - 1 \\ P_i^*(\underline{X}_i^k) & \text{if } 0 \le P_i^*(\underline{X}_i^k) < \beta \quad i = k, k + 1, \cdots, \\ \beta & \text{if } \beta \le P_i^*(\underline{X}_i^k) \qquad i = k, k + 1, \cdots, \end{cases}$$

Theorem 3) For the problem defined in this section $\underline{\varphi}^k = (\varphi_1^k, \varphi_2^k, \cdots)$ is uniformly asymptotically optimal of k^{th} order for $k = 1, 2, \cdots$, and $\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta}) \le B(k, n) = O\left(\frac{\log^k n}{n^{1/\text{[illegible]}}}\right)$.

Proof: We shall show the conditions of Theorem 2) are satisfied.

Fix k. Fix $\underline{\theta}$.

Recall $Q_i(\underline{x}_k) = \frac{1}{i - k + 1} \sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell) \quad i = k, k + 1, \cdots$

We observe that there exists a set R_i in k dimensional space such that $P[\underline{X}_i^k \in R_i] = 1$ and $\underline{x}_k \in R_i \Rightarrow g(x_k)Q_i(\underline{x}_k) > 0$. We then have $\underline{x}_k \in R_i \Rightarrow g(x_k + 1)Q_i(\underline{x}_k) > 0$. $\forall\, \underline{x}_k \in R_i$ we know from definition 7) that $\forall\, i \ge k$:

$$\psi_i^k(\underline{x}_k) = \frac{\sum_{j=k}^{i} \theta_j \prod_{\ell=1}^{k} \theta_{j-k+\ell}^{x_\ell} h(\theta_{j-k+\ell}) g(x_\ell)}{\sum_{j=k}^{i} \prod_{\ell=1}^{k} \theta_{j-k+\ell}^{x_\ell} h(\theta_{j-k+\ell}) g(x_\ell)}$$

$$= \frac{g(x_k) Q_i(x_1, x_2, \dots, x_{k-1}, x_k + 1)}{g(x_k + 1) Q_i(\underline{x}_k)}$$

We now consider the Y_j's. Since the X_j's $j = 1, \dots, i$ are independent it is clear that $\forall \underline{x}_k$, $\ell = 1, 2, \dots, k$, the random variables $Y_\ell(\underline{x}_k)$, $Y_{\ell+k}(\underline{x}_k)$, $Y_{\ell+2k}(\underline{x}_k)$, ... are mutually conditionally independent given $\underline{X}_i^k = \underline{x}_k$. Also:

$$E[Y_j(\underline{x}_k) | \underline{X}_i^k = \underline{x}_k] = P[\underline{X}_j^k = \underline{x}_k | \underline{X}_i^k = \underline{x}_k]$$

$$= \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell) \qquad j = k, \dots, i - k$$

$$0 \le E[Y_j(\underline{x}_k) | \underline{X}_i^k = \underline{x}_k] \le 1 \qquad j = i - k + 1, \dots, i$$

Now $\forall \underline{x}_k \epsilon R_i$ $\forall \delta_i(\underline{x}_k, \underline{\theta}_i) \ge \dfrac{kg(x_k + 1)}{i - k + 1}$ $i = 2k, \dots$

we have, using lemma 3):

$$P[|g(x_k + 1)P_i(\underline{x}_k) - g(x_k + 1)Q_i(\underline{x}_k)| \geq \delta_i | \underline{X}_i^k = \underline{x}_k]$$

$$\leq P\left[|P_{i-k}(\underline{x}_k) - Q_{i-k}(\underline{x}_k)| + \left|\frac{1}{i - 2k + 1} \sum_{j=i-k+1}^{i} Y_j(\underline{x}_k)\right.\right.$$

$$\left. - \frac{1}{i - 2k + 1} \sum_{j=i-k+1}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)\right|$$

$$\left.\geq \frac{\delta_i(i - k + 1)}{g(x_k + 1)(i - 2k + 1)} \middle| \underline{X}_i^k = \underline{x}_k\right]$$

$$\leq P\left[|P_{i-k}(\underline{x}_k) - Q_{i-k}(\underline{x}_k)| \geq \frac{\delta_i(i - k + 1)}{g(x_k + 1)(i - 2k + 1)}\right.$$

$$\left. - \frac{k}{(i - 2k + 1)} \middle| \underline{X}_i^k = \underline{x}_k\right]$$

$$\leq 2k \exp\left\{- 2 \frac{(i - 2k + 1)^2}{k(i - k + 1)} \left[\frac{\delta_i(i - k + 1)}{g(x_k + 1)(i - 2k + 1)} - \frac{k}{(i - 2k + 1)}\right]^2\right\}$$

$$\leq 2k \exp\left\{\frac{4\delta_i}{g(x_k + 1)} - \frac{2\delta_i^2(i - k + 1)}{k\, g^2(x_k + 1)}\right\}$$

By a similar argument $\forall\, \underline{x}_k \epsilon R_i\ \forall\, \epsilon_i(\underline{x}_k, \theta_i) \geq \dfrac{kg(x_k)}{i - k + 1}$ we have:

$$P[|g(x_k)P_i(x_1, \ldots, x_{k-1}, x_k + 1) - g(x_k)Q_i(x_1, \ldots, x_{k-1}, x_k + 1)|$$

$$\geq \epsilon_i | \underline{X}_i^k = \underline{x}_k] \leq 2k \exp\left\{\frac{4\epsilon_i}{g(x_k)} - \frac{2\epsilon_i^2(i - k + 1)}{k\, g^2(x^k)}\right\}$$

We are now in a position to apply lemma 4) to obtain the functions ξ_i and ζ_i for condition a) of theorem 2). If we substitute in lemma 4)

$$W = \varphi_i^k(\underline{X}_i)$$

$$\mu_1 = g(x_k)Q_i(x_1, \ldots , x_{k-1}, x_k + 1)$$

$$\mu_2 = g(x_k + 1)Q_i(\underline{x}_k)$$

Then we have $\forall\ i = 2k, 2k + 1, \ldots$

$$P\left[|\varphi_i^k - \psi_i^k| \geq \frac{\epsilon_i + \beta\delta_i}{Q_i(\underline{x}_k)g(x_k + 1)}|\underline{X}_i^k = \underline{x}_k\right]$$

$$\leq 2k\left[e^{-\frac{2\epsilon_i^2(i-k+1)}{kg^2(x_k)} + \frac{4\epsilon_i}{g(x_k)}} + e^{-\frac{2\delta_i^2(i-k+1)}{kg^2(x_k+1)} + \frac{4\delta_i}{g(x_k+1)}}\right]$$

Recall the above inequality has been shown only for $\underline{x}_k \epsilon R_i$, $i \geq 2k$, $\delta_i \geq \frac{kg(x_k + 1)}{i - k + 1}$, $\epsilon_i \geq \frac{kg(x_k)}{i - k + 1}$. If any of these conditions should not hold we shall use the trivial inequality

$$P[|\varphi_i^k - \psi_i^k| \geq 0|\underline{X}_i^k = \underline{x}_k] \leq 1$$

Let: $\xi_i = \dfrac{\epsilon_i + \beta\, \delta_i}{Q_i(\underline{x}_k) g(x_k + 1)}$

$$\zeta_i = 2k \exp\left\{\frac{4\epsilon_i}{g(x_k)} + \frac{4\delta_i}{g(x_k+1)} - \frac{2(i-k+1)}{k}\left(\frac{\epsilon_i^2}{g^2(x_k)} + \frac{\delta_i^2}{g^2(x_k+1)}\right)\right\}.$$

We have now produced the inequalities for condition a) of theorem 2). We must now choose functions $\delta_i(\underline{x}_k, \underline{\theta}_i)$, $\epsilon_i(\underline{x}_k, \underline{\theta}_i)$ subject to the conditions that

$$\delta_i < \frac{k\, g(x_k + 1)}{i - k + 1} \Rightarrow \delta_i = 0$$

$$\epsilon_i < \frac{k\, g(x_k)}{i - k + 1} \Rightarrow \epsilon_i = 0$$

and show that for Q_i and some a_i condition b) of theorem 2) is satisfied.

We first prove the following lemma.

Lemma 6) For $\{p_\theta: \theta \in \Omega\}$ defined in this section $\forall\, k = 1, 2, \ldots$ $\forall\, \epsilon\ \ 0 < \epsilon < 1$ there exists a constant $M = M(\epsilon, k) < \infty$ such that $\forall\, \underline{\theta} \in \Omega^\infty$, $\forall\ \ n = k, k+1, \ldots,\ \ n \neq 1$

$$\sum_{i=k}^{n} P\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(X_{i-k+\ell}) < i^\epsilon\right] \leq M n^\epsilon \log^k n$$

Proof. We first consider the case $k = 1$. Let d be the smallest integer such that $g(d) \neq 0$. $\forall\, m = d, d+1, \ldots, \ \forall\, \theta \in \Omega$

$$\frac{\sum_{x=m}^{\infty} p_\theta(x)}{p_\theta(m)} = \sum_{x=m}^{\infty} \theta^{x-m} \frac{g(x)}{g(m)}$$

$$= \sum_{y=d}^{\infty} \theta^y h(\theta) g(y) \left(\frac{g(y+m)}{h(\theta) g(y) g(m)} \right) + \sum_{y=0}^{d-1} \theta^y \frac{g(y+m)}{g(m)}$$

$$\leq \frac{M''}{h(\beta)} \quad \text{where } M'' < \infty \text{ from condition iii) and the}$$

easily verified fact that h is a decreasing function.

Also it is clear that $\forall\, \underline{\theta}$ there exists a smallest non negative integer m_i such that $\sum_{j=1}^{i} p_{\theta_j}(m_i) < i^{\epsilon}$.

Let
$$I(a, b) = \begin{cases} 1 & \text{if } a < b \\ 0 & \text{otherwise} \end{cases} .$$

Then:

$$\sum_{i=1}^{n} P\left[\sum_{j=1}^{i} p_{\theta_j}(X_i) < i^{\epsilon} \right] = \sum_{i=1}^{n} \sum_{x=0}^{\infty} I\left[\sum_{j=1}^{i} p_{\theta_j}(x),\ i^{\epsilon} \right] p_{\theta_i}(x)$$

$$\leq \sum_{i=1}^{n} \sum_{x=m_i}^{\infty} p_{\theta_i}(x)$$

$$\leq \frac{M''}{h(\beta)} \sum_{i=1}^{n} p_{\theta_i}(m_i)$$

Now $\sum_{j=1}^{i} p_{\theta_j}(x) = \sum_{j=1}^{i} \theta_j^x h(\theta_j)g(x) \leq i\beta^x h(0)g(x)$ so that if

$\beta^x h(0)g(x) < i^{\epsilon-1}$ then $m_i \leq x$. But $\beta^x h(0)g(x) \geq i^{\epsilon-1}$

$$\Rightarrow x \log \beta + \log g(x) \geq (\epsilon - 1) \log i - \log h(0) .$$

If $\beta < 1$ then we have from condition iv) that

$$x \log \beta + b \log x \geq (\epsilon - 1) \log i - \log h(0)$$

$$\Rightarrow x \leq \frac{1 - \epsilon}{\log \frac{1}{\beta}} \left[\log i + \frac{\log h(0)}{1 - \epsilon} + \frac{b \log x}{1 - \epsilon} \right] .$$

$\Rightarrow$ there exists $M''' < \infty$ such that $x \leq M''' \log i$ for $i > 1$.

If $\beta \geq 1$ then we have from condition iv) that

$$x \log \beta - (x - b) \log x \geq (\epsilon - 1) \log i - \log h(0)$$

$$\Rightarrow x(\log x - \log \beta) \leq (1 - \epsilon) \log i + \log h(0) + b \log x .$$

But $x \geq \beta^2 \Rightarrow \log x - \log \beta \geq \log \beta$

so if $x \geq \beta^2 \neq 1$ then also $x \leq \frac{1 - \epsilon}{\log \beta} \left[\log i + \frac{\log h(0)}{1 - \epsilon} + \frac{b \log x}{1 - \epsilon} \right]$.

If $\beta = 1$ then if $x > 1$ also $x \leq \frac{1 - \epsilon}{\log x} \left[\log i + \frac{\log h(0)}{1 - \epsilon} + \frac{b \log x}{1 - \epsilon} \right]$.

In each case there exists $M''' < \infty$ such that $x \leq M''' \log i$ for $i > 1$.

Let M_i be the smallest integer such that $M_i > M''' \log i$ for $i \geq 1$. Then $\beta^{M_i} h(0) g(M_i) < i^{\epsilon-1}$ and hence $m_i \leq M_i$. Now let $I_\nu = \{i: i = 1, 2, \ldots, n \text{ and } m_i = \nu\}$ for $\nu = 1, 2, \ldots, M_n$. Let i_ν be the greatest integer in I_ν. Then

$$\sum_{i=1}^{n} P\left[\sum_{j=1}^{i} p_{\theta_j}(X_i) < i^\epsilon\right] \leq \frac{M''}{h(\beta)} \sum_{i=1}^{n} p_{\theta_i}(m_i)$$

$$= \frac{M''}{h(\beta)} \sum_{\nu=1}^{M_n} \sum_{i\in I_\nu} p_{\theta_i}(\nu)$$

$$\leq \frac{M''}{h(\beta)} \sum_{\nu=1}^{M_n} (i_\nu)^\epsilon \quad \text{from the definition of } m_i$$

$$\leq \frac{M'' M_n}{h(\beta)} n^\epsilon \leq M n^\epsilon \log n \quad .$$

This completes the proof for the case $k = 1$. We now consider the case $k \geq 2$.

$$\sum_{i=k}^{n} P\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(X_{i-k+\ell}) < i^\epsilon\right]$$

$$= \left[\sum_{i=k}^{n} \sum_{x_1=0}^{\infty} \cdots \sum_{x_{k-1}=0}^{\infty} \sum_{x_k=0}^{m_i-1} I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^\epsilon\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell)\right]$$

$$+ \left[\sum_{i=k}^{n} \sum_{x_1=0}^{\infty} \cdots \sum_{x_{k-1}=0}^{\infty} \sum_{x_k=m_i}^{\infty} I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^\epsilon\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell)\right]$$

where m_i is as defined for the case $k = 1$. We shall now consider the two bracketed parts of the right hand side of the last equation separately. The second bracketed expression is clearly seen to be less than or equal to $M^*n^\epsilon \log n$ for some $M^* < \infty$ by the argument used in the case $k = 1$. The first bracketed expression can now be broken up into k expressions, $k - 1$ of which are less than or equal to $M^*n^\epsilon \log n$ with the remaining expression being

$$\sum_{i=k}^{n} \sum_{x_1=0}^{m_i-1} \cdots \sum_{x_k=0}^{m_i-1} I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^\epsilon\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell) .$$

But for each $(k-1)$-tuple of possible values $(x_1, x_2, \ldots , x_{k-1})$ either $I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^\epsilon\right]$ is zero for all $x_k = 0, 1, \ldots , m_i - 1$ or there exists a non negative integer $a_{k,i}(x_1, \ldots , x_{k-1}) \leq m_i - 1$ such that the indicator is zero for $x_k = 0, 1, \ldots , a_{k,i} - 1$, and one for $x_k = a_{k,i}$. We may now use the same arguments as in the case $k = 1$, for those $(k-1)$-tuples $(x_1, \ldots , x_{k-1})$ such that the indicator function is not zero for $x_k = 0, 1, \ldots , m_i - 1$, to show that

$$\sum_{i=1}^{n} \sum_{x_k=0}^{m_i-1} I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^\epsilon\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell)$$

$$\leq \sum_{i=1}^{n} \sum_{x_k=a_{k,i}}^{\infty} \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell)$$

$$\leq \frac{M''}{h(\beta)} \sum_{i=1}^{n} p_{\theta_i}(a_{k,i}) \prod_{\ell=1}^{k-1} p_{\theta_{i-k+\ell}}(x_\ell)$$

$$\leq \frac{M'' \, M_n n^{\epsilon}}{h(\beta)} \; .$$

Thus:

$$\sum_{i=k}^{n} \sum_{x_1=0}^{m_i-1} \cdots \sum_{x_k=0}^{m_i-1} I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^{\epsilon}\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell)$$

$$\leq \sum_{i=k}^{n} \sum_{x_1=0}^{M_n} \cdots \sum_{x_{k-1}=0}^{M_n} \sum_{x_k=0}^{m_i-1} I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^{\epsilon}\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell)$$

$$\leq \sum_{x_1=0}^{M_n} \cdots \sum_{x_{k-1}=0}^{M_n} \frac{M'' \, M_n n^{\epsilon}}{h(\beta)}$$

$$\leq \frac{M''(M_n)^k n^{\epsilon}}{h(\beta)} \; .$$

Upon combining the above result with the previous k inequalities we have the desired conclusion, and the proof of the lemma is complete.

We turn now to the task of showing condition b) of theorem 2) is satisfied for a suitable choice of ξ_i and ζ_i. The theorem will then give us an upper bound for $\mathcal{E}_n^*(\underline{\varphi}^k, \underline{\theta})$ and we shall then see it has the claimed rate of convergence to zero. Recalling

$$Q_i = \frac{1}{i-k+1} \sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(X_{i-k+\ell}) \quad i = k, k+1, \dots, n$$

and defining $a_i = \left(\frac{1}{i}\right)^{1/4}$ we have from lemma 6) $\frac{1}{n} \sum_{i=k}^{n} P[Q_i < a_i] \leq \frac{M \log^k n}{n^{1/4}}$.

Earlier in the proof we showed we could take $\xi_i = \dfrac{\epsilon_i + \beta\delta_i}{Q_i g(X_i + 1)}$

and $$\zeta_i = 2k\left[e^{\frac{2\epsilon_i^2(i-k+1)}{kg^2(X_i)} + 4\frac{\epsilon_i}{g(X_i)}} + e^{\frac{2\delta_i^2(i-k+1)}{kg^2(X_k+1)} + \frac{4\delta_i}{g(X_k+1)}}\right]$$

for arbitrary functions δ_i and ϵ_i such that

$$\delta_i \neq 0 \Rightarrow \delta_i \geq \frac{kg(X_i + 1)}{i-k+1} \qquad \epsilon_i \neq 0 \Rightarrow \epsilon_i \geq \frac{kg(X_i)}{i-k+1}$$

Let i_o be the smallest integer such that $\dfrac{\log i_o}{\sqrt{i_o}} \geq \dfrac{k}{i_o - k + 1}$ and such that $i_o \geq 2k$.

Also let

$$\delta_i = \begin{cases} \dfrac{g(X_i + 1)Q_i \log i}{i^{1/4}} & \text{if } Q_i > \dfrac{1}{i^{1/4}} \quad i = i_o, i_o + 1, \dots \\ 0 & \text{otherwise} \end{cases}$$

$$\epsilon_i = \begin{cases} \dfrac{g(X_i)Q_i \log i}{i^{1/4}} & \text{if } Q_i > \dfrac{1}{i^{1/4}} \quad i = i_o, i_o + 1, \dots \\ 0 & \text{otherwise} \end{cases}$$

Then $\forall \; i \geq i_o$

$$E\left[\frac{\epsilon_i}{Q_i g(X_i + 1)} \Big| Q_i \geq \frac{1}{i^{1/4}}\right] P\left[Q_i \geq \frac{1}{i^{1/4}}\right]$$

$$= \frac{\log i}{i^{1/4}} E\left[\frac{g(X_i)}{g(X_i + 1)} \Big| Q_i \geq \frac{1}{i^{1/4}}\right] P\left[Q_i \geq \frac{1}{i^{1/4}}\right]$$

$$\leq M^* \frac{\log i}{i^{1/4}} \quad \text{since from condition ii)}$$

$$E\left[\frac{g(X_i)}{g(X_i + 1)}\right] < M^* .$$

Also $$E\left[\frac{\delta_i}{Q_i g(X_i + 1)} \Big| Q_i \geq \frac{1}{i^{1/4}}\right] P\left[Q_i \geq \frac{1}{i^{1/4}}\right] \leq \frac{\log i}{i^{1/4}} .$$

Thus $$E[\xi_i | Q_i \geq a_i] P[Q_i \geq a_i] \leq \frac{\log i}{i^{1/4}} [M^* + \beta] .$$

Using the definitions of δ_i and ϵ_i we have

$$\zeta_i = 4k \exp\left[\frac{4\, Q_i \log i}{i^{1/4}} - \frac{2\, Q_i^2 (i - k + 1) \log^2 i}{k\sqrt{i}}\right]$$

whenever $Q_i \geq \frac{1}{i^{1/4}}$. Thus

$$E\left[\zeta_i \middle| Q_i \geq \frac{1}{i^{1/4}}\right] P\left[Q_i \geq \frac{1}{i^{1/4}}\right] \leq 4k \exp\left[\frac{4 \log i}{i^{1/4}} - \frac{2(i - k + 1)\log^2 i}{ki}\right]$$

noting $Q_i \leq 1$.

Collecting terms we see that condition b) is satisfied and

$$\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta}) \leq \frac{(k - 1)B^2}{n} + \frac{2B}{n}\left[i_o - k + (M^* + \beta) \sum_{i=i_o}^{n} \frac{\log i}{i^{1/4}} + 4Bk \sum_{i=i_o}^{n} e^{-\frac{2(i-k+1)\log^2 i}{ki} + \frac{4 \log i}{i^{1/4}}}\right] + \frac{2B^2 M \log^k n}{n^{1/4}}$$

for all $\underline{\theta} \in \Omega^\infty$.

We have now given an upper bound for $\mathcal{E}_n^k(\underline{\varphi}, \underline{\theta})$ for all n, and it remains to find the rate at which this bound goes to zero. Examining the various components of the upper bound we have

$$\frac{(k - 1)B^2}{n} = O\left(\frac{1}{n}\right)$$

and: $$\frac{2B}{n}(i_o - k) = O\left(\frac{1}{n}\right)$$

$$\frac{2B}{n}(M^* + \beta) \sum_{i=i_o}^{n} \frac{\log i}{i^{1/4}} = O\left(\frac{\log n}{n^{1/4}}\right)$$

$$\frac{8B^2k}{n} \sum_{i=i_o}^{n} e^{-\frac{2(i-k+1)}{ki}\log^2 i + \frac{4 \log i}{i^{1/4}}} = O\left(\frac{1}{n}\right)$$

$$\frac{2B^2M \log^k n}{n^{1/4}} = O\left(\frac{\log^k n}{n^{1/4}}\right) .$$

Hence $$\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta}) = O\left(\frac{\log^k n}{n^{1/4}}\right) \text{ uniformly in } \underline{\theta}.$$

Q.E.D.

We observe that the sum of r independent identically distributed random variables, each with density function $p_\theta(x) = \theta^x h(\theta) g(x)$, has the density function $f_\theta(y) = \theta^y h^r(\theta) g^{(r)}(y)$, where $g^{(r)}$ is the r-fold convolution of g. Thus the density function of the sum has the same form, and if conditions ii), iii), and iv) are satisfied for $g^{(r)}(y)$ then theorem 3) may be used even if the original problem is modified to allow r independent observations for each θ_i, observing that the sum is sufficient for θ_i. For the geometric, negative binomial, and Poisson families $g^{(r)}$ satisfies the conditions.

B. The modified negative binomial distribution.

Let $P[X = x|\theta] = p_\theta(x) = \binom{a + x - 1}{x}\left(\frac{a}{a + \theta}\right)^a\left(\frac{\theta}{a + \theta}\right)^x$

$x = 0, 1, 2, \ldots \quad a > 0 \quad 0 \leq \theta \leq \beta < \infty$.

This reparameterization of the negative binomial is of interest for two reasons. First $\forall\, a, E[X|\theta] = \theta$ unlike the usual parameterization. Secondly the form of the decision procedure is different than that usually encountered.

In this and the following sections we shall not give as many details as in Section A. The method of proving asymptotic optimality is similar in each of these sections, and may be summarized as follows:

Since $\psi_i^k(\underline{x}_k) = \frac{Q_i^*(\underline{x}_k)}{Q_i(\underline{x}_k)}$ when $Q_i(\underline{x}_k) > 0$, we seek an estimator $\varphi_i^k(\underline{x}_k)$ which is for "most" $\underline{x}_k$ equal to a ratio $\frac{\hat{P}_i(\underline{x}_k)}{P_i(\underline{x}_k)}$, such that $E[\hat{P}_{i-k}(\underline{X}_i^k)|\underline{X}_i^k = \underline{x}_k] = Q_{i-k}^*(\underline{x}_k)$ and such that $E[P_{i-k}(\underline{X}_i^k)|\underline{X}_i^k = \underline{x}_k] = Q_{i-k}(\underline{x}_k)$. Then using the methods of Section A the functions ξ_i and ζ_i may be obtained for condition a) of theorem 2). It is then only necessary to show either condition b) or b') holds.

Let:

$$Y_j(\underline{x}_k) = \begin{cases} 1 & \text{if } \underline{x}_j^k = \underline{x}_k \\ 0 & \text{otherwise} \end{cases}$$

$$g(i, j) = \frac{a\binom{a+i-1}{i}}{\binom{a+i+j-1}{i+j}} \qquad \begin{array}{l} i = 0, 1, \dots \\ j = 1, 2, \dots \end{array}$$

$$Z_j(\underline{x}_k) = \begin{cases} g(x_k, t) & \text{if there exists } t = 1, 2, \dots \text{ such that} \\ & \underline{x}_j^k = (x_1, \dots, x_{k-1}, x_k + t) \\ 0 & \text{otherwise} \end{cases}$$

$$P_i(\underline{x}_k) = \frac{\sum_{j=k}^{i} Y_j(\underline{x}_k)}{i - k + 1}$$

$$\hat{P}_i(\underline{x}_k) = \frac{\sum_{j=k}^{i} Z_j(\underline{x}_k)}{i - k + 1}$$

$$P_i^*(\underline{x}_k) = \begin{cases} \dfrac{\hat{P}_i(\underline{x}_k)}{P_i(\underline{x}_k)} & \text{if } P_i(\underline{x}_k) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\varphi_i^k(\underline{X}_i) = \begin{cases} \frac{\beta}{2} & \text{if } i = 1, \dots, k-1 \\ P_i^*(\underline{X}_i^k) & \text{if } 0 \le P_i^*(\underline{X}_i^k) \le \beta \quad i = k, k+1, \dots \\ \beta & \text{if } \beta < P_i^*(\underline{X}_i^k) \quad i = k, k+1, \dots . \end{cases}$$

Using the above definitions we shall show that $\underline{\varphi}^k = (\varphi_1^k, \varphi_2^k, \dots)$ is uniformly asymptotically optimal of k^{th} order and $\mathcal{E}_n^k(\underline{\varphi}^k, \underline{\theta}) \le B(k, n) = O\left(\frac{\log^k n}{n^{1/4}}\right)$.

Recall $$Q_i(\underline{x}_k) = \frac{1}{i-k+1} \sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell)$$

$$Q_i^*(\underline{x}_k) = \frac{1}{i-k+1} \sum_{j=k}^{i} \theta_j \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell) \quad .$$

We observe there exists a set R_i of k dimensional vectors such that $P[\underline{X}_i^k \epsilon R_i] = 1$ and $\underline{x}_k \epsilon R_i \Rightarrow Q_i(\underline{x}_k) > 0$. Thus for $i \ge k$ and $\underline{x}_k \epsilon R_i$, $\psi_i^k(\underline{x}_k) = \frac{Q_i^*(\underline{x}_k)}{Q_i(\underline{x}_k)}$. For $i \ge 2k$, $\underline{x}_k \epsilon R_i$, $j = k, \dots, i-k$ we have

$$E[Y_j(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k] = \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell) \quad .$$

We also have

$$E[Z_j(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k] = \sum_{t=1}^{\infty} g(x_k, t) p_{\theta_j}(x_k + t) \prod_{\ell=1}^{k-1} p_{\theta_{j-k+\ell}}(x_\ell) \; .$$

But

$$\sum_{t=1}^{\infty} g(x, t) p_\theta(x + t) = \sum_{t=1}^{\infty} a \binom{a + x - 1}{x} \left(\frac{a}{a + \theta}\right)^a \left(\frac{\theta}{a + \theta}\right)^{x+t}$$

$$= a p_\theta(x) \sum_{t=1}^{\infty} \left(\frac{\theta}{a + \theta}\right)^t$$

$$= \theta p_\theta(x) \; .$$

Hence:

$$E[P_i(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k]$$

$$= \frac{i - 2k + 1}{i - k + 1} Q_{i-k}(\underline{x}_k) + \frac{1}{i - k + 1} \sum_{j=i-k+1}^{i} E[Y_j(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k]$$

$$E[\hat{P}_i(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k]$$

$$= \frac{i - 2k + 1}{i - k + 1} Q^*_{i-k}(\underline{x}_k) + \frac{1}{i - k + 1} \sum_{j=i-k+1}^{i} E[Z_j(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k] \; .$$

The remainder of the proof follows that in Section A).

C. The binomial distribution.

Let $P[X = x|\theta] = p_\theta(x) = \binom{a}{x}\theta^x(1-\theta)^{a-x}$ $x = 0, 1, \ldots, a$ where a is a known positive integer and $0 \leq \theta \leq 1$. For this family it is necessary to modify slightly the definition of asymptotic optimality. Robbins [5] and others have demonstrated why this modification is necessary. Let $R_{k,a}(\underline{\theta}_n)$ be the k^{th} standard as defined in definition 3) with the parameter value a. We shall develop a procedure $\underline{\varphi}^k$ such that

$$\text{c)} \qquad \overline{\lim_{n\to\infty}} \{\sup_{\underline{\theta}} [R_n(\underline{\varphi}^k, \underline{\theta}) - R_{k,a-1}(\underline{\theta}_n)]\} \leq 0 .$$

Such a procedure will be said to have property c). In addition we shall show $R_n(\underline{\varphi}^k, \underline{\theta}) - R_{k,a-1}(\underline{\theta}_n) \leq B(k, n) = O\left(\frac{\log n}{n^{1/4}}\right)$ uniformly in $\underline{\theta}$.

We shall first exhibit a procedure having property c). We shall then introduce a new procedure which not only has property c) but for most $\underline{\theta}$ actually improves upon the original procedure at each stage and produces strict inequality in equation c).

We first assume that corresponding to every observation X we have available the related observation X' which would have resulted had we observed a binomial random variable with parameter $a - 1$. For example, if X is the number of successes in a independent Bernoulli trials with probability θ of success, then X' is the number of successes in the first $a - 1$ of these a trials. While in most situations this assumption will hold, we shall see later that it will not be needed.

Let:

$$\underline{X}_{i,\nu}^{k} = \begin{cases} (X'_{i-k+1}, \dots, X'_{i}) & \text{if } \nu = 0 \\ (X'_{i-k+1}, \dots, X'_{i-1}, X_i) & \text{if } \nu = 1 \end{cases}$$

$$Y_j(\underline{x}_k) = \begin{cases} 1 \text{ if } \underline{X}_{j,0}^{k} = \underline{x}_k \\ 0 \text{ elsewhere} \end{cases}$$

$$Z_j(\underline{x}_k) = \begin{cases} x_k + 1 & \text{if } \underline{X}_{j,1}^{k} = (x_1, \dots, x_{k-1}, x_k + 1) \\ 0 & \text{otherwise} \end{cases}$$

$$P_i(\underline{x}_k) = \frac{\sum_{j=k}^{i} Y_j(\underline{x}_k)}{i - k + 1}$$

$$\hat{P}_i(\underline{x}_k) = \frac{\sum_{j=k}^{i} Z_j(\underline{x}_k)}{a(i - k + 1)}$$

$$P_i^*(\underline{x}_k) = \begin{cases} \dfrac{\hat{P}_i(\underline{x}_k)}{P_i(\underline{x}_k)} & \text{if } P_i(\underline{x}_k) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\varphi_i^k(\underline{X}_i) = \begin{cases} \frac{1}{2} & \text{if } i = 1, \dots, k-1 \\ P_i^*(\underline{X}_{i,0}^k) & \text{if } 0 \le P_i^*(\underline{X}_{i,0}^k) \le 1 \quad i = k, k+1, \dots \\ 1 & \text{if } 1 < P_i^*(\underline{X}_{i,0}^k) \quad i = k, k+1, \dots \end{cases}$$

We shall now show $\underline{\varphi}^k = (\varphi_1^k, \varphi_2^k, \dots)$ has property c). We shall proceed as in the previous examples, noting that theorem 2) is still true when the property of asymptotic optimality is replaced by property c).

$\forall\ i \ge 2k$ let:

$$Q_i(\underline{x}_k) = \frac{1}{i-k+1} \sum_{j=k}^{i} \prod_{\ell=1}^{k} \binom{a-1}{x_\ell} \theta_{j-k+\ell}^{x_\ell} (1-\theta_{j-k+\ell})^{a-1-x_\ell}$$

R_i be a set of k dimensional vectors such that

$P[\underline{X}_{i,0}^k \in R_i] = 1$ and $\underline{x}_k \in R_i \Rightarrow Q_i(\underline{x}_k) > 0$.

For $\underline{x}_k \in R_i$ we have for the parameter $a - 1$:

$$\psi_i^k(\underline{x}_k) = \frac{\frac{1}{i-k+1} \sum_{j=k}^{i} \theta_j \prod_{\ell=1}^{k} \binom{a-1}{x_\ell} \theta_{j-k+\ell}^{x_\ell} (1-\theta_{j-k+\ell})^{a-1-x_\ell}}{Q_i(\underline{x}_k)} .$$

Now for $j = k, \ldots, i - k$

$$E[Y_j(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k] = \prod_{\ell=1}^{k} \binom{a - 1}{x_\ell} \theta_{j-k+\ell}^{x_\ell} (1 - \theta_{j-k+\ell})^{a-1-x_\ell}$$

$$E[A_j(\underline{X}_i^k) | \underline{X}_i^k = \underline{x}_k]$$

$$= (x_k + 1) \binom{a}{x_k + 1} \theta_j^{x_k+1} (1 - \theta_j)^{a-x_k-1} \prod_{\ell=1}^{k-1} \binom{a - 1}{x_\ell} \theta_{j-k+\ell}^{x_\ell} (1 - \theta_{j-k+\ell})^{a-1-x_\ell}$$

$$= a\theta_j \prod_{\ell=1}^{k} \binom{a - 1}{x_\ell} \theta_{j-k+\ell}^{x_\ell} (1 - \theta_{j-k+\ell})^{a-1-x_\ell} .$$

Thus arguing as in Section A)

$$P\left[|\varphi_i^k(\underline{X}_i) - \psi_i^k(\underline{x}_k)| \geq \frac{1}{Q_i(\underline{x}_k)} (\varepsilon_i + a\delta_i) | \underline{X}_{i,0}^k = \underline{x}_k \right] \leq 4ke^{-\frac{2}{k}(i-k+1)\varepsilon_i^2 + 4\delta_i}$$

for $\delta_i \geq \frac{k}{i - k + 1}$.

We now look for an upper bound for the quantity $\sum_{i=k}^{n} P\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+1}}(X_{i-k+\ell}) < i^\epsilon \right]$. We shall show an upper bound is $(a + 1)^k n^\epsilon$.

$$\sum_{i=k}^{n} P\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+1}}(X_{i-k+\ell}) < i^\epsilon \right] =$$

$$= \sum_{x_1=0}^{a} \cdots \sum_{x_k=0}^{a} \sum_{i=k}^{n} I\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^\epsilon\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell) .$$

For any fixed $\underline{x}_k$, $x_\ell = 0, 1, \ldots, a$ $\ell = 1, 2, \ldots, k$. There exists a subsequence of integers $\{i_\nu\}$ (possibly finite) such that $\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell) < i^\epsilon \Leftrightarrow i = i_\nu$ for some $\nu = 1, 2, \ldots$. Let $i_o = 0$. But for all $n \geq k$ there exists a ν_n such that $i_\nu \leq n \Rightarrow \nu \leq \nu_n$ and:

$$\sum_{i=k}^{n} P\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} p_{\theta_{j-k+\ell}}(x_\ell),\ i^\epsilon\right] \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell)$$

$$\leq \sum_{i=k}^{i_{\nu_n}} \prod_{\ell=1}^{k} p_{\theta_{i-k+\ell}}(x_\ell) < i_{\nu_n}^\epsilon \leq n^\epsilon .$$

Since there are $(a + 1)^k$ different $\underline{x}_k$ to sum over, we have shown the claimed upper bound holds.

It is now easy to show, using an argument similar to that in Section A, that an upper bound for $R_n(\underline{\varphi}, \underline{\alpha})$ is

$$\frac{k - 1}{n} + \frac{2}{n}(i_o - k) + \frac{4}{n} \sum_{i=i_o}^{n} \frac{\log i}{n^{1/4}}$$

$$+ \frac{8k}{n} \sum_{i=i_o}^{n} \exp\left[\frac{4 \log i}{i^{1/4}} - \frac{2(i - k + 1)}{ki} \log^2 i\right]$$

$$+ \frac{2(a + 1)^k}{n^{1/4}}$$

and clearly this upper bound is uniformly of the order $\frac{\log n}{n^{1/4}}$. The desired conclusion follows.

In the above procedure we chose at times to neglect the results of the "a^{th} trial" in many of the observations. This choice of which information to neglect was quite arbitrary, and it is easily seen that the above proof does not depend on which trial's information was neglected. We may thus conclude that if in some situation the related observation X' is not obtainable, we may construct a new X' which will do as well. An example will illustrate. Suppose $a = 17$ and $X = 10$. With the aid of some random device we let

$$X' = \begin{cases} 9 & \text{with probability } 10/17 \\ 10 & \text{with probability } 7/17 \end{cases} .$$

This X' will work as well as the original X'.

We shall now exhibit a procedure $\tilde{\underline{\varphi}}^k$ which improves upon $\underline{\varphi}^k$. At the i^{th} stage of the decision problem we could have defined φ_i^k in any one of several different ways, depending on what information we chose to neglect. For fixed $i \geq k$ and $\underline{x}_k$ there are a^k ways to define $\underline{X}^k_{j,0}$ and hence $Y_j(\underline{x}_k)$. Thus there are $a^{k(i-k+1)}$ ways to define $P_i(\underline{x}_k)$. Similarly there are a^{k-1} ways to define $Z_j(\underline{x}_k)$ and hence $a^{(k-1)(i-k+1)}$ ways to define $\hat{P}_i(\underline{x}_k)$. Thus there are $a^{(2k-1)(i-k+1)}$ ways to define $P_i^*(\underline{x}_k)$. From the above, and observing the $\underline{X}_{i,0}^k$ used in $P_i^*(\underline{X}_{i,0}^k)$ could have a^k different definitions, we see there are at least $a^{(2k-1)(i-k+1)+k}$ ways to define the random variable $\varphi_i(\underline{X}_i)$. Most of these different definitions will result in essentially the same estimator for large i. We may obtain an improved procedure, however, by considering some of them.

We define $\underline{X}_{i,(u)}^k$ $u = 1, \ldots, a^k$ as follows: Let $u = 1, \ldots, a^k$ be an indexing of the a^k distinct k-tuples each of whose elements are integers from the set $\{1, 2, \ldots, a\}$. Let the u^{th} k-tuple be $(t_{u,1}, \ldots, t_{u,k})$. Let $X^{(\ell)}$ $\ell = 1, \ldots, a$ be the random variable derived from X by not counting the result of the ℓ^{th} trial. For example, $X^{(a)}$ equals the previously defined X'. We now define $\underline{X}_{i,(u)}^k$ to be the random vector $(X_{i-k+1}^{(t_{u,1})}, X_{i-k+2}^{(t_{u,2})}, \ldots, X_i^{(t_{u,k})})$.

$$\text{Let: } \varphi_{i,u}^k(\underline{X}_i) = \begin{cases} \frac{1}{2} & \text{if } i = 1, \ldots, k-1 \\ P_i^*(\underline{X}_{i,(u)}^k) & \text{if } 0 \le P_i^*(\underline{X}_{i(u)}^k) \le 1 \quad i = k, k+1, \ldots \\ 1 & \text{if } 1 < P_i^*(\underline{X}_{i(u)}^k) \quad i = k, k+1, \ldots \end{cases}$$

Let: $R(\varphi, \underline{\theta}_i) = E[(\varphi(X_i) - \theta_i)^2]$

$$\tilde{\varphi}_i^k(\underline{X}_i) = \frac{1}{a^k} \sum_{u=1}^{a^k} \varphi_{i,u}^k(\underline{X}_i)$$

$$\underline{\tilde{\varphi}}^k = (\tilde{\varphi}_1^k, \tilde{\varphi}_2^k, \ldots)$$

We shall now show $\forall \underline{\theta}, \forall i \; R(\tilde{\varphi}_i^k, \underline{\theta}_i) \leq R(\varphi_i^k, \underline{\theta}_i)$ where φ_i^k is as previously defined. We first observe that the $\varphi_{i,u}^k(\underline{X}_i)$ are identically distributed as $\varphi_i^k(\underline{X}_i)$. Then, supressing the superscript k, we have

$$\begin{aligned}
R(\tilde{\varphi}_i, \underline{\theta}_i) &= E[(\tilde{\varphi}_i(\underline{X}_i) - \theta_i)^2] \\
&= E[\tilde{\varphi}_1^2] - 2\theta_i E[\tilde{\varphi}_1] + \theta_i^2 \\
&= \frac{1}{a^{2k}} \left[\sum_{u=1}^{a^k} E[\varphi_{i,u}^2] + 2 \sum_{u<v} E[\varphi_{i,u}\varphi_{i,v}] \right] - \frac{2\theta_i}{a^k} \sum_{u=1}^{a^k} E[\varphi_{i,u}] + \theta_i^2 \\
&= E[(\varphi_i - \theta_i)^2] - \left[\frac{a^k - 1}{a^k} E[\varphi_i^2] - \frac{2}{a^{2k}} \sum_{u<v} E[\varphi_{i,u} \; \varphi_{i,v}] \right] \\
&= R(\varphi_i, \underline{\theta}_i) - \frac{1}{a^{2k}} \sum_{u<v} \left[E[\varphi_{i,u}^2] - 2E[\varphi_{i,u} \; \varphi_{i,v}] + E[\varphi_{i,v}^2] \right] \\
&= R(\varphi_i, \underline{\theta}_i) - \frac{1}{a^{2k}} \sum_{u<v} E[(\varphi_{i,u} - \varphi_{i,v})^2] \\
&\leq R(\varphi_i, \underline{\theta}_i)
\end{aligned}$$

we notice strict inequality holds unless $P[\varphi_{i,u}(\underline{X}_i) = \varphi_{i,v}(\underline{X}_i)] = 1$ for all $u, v \;\; 1 \leq u , v \leq a^k$. Thus we have strict inequality holding unless $a = 1$ or unless $\underline{\epsilon}_i^k$ is composed of 1's and 0's. To investigate the asymptotic properties of $\tilde{\varphi}$ we observe:

$$\overline{\lim_{n\to\infty}} \left\{ \frac{1}{n} \sum_{i=1}^{n} R(\tilde{\varphi}_i, \underline{\theta}_i) - R_{k,a-1}(\underline{\theta}_n) \right\}$$

$$= \overline{\lim_{n\to\infty}} \left\{ \frac{1}{n} \sum_{i=1}^{n} [R(\tilde{\varphi}_i, \underline{\theta}_i) - R(\varphi_i, \underline{\theta}_i)] + \frac{1}{n} \sum_{i=1}^{n} R(\varphi_i, \underline{\theta}_i) - R_{k,a-1}(\underline{\theta}_n) \right\}$$

$$\leq \overline{\lim_{n\to\infty}} \left\{ \frac{1}{n} \sum_{i=1}^{n} [R(\tilde{\varphi}_i, \underline{\theta}_i) - R(\varphi_i, \underline{\theta}_i)] \right\} \text{ since } \varphi \text{ has property c)}$$

$$= \overline{\lim_{n\to\infty}} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \frac{1}{a^{2k}} \sum_{u<v} E[(\varphi_{i,u} - \varphi_{i,v})^2] \right\}$$

$$\leq 0$$

Hence $\tilde{\varphi}$ has property c). In order for strict inequality to hold it is sufficient that there exists $\epsilon > 0$ such that, with the possible exception of a finite number of values of i, $\sum_{u<v} E[|\varphi_{i,u} - \varphi_{i,v}|] \geq \epsilon$. If $a = 1$ this condition is never satisfied. For $a > 1$, however, and for a large class of $\underline{\theta}$ such an ϵ will exist. Let Ω^* be the set of $\underline{\theta}$ such that $\forall\ \underline{\theta} \in \Omega^*$ there exist ξ_1, ξ_2 such that $0 < \xi_1 \leq \theta_i \leq \xi_2 < 1$ for all but finitely many i and such that the first order empirical distribution function of $\underline{\theta}_n$ does not tend in the limit to the distribution function of a degenerate random variable. It may then be shown that $a > 1$ $\underline{\theta} \in \Omega^*$ implies

$\sum_{u<v} E[|\varphi_{i,u} - \varphi_{i,v}|] \geq \epsilon > 0$ for all except possibly a finite number of i.

Since the sum of independent identically distributed binomial random variables is again a binomial random variable, and since the sum is a sufficient statistic for θ, it is clear the methods of this section can be applied to the case of r independent observations for each θ_i.

D. The normal distribution

Let
$$P[X \leq x \mid \theta] = \int_{-\infty}^{x} p_\theta(t)\,dt$$
$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{t-\theta}{\sigma}\right)^2} dt \qquad -\infty < x < \infty$$

for $\sigma > 0$ and $-\infty < \alpha \leq \theta \leq \beta < \infty$.

For the present we assume σ is known, although later we shall modify this assumption somewhat. As we shall see, the estimation procedure in the continuous case is similar to that in the discrete case. Without loss of generality we take $\sigma = 1$. We fix $k \geq 1$.

Let $\{c_i, i = 1, 2, \ldots\}$ be a sequence of positive numbers such that $\lim_{i\to\infty} c_i \log i = 0$ and $\lim_{i\to\infty} i(c_i)^{4(k+1)} = \infty$.

$\underline{e}_k$ be the k dimensional vector consisting of zeros for all components except for the k^{th} which is equal to one.

$$Y_{j,i}(\underline{y}_k) = \begin{cases} 1 & \text{if } y_\ell - c_i < X_{j-k+\ell} < y_\ell + c_i \quad \ell = 1, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

$$f_i(\underline{y}_k) = \frac{\sum_{j=k}^{i-k} Y_{j,i}(\underline{y}_k)}{(i - k + 1)(2c_i)^k}$$

$$g_i(\underline{y}_k) = \frac{f_i(\underline{y}_k + c_i\underline{e}_k) - f_i(\underline{y}_k - c_i\underline{e}_k)}{2c_i}$$

$$P_i^*(\underline{y}_k) = \begin{cases} y_k + \dfrac{g_i(\underline{y}_k)}{f_i(\underline{y}_k)} & \text{if } f_i(\underline{y}_k) > 0 \text{ and } i = k, k+1, \ldots \\ y_k & \text{otherwise} \end{cases}$$

$$\varphi_i^k(\underline{X}_i) = \begin{cases} \alpha & \text{if } P_i^*(\underline{X}_i^k) \le \alpha \\ P_i^*(\underline{X}_i^k) & \text{if } \alpha < P_i^*(\underline{X}_i^k) < \beta \\ \beta & \text{if } \beta \le P_i^*(\underline{X}_i^k) \end{cases} .$$

We shall now prove the decision procedure $\underline{\varphi}^k = (\varphi_1^k, \varphi_2^k, \ldots)$ is uniformly asymptotically optimal of k^{th} order.

We shall show the conditions of the corollary to theorem 2) are satisfied.

Let: $$Q_i(\underline{y}_k) = \sum_{j=k}^{i} \left(\frac{1}{2\pi}\right)^{\frac{k}{2}} e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2}$$

$$Q_i'(\underline{y}_k) = \frac{\partial Q_i(\underline{y}_k)}{\partial y_k}$$

$$Q_i^*(\underline{y}_k) = \sum_{j=k}^{i} \theta_j \left(\frac{1}{2\pi}\right)^{\frac{k}{2}} e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2}.$$

Now for any $\underline{y}_k$,

$$\psi_i^k(\underline{y}_k) = \frac{Q_i^*(\underline{y}_k)}{Q_i(\underline{y}_k)}.$$

But $$Q_i'(\underline{y}_k) = -\sum_{j=k}^{i} (y_k - \theta_j)\left(\frac{1}{2\pi}\right)^{\frac{k}{2}} e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2}$$

$$= - y_k Q_i(\underline{y}_k) + Q_i^*.$$

Hence $$\psi_i^k(\underline{y}_k) = y_k + \frac{Q_i'(\underline{y}_k)}{Q_i(\underline{y}_k)}.$$

Let: $m_i(\underline{y}_k) = E[f_i(\underline{X}_i^k) | \underline{X}_i^k = \underline{y}_k]$

$$z_i(\underline{y}_k) = \frac{\sum_{j=k}^{i-k}\left[e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y^*_{\ell,j}-\theta_{j-k+\ell})^2} - e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2}\right]}{(i-k+1)(2\pi)^{k/2}}$$

for some $y^*_{k,j}$ such that $y_\ell - c_i < y^*_{\ell,j} < y_\ell + c_i$ $\ell = 1, \ldots, k$. What particular $\underline{y}^*_{k,j}$ is intended will be clear from the context.

Then:

$$m_i(\underline{y}_k) = \frac{1}{(i-k+1)}\sum_{j=k}^{i-k}\prod_{\ell=1}^{k}\frac{1}{2c_i}\int_{y_\ell-c_i}^{y_\ell+c_i}\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(t-\theta_{j-k+\ell})^2}\,dt$$

$$= \frac{Q_{i-k}(\underline{y}_k)}{i-k+1} + z_i(\underline{y}_k)$$

where the $\underline{y}^*_{k,j}$ in z_i are those vectors whose components arise from use of the mean value theorem.

Now: for $\forall\ i = 2k, 2k+1, \ldots,\ y_k \epsilon R^k,\ \underline{\theta} \epsilon\ \Omega^\infty$

$$P\left[\left|f_i(\underline{X}_i^k) - \frac{Q_i(\underline{X}_i^k)}{i - k + 1}\right| \geq \delta_i \middle| \underline{X}_i^k = \underline{y}_k\right]$$

$$\leq P\left[\frac{i - k + 1}{i - 2k + 1}\left|f_i(\underline{y}_k) - m_i(\underline{y}_k)\right| \geq \frac{i - k + 1}{i - 2k + 1}\left(\delta_i - |z_i| - \frac{k}{i - k + 1}\right)\right]$$

but from lemma 3) this probability is

$$\leq 2k \exp\left\{- 2\,\frac{(i - k + 1)}{k}\left(\delta_i - |z_i| - \frac{k}{i - k + 1}\right)^2 (2c_i)^{2k}\right\}$$

$$\leq 2k \exp\left\{2^{2(k+1)} c_i^{2k} \delta_i - 2^{2k+1}\left(\frac{i - k + 1}{k}\right) c_i^{2k} (\delta_i - |z_i|)^2\right\}$$

provided $\delta_i - |z_i| - \frac{k}{i - k + 1} \geq 0$.

In a similar manner:

Let:

$$z_i'(\underline{y}_k) = \frac{z_i(\underline{y}_k + c_i\underline{e}_k) - z_i(\underline{y}_k - c_i\underline{e}_k)}{2c_i} \quad ,$$

$$q_i(\underline{y}_k) = \frac{1}{i - k + 1}\left(\frac{Q_{i-k}(\underline{y}_k + c_i\underline{e}_k) - Q_{i-k}(\underline{y}_k - c_i\underline{e}_k)}{2c_i} - Q_{i-k}'(\underline{y}_k)\right) .$$

Then

$$E[g_i(\underline{X}_i^k) | \underline{X}_i^k = \underline{y}_k] = \frac{m_i(\underline{y}_k + c_i\underline{e}_k) - m_i(\underline{y}_k - c_i\underline{e}_k)}{2c_i}$$

$$= \frac{Q_{i-k}'(\underline{y}^k)}{i - k + 1} + q_i(\underline{y}_k) + z_i'(\underline{y}_k) .$$

Hence:

$$P\left[\left|g_i(\underline{X}_i^k) - \frac{Q_i'(\underline{X}_i^k)}{i-k+1}\right| \geq \epsilon_i \middle| \underline{X}_i^k = \underline{y}_k\right]$$

$$\leq 2k \exp\left\{-\frac{2(i-k+1)}{k}\left(\epsilon_i - |z_i'| - |q_i| - \frac{k}{i-k+1}\right)^2 \frac{(2c_i)^{2(k+1)}}{4}\right\}$$

$$\leq 2k \exp\left\{(2c_i)^{2(k+1)}\epsilon_i - 2^{2k+1}c_i^{2(k+1)}\frac{i-k+1}{k}(\epsilon_i - |z_i'| - |q_i|)^2\right\}.$$

$$\text{provided} \quad \epsilon_i - |z_i'| - |q_i| - \frac{k}{i-k+1} \geq 0$$

Thus, using an argument similar to that used in proving lemma 4) and letting $B = \max[|\alpha|, |\beta|]$, we have:

$$P\left[|\varphi_i(\underline{X}_i) - \psi_i(\underline{X}_i)| > \frac{i-k+1}{Q_i(\underline{y}_k)}(\epsilon_i + B\delta_i + |y_k|\delta_i) \middle| \underline{X}_i^k = \underline{y}_k\right]$$

$$\leq 2k \exp\left\{2^{2(k+1)}c_i^{2k}\delta_i - 2^{2k+1}\left(\frac{i-k+1}{k}\right)c_i^{2k}(\delta_i - |z_i|)^2\right\}$$

$$+ 2k \exp\left\{(2c_i)^{2(k+1)}\epsilon_i - 2^{2k+1}c_i^{2(k+1)}\frac{i-k+1}{k}(\epsilon_i - |z_i'| - |q_i|)^2\right\}$$

$$\text{provided} \quad \delta_i - |z_i| - \frac{k}{i-k+1} \geq 0$$

$$\epsilon_i - |z_i'| - |q_i| - \frac{k}{i-k+1} \geq 0 .$$

To complete the proof it is sufficient to show that for some functions $\epsilon_i(\underline{y}_k, \underline{\theta}_i)$ and $\delta_i(\underline{y}_k, \underline{\theta}_i)$, such that either $\epsilon_i \geq |z_i'| + |q_i| + \frac{k}{i - k + 1}$ or $\epsilon_i = 0$ and either $\delta_i \geq |z_i| + \frac{k}{i - k + 1}$ or $\delta_i = 0$, the following limits hold uniformly in $\underline{\theta}$.

$$\text{i)} \quad \lim_{i \to \infty} E\left[\frac{i - k + 1}{Q_i(\underline{X}_i^k)} \epsilon_i(\underline{X}_i^k)\right] = 0$$

$$\text{ii)} \quad \lim_{i \to \infty} E\left[\frac{i - k + 1}{Q_i(\underline{X}_i^k)} (B + |X_i|)\delta_i(\underline{X}_i^k)\right] = 0$$

$$\text{iii)} \quad \lim_{i \to \infty} E\left[e^{-ic_i^{2k}(\delta_i(\underline{X}_i^k) - |z_i(\underline{X}_i^k)|)^2} \middle| \delta_i - |z_i| \geq \frac{k}{(i - k + 1)^{1/4}}\right] = 0$$

$$\text{iv)} \quad \lim_{i \to \infty} E\left[e^{-ic_i^{2(k+1)}(\epsilon_i(\underline{X}_i^k) - |z_i'(\underline{X}_i^k)| - |q_i(\underline{X}_i^k)|)^2} \middle| \epsilon_i - |z_i'| - |q_i| \geq \frac{k}{(i - k + 1)^{1/4}}\right] = 0$$

$$\text{v)} \quad \lim_{i \to \infty} P\left[\delta_i(\underline{X}_i^k) - |z_i(\underline{X}_i^k)| < \frac{k}{(i - k + 1)^{1/4}}\right] = 0$$

$$\text{vi)} \quad \lim_{i \to \infty} P\left[\epsilon_i(\underline{X}_i^k) - |z_i'(\underline{X}_i^k)| - |q_i(\underline{X}_i)| < \frac{k}{(i - k + 1)^{1/4}}\right] = 0$$

Let

$$\delta_i = \begin{cases} \dfrac{Q_i}{(i-k+1)\log i} & \text{if } \dfrac{Q_i}{(i-k+1)\log i} \geq |z_i| + \dfrac{k}{(i-k+1)^{1/4}} \\ 0 & \text{otherwise} \end{cases}$$

$$\epsilon_i = \begin{cases} \dfrac{Q_i}{(i-k+1)\log i} & \text{if } \dfrac{Q_i}{(i-k+1)\log i} \geq |z_i'| + |q_i| + \dfrac{k}{(i-k+1)^{1/4}} \\ 0 & \text{otherwise} \end{cases}$$

Since there exists $J < \infty$ such that $E[|X|] < U$ for all $\theta \epsilon \Omega$, the first two limits hold uniformly. The third and fourth also hold, recalling $\lim_{i \to \infty} i c_i^{4(k+1)} = \infty$.

Since $\forall \eta > 0$ there exists $S(\eta) < \infty$ such that $P[|X| > S] < \eta$ for all $\theta \epsilon \Omega$, we have:

$$P\left[\delta_i - |z_i| < \frac{k}{(i-k+1)^{1/4}}\right] \leq P\left[\frac{Q_i}{(i-k+1)\log i} - |z_i|\right.$$

$$\left. < \frac{k}{(i-k+1)^{1/4}} \,\middle|\, |X_{i-k+\ell}| \leq S \quad \ell = 1, \ldots, k\right] + 1 - (1-\eta)^k =$$

$$= P\left[\frac{\sum_{j=k}^{i} \left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}\sum_{\ell=1}^{k}(X_{i-k+\ell}-\theta_{j-k+\ell})^2}}{i - k + 1} < \frac{k \log i}{(i - k + 1)^{1/4}} + |z_i| \log i \;\middle|\; |X_{i-k+\ell}| \leq S \quad \ell = 1, \ldots, k \right] + 1 - (1 - \eta)^k$$

$$\leq P\left[\left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}k(S+B)^2} < \frac{k \log i}{(i - k + 1)^{1/4}} + |z_i| \log i \;\middle|\; |X_{i-k+\ell}| \leq S \right]$$

$$+ 1 - (1 - \eta)^k.$$

Clearly the fifth limit will hold if we show $\lim_{i\to\infty} |z_i(\underline{y}_k)| \log i = 0$ uniformly in $\theta \in \Omega$ and $|y_\ell| \leq S$ $\ell = 1, \ldots, k$. By a similar argument, to prove the sixth limit holds we need only to show $\lim_{i\to\infty} (|z_i'(\underline{y}_k) + |q_i(\underline{y}_k)|) \log i = 0$ uniformly for $\theta \in \Omega$ and $|y_\ell| \leq S$ $\ell = 1, \ldots, k$.

We first consider $\lim_{i\to\infty} |z_i(\underline{y}_k)| \log i$. For $y_\ell - c_i < y_\ell^* < y_\ell + c_i$ $\ell = 1, \ldots, k;\ \forall\ j = k, \ldots, i$

$$\left| e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell^*-\theta_{j-k+\ell})^2} - e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2} \right|$$

$$= \left| e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell^*-\theta_{j-k+\ell})^2} \right| \left| 1 - e^{-\frac{1}{2}\sum_{\ell=1}^{k}[(y_\ell-\theta_{j-k+\ell})^2-(y_\ell^*-\theta_{j-k+\ell})^2]} \right|.$$

But

$$|(y_\ell - \theta_{j-k+\ell})^2 - (y_\ell^* - \theta_{j-k+\ell})^2| \le |y_\ell - y_\ell^*||y_\ell + y_\ell^* - 2\theta_{j-k+\ell}|$$

$$\le c_i(2S + 2B + c_i) \text{ for all } \theta_{j-k+\ell} \in \Omega .$$

Now $\quad |x| < 1 \Rightarrow |1 - e^x| = |\sum_{\nu=1}^{\infty} \frac{x^\nu}{\nu!}| \le |x| \; (e - 1)$.

Thus

$$|z_i(\underline{y}_k)| = \left| \frac{\sum_{j=k}^{i-k} \left[e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y^*_{\ell,j}-\theta_{j-k+\ell})^2} - e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2} \right]}{(i - k + 1)(2\pi)^{k/2}} \right|$$

$$\le \frac{1}{(2\pi)^{k/2}} c_i(S + B + c_i)(e - 1) k \text{ for all } \underline{\theta} \text{ and } i \text{ suffi-}$$

ciently large.

But since $\lim\limits_{i\to\infty} c_i \log i = 0$ we have shown $\lim\limits_{i\to\infty} |z_i| \log i = 0$ for all $\underline{\theta}$ and $|y_\ell| \le S \quad \ell = 1, \ldots , k.$

We now consider $\lim\limits_{i\to\infty} |z_i'| \log i$. This limit is more difficult to evaluate. By examining the argument leading to the consideration of this limit it is clear that we need only show $\lim\limits_{i\to\infty} |z_i'| \log i = 0$ for $\underline{y}_k$ such that $|y_\ell| \le S$ for $\ell = 1, \ldots , k$ and $y_k \ne \theta_j$ for $j = k, k + 1, \ldots$. We shall use lemma 5). It is clear that there exists $M < \infty$ such that $C_M = (-\infty, \infty)$ for all $\theta \in \Omega$. Thus we have for $i \ge k$ and $\underline{y}_k$ such that $|y_\ell| \le S$ and $y_k \ne \theta_j \quad \ell = 1, \ldots , k \quad j = k, k + 1, \ldots$

$$|z_i'(\underline{y}_k)| = \left| \sum_{j=k}^{i-k} \frac{e^{-\frac{1}{2}\sum_{\ell=1}^{k-1}(y_\ell^*-\theta_{j-k+\ell})^2}\left(e^{-\frac{1}{2}(y_k^{**}-\theta_j)^2} - e^{-\frac{1}{2}(y_k^{***}-\theta_j)^2}\right)}{2c_i(i \quad k+1)(2\pi)^{k/2}}\right.$$

$$\left. - \sum_{j=k}^{i-k} \frac{e^{-\frac{1}{2}\sum_{\ell=1}^{k-1}(y_\ell-\theta_{j-k+\ell})^2}\left(e^{-\frac{1}{2}(y_k+c_i-\theta_j)^2} - e^{-\frac{1}{2}(y_k-c_i-\theta_j)^2}\right)}{2c_i(i-k+1)(2\pi)^{k/2}}\right|$$

where $y_\ell - c_i < y_\ell^* < y_\ell + c_i \quad y_k^{**} = y_k + 2p_{i,j}c_i$

$y_k^{***} = y_k - 2p_{2,j}c_i$

$$|p_{\nu,j} - \tfrac{1}{2}| < \frac{4\sqrt{2\pi}\ M\ e^{\frac{1}{2}(B+3)^2}}{3|y_k - \theta_j|}\, c_i \qquad \nu = 1,\ 2$$

and $$|z_i'| \leq \sum_{j=k}^{i-k} \frac{e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2}}{(i-k+1)(2\pi)^{k/2}2c_i}\left| e^{-\frac{1}{2}\sum_{\ell=1}^{k-1}(y_\ell^*-y_\ell)(y_\ell^*+y_\ell-2\theta_{j-k+\ell})}\right.$$

$$\left(e^{-\frac{1}{2}(y_k^{**}-y_k)(y_k^{**}+y_k-2\theta_j)} - e^{-\frac{1}{2}(y_k^{***}-y_k)(y_k^{***}+y_k-2\theta_j)}\right)$$

$$\left. - e^{-\frac{1}{2}(c_i^2+2c_iy_k-2c_i\theta_j)}\left(1 - e^{2c_i(y_k-\theta_j)}\right)\right|$$

We shall now consider various parts of the right hand side of the above inequality.

$\forall\, j = k, \ldots, i$ and i sufficiently large:

a) Let $\zeta_{i,j} = \sum_{\nu=1}^{\infty} \frac{\left[-\frac{1}{2}\sum_{\ell=1}^{k-1}(y_\ell^* - y_\ell)(y_\ell^* + y_\ell - 2\theta_{j-k+\ell})\right]^\nu}{\nu!}$

then $|\zeta_{1,j}| \le c_i(e - 1)(S + B + c_i)(k - 1)$

and $e^{-\frac{1}{2}\sum_{\ell=1}^{k-1}(y_\ell^*-y_\ell)(y_\ell^*+y_\ell-2\theta_{j-k+\ell})} = 1 + \zeta_{1,j}$.

b) Let $\zeta_{2,j} = \sum_{\nu=1}^{\infty} \frac{(-\frac{1}{2}c_i^2 - c_i y_k + c_i\theta_j)^\nu}{\nu!}$

then $|\zeta_{2,j}| \le c_i(e - 1)(S + B + c_i)$

and $e^{-\frac{1}{2}(c_i^2+2c_i y_k-2c_i\theta_j)} = 1 + \zeta_{2,j}$.

c) Let $\zeta_{3,j} = \sum_{\nu=1}^{\infty} \frac{c_i^\nu[2(y_k - \theta_j)]^{\nu+1}}{(\nu + 1)!}$

then $|\zeta_{3,j}| \le c_i(e^{2(B+S)} - 1)$

and $\frac{1 - e^{2c_i(y_k-\theta_j)}}{c_i} = -2(y_k - \theta_j) - \zeta_{3,j}$.

d) Let $$\zeta_{4,j} = \sum_{\nu=2}^{\infty} \frac{[-\frac{1}{2}(y_k^{**} - y_k)(y_k^{**} + y_k - 2\theta_j)]^\nu}{c_i \nu!}$$

$$- \sum_{\nu=2}^{\infty} \frac{[-\frac{1}{2}(y_k^{***} - y_k)(y_k^{***} + y_k - 2\theta_j)]^\nu}{c_i \nu!}$$

then $|\zeta_{4,j}| \leq 2c_i(e^{(S+B+c_i)} - 1)$

and $$\frac{e^{-\frac{1}{2}(y_k^{**}-y_k)(y_k^{**} + y_k-2\theta_j)} - e^{-\frac{1}{2}(y_k^{***}-y_k)(y_k^{***}+y_k-2\theta_j)}}{c_i}$$

$$= \frac{-(y_k^{**} - y_k)(y_k^{**}+ y_k - 2\theta_j) + (y_k^{***} - y_k)(y_k^{***}+ y_k - 2\theta_j)}{2c_i} + \zeta_{4,j} .$$

e) Let $\zeta_{5,j} = 2c_i(p_{1,j}^2 - p_{2,j}^2)$

then $|\zeta_{5,j}| \leq 2c_i$

and $$\frac{(y_k^{**} - y_k)(y_k^{**} + y_k - 2\theta_j) - (y_k^{***} - y_k)(y_k^{***} + y_k - 2\theta_j)}{2c_i}$$

$$= \frac{(y_k^{**})^2 - (y_k^{***})^2 - 2\theta_j(y_k^{**} - y_k^{***})}{2c_i}$$

$$= \frac{(y_k + 2p_{1,j}c_i)^2 - (y_k - 2p_{2,j}c_i)^2 - 4\theta_j c_i(p_{1,j} + p_{2,j})}{2c_i}$$

$$= \frac{4c_i y_k(p_{1,j} + p_{2,j}) + 4c_i^2(p_{1,j}^2 - p_{2,j}^2) - 4\theta_j c_i(p_{1,j} + p_{2,j})}{2c_i}$$

$$= 2(y_k - \theta_j)(p_{1,j} + p_{2,j}) + \zeta_{5,j} .$$

f) Let $\zeta_{6,j} = p_{1,j} + p_{2,j} - 1$

then $|\zeta_{6,j}| \leq |p_{1,j} - \frac{1}{2}| + |p_{2,j} - \frac{1}{2}| \leq c_i \frac{8\sqrt{2\pi}}{3|y_k - \theta_j|} e^{\frac{1}{2}(B+S)^2}$

and $2(y_k - \theta_j)(p_{1,j} + p_{2,j}) = 2(y_k - \theta_j)(1 + \zeta_{6,j})$

$$= 2(y_k - \theta_j) + \zeta_{7,j}$$

where $|\zeta_{7,j}| \leq c_i \frac{16\sqrt{2\pi}}{3} e^{\frac{1}{2}(B+S)^2}$.

Using the above six results we have for sufficiently large i:

$$|z_i'| \leq \sum_{j=k}^{i-k} \frac{|(1+\zeta_{1,j})(\zeta_{4,j} - \zeta_{5,j} - 2(y_k - \theta_j) - \zeta_{7,j}) + (1+\zeta_{2,j})(2(y_k - \theta_j) + \zeta_{3,j})|}{2(2\pi)^{k/2}(i - k + 1)}$$

$\leq M^* c_i$ where M^* is some finite constant independent of $\underline{\theta}$, i, and $\underline{y}_k$.

Hence $\lim_{i \to \infty} |z_i'| \log i = 0$ uniformly in $\underline{\theta}$ and $\underline{y}_k$ such that $|y_\ell| \leq S$ $\ell = 1, \ldots, k$ and $y_k \neq \theta_j$ $j = k, k+1, \ldots$

We now consider $\lim_{i \to \infty} |q_i(\underline{y}_k)| \log i$ for $|y_\ell| \leq S$ $\ell = 1, \ldots, k$

$$q_i(\underline{y}_k) = \frac{\sum_{j=k}^{i-k} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\sum_{\ell=1}^{k-1}(y_\ell-\theta_{j-k+\ell})^2} \left(e^{-\frac{1}{2}(y_k+c_i-\theta_j)^2} - e^{-\frac{1}{2}(y_k-c_i-\theta_j)^2} \right)}{2(i-k+1)c_i}$$

$$+ \frac{1}{i-k+1} \sum_{j=k}^{i-k} (y_k - \theta_j) \left(\frac{1}{2\pi}\right)^{k/2} e^{-\frac{1}{2}\sum_{\ell=1}^{k}(y_\ell-\theta_{j-k+\ell})^2}$$

So

$$|q_i(\underline{y}_k)| \le \frac{1}{i-k+1} \sum_{j=k}^{i-k} \left| \frac{e^{-\frac{1}{2}(y_k+c_i-\theta_j)^2} - e^{-\frac{1}{2}(y_k-c_i-\theta_j)^2}}{2c_i} + (y_k-\theta_j)e^{-\frac{1}{2}(y_k-\theta_j)^2} \right|$$

$$\le \frac{1}{i-k+1} \sum_{j=k}^{i-k} \left| \frac{e^{-\frac{1}{2}(c_i^2+2c_iy_k-2c_i\theta_j)} - e^{-\frac{1}{2}(c_i^2-2c_iy_k+2c_i\theta_j)}}{2c_i} + (y_k-\theta_j) \right| .$$

Let $$\zeta_j = \sum_{\nu=2}^{\infty} \frac{[-c_i(\frac{c_i}{2}+y_k-\theta_j)]^\nu}{2c_i\nu!} - \sum_{\nu=2}^{\infty} \frac{[-c_i(\frac{c_i}{2}-y_k+\theta_j)]^\nu}{2c_i\nu!}$$

then $|\zeta_j| \le 2c_i(e^{B+S+c_i} - 1)$

and $$\frac{e^{-\frac{1}{2}(c_i^2+2c_iy_k-2c_i\theta_j} - e^{-\frac{1}{2}(c_i^2-2c_iy_k+2c_i\theta_j)}}{2c_i} = - y_k + \theta_j + \zeta_j .$$

Hence $|q_i(\underline{y}_k| \leq 2c_i(e^{B+S+c_i} - 1)$ for all $\underline{\theta}$, $\underline{y}_k$ such that $|y_\ell| \leq S$;

and $\lim_{i\to\infty} |q_i(\underline{y}_k)| \log i = 0$ uniformly in $\underline{\theta}$ as desired.

This completes the proof that the decision procedure $\underline{\varphi}^k$ is uniformly asymptotically optimal of k^{th} order. $\underline{\varphi}^k$ was defined for the case $\sigma = 1$. If we had kept arbitrary σ then $\underline{\varphi}^k$ would have been defined in the same manner except tnat $g_i(\underline{y}_k)$ would have been defined as $\frac{\sigma^2[f_i(\underline{y}_k + c_i\underline{e}_k) - f_i(\underline{y}_k - c_i\underline{e}_k)]}{2c_i}$. If we relax the assumption σ known to the assumption σ unknown but equal for all observations, then it may be shown that if $\hat{\sigma}_i^2$ is an estimate which converges in probability to σ^2 uniformly in $\underline{\theta}$, we may replace σ^2 with $\hat{\sigma}_i^2$ in the definition of $g_i(\underline{y}_k)$, and the resulting decision procedure is still uniformly asymptotically optimal.

If the problem is modified to allow r independent observations for each θ_i, then since the sum of these r observations is sufficient for θ_i and also normally distributed, the above procedure will still apply. We note in this case that if the common variance is unknown, then for each i the usual estimate $\hat{\sigma}_i^2$ is independent of θ_i and $\frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_i^2$ is a consistent estimate for σ^2.

E. The gamma distribution.

$$\text{Let} \quad P[X \le x \mid \theta] = \int_0^x p_\theta(t)\,dt$$

$$= \int_0^x \frac{\theta}{\Gamma(a)} (\theta t)^{a-1} e^{-\theta t}\,dt \qquad 0 < x < \infty$$

for $a > 0$, $0 < \alpha \le \theta \le \beta < \infty$.

We assume a is known, and fix $k \ge 1$.

Let $\{c_i\}$, $\underline{s}_k$, $Y_{j,i}$, f_i, and g_i be defined as in Section D.

Let:

$$P_i^*(\underline{y}_k) = \begin{cases} \dfrac{a-1}{y_k} - \dfrac{g_i(\underline{y}_k)}{f_i(\underline{y}_k)} & \text{if } f_i(\underline{y}_k) > 0 \text{ and } i = k, k+1, \ldots \\ \\ \dfrac{a}{y_k} & \text{otherwise} \end{cases}$$

$$\varphi_i^k(\underline{X}_i) = \begin{cases} \alpha & \text{if } P_i^*(\underline{X}_i^k) \le \alpha \\ P_i^*(\underline{X}_i^k) & \text{if } \alpha < P_i^*(\underline{X}_i^k) < \beta \\ \beta & \text{if } \beta \le P_i^*(\underline{X}_i^k) \end{cases} .$$

We shall prove the decision procedure $\underline{\varphi}^k = (\varphi_1^k, \varphi_2^k, \ldots)$ is uniformly asymptotically optimal of k^{th} order.

We shall show the conditions of the corollary to theorem 2) are satisfied. Since much of the argument is similar to that in the normal case we shall omit many of the intermediate steps.

Let: $$Q_i(\underline{y}_k) = \sum_{j=k}^{i} \prod_{\ell=1}^{k} \frac{\theta^a_{j-k+\ell}}{\Gamma(a)} y_\ell^{a-1} e^{-y_\ell \theta_{j-k+\ell}}$$

$$Q_i'(\underline{y}_k) = \frac{\partial Q_i(\underline{y}_k)}{\partial y_k}$$

$$Q_i^*(\underline{y}_k) = \sum_{j=k}^{i} \theta_j \prod_{\ell=1}^{k} \frac{\theta^a_{j-k+\ell}}{\Gamma(a)} y_\ell^{a-1} e^{-y_\ell \theta_{j-k+\ell}} .$$

Then for all $\underline{y}_k$ such that $y_\ell > 0$ $\ell = 1, \ldots, k$,

$$\psi_i^k(\underline{y}_k) = \frac{Q_i^*(\underline{y}_k)}{Q_i(\underline{y}_k)} .$$

But

$$Q_i'(\underline{y}_k) = \sum_{j=k}^{i} \left[\prod_{\ell=1}^{k-1} \frac{\theta^a_{j-k+\ell}}{\Gamma(a)} y_\ell^{a-1} e^{-y_\ell \theta_{j-k+\ell}} \right] \frac{\theta^a_j}{\Gamma(a)} e^{-y_k \theta_j} \left[(a-1) y_k^{a-2} - \theta_j y_k^{a-1} \right]$$

$$= \frac{a-1}{y_k} Q_i(\underline{y}_k) - Q_i^*(\underline{y}_k).$$

Hence $\psi_i^k(\underline{y}_k) = \dfrac{a-1}{y_k} - \dfrac{Q_i'(\underline{y}_k)}{Q_i(\underline{y}_k)}$

Let: $m_i(\underline{y}_k) = E[f_i(\underline{X}_i^k) | \underline{X}_i^k = \underline{y}_k]$

$$z_i(\underline{y}_k) = \frac{1}{i-k+1} \sum_{j=k}^{i-k} \left[\prod_{\ell=1}^{k} \frac{\theta_{j-k+\ell}^{a}}{\Gamma(a)} (y^*_{\ell,j})^{a-1} e^{-y^*_{\ell,j}\theta_{j-k+\ell}} - \prod_{\ell=1}^{k} \frac{\theta_{j-k+\ell}^{a}}{\Gamma(a)} (y_\ell)^{a-1} e^{-y_\ell \theta_{j-k+\ell}} \right]$$

for $\underline{y}^*_{k,j}$ such that $y_\ell - c_i < y^*_{\ell,j} < y_\ell + c_i$.

Then: $m_i(\underline{y}_k) = \dfrac{Q_{i-k}(\underline{y}_k)}{i-k+1} + z_i(\underline{y}_k)$ for suitable $y^*_{\ell,j}$,

and for $i = k, k+1, \ldots$ $\underline{y}_k$ such that $y_\ell > 0$ $\ell = 1, \ldots, k$ $\underline{\theta} \epsilon \Omega^\infty$

$$P\left[\left| f_i(\underline{X}_i^k) - \frac{Q_i(\underline{X}_i^k)}{i-k+1} \right| \geq \delta_i \ \left| \underline{X}_i^k = \underline{y}_k \right.\right]$$

$$\leq 2k \exp\left\{ B2^{2(k+1)} c_i^{2k} \delta_i - 2^{2k+1} \frac{i-k+1}{k} c_i^{2k} (\delta_i - |z_i|)^2 \right\}$$

provided $\delta_i - |z_i| - \dfrac{kB}{i-k+1} \geq 0$

where $B = \max\left\{ \left[3\dfrac{(a-1)^{a-1}}{\Gamma(a)} e^{-(a-1)} \right]^k, 1 \right\}$.

Let: $z_i'(\underline{y}_k) = \dfrac{z_i(\underline{y}_k + c_i\underline{e}_k) - z_i(\underline{y}_k - c_i\underline{e}_k)}{2c_i}$

$$q_i(\underline{y}_k) = \frac{1}{i - k + 1}\,\frac{Q_{i-k}(\underline{y}_k + c_i\underline{e}_k) - Q_{i-k}(\underline{y}_k - c_i\underline{e}_k)}{2c_i} - Q'_{i-k}(\underline{y}_k)\ .$$

Then: $E[g_i(\underline{X}_i^k) \mid \underline{X}_i^k = \underline{y}_k] = \dfrac{Q'_{i-k}(\underline{y}_k)}{i - k + 1} + q_i(\underline{y}_k) + z_i'(\underline{y}_k)$

and for $i = k,\ k + 1, \ldots$ $\underline{y}_k$ such that $y_\ell > 0$ $\ell = 1, \ldots, k$ $\underline{\theta} \in \Omega$

$$P\left[\left|g_i(\underline{X}_i^k) - \frac{Q_i'(\underline{X}_i^k)}{i - k + 1}\right| \geq \epsilon_i \middle| \underline{X}_i^k = \underline{y}_k\right]$$

$$\leq 2k \exp\left\{B^*(2c_i)^{2(k+1)}\epsilon_i - 2^{2k+1}c_i^{2(k+1)}\,\frac{i - k + 1}{k}\,(\epsilon_i - |z'| - |q_i|)^2\right\}$$

provided $\epsilon_i - |z_i'| - |q_i| - \dfrac{kB^*}{i - k + 1} \geq 0$

where $B^* < \infty$ is such that $\max\limits_{x,y>0} (p_\theta(x))^{k-1} \left.\dfrac{\partial p_\theta(t)}{\partial t}\right|_{t=y} \leq B^*$.

Thus we have:

$$P\left[|\varphi_i(\underline{X}_i) - \psi_i(\underline{X}_i)| \geq \frac{i - k + 1}{Q_i(\underline{y}_k)}\left(\epsilon_i + \beta\delta_i + \frac{|a - 1|}{y_k}\delta_i\right) \middle| \underline{X}_i^k = \underline{y}_k\right]$$

$$\leq 2k \exp\left\{B2^{2(k+1)}c_i^{2k}\delta_i - 2^{2k+1}\,\frac{i - k + 1}{k}\,c_i^{2k}(\delta_i - |z_i|)^2\right\}$$

$$+\ 2k \exp\left\{B^*(2c_i)^{2(k+1)}\epsilon_i - 2^{2k+1}c_i^{2(k+1)}\,\frac{i - k + 1}{k}(\epsilon_i - |z'| - |q_i|)^2\right\}$$

provided $\delta_i - |z_i| - \frac{kP}{i - k + 1} \geq 0$

$$\epsilon_i - |z_i'| - |q_i| - \frac{kB^*}{i - k + 1} \geq 0 .$$

To complete the proof it is sufficient to show that, for appropriate δ_i and ϵ_i, the limits i), iii), iv), v) and vi) listed in Section D hold uniformly in $\underline{\theta}$, and that

ii') $\lim_{i\to\infty} E\left[\frac{i - k + 1}{Q_i(\underline{X}_i^k)} (\beta + \frac{1}{X_i})\delta_i(\underline{X}_i^k)\right] = 0$ uniformly in $\underline{\theta}$.

Let:

$$\delta_i(\underline{y}_k) = \begin{cases} \dfrac{y_k Q_i(\underline{y}_k)}{(i - k + 1)\log i} & \text{if } \dfrac{y_k Q_i(\underline{y}_k)}{(i - k + 1)\log i} \geq |z_i| + \dfrac{kB}{(i - k + 1)^{1/4}} \\ 0 & \text{otherwise} \end{cases}$$

$$\epsilon_i(\underline{y}_k) = \begin{cases} \dfrac{Q_i(\underline{y}_k)}{(i - k + 1)\log i} & \text{if } \dfrac{Q_i}{(i - k + 1)\log i} \geq |z_i'| + |q_i| + \dfrac{k B^*}{(i - k + 1)^{1/4}} \\ 0 & \text{otherwise} \end{cases}$$

Then limits i), iii), and iv) clearly hold. Since $E[X] \leq \frac{a}{\alpha}$ for all $\theta \epsilon \Omega$ limit ii') also holds. $\forall \eta > 0$ there exist $s(\eta)$, $S(\eta)$ such that $0 < s < S < \infty$, $P[X < s] < \frac{\eta}{2}$ and $P[X > S] < \frac{\eta}{2}$ for all $\theta \epsilon \Omega$. Hence it remains only to show:

$\lim_{i \to \infty} |z_i(\underline{y}_k)| \log i = 0$ uniformly in $\underline{\theta}$ and $\underline{y}_k$ such that

$s \leq y_\ell \leq S \quad \ell = 1, \ldots, k$

$\lim_{i \to \infty} |z_i'(\underline{y}_k)| \log i = 0$ uniformly in $\underline{\theta}$ and $\underline{y}_k$ such that

$s \leq y_\ell \leq S$ and $y_k \neq \frac{a-1}{\theta_j}$

$\ell = 1, \ldots, k \quad j = k, k+1, \ldots$

$\lim_{i \to \infty} |q_i(\underline{y}_k)| \log i = 0$ uniformly in $\underline{\theta}$ and $\underline{y}_k$ such that

$s \leq y_\ell \leq S \quad \ell = 1, \ldots, k$.

Now

$$|z_i(\underline{y}_k)| \leq \frac{1}{i-k+1} \sum_{j=k}^{i-k} \frac{\left(\prod_{\ell=1}^{k} \theta_{j-k+\ell}^{a}\right)}{(\Gamma(a))^k} \left| \prod_{\ell=1}^{k} (y^*_{\ell,j})^{a-1} e^{-y^*_{\ell,j}\theta_{j-k+\ell}} - \prod_{\ell=1}^{k} y_\ell^{a-1} e^{-y_\ell \theta_{j-k+\ell}} \right|$$

$\forall\ j = k, \ldots, i - k$

Let $\zeta_{1,\ell} = 1 - \dfrac{y_\ell^*}{y_\ell}$

then $y_\ell - c_i < y_\ell^* < y_\ell + c_i \Rightarrow 1 - \dfrac{c_i}{s} < \dfrac{y_\ell^*}{y_\ell} < 1 + \dfrac{c_i}{s}$

$$\Rightarrow |\zeta_{1,\ell}| < \frac{c_i}{s}$$

so that $\left|\dfrac{\prod_{\ell=1}^{k} y_\ell^*}{\prod_{\ell=1}^{k} y_\ell}\right|^{a-1} = \prod_{\ell=1}^{k} |1 - \zeta_{1,\ell}|^{a-1} = 1 + \zeta_1$

where, for i so large that $c_i < s$,

$$|\zeta_1| \leq 1 + 2^{k(a-1)} \frac{c_i}{s} .$$

Let $\zeta_2 = \displaystyle\sum_{\nu=1}^{\infty} \frac{\left(\sum_{\ell=1}^{k} (y_\ell^* - y_\ell)\theta_{j-k+\ell}\right)^{\nu}}{\nu!}$

then $|\zeta_2| \leq c_i(e^{k\beta} - 1)$ for $|c_i| < 1$

and $e^{-\sum_{\ell=1}^{k} (y_\ell^* - y_\ell)\theta_{j-k+\ell}} = 1 + \zeta_2$

Thus

$$\left|\prod_{\ell=1}^{k} (y_\ell^*)^{a-1}\, e^{-y_\ell^*\theta_{j-k+\ell}} - \prod_{\ell=1}^{k} y_\ell^{a-1}\, e^{-y_\ell\theta_{j-k+\ell}}\right|$$

$$= \prod_{\ell=1}^{k} y_\ell^{a-1}\, e^{-y_\ell\theta_{j-k+\ell}} \left|\prod_{\ell=1}^{k} \left(\frac{y_\ell^*}{y_\ell}\right)^{a-1} e^{-(y_\ell^*-y_\ell)\theta_{j-k+\ell}} - 1\right|$$

$$\leq S^{a-1}|(1 - \zeta_1)(1 + \zeta_2) - 1|$$

$$\leq S^{a-1}(|\zeta_1| + |\zeta_2| + |\zeta_1\zeta_2|)$$

so that $\lim\limits_{i\to\infty} |z_i(\underline{y}_k)| \log i = 0$ uniformly.

We now consider $\lim\limits_{i\to\infty} |z_i'(\underline{y}_k)| \log i$. As in the normal case we shall use lemma 5).

We have:

$$|z_i'(\underline{y}_k)| \leq \frac{1}{(i - k + 1)} \sum_{j=k}^{i-1} \frac{\prod_{\ell=1}^{k} \theta^a_{j-k+\ell}}{(\Gamma(a))^k} \left| \frac{\left[\prod_{\ell=1}^{k-1} (y^*_{\ell,j})^{a-1}\, e^{-y^*_{\ell,j}\theta_{j-k+\ell}}\right]}{} \right.$$

$$\frac{\left[(y^{**}_{k,j})^{a-1}\, e^{-y^{**}_{k,j}\theta_j} - (y^{***}_{k,j})^{a-1}\, e^{-y^{***}_{k,j}\theta_j}\right]}{2c_i}$$

$$\left. - \frac{\left[\prod_{\ell=1}^{k-1} y_\ell^{a-1}\, e^{-y_\ell\theta_{j-k+\ell}}\right]\left[(y_k + c_i)^{a-1}\, e^{-(y_k+c_i)\theta_j} - (y_k - c_i)^{a-1} e^{-(y_k-c_i)\theta_j}\right]}{2c_i} \right| \leq$$

$$\leq \frac{S^{k(a-1)}\beta^{ka}}{(i-k+1)(\Gamma(a))^k} \sum_{j=k}^{i-1} \left| \left(\prod_{\ell=1}^{k-1} \left(\frac{y^*_{\ell,j}}{y_\ell} \right)^{a-1} e^{-(y^*_{\ell,j} - y_\ell)\theta_{j-k+\ell}} \right) \right.$$

$$\left[\frac{\left(\frac{y^{**}_{k,j}}{y_k} \right)^{a-1} - \left(\frac{y^{***}_{k,j}}{y_k} \right)^{a-1} e^{-(y^{***}_{k,j} - y^{**}_{k,j})\theta_j}}{2c_i} \right]$$

$$\left. - e^{(y_k + c_i - y^{**}_{k,j})\theta_j} \left[\frac{\left(\frac{y_k + c_i}{y_k} \right)^{a-1} - \left(\frac{y_k - c_i}{y_k} \right)^{a-1} e^{2c_i\theta_j}}{2c_i} \right] \right|$$

where: $y^{**}_{k,j} = y_k + 2p_{1,j}c_i \qquad y^{***}_{k,j} = y_k - 2p_{2,j}c_i$

$$(p_{\nu,j} - \tfrac{1}{2}) \leq \frac{4Mc_i}{3|p'_\theta(y_k)|} \leq \frac{4M\Gamma(a)e^{aS}c_i}{3S^{a-2}|a - 1 - \theta_j y_k|} \qquad \nu = 1, 2 .$$

We shall now consider various parts of the preceding inequality for some fixed j $k \leq j \leq i - k$. To simplify the notation we let $y = y_{k,j}$ $c = c_i$ $\theta = \theta_j$ in what follows. We assume sufficiently large i.

a) Let $\zeta_{1,\ell} = 1 - \frac{y^*_{\ell,j}}{y_\ell}$

then $|\zeta_{1,\ell}| \leq \frac{c}{s}$

and $$\prod_{\ell=1}^{k-1}\left(\frac{y^*_{\ell,j}}{y_\ell}\right)^{a-1} = 1 + \zeta_1$$

where $|\zeta_1| \leq c2^{(k-1)(a-1)}$.

b) Let $$\zeta_2 = \sum_{\nu=2}^{\infty} \frac{\binom{a-1}{\nu}\left(\frac{2p_1c}{\nu}\right)^{\nu}}{\nu!}$$

then $$|\zeta_2| \leq c^2 \frac{2^{a+1}}{s^2}$$

and $$\left(\frac{y^{**}}{y}\right)^{a-1} = \left(1 + \frac{2p_1c}{y}\right)^{a-1} = 1 + \frac{2(a-1)p_1c}{y} + \zeta_2 \; .$$

c) Let $$\zeta_3 = \sum_{\nu=2}^{\infty} \frac{\binom{a-1}{\nu}\left(-\frac{2p_2c}{\nu}\right)^{\nu}}{\nu!}$$

then $$|\zeta_3| \leq c^2 \frac{2^{a+1}}{s^2}$$

and $$\left(\frac{y^{***}}{y}\right)^{a-1} = \left(1 - \frac{2p_2c}{y}\right)^{a-1} = 1 - \frac{2(a-1)p_2c}{y} + \zeta_3 \; .$$

d) Let $$\zeta_4 = \sum_{\nu=2}^{\infty} \frac{\binom{a-1}{\nu}\left(\frac{c}{y}\right)^{\nu}}{\nu!}$$

then $$|\zeta_4| \leq c^2 \frac{2^{a-1}}{s^2}$$

and $\left(\frac{y + c}{y}\right)^{a-1} = 1 + \frac{(a - 1)c}{y} + \zeta_4$.

e) Let $\zeta_5 = \sum_{\nu=2}^{\infty} \binom{a - 1}{\nu} \frac{\left(- \frac{c}{y}\right)^{\nu}}{\nu!}$

then $|\zeta| \leq c^2 \frac{2^{a-1}}{s^2}$

and $\left(\frac{y - c}{y}\right)^{a-1} = 1 - \frac{(a - 1)c}{y} + \zeta_5$.

f) Let $\zeta_6 = \sum_{\nu=2}^{\infty} \frac{[2c(p_1 + p_2)\theta]^{\nu}}{\nu!}$

then $|\zeta_6| \leq c^2(e^{4\beta} - 1)$

and $e^{-(y^{***}-y^{**})\theta} = e^{2c(p_1+p_2)\theta} = 1 + 2c(p_1 + p_2)\theta + \zeta_6$.

g) Let $\zeta_7 = \sum_{\nu=1}^{\infty} \frac{[(2p_1 - 1)c\theta]^{\nu}}{\nu!}$

then $|\zeta_7| \leq c(e^{\beta} - 1)$

and $e^{-(y+c-y^{**})\theta} = e^{(2p_1-1)c\theta} = 1 + \zeta_7$.

h) Let $\zeta_8 = \sum_{\nu=2}^{\infty} \frac{(2c\theta)^{\nu}}{\nu!}$

then $|\zeta_8| \leq c^2(e^{2\beta} - 1)$

and $e^{2c\theta} = 1 + 2c\theta + \zeta_8$.

i) Let $\zeta_9 = \sum_{\nu=1}^{\infty} \dfrac{\left[\sum_{\ell=1}^{k-1} (y^*_{\ell,j} - y_\ell)\theta_{j-k+\ell}\right]^\nu}{\nu!}$

then $|\zeta_9| \leq c(e^{(k-1)\beta} - 1)$

and $\prod_{\ell=1}^{k-1} e^{-(y^*_{\ell,j} - y_\ell)\theta_{j-k+\ell}} = 1 + \zeta_9$.

Thus

$$\left|\left(\prod_{\ell=1}^{k-1}\left(\frac{y_{\ell,j}}{y_\ell}\right)^{a-1} e^{-(y^*_{\ell,j}-y_\ell)\theta_{j-k+\ell}}\right)\left[\frac{\left(\frac{y^{**}}{y}\right)^{a-1} - \left(\frac{y^{***}}{y}\right)^{a-1} e^{-(y^{***}-y^{**})\theta}}{2c}\right]\right.$$

$$\left. - e^{-(y+c-y^{**})\theta}\left[\frac{\left(\frac{y+c}{y}\right)^{a-1} - \left(\frac{y-c}{y}\right)^{a-1} e^{2c\theta}}{2c}\right]\right|$$

$$= \left|(1+\zeta_1)(1+\zeta_9)\left[\frac{1+\frac{2(a-1)p_1c}{y} + \zeta_2 - \left(1 - \frac{2(a-1)p_2c}{y} + \zeta_3\right)(1+2c(p_1+p_2)\theta + \zeta_6)}{2c}\right]\right.$$

$$\left. - (1+\zeta_7)\frac{\left[1 + \frac{(a-1)c}{y} + \zeta_4 - \left(1 - \frac{(a-1)c}{y} + \zeta_5\right)(1 + 2c\theta + \zeta_8)\right]}{2c}\right| =$$

$= |(p_1 + p_2 - 1)(\frac{a-1}{y} - \theta) + \zeta|$ where $|\zeta| < cM^*$ for some finite constant M^* independent of $\underline{y}_k$, $\underline{\theta}$, or j.

$\leq |p_1 - \frac{1}{2}|\,|\frac{a-1}{y} - \theta| + |p_2 - \frac{1}{2}|\,|\frac{a-1}{y} - \theta| + |\zeta|$

$\leq cM^{**}$ for some finite constant M^{**} independent of $\underline{y}_k$, $\underline{\theta}$, or j.

Hence $\lim_{i\to\infty} |z_i'(\underline{y}_k)| \log i = 0$ uniformly in $\underline{\theta}$ and $\underline{y}_k$ for the set of $\underline{y}_k$ of interest.

It now remains only to show $\lim_{i\to\infty} |q_i(\underline{y}_k)| \log i = 0$ uniformly for $\underline{y}_k$ such that $s < y_\ell < S$ $\ell = 1, \ldots, k$. But

$$|q_i(\underline{y}_k)| \leq \frac{1}{i-k+1} \sum_{j=k}^{i-k} \left(\prod_{\ell=1}^{k-1} \frac{\theta^a_{j-k+\ell}}{\Gamma(a)} y_\ell^{a-1} e^{-y_\ell \theta_{j-k+\ell}} \right) \frac{\theta_j^a}{\Gamma(a)}$$

$$e^{-y_k\theta_j} \left| \frac{\left(\frac{y_k + c_i}{y_k}\right)^{a-1} e^{-c_i\theta_j} - \left(\frac{y_k - c_i}{y_k}\right)^{a-1} e^{c_i\theta_j}}{2c_i} - \left[\frac{(a-1)}{y_k} - \theta_j\right] \right| .$$

So it is sufficient to show that $s < y < S$ $\theta \in \Omega$

$$\Rightarrow \left| \frac{\left(\frac{y+c}{y}\right)^{a-1} e^{-c\theta} - \left(\frac{y-c}{y}\right)^{a-1} e^{c\theta}}{2c} - \frac{a-1}{y} + \theta \right| < cM$$

where M is some finite constant independent of θ, y, or c. Using parts d), e) and slight modifications of h) in the proof of the previous limit we have:

$$\left|\frac{\left(\frac{y+c}{y}\right)^{a-1}e^{-c\theta} - \left(\frac{y-c}{y}\right)^{a-1}e^{c\theta}}{2c} - \frac{a-1}{y} + \theta\right|$$

$$= \left|\frac{\left(1 + \frac{(a-1)c}{y} + \zeta_4\right)(1 - c\theta + \zeta_8') - \left(1 - \frac{(a-1)c}{y} + \zeta_5\right)(1 + c\theta + \zeta_8'')}{2c} - \frac{a-1}{y} + \theta\right|$$

$\leq cM$ as desired. This completes the proof that the decision procedure φ^k is uniformly asymptotically optimal of k^{th} order.

We note, as in the previous sections, that if the problem is modified to allow r independent observations for each θ_i, then the sum of these observations may be used to obtain an asymptotically optimal decision procedure of k^{th} order.

We now consider the case in which the other parameter is unknown, that is $p_\theta(x) = \frac{\lambda^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\lambda x}$ for known λ. It may be shown that if

$$P_i^*(\underline{y}_k) = \begin{cases} 1 + \left(\lambda + \frac{g_i(\underline{y}_k)}{f_i(\underline{y}_k)} y_k\right) & \text{if } f_i(\underline{y}_k) > 0 \text{ and } i = k, k+1, \ldots \\ 1 + \lambda y_k & \text{otherwise} \end{cases}$$

and all other definitions in this section are unchanged, then the resulting φ^k is uniformly asymptotically optimal of k^{th} order.

F. The non-parametric case.

We now consider the following problem. Let $\mathfrak{F} = \{p_\theta(\cdot): \theta \in \Omega\}$ be a class of probability mass functions, each of which assigns probability one to a specified denumerable class $\mathfrak{X} = \{x\}$ of real numbers. Ω is an arbitrary index set. We assume for each θ in Ω that $p_\theta(\cdot)$ is completely specified. Let $h(\cdot)$ be a real valued function on $\mathfrak{X}$. Let $\lambda(\theta) = E[h(X)|\theta]$. We assume $E[h^2(X)|\theta] \leq B < \infty$ for all θ in Ω. For some unknown $\theta_j \in \Omega$ we observe $r+1$ independent identically distributed random variables $X_{j,1}, \ldots, X_{j,r+1}$ with $P[X_{j,s} = x] = p_{\theta_j}(x)$ $s = 1, \ldots, r+1$ $x \in \mathfrak{X}$. We wish to estimate $\lambda(\theta_j)$ on the basis of these observations. For example, if $h(x) = x$ then we are estimating $E[X|\theta]$. If φ_j is our estimate we suffer a loss of $(\varphi_j - \lambda(\theta_j))^2$. We now assume we are faced with a sequence of such decisions. In other words a sequence $\{\theta_j: j = 1, 2, \ldots\}$ is selected from Ω^∞. For each θ_j we have $r+1$ observations and we may use $\underline{\underline{X}}_j$ to estimate θ_j, where $\underline{\underline{X}}_j$ is the $j \times (r+1)$ matrix of observations $(X_{s,t})$. Johns [3] has considered this problem under the assumption that each θ_j is an independent observation of an Ω-valued random variable Θ with unknown a priori probability measure G defined over a suitable σ-algebra of subsets of Ω. We shall consider the case in which the sequence $\{\theta_j\}$ is arbitrarily chosen.

As in the previous cases we need a standard to use in evaluating a particular decision procedure. For any $\underline{\theta}_n \in \Omega^n$ we form the k^{th} order empirical probability measure G_n^k such that for any sets $\Omega_1, \ldots, \Omega_k$ in the σ-algebra,

$G_n^k(\Omega_1, \ldots, \Omega_k) = \frac{1}{n-k+1}$ [number of i $(k \leq i \leq n)$ such that $\theta_{i-k+\ell} \in \Omega_\ell$ $\ell = 1, \ldots, k$]. If we assume for $i \leq k \leq m$ that $\{\Theta_i, i = 1, \ldots, m\}$ is a sequence of random variables taking values in Ω with $\underline{\Theta}_m^k$ having any k-dimensional probability measure G^k, and Θ_m is independent of $\Theta_1, \ldots, \Theta_{m-k}$, then the Bayes estimate for $\lambda(\Theta_m)$ is $E[\lambda(\Theta_m)|\underline{\underline{X}}_m^k]$ and the Bayes risk is $R_{r+1}(G^k) = E\{[\lambda(\Theta_m) - E[\lambda(\Theta_m)|\underline{\underline{X}}_m^k]]^2\}$, where $\underline{\underline{X}}_m^k$ is the $k \times (r+1)$ matrix consisting of the last k rows of $\underline{\underline{X}}_m$ and the subscript $r+1$ in the Bayes risk refers to the number of observations for each parameter value. We now take as our standard $R_{k,r}(\underline{\theta}_n) = R_r(G_n^k)$, and seek a procedure $\underline{\varphi}^k$ such that

$$\text{F)}\quad \overline{\lim_{n\to\infty}} \left\{ \sup_{\underline{\theta}} \left[\frac{1}{n} \sum_{i=1}^{n} E[(\varphi_i^k(\underline{\underline{X}}_i) - \lambda(\theta_i))^2] - R_{k,r}(\underline{\theta}_n) \right] \right\} \leq 0 .$$

We observe that $R_{k,r+1}(\underline{\theta}_n)$ is not a desirable standard since if $\mathfrak{F}$ is the class of binomial densities, for example, then, as mentioned earlier, $R_{k,r+1}(\underline{\theta}_n)$ could not be achieved.

We observe that theorems 1) and 2) are still valid in this case when $R_{k,r}(\underline{\theta}_n)$ is substituted for $R_k(\underline{\theta}_n)$ and property F) for the property of uniformly asymptotically optimal of k^{th} order.

We define:

$$A^k = \begin{pmatrix} x_{1,1}, & \cdots, & x_{1,r} \\ \vdots & & \vdots \\ x_{k,1}, & \cdots, & x_{k,r} \end{pmatrix}$$ where $x_{s,t}$ is an arbitrary real number.

$A^k_{(q)}$ $q = 1, 2, \ldots, m(A^k)$ to be the $m(A^k)$ distinct matrices obtained from A^k by independently permuting the elements within each row. Clearly $1 \leq m(A^k) \leq (r!)^k$.

$$\underline{\underline{X}}^k_{j,r} = \begin{pmatrix} X_{j-k+1,1} & X_{j-k+1,2} & \cdots & X_{j-k+1,r} \\ X_{j-k+2,1} & X_{j-k+2,2} & \cdots & X_{j-k+2,r} \\ \vdots & \vdots & & \vdots \\ X_{j,1} & X_{j,2} & \cdots & X_{j,r} \end{pmatrix}$$

$$M_j(A^k) = \begin{cases} 1 & \text{if there exists } q = 1, \ldots, m(A^k) \text{ such that} \\ & \underline{\underline{X}}^k_{j,r} = A^k_{(q)} \quad j = k, k+1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

$$Z_j(A^k) = M_j(A^k)h(X_{j,r+1}) \quad j = k, k+1, \ldots$$

$$P_i(A^k) = \frac{\sum_{j=k}^{i} M_j(A^k)}{m(A^k)(i - k + 1)}$$

$$\hat{P}_i(A^k) = \frac{\sum_{j=k}^{i} Z_j(A^k)}{m(A^k)(i - k + 1)}$$

$$P_i^*(A^k) = \begin{cases} \dfrac{\hat{P}_i(A^k)}{P_i(A^k)} & \text{if } P_i(A^k) > 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$\varphi_i^k(\underline{X}_i) = \begin{cases} \dfrac{1}{r+1} \sum_{s=1}^{r+1} h(X_{i,s}) & \text{if } i = 1, \ldots, k-1 \\ -B^{1/2} & \text{if } P_i^*(\underline{X}_{i,r}^k) \leq -B^{1/2} \\ & \quad i = k, k+1, \ldots \\ P_i^*(\underline{X}_{i,r}^k) & \text{if } -B^{1/2} < P_i^*(\underline{X}_{i,r}^k) < B^{1/2} \\ & \quad i = k, k+1, \ldots \\ B^{1/2} & \text{if } B^{1/2} \leq P_i^*(\underline{X}_{i,r}^k) \end{cases}$$

We shall prove $\underline{\varphi}^k = (\varphi_1^k, \varphi_2^k, \ldots)$ has property F) provided the following condition on $\mathfrak{F}$ is satisfied. $\forall\, \epsilon\ 0 < \epsilon < 1$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=k}^{n} P\left[\sum_{j=k}^{i} \prod_{\ell=1}^{k} \prod_{s=1}^{r} p_{\theta_{j-k+\ell}}(X_{i-k+\ell,s}) < i^{\epsilon}\right] = 0 \text{ uniformly in } \underline{\theta}.$$

where $X_{i-k+\ell,s}$ has probability mass function $p_{\theta_{i-k+\ell}}(\cdot)$. Such a condition is satisfied, for example, if $\theta\epsilon\,\Omega\ x\epsilon\mathscr{X} \Rightarrow p_{\theta}(x) \geq \eta(x) > 0$, since in this case

$$P\left[\sum_{j=1}^{i} \prod_{\ell=1}^{k} \prod_{s=1}^{r} p_{\theta_{j-k+\ell}}(X_{i-k+\ell}) < i^{\epsilon}\right]$$

$$= \sum_{x_i\epsilon\mathscr{X}} \cdots \sum_{x_{kr}\epsilon\mathscr{X}} I\left[\sum_{j=1}^{i} \prod_{\ell=1}^{k} \prod_{s=1}^{r} p_{\theta_{j-k+\ell}}(x_{(\ell-1)r+s}),\ i^{\epsilon}\right] \prod_{\ell=1}^{k} \prod_{s=1}^{r} p_{\theta_{i-k+\ell}}(x_{(\ell-1)r+s})$$

$$\leq \sum_{x_1\epsilon\mathscr{X}} \cdots \sum_{x_{kr}\epsilon\mathscr{X}} I\left[\prod_{\ell=1}^{k} \prod_{s=1}^{r} \eta(x_{(\ell-1)r+s}),\ \frac{1}{i^{1-\epsilon}}\right] \prod_{\ell=1}^{k} \prod_{s=1}^{r} p_{\theta_{i-k+\ell}}(x_{(\ell-1)r+s})$$

where, as before, $I(a, b) = \begin{cases} 1 & \text{if } a < b \\ 0 & \text{otherwise} . \end{cases}$

But $\forall\ \delta > 0$ there exists a set $\mathscr{X}_{\delta} \subset \mathscr{X}$ such that $\mathscr{X}_{\delta}$ has only a finite number N_{δ} of elements and $P[x\epsilon\mathscr{X}_{\delta}] > 1 - \delta$. Hence:

$$\sum_{x_1\in\mathfrak{X}} \cdots \sum_{x_{kr}\in\mathfrak{X}} I\left[\prod_{\ell=1}^{k}\prod_{s=1}^{r} \eta(x_{(\ell-1)r+s}), \frac{1}{i^{1-\epsilon}}\right] \prod_{\ell=1}^{k}\prod_{s=1}^{r} p_{\theta_{i-k+\ell}}(x_{(\ell-1)r+s})$$

$$\leq \sum_{x_1\in\mathfrak{X}_\delta} \cdots \sum_{x_{kr}\in\mathfrak{X}_\delta} I\left[\prod_{\ell=1}^{k}\prod_{s=1}^{r} \eta(x_{(\ell-1)r+s}), \frac{1}{i^{1-\epsilon}}\right] + 1 - (1-\delta)^{kr}$$

$$\leq 1 - (1-\delta)^{kr} \quad \text{for} \quad i \geq i_o \quad \text{where} \quad \frac{1}{(i_o)^{1-\epsilon}} < \min_{x\in\mathfrak{X}_\delta} [\eta(x)]^{kr} .$$

Since δ was arbitrary the result quickly follows. This is not the only case, however, in which the condition is satisfied, as was seen in Sections A and C.

The proof that $\underline{\varphi}^k$ has property F) follows the same general lines as in our other examples. $\forall\, i = k, k+1, \ldots$

Let: $$Q_i(A^k) = \frac{1}{i-k+1}\sum_{j=k}^{i}\prod_{\ell=1}^{k}\prod_{s=1}^{r} p_{\theta_{j-k+\ell}}(x_{\ell,s})$$

$$Q_i^*(A^k) = \frac{1}{i-k+1}\sum_{j=k}^{i}\sum_{x\in\mathfrak{X}} h(x)p_{\theta_j}(x)\prod_{\ell=1}^{k}\prod_{s=1}^{r} p_{\theta_{j-k+\ell}}(x_{\ell,s})$$

R_i be a set of $k \times r$ matrices such that $P[\underline{X}^k_{i,r}\in R_i] = 1$ and $A^k\in R_i \Rightarrow Q(A^k) > 0$

Then for $A^k \epsilon R_i$ we have that one version of

$E[\lambda(\Theta_m)|\underset{=}{X}^k_{i,r} = A^k]$ is equal to $\psi^k_{i,r}(A^k) = \dfrac{Q^*_i(A^k)}{Q_i(A^k)}$ when Θ_m has probability measure G^k_i.

But

$$E[M_j(\underset{=}{X}^k_{i,r})|\underset{=}{X}^k_{i,r} = A^k] = \sum_{q=1}^{m(A^k)} \prod_{\ell=1}^{k} \prod_{s=1}^{r} p_{\theta_{j-k+\ell}}(x_{\ell,s}) \quad j = k, \ldots , i - k$$

and

$$E[Z_j(\underset{=}{X}^k_{i,r})|\underset{=}{X}^k_{i,r} = A^k] = E[M_j(A^k)]E[h(X_{j,r+1})]$$

$$= m(A^k) \prod_{\ell=1}^{k} \prod_{s=1}^{r} p_{\theta_{j-k+\ell}}(x_{\ell,s}) \sum_{x\epsilon\mathfrak{X}} h(x)p_{\theta_j}(x)$$

$$j = k, \ldots , i - k .$$

Hence, arguing as in Section A,

$$P[|P_i(X^k_{i,r}) - Q_i(X^k_{i,r})| > \delta_i(A^k)|\underset{=}{X}^k_{i,r} = A^k] \leq 2ke^{4\delta_i} e^{-2\frac{i-k+1}{k}\delta_i^2}$$

rovided $\delta_i \geq \dfrac{k}{i - k + 1}$.

Since Z_j is not bounded we are unable to use lemma 3). We can, however, use a simple Chebyshev bound, observing that $\mathrm{Var}[Z_j(A^k)] \leq E[h^2(X_{j,r+1})] \leq B$; and hence using an argument similar to that in the proof of lemma 3) we have:

$$P[|\hat{P}_i(X^k_{i,r} - Q_i(X^k_{i,r})| > \delta_i(A^k)|\underline{\underline{X}}^k_{i,r} = A^k]$$

$$\leq \frac{k^2 B}{(i-k+1)\left|\delta_i - \frac{k}{i-k+1}\right|^2} \quad \text{provided} \quad \delta_i \geq \frac{k}{i-k+1} .$$

It thus follows, using a modified version of lemma 4), that: $\forall\ A^k \epsilon R_i$

$$P[|\varphi_i(\underline{\underline{X}}_i) - \psi_{i,r}(\underline{\underline{X}}^k_{i,r})| \geq \frac{(\delta_i + B^{1/2}\delta_i)}{Q_i(A^k)}|\underline{\underline{X}}^k_{i,r} = A_k]$$

$$\leq 2ke^{4\delta_i} e^{-2\frac{i-k+1}{k}\delta_i^2} + \frac{k^2 B}{(i-k+1)(\delta_i - \frac{k}{i-k+1})^2}$$

provided $\delta_i \geq \frac{k}{i-k+1}$

Thus we may take

$$\xi_i = \begin{cases} \dfrac{\delta_i + B^{1/2}\delta_i}{Q_i} & \text{if } \delta_i \geq \dfrac{k}{i-k+1} \\ 0 & \text{otherwise} \end{cases}$$

$$\zeta_i = \begin{cases} 2ke^{4\delta_i}e^{-2\frac{i-k+1}{k}\delta_i^2} + \dfrac{k^2 B(i-k+1)}{[(i-k+1)\delta_i - k]^2} & \text{if } \delta_i \geq \dfrac{k}{i-k+1} \\ 1 & \text{otherwise} \end{cases}$$

Let $a_i = \left(\frac{1}{i}\right)^{1/4}$

$$\delta_i = \begin{cases} \dfrac{Q_i \log i}{i^{1/4}} & \text{if } Q_i \geq \dfrac{k\, i^{1/4}}{(i - k + 1)\log i} \\ 0 & \text{otherwise} \end{cases} .$$

Then clearly $\lim_{n\to\infty} \frac{1}{n} \sum_{i=k}^{n} E[\xi_i + \zeta_i | Q_i \geq a_i] = 0$ uniformly in $\underline{\theta}$.
Since our assumed condition assures that $\lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=k}^{n} P[Q_i < a_i]\right] = 0$ uniformly in $\underline{\theta}$, theorem 2) is satisfied, and $\underline{\varphi}^k$ has property F).

We observe that in this case, as in the binomial, the choice of which information to neglect at the i^{th} stage was arbitrary. In particular $\underline{\underline{X}}^k_{i,r}$ could have been defined in any one of $(r + 1)^k$ ways. We thus could have defined $(r + 1)^k$ essentially different estimators $\varphi_{i,u}$, each of which would have the desired properties. As in the binomial case it may be shown that $\tilde{\varphi}_i = \frac{1}{(r + 1)^k} \sum_{u=1}^{(r+1)^k} \varphi_{i,u}$ is an improved estimate.

We note that if $\mathfrak{F}$ had been defined as a class of absolutely continuous distribution functions, a similar decision procedure could have been derived. As in the normal and gamma examples, a sequence $\{c_i\}$ would allow us to treat this continuous case as we did the discrete case, using lemma 5) to show the appropriate limits hold.

G. The empirical Bayes problem.

We now consider a modification of our decision problem in which the sequence $\{\theta_i\}$ is not an arbitrary sequence, but is instead a sequence of observations of random variables. If these random variables are independent and identically distributed then the problem has been called the empirical Bayes problem. Many fine articles have been written on this problem and the results obtained have inspired this paper. We shall here, however, consider a more general form of the problem. Instead of assuming the θ_i to be independent observations of a random variable Θ, we assume the sequence $\{\theta_i\}$ to be a realization of a stochastic process $\{\Theta_i: \ i = 1, 2, \dots \}$ which is strictly stationary of order k. In other words for any k positive integers $i_1, i_2, \dots, i_k$ and any positive integer j the k dimensional random vectors $(\Theta_{i_1}, \Theta_{i_2}, \dots, \Theta_{i_k})$ and $(\Theta_{i_1+j}, \Theta_{i_2+j}, \dots, \Theta_{i_k+j})$ are identically distributed. In particular, we suppose that $\forall\ i = k, k+1, \dots$ the vector $(\Theta_{i-k+1}, \dots, \Theta_i)$ has distribution function $G^k(\underline{y}_k)$. Thus if $G^k(\underline{y}_k) = \prod_{\ell=1}^{k} G(y_\ell)$ for some G we would have the standard empirical Bayes case.

If G^k were known and if Θ_i were distributed independently of $(\Theta_1, \Theta_2, \ldots, \Theta_{i-k})$ then the standard Bayes argument would yield $\Lambda = E[\Theta_i | \underline{X}_i^k]$ as an estimate for θ_i which minimizes the expected loss and achieves the Bayes risk $R(G^k)$. Even if Θ_i were not distributed independently of $(\Theta_1, \Theta_2, \ldots, \Theta_{i-k})$ Λ might still be a "good" estimate, and the risk $R(G^k)$ a reasonable risk to attain. We shall show that any procedure which is asymptotically optimal of k^{th} order (derived under the assumption of an arbitrary $\underline{\theta}$) will also achieve asymptotically an average risk less than or equal to $R(G^k)$. To be more precise, we shall show the following:

Let: Ω be a bounded interval of the real line.

$\{\Theta_i, i = 1, 2, \ldots\}$ be a strictly stationary stochastic process of order k.

G^n be the joint distribution function of $(\Theta_1, \ldots, \Theta_n)$ $n = 1, 2, \ldots$

$\mathcal{G}$ be the class of all possible sequences of distribution functions $\{G^n; n = 1, 2, \ldots\}$ such that G^n is the n dimensional marginal distribution obtained from G^{n+1}, G^n satisfies the above definitions, and G^n puts probability one on Ω^n for all n.

$R(\varphi_i, G^i)$ be the risk of using the estimate φ_i for θ_i when the vector $\underline{\Theta}_i$ is distributed according to G^i. This risk depends, of course, on the class $\mathfrak{F} = \{p_\theta(\cdot): \theta\epsilon\Omega\}$.

$R(\underline{\varphi}_n, G^n) = \frac{1}{n} \sum_{i=1}^{n} R(\varphi_i, G^i)$ where G^i is the i dimensional marginal distribution obtained from G^n.

We now state and prove a generalization of a theorem by Samuel [11].

Theorem 4) Let $\mathfrak{F} = \{p_\theta(\cdot): \theta\epsilon\Omega\}$ be a class of distribution functions. Let $\underline{\varphi}^k$ be a decision procedure which is asymptotically optimal of k^{th} order for $\mathfrak{F}$. Then

$$\overline{\lim_{n \to \infty}} R(\underline{\varphi}_n^k, G^n) \leq R(G^k)$$

for all $\{G^n\}\epsilon\mathcal{G}$. If $\underline{\varphi}^k$ is uniformly asymptotically optimal then the above inequality becomes

$$\overline{\lim_{n \to \infty}} \left\{ \sup_{\{G^n\}\epsilon\mathcal{G}} [R(\underline{\varphi}_n^k, G^n) - R(G^k)] \right\} \leq 0 .$$

Proof:

Let $R(\varphi_i^k, \underline{\theta}_i) = E[(\varphi_i^k(\underline{X}_i^k) - \theta_i)^2]$. Then:

$$R(\underline{\varphi}_n^k, G^n) = \frac{1}{n}\sum_{i=1}^{n} R(\varphi_i^k, G^i) = \frac{1}{n}\sum_{i=1}^{n} E[(\varphi_i^k(\underline{X}_i) - \Theta_i)^2]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\{E[(\varphi_i^k(\underline{X}_i) - \Theta_i)^2)|\underline{\Theta}_i]\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[R(\varphi_i^k, \underline{\Theta}_i)] .$$

For the remainder of the proof we shall let $E_1[\cdot]$ represent expectation where $\forall\ i \geq k$ $\Theta_{i-k+1}, \ldots, \Theta_i$ have a priori distribution function G^k, and $E_2[\cdot]$ represent expectation where $\Theta_{i-k+1}, \ldots, \Theta_i$ have a priori distribution function G_n^k, the k^{th} order empirical distribution function generated by $\theta_1, \theta_2, \ldots, \theta_n$. We now let $\Lambda(\underline{X}_j^k)$ be a form of $E_1[\Theta_j|\underline{X}_j^k]$ and $\psi(\underline{X}_j^k)$ be a form of $E_2[\Theta_j|\underline{X}_j^k]$. Then Λ achieves the risk $R(G^k)$ and ψ the risk $R_k(\underline{\theta}_n)$. We observe $\forall\ \underline{\theta}_k$

$$E_1[(\Lambda(\underline{X}_j^k) - \Theta_j)^2|\Theta_j = \underline{\theta}_k] = E_2[(\Lambda(\underline{X}_j^k) - \Theta_j)^2|\Theta_j = \underline{\theta}_k] .$$

We call this common value $L(\underline{\theta}_k)$. Then

$$R_k(\underline{\theta}_n) = E_2[(\psi(\underline{X}_k) - \Theta_k)^2]$$

$$\leq E_2[(\Lambda(\underline{X}_k) - \Theta_k)^2]$$

$$= E_2\{E_2[(\Lambda(\underline{X}_k) - \Theta_k)^2 | \underline{\Theta}_k]\}$$

$$= \frac{1}{n - k + 1} \sum_{i=k}^{n} L(\theta_{i-k+1}, \ldots, \theta_i) .$$

Hence:

$$E_1[R_k(\underline{\Theta}_n)] \leq \frac{1}{n - k + 1} \sum_{i=k}^{n} E_1[L(\Theta_{i-k+1}, \ldots, \Theta_i)]$$

$$= \frac{1}{n - k + 1} \sum_{i=k}^{n} R(G^k) = R(G^k) .$$

But since $\underline{\varphi}^k$ is asymptotically optimal we have

$$E_1 \left\{ \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{i=1}^{n} R(\varphi_i^k, \underline{\Theta}_i) - R_k(\underline{\Theta}_n) \right\} \leq 0$$

and hence, since our losses are bounded,

$$\overline{\lim_{n\to\infty}} \left[\frac{1}{n} \sum_{i=1}^{n} E_1[R(\varphi_i^k, \underline{\Theta}_i)] - E_1[R_k(\underline{\Theta}_n)] \right] \leq 0$$

$$\Rightarrow \overline{\lim_{n\to\infty}} \left[R(\underline{\varphi}_n^k, G^n) - R(G^k) \right] \leq 0 .$$

The proof of the second part of the theorem follows immediately.

Q.E.D.

Corollary.

If in addition to the assumptions of theorem 4) we add the condition that $\forall\, j = k+1, k+2, \ldots$ Θ_j is distributed independently of the vector $\underline{\Theta}_{j-k}$ then the two conclusions of the theorem may be replaced by

$$\lim_{n\to\infty} R(\underline{\varphi}_n^k, G^n) = R(G^k)$$

and

$$\lim_{n\to\infty} R(\underline{\varphi}_n^k, G^n) = R(G^k) \quad \text{uniformly for} \quad \{G_n\} \in \mathcal{A}$$

respectively.

Proof:

To prove both parts of the corollary it is sufficient to show $\underline{\lim}_{n\to\infty} R(\underline{\varphi}_n^k, G^n) \geq R(G^k)$. But since Θ_j is independent of $\underline{\Theta}_{j-k}$, $R(G^k)$ is the minimum risk that can be attained by any estimate of θ_j. Hence $R(\varphi_i^k, G^i) \geq R(G^k)$ for all $i \geq k$. It may be shown that $i < k \Rightarrow R(G^i) \geq R(G^k)$ so that $R(\varphi_i^k, G^i) \geq R(G^i) \geq R(G^k)$ for all $i < k$. Thus $R(\underline{\varphi}_n, G^n) \geq R(G^k)$ for all n so that $\underline{\lim}_{n\to\infty} R(\underline{\varphi}_n, G^n) \geq R(G^k)$ as desired.

Q.E.D.

BIBLIOGRAPHY

[1] Barndorff-Nielsen, O. "On the Empirical Bayes Methodology for Stochastically Dependent Parameters," unpublished notes, 1964.

[2] Hoeffding, W. "Probability Inequalities for Sums of Bounded Random Variables," Journal of the American Statistical Association, 58, 13-30, 1963.

[3] Johns, M. V. Jr. "Non-parametric Empirical Bayes Procedures," Annals of Mathematical Statistics, 28, 649-669, 1957.

[4] ———. "An Empirical Bayes Approach to Non-parametric Two Way Classification." Studies in Item Analysis and Prediction (ed. by Solomon, H.). Stanford University Press, 221-232, 1961.

[5] ———. "The Parametric and Non-parametric Compound Decision Problems in the Sequence Case," unpublished, 1964.

[6] Robbins, H. "An Empirical Bayes Approach to Statistics." Proceedings Third Berkeley Symposium Mathematical Statistics and Probability. University of California Press, 157-163. 1955.

[7] ———. "The Empirical Bayes Approach to Testing Statistical Hypotheses," Review of the International Statistical Institute, 31, 195-208, 1963.

[8] ————. "The Empirical Bayes Approach to Statistical Decision Problems," Annals of Mathematical Statistics, 35, 1-20, 1964.

[9] Samuel, E. "Asymptotic Solutions of the Sequential Compound Decision Problem," Annals of Mathematical Statistics, 34, 1079-1094, 1963.

[10] ————. "An Empirical Bayes Approach to the Testing of Certain Parametric Hypotheses," Annals of Mathematical Statistics, 34, 1370-1385, 1963.

[11] ————. "On the Sequential Compound Decision Problem in the Fixed Case and the Opponent Case," unpublished, 1964.