# ASYMPTOTIC MINIMAX SOLUTIONS OF SEQUENTIAL POINT ESTIMATION PROBLEMS

A. WALD
COLUMBIA UNIVERSITY

## 1. Introduction

Exact minimax solutions for sequential point estimation problems are, in general, very difficult to obtain. As far as the author is aware, such solutions are known, at present, only in two special cases: (1) in estimating the mean of a normal distribution with known variance (see Wolfowitz [1]) and (2) in estimating the mean of a rectangular distribution with unit range (see Wald [2]). The solution in the first case coincides with the classical nonsequential one, while the solution in the second case is truly sequential.

In this note, we shall derive an asymptotic minimax solution for a general class of point estimation problems. The point estimation problem considered here may be stated as follows: Let $\{X_i\}$ $(i = 1, 2, \ldots, \text{ad inf.})$ be a sequence of independently and identically distributed chance variables. Let $F(u|\theta)$ be the common distribution function involving an unknown parameter $\theta$, that is, $Pr\{X < u\} = F(u|\theta)$. We shall assume that $F(u|\theta)$ admits a density function $f(u|\theta)$. A sequential point estimation procedure $T$ can be defined in terms of two sequences of functions $\{\varphi(x_1, \ldots, x_m)\}$ and $\{t(x_1, \ldots, x_m)\}$ $(m = 1, 2, \ldots, \text{ad inf.})$ where $\varphi(x_1, \ldots, x_m)$ can take only the values 0 and 1. The estimation procedure is then given as follows: Let $x_i$ denote the observed value of $X_i$. We continue taking observations as long as $\varphi(x_1, \ldots, x_m) = 0$. At the first time when $\varphi(x_1, \ldots, x_m) = 1$, we stop experimentation and estimate the unknown parameter value by $t(x_1, \ldots, x_m)$. We shall assume that the cost of experimentation is proportional with the number of observations. Let $c$ denote the cost of a single observation and let the loss due to estimating the true parameter value $\theta$ by $t$ be given by $(t - \theta)^2$.

Let $\nu(\theta, T)$ denote the expected number of observations when $\theta$ is the true parameter value and the estimation procedure $T$ is adopted. Furthermore, let $\rho(\theta, T)$ be the expected value of $(t - \theta)^2$ when $\theta$ is true and $T$ is adopted. This expected value is given by

$$(1.1)\quad \rho(\theta, T) = \sum_{m=1}^{\infty} \int_{R_m} [t(x_1, \ldots, x_m) - \theta]^2 f(x_1|\theta) \ldots \times f(x_m|\theta)\, dx_1 \ldots dx_m$$

where $R_m$ is the totality of all sample points $(x_1, \ldots, x_m)$ for which $\varphi_i(x_1, \ldots, x_i) = 0$

This research was done under contract with the Office of Naval Research.

when $i < m$ and $\varphi_m(x_1, \ldots, x_m) = 1$. The risk when $\theta$ is true, $T$ is adopted, and $c$ is the cost of a single observation is then given by

$$(1.2)\qquad r(\theta, T, c) = \rho(\theta, T) + c\nu(\theta, T) .$$

An estimation procedure $T_0$ is said to be a minimax solution for a given value of $c$ if

$$(1.3)\qquad \sup_\theta r(\theta, T_0, c) \leqq \sup_\theta r(\theta, T, c) \text{ for all } T .$$

The symbol $\sup_\theta$ means supremum with respect to $\theta$.

For every positive $c$, let $T_c^0$ be an estimation procedure. We shall say that $T_c^0$ is an asymptotic minimax solution if

$$(1.4)\qquad \lim_{c=0} \frac{\sup_\theta r(\theta, T_c^0, c)}{\inf_T \sup_\theta r(\theta, T, c)} = 1 .$$

The symbol $\inf_T$ means infimum with respect to $T$. Clearly, if $T_c^0$ is an asymptotic minimax solution, for practical purposes $T_c^0$ may be regarded as a minimax solution when $c$ is sufficiently small.

For any relation $H$, the symbol $Pr\{H | \theta\}$ will denote the probability that $H$ holds when $\theta$ is the true value of the parameter. Furthermore, for any chance variable $y$, the symbol $E(y | \theta)$ will denote the expected value of $y$ when $\theta$ is the true parameter value. Let

$$(1.5)\qquad d(\theta) = E\left[\left(\frac{\partial \log f(x | \theta)}{\partial \theta}\right)^2 | \theta\right]$$

and let

$$(1.6)\qquad d_0 = \inf_\theta d(\theta) .$$

The main result of this paper is that under certain regularity conditions the estimation procedures $T_c^0$ and $T_c^1$ are asymptotic minimax solutions where $T_c^0$ and $T_c^1$ are defined as follows:

*Estimation rule* $T_c^0$: Take $N_c$ observations and estimate $\theta$ by the maximum likelihood estimate $\hat{\theta}_{N_c}$ based on the first $N_c$ observations where $N_c$ is the smallest integer $\geqq 1/\sqrt{cd_0}$.

*Estimation rule* $T_c^1$: Stop experimentation for the smallest positive integral value of $n$ for which

$$(1.7)\qquad \frac{1}{n d(\hat{\theta}_n)} - \frac{1}{(n+1) d(\hat{\theta}_n)} \leqq c$$

and estimate $\theta$ by $\hat{\theta}_n$. Here $\hat{\theta}_n$ denotes the maximum likelihood estimate of $\theta$ based on the first $n$ observations.

Although $T_c^0$ and $T_c^1$ both are asymptotic minimax solutions, $T_c^1$ seems to be preferable to $T_c^0$ for small $c$, since

$$1.8)\qquad \lim_{c=0} \frac{r(\theta, T_c^1, c)}{r(\theta, T_c^0, c)} < 1 \text{ for any } \theta \text{ for which}$$

$$d(\theta) > d_0 , \tag{1.9}$$

as will be seen later.

## 2. Regularity assumptions

In what follows, for any chance variable $y$, the symbol $\sigma^2(y|\theta)$ will denote the variance of $y$ when $\theta$ is the true parameter value. The symbol $n$ will be used to denote the number of observations required for the estimation procedure, that is, $n$ is the smallest integer for which $\varphi_n(x_1, \ldots, x_n) = 1$. Wolfowitz [3] has shown that under some weak regularity conditions the following inequality holds for any estimation procedure $T$:

$$\sigma^2 [t(X_1, \ldots, X_n) \mid \theta] \geqq \frac{\left(1 + \dfrac{\partial b(\theta, T)}{\partial \theta}\right)^2}{\nu(\theta, T)\, d(\theta)} \tag{2.1}$$

where

$$b(\theta, T) = E[t(X_1, \ldots, X_n) - \theta | \theta] . \tag{2.2}$$

Since we shall make use of the above inequality, we shall postulate the following assumption:

ASSUMPTION 2.1. *The regularity conditions postulated by Wolfowitz* [3] *to insure the validity of* (2.1) *are fulfilled.*

In addition to the above assumption, we shall make the following assumptions:

ASSUMPTION 2.2. *The domain of $\theta$ is an open (finite or infinite interval) interval of the real axis.*

ASSUMPTION 2.3. *$d(\theta)$ is a continuous function of $\theta$ and there exists a value $\theta_0$ for which $d(\theta_0) = d_0 = \min_\theta d(\theta)$.*

ASSUMPTION 2.4. *For any positive integer $N$ and for any $\theta$ let $Z_N(\theta) = \sqrt{N}(\hat{\theta}_N - \theta)$. The following limit relation holds:*

$$\lim_{N=\infty} Pr\{Z_N(\theta) \sqrt{d(\theta)} < \lambda \mid \theta\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-u^2/2} du$$

*uniformly in $\lambda$ and $\theta$.*

ASSUMPTION 2.5. *$E[Z_N^{2+\delta}(\theta)|\theta]$ is a bounded function of $\theta$ and $N$ for some positive $\delta$.*

It is well known that assumption 2.4 holds under rather general conditions (see, for example, [4, p. 430]). The above assumptions can no doubt be weakened, but for the sake of simplicity the author has not attempted to do so here.

## 3. Proof that $T_c^0$ is an asymptotic minimax solution

It follows from (1.2) and (2.1) that

$$r(\theta, T, c) \geqq b^2(\theta, T) + \frac{\left(1 + \dfrac{\partial b(\theta, T)}{\partial \theta}\right)^2}{\nu(\theta, T)\, d(\theta)} + c\nu(\theta, T) . \tag{3.1}$$

Taking the minimum with respect to $\nu$, we obtain from (3.1)

$$r(\theta, T, c) \geqq b^2(\theta, T) + \frac{2\sqrt{c}\left|1+\frac{\partial b}{\partial \theta}\right|}{\sqrt{d(\theta)}}. \tag{3.2}$$

Consider a fixed finite and closed interval $I$ of the $\theta$-axis. Let $l(I)$ be the length of $I$. If $\frac{\partial b}{\partial \theta} \leqq -\epsilon$ $(0 < \epsilon)$ for all $\theta$ in $I$, then

$$\sup_{\theta \in I} r(\theta, T, c) \geqq \sup_{\theta \in I} \rho(\theta, T) \geqq \sup_{\theta \in I} b^2(\theta, T) \geqq \frac{\epsilon^2 l^2(I)}{4}. \tag{3.3}$$

If $\frac{\partial b}{\partial \theta} > -\epsilon$ for some $\theta$ in $I$, it follows from (3.2) that

$$\sup_{\theta \in I} r(\theta, T, c) \geqq \frac{2\sqrt{c}(1-\epsilon)}{\sqrt{\max_{\theta \in I} d(\theta)}}. \tag{3.4}$$

Let

$$\epsilon(I, c) = \frac{\sqrt{8\sqrt[4]{c}}}{l(I)\sqrt[4]{\max_{\theta \in I} d(\theta)}}. \tag{3.5}$$

Clearly,

$$\frac{\epsilon^2(I, c)\, l^2(I)}{4} = \frac{2\sqrt{c}}{\sqrt{\max_{\theta \in I} d(\theta)}}. \tag{3.6}$$

Since the right hand member of (3.6) is greater than the right hand member of (3.4), it follows from (3.3) and (3.4) that

$$\sup_{\theta \in I} r(\theta, T, c) \geqq \frac{2\sqrt{c}[1-\epsilon(I, c)]}{\sqrt{\max_{\theta \in I} d(\theta)}}. \tag{3.7}$$

Let $\Omega$ denote the whole parameter space. It then follows from (3.7) that

$$\sup_{\theta \in \Omega} r(\theta, T, c) \geqq \sup_{I} \frac{2\sqrt{c}[1-\epsilon(I, c)]}{\sqrt{\max_{\theta \in I} d(\theta)}}. \tag{3.8}$$

Let $\theta_0$ be a value of $\theta$ for which $d(\theta_0) = \min_{\theta} d(\theta) = d_0$. The existence of such a value is postulated in assumption 2.3. Let $I_0$ be the closed interval of length $l_0$ and midpoint $\theta_0$. We then obtain from (3.8)

$$\sup_{\theta \in \Omega} r(\theta, T, c) \geqq \frac{2\sqrt{c}[1-\epsilon(I_0, c)]}{\sqrt{\max_{\theta \in I_0} d(\theta)}}. \tag{3.9}$$

Since by assumption 2.3 the function $d(\theta)$ is continuous in $\theta$, there exists a positive $\delta$, say $\delta_{l_0}$ (depending on $l_0$) such that

$$\max_{\theta \in I_0} d(\theta) \leqq d_0 + \delta_{l_0} \tag{3.10}$$

and

(3.11) $$\lim_{l_0=0} \delta_{l_0} = 0 .$$

We then obtain from (3.9)

(3.12) $$\sup_{\theta \in \Omega} r(\theta, T, c) \geqq \frac{2\sqrt{c}\,[1-\epsilon(I_0, c)]}{\sqrt{d_0+\delta_{l_0}}} .$$

Since the right hand member of (3.12) does not depend on $T$, we obtain

(3.13) $$\inf_T \sup_\theta r(\theta, T, c) \geqq \frac{2\sqrt{c}\,[1-\epsilon(I_0, c)]}{\sqrt{d_0+\delta_{l_0}}} .$$

For fixed $l_0$, we have

(3.14) $$\lim_{c=0} \epsilon(I_0, c) = 0 .$$

Hence, it follows from (3.13) that

(3.15) $$\liminf_{c=0} \frac{\sqrt{d_0+\delta_{l_0}}}{2\sqrt{c}} \inf_T \sup_\theta r(\theta, T, c) \geqq 1 .$$

Since $\delta_{l_0}$ can be made arbitrarily small by choosing $l_0$ sufficiently small, and since $d_0 > 0$, we obtain

(3.16) $$\liminf_{c=0} \frac{\inf_T \sup_\theta r(\theta, T, c)}{2\sqrt{\dfrac{c}{d_0}}} \geqq 1 .$$

We shall now show that

(3.17) $$\lim_{c=0} \frac{\sup_\theta r(\theta, T_c^0, c)}{2\sqrt{\dfrac{c}{d_0}}} = 1 .$$

Clearly, for the estimation procedure $T_c^0$ defined in section 1 we have

(3.18) $$N_c r(\theta, T_c^0, c) = N_c E[(\hat\theta_{N_c} - \theta)^2 \mid \theta] + cN_c^2 .$$

Let $\{\theta_N\}$ ($N = 1, 2, \ldots$, ad inf.) be any sequence of parameter points. It follows from assumption 2.4 that the distribution of $Z_N(\theta_N)\sqrt{d(\theta_N)}$, as $N \to \infty$, converges to the normal distribution with zero mean and unit variance. Hence, the Helly-Bray theorem [7, p. 31] and assumption 2.5 give

(3.19) $$\lim_{N=\infty} E[Z_N^2(\theta_N) \mid \theta_N]\, d(\theta_N) = 1 .$$

From this it follows that

(3.20) $$\lim_{c=0} N_c E(\hat\theta_{N_c} - \theta)^2 = \frac{1}{d(\theta)}$$

uniformly in $\theta$. Hence, because of (3.18), we have

(3.21) $$\lim_{c=0} N_c r(\theta, T_c^0, c) = \frac{1}{d(\theta)} + \frac{1}{d_0}$$

uniformly in $\theta$. Hence,

$$\lim_{c=0} N_c \sup_\theta r(\theta, T_c^0, c) = \sup_\theta \frac{1}{d(\theta)} + \frac{1}{d_0} = \frac{2}{d_0}. \tag{3.22}$$

Since $N_c = 1/\sqrt{cd_0}$, (3.17) is an immediate consequence of (3.22). From (3.16) and (3.17) it follows that

$$\lim_{c=0} \frac{\inf_T \sup_\theta r(\theta, T, c)}{2\sqrt{\frac{c}{d_0}}} = 1. \tag{3.23}$$

Equations (3.17) and (3.23) imply that $T_c^0$ is an asymptotic minimax solution.

## 4. Limiting distribution of the maximum likelihood estimate when the number of observations is determined by a sequential rule

In order to study the risk function associated with the estimation procedure $T_c^1$, it will be necessary to obtain the limiting distribution of $\sqrt{n}\,(\hat{\theta}_n - \theta)$ when $n$ is determined by a sequential rule.

For any positive value $c$, let $\{\varphi_c(x_1, \dots, x_m)\}$ $(m = 1, 2, \dots, \text{ad inf.})$ be a sequence of functions which can take only the values 0 and 1. Let $n_c$ be the smallest positive integer for which

$$\varphi_c(x_1, \dots, x_m) = 0 \text{ for } m < n_c$$

and

$$\varphi_c(x_1, \dots, x_{n_c}) = 1.$$

We shall make the following assumptions:

ASSUMPTION 4.1. *There exist a function $N(c, \theta)$ of $c$ and $\theta$, and a positive function $\epsilon(c)$ of $c$ such that*

$$\lim_{c=0} N(c, \theta) = \infty \text{ uniformly in } \theta, \tag{4.1}$$

$$\lim_{c=0} \epsilon(c) = 0, \tag{4.2}$$

*and*

$$\lim_{c=0} Pr\{N - \epsilon N \leqq n_c < N + \epsilon N \mid \theta\} = 1 \tag{4.3}$$

*uniformly in $\theta$.*

ASSUMPTION 4.2. *The derivatives* $\frac{\partial \log f(x \mid \theta)}{\partial \theta}$ *and* $\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2}$ *exist.*

ASSUMPTION 4.3. *For some positive $\delta$,* $E\left[\left(\frac{\partial \log f(x \mid \theta)}{\partial \theta}\right)^{2+\delta} \mid \theta\right]$ *is a bounded function of $\theta$.*

ASSUMPTION 4.4.

$$E\left[\left(\frac{\partial \log f(x \mid \theta)}{\partial \theta}\right)^2 \mid \theta\right] = d(\theta)$$

*has a positive lower bound and is uniformly continuous in $\theta$.*

For any positive $\rho$, let

$$h(x, \theta, \rho) = \sup_{\theta'} \left| \left(\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2}\right)_{\theta=\theta'} \right| \tag{4.4}$$

where $\theta'$ is restricted to values in the closed interval $|\theta - \rho, \theta + \rho|$. Furthermore, let

$$(4.4a) \qquad h_1(x, \theta, \rho) = \sup_{\theta'} \left[ \left( \frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2} \right)_{\theta = \theta'} \right]$$

and

$$(4.4b) \qquad h_2(x, \theta, \rho) = \inf_{\theta'} \left[ \left( \frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2} \right)_{\theta = \theta'} \right]$$

when $|\theta' - \theta| \leqq \rho$.

ASSUMPTION 4.5. *$E[h(x, \theta, \rho) \mid \theta]$ is a bounded function of $\theta$ for some positive $\rho$ and*

$$\lim_{\rho = 0} E\left[h_i(x, \theta, \rho) \mid \theta\right] = E\left[\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2} \mid \theta\right] = -d(\theta),$$

*uniformly in $\theta$.*

For any $\theta$, any positive integer $m$ and for any positive $\delta$, let $Q_{\theta,m,\delta}$ denote the event that

$$(4.5) \qquad |\hat{\theta}_k - \theta| \leqq \delta \quad \text{for all } k \geqq m.$$

ASSUMPTION 4.6. *For any positive $\delta$, we have*

$$(4.6) \qquad \lim_{m = \infty} Pr\{Q_{\theta, m, \delta} \mid \theta\} = 1$$

*uniformly in $\theta$.*[1]

We shall prove the following theorem:

THEOREM 4.1. *If assumptions 4.1 to 4.6 hold, then*

$$(4.7) \quad \lim_{c = 0} Pr\{\sqrt{n_c}\,(\hat{\theta}_{n_c} - \theta)\sqrt{d(\theta)} < \lambda \mid \theta\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-u^2/2} du$$

*uniformly in $\lambda$ and $\theta$.*

PROOF. Taylor expansion of $\sum_{\alpha=1}^{n_c} \frac{\partial \log f(x_\alpha \mid \theta)}{\partial \theta}$ at $\theta = \hat{\theta}_{n_c}$ gives

$$(4.8) \quad \sum_{\alpha=1}^{n_c} \frac{\partial \log f(x_\alpha \mid \theta)}{\partial \theta} = \sum_{\alpha=1}^{n_c} \frac{\partial \log f(x_\alpha \mid \hat{\theta}_{nc})}{\partial \theta} + (\theta - \hat{\theta}_{n_c}) \sum_{\alpha=1}^{n_c} \frac{\partial^2 \log f(x_\alpha \mid \bar{\theta}_c)}{\partial \theta^2}$$

where $\bar{\theta}_c$ lies between $\hat{\theta}_{n_c}$ and $\theta$ and $\frac{\partial^i \log f(x \mid \theta^*)}{\partial \theta^i}$ denotes the value of $\frac{\partial^i \log f(x \mid \theta)}{\partial \theta^i}$ at $\theta = \theta^*$. Since the first term on the right hand side of (4.8) is zero, we obtain

$$(4.9) \qquad \sum_{\alpha=1}^{n_c} \frac{\partial \log f(x_\alpha \mid \theta)}{\partial \theta} = -(\hat{\theta}_{n_c} - \theta) \sum_{\alpha=1}^{n_c} \frac{\partial^2 \log f(x_\alpha \mid \bar{\theta}_c)}{\partial \theta^2}.$$

[1] This assumption states that the maximum likelihood estimate converges strongly to the true parameter point $\theta$ and that this convergence is uniform in $\theta$. The strong convergence of the maximum likelihood estimate was proved under very general conditions (see, for example, [5] and [6]). The uniformity of this convergence in $\theta$ can also be proved under some slight additional regularity conditions, by making use of a result by Chung [8] concerning the uniformity of the strong law of large numbers.

Hence,

$$(4.10)\qquad \frac{\frac{1}{\sqrt{N}}\left(\sum_{a=1}^{n_c}\frac{\partial \log f(x_a\mid\theta)}{\partial\theta}\right)\sqrt{d(\theta)}}{-\frac{1}{N}\sum_{a=1}^{n_c}\frac{\partial^2\log f(x_a\mid\bar\theta_c)}{\partial\theta^2}} = \sqrt{N}\,(\hat\theta_{n_c}-\theta)\,\sqrt{d(\theta)}\,.$$

Let $\{\eta_i\}$ ($i = 1, 2, \ldots$, ad inf.) be a sequence of positive numbers such that $\lim_{i=\infty} \eta_i = 0$. It follows from assumption 4.6 that there exists a sequence $\{k_i\}$ ($i = 1, 2, \ldots$, ad inf.) of positive integers such that $\lim_{i=\infty} k_i = \infty$ and

$$(4.11)\qquad \lim_{i=\infty} Pr\{Q_{\theta,k_i,\eta_i}\mid\theta\} = 1 \text{ uniformly in } \theta\,.$$

For any positive $k \geqq k_1$, let $\rho_k = \eta_i$ where $i$ is the largest positive integer for which $k \geqq k_i$. Clearly, we have

$$(4.12)\qquad \lim_{k=\infty}\rho_k = 0$$

and

$$(4.13)\qquad \lim_{k=\infty} Pr\{Q_{\theta,k,\rho_k}\mid\theta\} = 1 \text{ uniformly in } \theta\,.$$

It follows from (4.13) and assumption 4.1 that

$$(4.13a)\qquad \lim_{c=0} Pr\{\,|\,\hat\theta_{n_c}-\theta\,| < \rho_{[N-\epsilon N]}\mid\theta\} = 1$$

uniformly in $\theta$. The symbol $[a]$ denotes the smallest integer $\geqq a$. Since $\bar\theta_c$ lies between $\hat\theta_{n_c}$ and $\theta$, the above equation gives

$$(4.14)\qquad \lim_{c=0} Pr\{\,|\,\bar\theta_c-\theta\,|\leqq \rho_{[N-\epsilon N]}\mid\theta\}\text{ uniformly in } \theta\,.$$

Let $n_c^*$ be defined as follows:

$$(4.15)\qquad \begin{aligned} n_c^* &= n_c \text{ when } N-\epsilon N \leqq n_c \leqq N+\epsilon N\\ n_c^* &= [N-\epsilon N] \text{ when } n_c < N-\epsilon N\\ n_c^* &= [N+\epsilon N] \text{ when } n_c > N+\epsilon N\,. \end{aligned}$$

For any sequence $\{u_i\}$ of chance variables the symbol

$$\operatorname*{plim}_{i=\infty}(u_i\mid\theta) = \lambda$$

will mean that $\lim_{i=\infty} Pr\{|u_i-\lambda| > \rho\,|\,\theta\} = 0$ for any $\rho > 0$. It follows immediately from assumption 4.1 that

$$(4.16)\qquad \operatorname*{plim}_{c=0}\left[\frac{1}{\sqrt{N}}\left(\sum_{a=1}^{n_c}\frac{\partial\log f(X_a\mid\theta)}{\partial\theta}-\sum_{a=1}^{n_c^*}\frac{\partial\log f(X_a\mid\theta)}{\partial\theta}\right)\Big|\,\theta\right]=0$$

and

$$(4.17)\qquad \operatorname*{plim}_{c=0}\left[\frac{1}{N}\left(\sum_{a=1}^{n_c}\frac{\partial^2\log f(X_a\mid\bar\theta_c)}{\partial\theta^2}-\sum_{a=1}^{n_c^*}\frac{\partial^2\log f(X_a\mid\bar\theta_c)}{\partial\theta^2}\right)\Big|\,\theta\right]=0$$

uniformly in $\theta$. Clearly,

$$(4.18)\qquad \left|\sum_{\alpha=n_c^*+1}^{[N+\epsilon N]} \frac{\partial^2 \log f(x_\alpha \mid \bar{\theta}_c)}{\partial \theta^2}\right| \leqq \sum_{\alpha=[N-\epsilon N]}^{[N+\epsilon N]} h(x_\alpha, \theta, \rho_{[N-\epsilon N]})$$

holds when $|\bar{\theta}_c - \theta| \leqq \rho_{[N-\epsilon N]}$. Because of (4.12) and assumption 4.1, we have

$$(4.19)\qquad \lim_{c=0} (\rho_{[N-\epsilon N]}) = 0 \text{ uniformly in } \theta .$$

It follows from assumption 4.5 that for some positive $\rho$

$$(4.20)\qquad \lim_{c=0} \frac{1}{N} E\Big[\sum_{\alpha=[N-\epsilon N]}^{[N+\epsilon N]} h(X_\alpha, \theta, \rho) \mid \theta\Big] = 0 \text{ uniformly in } \theta .$$

Hence, because of (4.18), (4.19), and (4.14), we have

$$(4.21)\qquad \operatorname*{plim}_{c=0} \Big[\frac{1}{N} \sum_{\alpha=n_c^*+1}^{[N+\epsilon N]} \frac{\partial^2 \log f(X_\alpha \mid \bar{\theta}_c)}{\partial \theta^2} \Big| \theta\Big] = 0$$

uniformly in $\theta$.

Since

$$E\Big\{\Big[\sum_{\alpha=n_c^*+1}^{[N+\epsilon N]} \frac{\partial \log f(X_\alpha \mid \theta)}{\partial \theta}\Big]^2 \Big| \theta\Big\} = d(\theta) E\{([N+\epsilon N] - n_c^*) \mid \theta\}$$

it follows easily from assumption 4.1 that

$$(4.22)\qquad \operatorname*{plim}_{c=0} \Big[\frac{1}{\sqrt{N}} \sum_{\alpha=n_c^*+1}^{[N+\epsilon N]} \frac{\partial \log f(X_\alpha \mid \theta)}{\partial \theta} \Big| \theta\Big] = 0$$

uniformly in $\theta$. We shall now show that

$$(4.23)\qquad \operatorname*{plim}_{c=0} \Big[\frac{1}{N} \sum_{\alpha=1}^{[N+\epsilon N]} \frac{\partial^2 \log f(X_\alpha \mid \bar{\theta}_c)}{\partial \theta^2} \Big| \theta\Big] = -d(\theta)$$

uniformly in $\theta$.

Clearly,

$$(4.24)\qquad \frac{1}{N} \sum_{\alpha=1}^{[N+\epsilon N]} h_2(x_\alpha, \theta, \rho_{[N-\epsilon N]}) \leqq \frac{1}{N} \sum_{\alpha=1}^{[N+\epsilon N]} \frac{\partial^2 \log f(x_\alpha \mid \bar{\theta}_c)}{\partial \theta^2}$$
$$\leqq \frac{1}{N} \sum_{\alpha=1}^{[N+\epsilon N]} h_1(x_\alpha, \theta, \rho_{[N-\epsilon N]})$$

whenever $|\bar{\theta}_c - \theta| \leqq \rho_{[N-\epsilon N]}$. Hence, because of (4.14), equation (4.23) is proved if we can show that

$$(4.25)\qquad \operatorname*{plim}_{c=0} \Big[\frac{1}{N} \sum_{\alpha=1}^{[N+\epsilon N]} h_i(X_\alpha, \theta, \rho_{[N-\epsilon N]} \mid \theta\Big] = -d(\theta)$$

uniformly in $\theta$. But this follows from (4.19) and assumption 4.5. Thus, (4.23) is proved.

We obtain from (4.10), (4.16), (4.17), (4.21), (4.22), and (4.23) that

$$\frac{\frac{1}{\sqrt{N}}\left(\sum_{\alpha=1}^{N+\epsilon N}\frac{\partial \log f(x_\alpha \mid \theta)}{\partial \theta}\right)\sqrt{d(\theta)}+\xi_c}{d(\theta)+\zeta_c}=\sqrt{N}\,(\hat{\theta}_{n_c}-\theta)\,\sqrt{d(\theta)} \tag{4.26}$$

where

$$\operatorname*{plim}_{c=0}\,(\xi_c \mid \theta)=\operatorname*{plim}_{c=0}\,(\zeta_c \mid \theta)=0 \tag{4.27}$$

uniformly in $\theta$.

It follows from assumption 4.3 and the central limit theorem that

$$\lim_{c=0} Pr\left\{\frac{1}{\sqrt{N}}\sum_{\alpha=1}^{[N+\epsilon N]}\frac{\partial \log f(X_\alpha \mid \theta)}{\partial \theta}\,\frac{1}{\sqrt{d(\theta)}}<v \,\Big|\, \theta\right\} =\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{v} e^{-u^2/2}du \tag{4.28}$$

uniformly in $v$ and $\theta$.

Since according to assumption 4.4 $d(\theta)$ has a positive lower bound, theorem 4.1 follows easily from (4.26), (4.27), (4.28) and assumption 4.1.

## 5. Proof that $T_c^1$ is an asymptotic minimax solution and that (1.8) holds

Assumptions 4.2 to 4.6 are assumptions concerning $f(x \mid \theta)$ only. If these assumptions hold, it is not difficult to verify that assumption 4.1 is fulfilled for the sequential procedure $T_c^1$, where $N(c, \theta) = 1/\sqrt{cd(\theta)}$. In fact, it follows from the boundedness of $d(\theta)$ that

$$\lim_{c=0} n_c=\infty\,. \tag{5.1}$$

From this and assumption 4.6 it follows that for any $\delta > 0$

$$\lim_{c=0} Pr\left\{n_c \text{ is included in}\left[\inf_{|\theta'-\theta|\leq\delta}\left(\frac{1}{\sqrt{cd(\theta')}}-1\right),\ \sup_{|\theta'-\theta|\leq\delta}\left(\frac{1}{\sqrt{cd(\theta')}}+1\right)\right]\,\Big|\,\theta\right\}=1 \tag{5.2}$$

uniformly in $\theta$. Assumption 4.1 is a simple consequence of (5.1), (5.2) and assumption 4.4. Hence, theorem 4.1 yields

$$\lim_{c=0} Pr\left\{\sqrt{n_c}\,(\hat{\theta}_{n_c}-\theta)\,\sqrt{d(\theta)}<\lambda \mid \theta\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\lambda} e^{-u^2/2}du \tag{5.3}$$

uniformly in $\lambda$ and $\theta$. Clearly,

$$N(c, \theta)\, r(\theta, T_c^1, c) = N(c, \theta)\, E[(\theta_{n_c}-\theta)^2 \mid \theta] + N(c, \theta)\, cE(n_c \mid \theta)\,. \tag{5.4}$$

We shall make the additional assumption:

ASSUMPTION 5.1. $[N(c, \theta)]^{1+\delta/2}E[(\hat{\theta}_{n_c}-\theta)^{2+\delta} \mid \theta]$ *is a bounded function of* $c$ *and* $\theta$ *for some* $\delta > 0$.

Since

$$\operatorname{plim}\left(\frac{n_c}{N(c, \theta)} \,\Big|\, \theta\right)=1 \tag{5.5}$$

uniformly in $\theta$, it follows from (5.3) and assumption 5.1 that

$$\lim_{c=0} N(c, \theta) E[(\hat{\theta}_{n_c} - \theta)^2 | \theta] = \frac{1}{d(\theta)} \tag{5.6}$$

uniformly in $\theta$. Furthermore, it can easily be seen that

$$\lim_{c=0} N(c, \theta)\, cE(n_c | \theta) = \frac{1}{d(\theta)} \tag{5.7}$$

uniformly in $\theta$. Hence

$$\lim_{c=0} N(c, \theta)\, r(\theta, T_c^1, c) = \frac{2}{d(\theta)} \tag{5.8}$$

uniformly in $\theta$, or

$$\lim_{c=0} \frac{r(\theta, T_c^1, c)}{2\sqrt{\frac{c}{d(\theta)}}} = 1 \tag{5.9}$$

uniformly in $\theta$. This and (3.21) show that the following theorem holds:

THEOREM 5.1. *If assumptions* 4.2 *to* 4.6 *and* 5.1 *hold, and if* (3.21) *and* (3.23) *hold, then* $T_c^1$ *is an asymptotic minimax solution and* (1.8) *holds.*

Let $T_c^2$ be the estimation procedure defined as follows: Take first $m_c$ observations where $m_c = [1/\sqrt{cd_1}]$ and $d_1$ is the least upper bound of $d(\theta)$ with respect to $\theta$. Then take $n_c - m_c$ additional observations where $n_c = [1/\sqrt{cd(\hat{\theta}_{m_c})}]$. Estimate $\theta$ by $\theta_{n_c}$.

One can show in a similar way that if assumptions 4.2 to 4.6 hold, and if assumption 5.1 remains valid when $T_c^1$ is replaced by $T_c^2$, then

$$\lim_{c=0} \frac{r(\theta, T_c^2, c)}{2\sqrt{\frac{c}{d(\theta)}}} = 1 \tag{5.10}$$

uniformly in $\theta$. Thus, because of (5.9), we have

$$\lim_{c=0} \frac{r(\theta, T_c^2, c)}{r(\theta, T_c^1, c)} = 1 \tag{5.11}$$

uniformly in $\theta$.

## REFERENCES

[1] J. WOLFOWITZ, "Minimax estimates of the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 218–230.

[2] A. WALD, *Statistical Decision Functions*, Wiley, New York, 1950.

[3] J. WOLFOWITZ, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 215–230.

[4] A. WALD, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.

[5] J. L. DOOB, "Probability and statistics," *Trans. Amer. Math. Soc.*, Vol. 36 (1934), pp. 759–775.

[6] A. WALD, "Note on the consistency of the maximum likelihood estimate," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 595–601.

[7] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, Princeton, 1946.

[8] K. L. CHUNG, "The strong law of large numbers," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951, pp. 341–352.