71-29,169

FU, James Chuan, 1937-
ON SOME LARGE-SAMPLE METHODS IN STATISTICAL POINT ESTIMATION.

The Johns Hopkins University, Ph.D., 1971
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

ON SOME LARGE-SAMPLE METHODS IN STATISTICAL POINT ESTIMATION

by

James C. Fu

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

1971

TABLE OF CONTENTS

Abstract

Let X be a sample sapce, Θ be a parameter space which is an open interval of R^1, $\{X_n\}$ be a sequence of i.i.d. random variables with common density $f(x|\theta)$. An estimator θ_n^* is called a maximum probability estimator of θ with respect to the (prior) density $g(\theta)$ of a positive measure on Θ if $\prod_{i=1}^{n} f(x_i|\theta_n^*(x_1,\ldots,x_n)g(\theta_n^*(x_1,\ldots,x_n)) = \max_{\theta\in\Theta} \prod_{i=1}^{n} f(x_i|\theta)g(\theta)$ for all $n=1,2,\ldots$. In this thesis, we prove, under certain regularity conditions, that the estimator θ_n^* is asymptotically efficient in the sense that $\lim_{\epsilon\to 0}\lim_{n\to\infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|\theta_n^*-\theta| \geq \epsilon\} = -\frac{I(\theta)}{2}$, where $I(\theta)$ if Fisher's information. This proof also gives a direct method of verifying Bahadur's result [Ann. Math. Statist. 38 303-324, 1967] that the maximum likelihood estimator $\hat{\theta}_n$ is asymptotically efficient in the above sense. The existence, consistency and asymptotic distribution of the estimator θ_n^* are also discussed (for both one and k-dimensional parameters θ). Let T_n be any consistent estimator. For fixed $\epsilon > 0$, several inequalities for the quantity $\underline{\lim}_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\}$ are obtained. An upper bound for a measure of asymptotic relative efficiency, which compares performance of the sample mean with respect to the sample median when estimating the location parameter of a symmetric distribution is also obtained.

ACKNOWLEDGEMENT

The author is grateful to his advisor Professor Leon J. Gleser for his generous advice, continuous encouragement and support through all three years of the author's studies at The Johns Hopkins University, and for his constructive comments, valuable suggestions and help in correcting the author's English. His aid has made this thesis possible. The author also expresses his gratitude to Professor S. Schach for his valuable suggestions and careful reading of the dissertation; to Dennis Dixon, David Jacobs, M. Pagano, and J. Terry Smith for their help and friendship; to all of the faculty and staff at the Department of Statistics, The Johns Hopkins University.

The author is deeply thankful to his wife Wen-how for her constant encouragement and moral support and especially for delaying her Ph.D. program at Cornell University in order to support the author in his graduate study at The Johns Hopkins University. He also gives special thanks to his parents for their love and encouragement, and for making his education possible.

Finally the author wishes to express his appreciation to Mrs. Dorothy Hennegan and Mrs. Nancy Topper for their skillful typing of this dissertation.

1. Introduction

In a statistical point estimation problem, the goal is to use information obtained from a sample of observations to estimate an unknown parameter (or a function of an unknown parameter) of the probability distribution which governs the variability of sampling. Ideally, we would like to construct an estimator which has the property that with probability one a correct estimate is made of the true value θ of the unknown parameter. This goal is, of course, met only in quite trivial and meaningless cases. More realistically, we hope to find an estimator having highest possible probability of being "close" to θ.

1. A General Method for Constructing Point Estimators

Suppose that we have independent and identically distributed (i.i.d.) random observations $X_1, X_2, \ldots X_n$ having common distribution P_θ. Here we assume that each X_i is defined on a measure space $(\mathcal{X}, \beta)$, where $\mathcal{X}$ is any topological space, and β is a sigma-field of measurable sets. We assume that P_θ is a member of a class $\{P_\theta, \theta \in \Theta\}$ indexed by a point θ in k-dimensional Euclidean space, and that Θ is a subset of k-dimensional Euclidean space.

Further, we assume that the class $\{P_\theta, \theta \in \Theta\}$ is dominated by a σ-finite measure μ defined on β, so that each P_θ has a density (Radon-Nikodym derivative) $f(X|\theta) = dP_\theta/d\mu$ with respect to μ. The sample $X_1, X_2, \ldots, X_n$ is then defined on the Cartesian product space $(X^{(n)}, \beta^{(n)})$ with respect to the product probability measure $P_\theta^{(n)}$ which has density $\prod_{i=1}^{n} f(x_i|\theta)$ with respect to the product measure $\mu^{(n)}$.

With this probability background in mind, we define an estimator T_n of θ based on $X_1, X_2, \ldots, X_n$ to be a measurable function mapping $X^{(n)}$ into Θ. Since Θ is a subset of Euclidean k-dimensional space, we can measure distance by the usual Euclidean distance $|\theta - \theta'|$. One way of defining what we mean when we say that $T_n(X_1, X_2, \ldots, X_n)$ is "close" to θ is to choose a small $\epsilon > 0$, and to say that $T_n(X_1, X_2, \ldots, X_n)$ is "close" to θ if $|T_n(X_1, X_2, \ldots, X_n) - \theta| < \epsilon$. The probability $\alpha(T_n, \epsilon, \theta)$ that T_n is "close" to θ is then

$$\alpha(T_n, \epsilon, \theta) \equiv P_\theta\{|T_n(X_1, X_2, \ldots, X_n) - \theta| < \epsilon\} \quad . \qquad (1.1)$$

Let τ_n be the class of all estimators $T_n: X^{(n)} \to \Theta$. The estimator T_n^* has highest probability of being "close" to θ if

$$\alpha(T_n^*,\epsilon,\theta) = \sup_{T_n \epsilon \tau_n} \alpha(T_n,\epsilon,\theta) \quad . \tag{1.2}$$

We could define T_n^* to be a "best" estimator of θ if (1.2) holds for all $\theta \in \Theta$. Unfortunately, this definition of "best" (although clearly reasonable and meaningful) has the major disadvantage of almost never being satisfied in practical statistical problems. The estimator T_n^* satisfying (1.1) for a given ϵ and θ, may not satisfy (1.2) for other values of ϵ and θ; in other words, the optimal T_n^* under the criterion (1.2) depends on ϵ and θ.

To remove (or account for) the influence of θ in measuring goodness of estimators by means of the quantity $\alpha(T_n,\epsilon,\theta)$ in (1.1), we could consider weighting different values of θ by means of any of a class of measures defined on the Borel subsets of Θ. Assume that we have such a positive measure G which has a Radon-N. Kodym derivative $g(\theta)=dG/d\nu$ with respect to some standard σ-finite measure $\nu(\theta)$ (usually Lebesgue measure or counting measure). With respect to G, we might consider replacing the θ-specific measure (1.1) by the weighted probability

$$\alpha(T_n,\epsilon,g) = \int_{\Theta} \alpha(T_n,\epsilon,\theta)g(\theta)\, d\nu(\theta) \quad . \tag{1.3}$$

we can then define a <u>maximum probability estimator with respect to g and ε</u> to be an estimator T_n^* which maximizes (1.3) over all $T_n \in \tau_n$. If

$$\int_\Theta g(\theta)\, d\nu(\theta) < \infty, \tag{1.4}$$

then a maximum probability estimator with respect to g and ε is a Bayes estimator with respect to $g(\theta)/\int_\Theta g(\theta)\, d\nu(\theta)$ under the loss function

$$L_{\epsilon,n}(\theta,a) = \begin{cases} 1 & , \quad \text{if } |\theta-a| \geq \epsilon, \\ 0 & , \quad \text{if } |\theta-a| < \epsilon, \end{cases} \tag{1.5}$$

where $a \in \Theta$ is the action to estimate θ by a. When Θ is one-dimensional, such a Bayes estimator $T_n^* = T_n^*(X_1,X_2,\ldots,X_n)$ can be shown to satisfy

$$\int_{T_n^*-\epsilon}^{T_n^*+\epsilon} \prod_{i=1}^{n} f(x_i|\theta)g(\theta)d\nu(\theta) = \sup_{\substack{T_n \in \tau_n \\ T_n=T_n(X_1,X_2,\ldots,X_n)}} \int_{T_n-\epsilon}^{T_n+\epsilon} \prod_{i=1}^{n} f(x_i|\theta)g(\theta)d\nu(\theta) \tag{1.6}$$

for all $(X_1,X_2,\ldots,X_n) \in X^{(n)}$. Even if (1.4) does not hold, if

$$\int_{\mathcal{X}^{(n)}} \int_{\Theta} L_{\epsilon,n}(\theta, T_n(x_1,x_2,\ldots,x_n)) \prod_{i=1}^{n} f(x_i|\theta) g(\theta) d\nu(\theta) d\mu^{(n)}(x_1,x_2,\ldots,x_n) \tag{1.7}$$

is finite for some $T_n \epsilon \tau_n$, and Θ is one-dimensional, then the maximum probability estimator T_n^* with respect to g and ϵ satisfies (1.6). In the special case where $G(\theta)$ and $\nu(\theta)$ are Lebesgue measure on one-dimensional Euclidean space (so that $g(\theta) \equiv 1$) and (1.7) is finite for some $T_n \epsilon \tau_n$, we can see some similarity between the maximum probability estimator with respect to $g(\theta) \equiv 1$ and ϵ, the maximum likelihood estimator, and a somewhat different "maximum probability estimator" defined by Weiss and Wolfowitz (1966, 1967, 1968, 1970).

The maximum probability estimator T_n^* with respect to a $g(\theta)$ and a given $\epsilon > 0$ can usually be shown to exist, and is an intuitively meaningful estimator for many statistical problems. In the case when $g(\theta)$ satisfies (1.4) and Θ is one-dimensional, T_n^* can be seen to be the midpoint of a modal interval of length 2ϵ for the posterior distribution

$$h(\theta|x_1,x_2,\ldots,x_n) = \frac{g(\theta) \prod_{i=1}^{n} f(x_i|\theta)}{\int_{\Theta} g(\theta) \prod_{i=1}^{n} f(x_i|\theta) d\nu(\theta)} \quad . \tag{1.8}$$

Although this property of the maximum probability estimator T_n^* with respect to g and ϵ is appealing (and of theoretical importance since it relates point estimation to Bayesian fixed-width confidence intervals), calculation of the estimator is not always a simple task and the estimator may not have a convenient closed mathematical form. Further, the estimator T_n^* depends upon the constant $\epsilon > 0$ (and upon the measurement of distance used). Since which value of $\epsilon > 0$ to use is not always clear in practical statistical problems, we would like to somehow avoid an estimation procedure in which an explicit choice of ϵ must be made. Presumably $\epsilon > 0$ would be a small constant in most cases, so it is reasonable to consider the limit of maximum probability estimators $T_n^*(\epsilon)$ with respect to g and ϵ as ϵ tends to 0. If $T_n^*(\epsilon)$ is unique for each $\epsilon > 0$ and if $T_n^*(\epsilon)$ satisfies (1.6), then $T_n^*(\epsilon) \to \theta_n^*$ as $\epsilon \to 0$, where $\theta_n^* = \theta_n^*(x_1, x_2, \ldots, x_n)$ and

$$\prod_{i=1}^{n} f(x_i|\theta_n^*)g(\theta_n^*) = \sup_{\theta\epsilon\Theta} \prod_{i=1}^{n} f(x_i|\theta)g(\theta), \qquad (1.9)$$

all $x_1, x_2, \ldots, x_n \epsilon X^{(n)}$. When $g(\theta)$ Satisfies (1.4), $\theta_n^*(x_1, x_2, \ldots, x_n)$ is a mode of the posterior density $h(\theta|x_1, x_2, \ldots, x_n)$ defined in (1.8). We call $\theta_n^*(x_1, x_2, \ldots, x_n)$ satisfying (1.9) a <u>maximum probability estimator</u> with respect to $g(\theta)$. In many cases, θ_n^* is as easy or easier to compute than the maximum likelihood estimator $\hat{\theta}_n$. Of course, the maximum likelihood estimator is a maximum probability estimator with respect to $g(\theta) \equiv 1$, all θ .

It is important to note that the estimator θ_n^* is not the same as the maximum probability estimator of Weiss and Wolfowitz (1966, 1967, 1968, 1970). Their estimator is formed from $T_n^*(\epsilon)$ by choosing a particular useful sequence of ϵ-values $\{\epsilon_n\}$, $\epsilon_n \to 0$, chosen to give certain asymptotic properties to the resulting sequence of estimators $T_n^{*\prime}(\epsilon_n)$. Our estimator θ_n^* is formed (in many cases) for <u>fixed</u> n by taking $\lim_{\epsilon\to 0} T_n^*(\epsilon)$, or can be defined directly from (1.9).

In Chapter 2, we consider some of the asymptotic (as $n \to \infty$) properties of the maximum probability estimator θ_n^* with respect to g. In Chapter 2, Section 2, we find regularity conditions under which θ_n^* exists and is measurable $(\mathfrak{X}^{(n)},\beta^{(n)})$. We also prove, under additional regularity conditions, that θ_n^* is a strongly consistent estimator of θ ($\theta_n^* \xrightarrow{a.s.} \theta$). In Chapter 2, Section 3, we show (again under additional regularity conditions and under the assumption that Θ is a subset of the real line) that $\sqrt{n}(\theta_n^*-\theta) \xrightarrow{\mathcal{L}} N(0,1/I(\theta))$, where $I(\theta)$ is Fisher's information, that $\sqrt{n}(\theta_n^* - \hat{\theta}_n) \xrightarrow{p} 0$, and that $n^{\frac{1}{2}-\rho}(\theta_n^* - \hat{\theta}_n) \xrightarrow{a.s.} 0$, all ρ, $0 < \rho < \frac{1}{2}$. This shows that θ_n^* is a best asymptotic normal estimator in Fisher's classical sense. Finally, in Chapter 2, Section 4, we show that θ_n^* is asymptotically efficient in a sense of maximum rate of probability convergence around the true value θ defined by Bahadur (1960, 1967). As a result of the asymptotic convergence and asymptotic efficiency properties of maximum probability estimators shown in Chapter 2, and of the approximate Bayesian character of maximum probability estimators mentioned in the present chapter, we recommend that such estimators be considered and used in practical statistical problems. For large enough n, the choice of the weighting function $g(\theta)$ is irrelevant (provided $g(\theta)$ is positive everywhere and smooth enough - see Chapter 2, Sections 3 and 4). For small n, choice of $g(\theta)$ will influence the resulting estimates. In this dissertation, we do not intend to discuss choice of $g(\theta)$, since we feel that reasonable choices depend both upon the problem and upon prior judgements by the statistician.

2. Comparison of Point Estimators and Measures of Asymptotic Relative Efficiency

Let us now restrict consideration only to a one-dimensional parameter θ. Investigation of maximum probability estimation led us to question the classical criteria for determining the goodness of a point estimator. This classical theory concentrates on the biases and variances of point estimators, rather than the probability that these point estimators are "close" to the true value θ of the parameter. Certainly the latter criterion expresses the desires of the users of point estimators in finite (small) sample situations more closely than does the former criterion. Examples are easy to construct where estimators are close "on the average" (unbiased and with small variance), yet have rather small probability of being within ϵ, $\epsilon > 0$, of the true value θ of the parameter. Perhaps the reasons why classical theory concentrates upon comparison of point estimators by means of biases and variances rather than by probability concentrations about the true θ are that:

(i) Variances and biases are easier mathematically to work with in many problems than are probabilities,

(ii) The intuitive feeling, derived from analogies to mechanical models in physics (center of gravity, inertia), that variances and biases closely reflect probability concentrations.

(iii) The fact that the mean and variance of a normally distributed point estimator completely, determine the probability concentrations - if all estimators were normally distributed and unbiased, the estimator with smallest variance would also have the smallest probability concentration.

(iv) Many theorists argue that squared error loss is a reasonable local approximation to the cost of being in error when estimation is treated from a decision-theoretic point of view.

Since most applied statisticians are not willing to be explicit about their losses (even on an approximate basis), and since points (ii) and (iii) are easily shown to provide disastrously false intuitions for many statistical point estimation problems, it follows that if ways can be found to deal conveniently with probability concentrations in small sample situations, it would be worth while to replace classical criteria for point estimators with criteria based on probability concentrations. It is this view which led us to embark upon the analysis in Section 1 of this Chapter, and to propose maximum probability estimators as point estimators for practical statistical problems.

Large-sample theory for point estimators shows that many, if not most, methods of point estimation that have been proposed lean to consistent asymptotically normal (c.a.n.) estimators. Since probability concentrations about the mean for normal distributions are determined by the variance, it would seem that for c.a.n. estimators the variance $V(\theta)$ of the asymptotic normal distribution would determine the properties of the probability concentration about the true value θ of the parameter in large samples. This argument seems to have been the basis of the classical Fisherian theory of large sample efficiency of point estimators. Unfortunately, there are examples (Basu (1956)) which show that the asymptotic probability concentration of a c.a.n. estimator (sequence of estimators) about θ need not be related to the variance $V(\theta)$ of the limiting normal distribution. Hence, if probability concentration about the true θ is of central interest to us, we must deal with it directly, rather than taking the indirect rout of the classical asymptotic theory. If we do so, there is no reason to restrict attention to c.a.n. estimators.

For a fixed $\epsilon > 0$, Basu (1956) suggests using the rate with which $P_\theta\{|T_n(x_1,x_2,\ldots,x_n)-\theta| \geq \epsilon\}$ tends to zero as $n \to \infty$ to measure the asymptotic accuracy of a consistent sequence $\{T_n\}$ of estimators of θ. However, this rate depends in general not only on θ, but also on the somewhat arbitrary constant $\epsilon > 0$. Further, this criterion is very difficult to apply in most cases.

Bahadur (1960, 1967) modifies Basu's criterion somewhat. He suggests considering the quantity

$$\underline{\lim}_{\epsilon\to\infty}\ \underline{\lim}_{n\to\infty}\ \frac{1}{n\epsilon^2}\log\ P_\theta\{|T_n-\theta| \geq \epsilon\} \tag{2.1}$$

as a measure of the asymptotic performance of a consistent sequence of estimators $\{T_n\}$. Under certain regularity conditions, he shows (1960, 1967) that for any consistent estimator $\{T_n\}$, the quantity (2.1) is bounded below by $-I(\theta)/2$, where $I(\theta)$ is Fisher's information. Further, under much more severe regularity conditions, he shows that

$$\lim_{\epsilon\to\infty}\ \lim_{n\to\infty}\ \frac{1}{n\epsilon^2}\log P_\theta\{|\hat{\theta}_n-\theta| \geq \epsilon\} = -I(\theta)/2, \tag{2.2}$$

thus proving an optimality property for the maximum likelihood estimator $\{\hat{\theta}_n\}$ in terms of the asymptotic measure of probability concentration (2.1). As mentioned already in Section 1, in Chapter 2, Section 4, we prove a similar optimality property for maximum probability estimators $\{\theta_n^*\}$.

In Chapter 3 of this dissertation, we investigate several topics related to Bahadur's measure (2.1). First in Section 2 of Chapter 3, we show that for any consistent sequence of estimators (and under virtually no regularity conditions whatsoever),

$$\varliminf_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} \geq - \inf_{\theta'} \{K(\theta',\theta):|\theta'-\theta| > \epsilon\}, \tag{2.3}$$

where $K(\theta',\theta)$ is the Kullback-Leibler information obtained from the measures $P_{\theta'}$ and P_θ. We also show that if each T_n has a density $f_n(t|\theta)$, all θ, all n, and if for all θ', $\theta \in \Theta$,

$$P_{\theta'}\{\lim_{n\to\infty} \frac{1}{n} \log \frac{f_n(T_n(x_1,x_2,\ldots,x_n)|\theta')}{f_n(T_n(x_1,x_2,\ldots,x_n)|\theta\ )} = R(\theta',\theta|\theta')\} = 1$$

where $R(\theta',\theta|\theta')$ is a constant depending upon θ' and θ, then

$$\varliminf_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} \geq - \inf_{\theta'} \{R(\theta',\theta|\theta'):|\theta'-\theta| > \epsilon\} . \tag{2.4}$$

The result (2.3) enables us to directly prove Bahadur's result (2.2). The result (2.4) enables us to bound a measure of asymptotic relative efficiency for two sequences of consistent estimators $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ which we discuss in Chapter 3, Section 3. This measure of asymptotic relative efficiency is based upon the ratio of quantities $\varliminf_{\epsilon\to\infty} \varliminf_{n\to\infty} (n\epsilon^2)^{-1} \log P_\theta\{|T_n^{(i)}-\theta| \geq \epsilon\}$, $i = 1, 2$. In Section 4 of Chapter 3, we use this measure of asymptotic relative efficiency (and the bound (2.4)) to compare the performances of the sample mean and sample median when estimating the location parameter θ of a symmetric distribution.

2. ASYMPTOTIC PROPERTIES OF A CERTAIN ESTIMATOR RELATED TO THE MAXIMUM LIKELIHOOD ESTIMATOR

1. Introduction.

Let $\mathfrak{X}$ be a sample space (topological space) with a sigma-field $\mathcal{B}$ of subsets, and a σ-finite measure μ defined on $\mathcal{B}$. Let $\{P_\theta: \theta \in \Theta\}$ be a collection of probability measures defined on $\mathcal{B}$ and dominated by μ. Finally, assume that Θ is a subset of k-dimensional Euclidean space.

We observe independent, identically distributed (i.i.d.) random elements $X_1, X_2, \ldots,$ where each X_i has common probability measure P_θ. Let $f(x|\theta)$ be the density (Radon-Nikodym derivative) of P_θ with respect to μ; i.e., for any set $A \in \mathcal{B}$,

$$P_\theta(A) = \int_A f(x|\theta)d\mu(x).$$

By an estimator $T = \{T_n\}$ of θ. we mean a sequence $\{T_n\}$ of real vector-valued measurable functions such that for each n, T_n is a function only of $X_1, X_2, \ldots, X_n$. For simplicity of notation, when we speak of the properties of an estimator $T_n(\xi)$, where $\xi = (X_1, X_2, \ldots)$, we mean the properties of the sequence $\{T_n(\xi)\} = \{T_n(X_1, X_2, \ldots, X_n)\}$. An estimator $T_n(\xi)$ is said to be <u>asymptotically normal</u> if for every $\theta \in \Theta$ there exists a sequence $\{\Sigma_n(\theta)\}$ of $k \times k$ positive semi-definite matrices such that

$$\Sigma_n(\theta)\ (T_n(\xi) - \theta) \xrightarrow{\mathcal{L}} Z \quad \text{as} \quad n \to \infty, \tag{1.1}$$

where $\xrightarrow{\mathcal{L}}$ means convergence in law, and Z has the standard (zero mean vector, identity covariance matrix) k-variate normal distribution. An estimator $T_n(\xi)$ is said to be a <u>consistent estimator</u> of θ if for every $\theta \in \Theta$,

$$T_n(\xi) \xrightarrow{p} \theta, \quad \text{as} \quad n \to \infty, \tag{1.2}$$

where $\xrightarrow{p}$ means convergence in probability. An estimator $T_n(\xi)$ is said to be a <u>strongly consistent estimator</u> of θ if for every $\theta \in \Theta$,

$$T_n(\xi) \xrightarrow{a.s.} \theta \quad \text{as} \quad n \to \infty, \tag{1.3}$$

where $\xrightarrow{a.s.}$ means almost sure convergence.

An estimator $\hat{\theta}_n(\xi)$ is said to be a maximum likelihood estimator (M.L.E.) of θ if the equation

$$\prod_{i=1}^{n} f\left(X_i \mid \hat{\theta}_n(\xi)\right) = \max_{\theta \in \Theta} \prod_{i=1}^{n} f(X_i \mid \theta) \ , \qquad (1.4)$$

is satisfied almost surely for all $n = 1,2,\ldots$. Let ν be any positive σ-finite measure on Θ (a general prior distribution for the parameter), and suppose that ν has a Radon-Nikodym derivative $g = d\nu/d\theta$ with respect to Lebesgue measure on Θ. An estimator $\theta_n^*(\xi)$ is defined to be a maximum probability estimator (M.P.E.) of θ with respect to the prior density g if the equation

$$\prod_{i=1}^{n} f\left(X_i \mid \theta_n^*(\xi)\right) g\left(\theta_n^*(\xi)\right) = \max_{\theta \in \Theta} \prod_{i=1}^{n} f(X_i \mid \theta) g(\theta), \qquad (1.5)$$

is satisfied almost surely (with respect to P_θ) for $n = 1,2,\ldots;$ all $\theta \in \Theta$.

The classical theory of estimation from large samples has been concerned mainly with consistent, asymptotically normal (c.a.n.) estimators, and with the classical criterion of goodness of an estimator based on the asymptotic covariance matrix. However, it is well known that the asymptotic covariance matrix has a very weak relation to the actual concentration (probability concentration) of the distribution of the estimator in the neighborhood of the true value of θ. Basu (1956) gives an example for one-dimensional Θ that brings out the fact that an estimator may not be efficient in the classical sense, but yet may have higher probability concentration than any efficient estimator in a neighborhood of the true value of θ. Bahadur (1960,1967) in the same one-dimensional-Θ context has shown, under certain general regularity conditions, that the M.L.E. $\hat{\theta}_n$ has an optimal rate of convergence to the true value of θ in the sense that for any consistent estimator $T_n(\xi)$,

$$\underline{\lim}_{\epsilon \to 0} \underline{\lim}_{n \to \infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|T_n(\xi) - \theta| \geq \epsilon\} \geq -\frac{I(\theta)}{2},$$

and

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|\hat{\theta}_n(\xi) - \theta| \geq \epsilon\} = -\frac{I(\theta)}{2},$$

where (for one-dimensional θ) $I(\theta)$ is Fisher's information. In words, the above two equations mean that if T_n is any consistent estimator, then $P_\theta\{|T_n(\xi) - \theta| \geq \epsilon\}$ cannot tend to zero at an exponential rate faster than $\exp\{-\frac{1}{2} n\epsilon^2 I(\theta)\}$, while the M.L.E. $\hat{\theta}_n(\xi)$ has probability $P_\theta\{|\hat{\theta}_n(\xi) - \theta| \geq \epsilon\}$ which tends to zero at nearly this optimal rate.

In the present report, we prove that the maximum probability estimator $\theta_n^*(\xi)$ with respect to a prior density $g(\theta)$ is a strongly consistent estimator. In the case where Θ is one-dimensional, we also prove that the M.P.E. $\theta_n^*(\xi)$ is asymptotically normal, asymptotically efficient in the classical sense, and that

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|\theta_n^*(\xi) - \theta| \geq \epsilon\} = -\frac{I(\theta)}{2}$$

for all θ nearly irrespective of the prior density $g(\theta)$.

2. Existence and Consistency.

In this section, we prove that the M.P.E. θ_n^* with respect to a prior density $g(\theta)$ exists and is a strongly consistent estimator of θ. Our exposition of these results breaks naturally into two parts: (i) a proof of the existence and measurability of θ_n^* (Theorem 2.1), and (ii) a proof that $\{\theta_n^*\}$ is strongly consistent (Theorem 2.2).

We start by establishing sufficient conditions for the existence and measurability of θ_n^* for a given sample size $n \geq 1$. We have assumed in Section 1 that we observe i.i.d. observations $X_1, X_2, \ldots, X_n$, and that each X_i takes values in the measure space $(\mathcal{X}, \beta)$ according to the probability measure P_θ, where θ is a point in a subset Θ of Euclidean k-dimensional space. Hence $(X_1, X_2, \ldots, X_n)$ takes values in the product measure space $(\mathcal{X}^{(n)}, \beta^{(n)})$ according to the product probability measure $P_\theta^{(n)}$, $\theta \in \Theta$.

Since Θ is a subset of Euclidean k-dimensional space, its closure $\bar{\Theta}$ can be represented as the union $\bigcup_{t=1}^{\infty} K_t$ of a countable number of compact subsets $\{K_t: \ t = 1,2,\ldots\}$, where for every

$t = 1,2,\ldots,\ K_t \subseteq K_{t+1} \subseteq \bar{\Theta}$. Such a representation is by no means unique, but to prove that one such representation exists, note that the entire k-dimensional space $\mathcal{E}^{(k)}$ can be written as the union of the closed k-dimensional spheres $S_t = \{y\colon y \in \mathcal{E}^{(k)},\ |y| \leq t\}$, $t = 1,2,\ldots$ Let $K_t = \bar{\Theta} \cap S_t$. Then since S_t is compact and $\bar{\Theta}$ is closed, K_t is compact. Further, $\bar{\Theta} = \bigcup_{t=1}^{\infty} (S_t \cap \bar{\Theta}) = \bigcup_{t=1}^{\infty} K_t$ and

$K_t = S_t \cap \bar{\Theta} \subseteq S_{t+1} \cap \bar{\Theta} = K_{t+1}$.

In Section 1, it has been assumed that the collection $\{P_\theta,\ \theta \in \Theta\}$ of probability measures on (X,β) is dominated by the σ-finite measure μ, and that $dP_\theta/d\mu = f(x|\theta)$ is the Radon-Nikodym derivative of P_θ with respect to μ. Hence, the product measure $P_\theta^{(n)}$ has Radon-Nikodym derivative

$$L_n\Big((x_1,x_2,\ldots,x_n)|\theta\Big) = L_n(\underline{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \qquad (2.1)$$

with respect to the product measure $\mu^{(n)}$ on $\left(X^{(n)},\beta^{(n)}\right)$. We have also assumed that there is a non-negative σ-finite measure ν defined on Θ which has a Radon-Nikodym derivative $g(\theta) = d\nu/d\theta$ with respect to k-dimensional Lebsegue measure. Let

$$H_n(\underline{x}|\theta) = H_n\Big((x_1,x_2,\ldots,x_n)|\theta\Big) \equiv g(\theta)L_n(\underline{x}|\theta). \qquad (2.2)$$

We propose the following conditions:

Condition 2.1. For each $\theta \in \Theta$, $L_n(\underline{x}|\theta)$ is $\beta^{(n)}$-measurable (measurable with respect to $\left(X^{(n)},\beta^{(n)}\right)$). 2/

Condition 2.2. For all $\underline{x} \in X^{(n)}$ (except perhaps for a set of $\underline{x}$-values having $\mu^{(n)}$-measure equal to zero), $L_n(\underline{x}|\theta)$ is continuous in θ for all $\theta \in \Theta$, and can be extended to a function which is continuous over all $\theta \in \overline{\Theta}$. That is, for any sequence $\{\theta_j\} \subset \Theta$ such that $\lim_{j\to\infty} \theta_j = \theta_o$ exists, $\lim_{j\to\infty} L_n(\underline{x}|\theta_j)$ exists and has a value dependent on the sequence $\{\theta_j\}$ only through its limit $\theta_o \in \overline{\Theta}$. If $\theta_o \in \Theta$, this limit is $L_n(\underline{x}|\theta_o)$.

2/ We have called $L_n(\underline{x}|\theta)$ the Radon-Nikodym derivative of $P_\theta^{(n)}$ with respect to $\mu^{(n)}$. The Radon-Nikodym theorem asserts that there exists a $\beta^{(n)}$-measurable version of the Radon-Nikodym derivative, but this version of the Radon-Nikodym derivative need not satisfy Conditions 2.2 and 2.4. Condition 2.1 asserts that there is a version satisfying Conditions 2.2 and 2.4 which is also $\beta^{(n)}$-measurable for each $\theta \in \Theta$. (See also Wald (1949, pp. 596-7).

If $\theta_o \epsilon \bar{\Theta} \sim \Theta$, we denote the limit by $L_n(\underline{x}|\theta_o)$ for typographical convenience. Hence, $L_n(\underline{x}|\theta)$ in its extended definition is for all $\underline{x} \epsilon X^{(n)}$ now a continuous function of θ for all $\theta \epsilon \bar{\Theta}$.

Condition 2.3. The prior density $g(\theta)$ is continuous in θ for all $\theta \epsilon \Theta$ and can be extended to a continuous function of θ for all $\theta \epsilon \bar{\Theta}$. Further, $g(\Theta) \equiv \sup_{\theta \epsilon \Theta} g(\theta) < \infty$.

Condition 2.4. There exists an increasing sequence $\{K_t, t=1,2,\ldots.\}$ of compact subsets of $\bar{\Theta}$ such that $\bar{\Theta} = \bigcup_{t=1}^{\infty} K_t$ and $\lim_{t\to\infty} \sup_{\theta \epsilon K_t^c \cap \bar{\Theta}} L_n(\underline{x}|\theta) = 0$ for all $\underline{x} \epsilon X^{(n)}$ (except perhaps for a set of $\underline{x}$-values having $\mu^{(n)}$-measure equal to zero).

Condition 2.1 implies that $L_n(\underline{x}|\theta)g(\theta) = H_n(\underline{x}|\theta)$ is $\beta^{(n)}$-measurable for each $\theta \epsilon \Theta$. Since Θ is dense in $\bar{\Theta}$, it follows that for every $\theta \epsilon \bar{\Theta} \sim \Theta$ there exists a sequence $\{\theta_j\} \subset \Theta$ such that $\lim_{j\to\infty} \theta_j = \theta$. However, conditions 2.2 and 2.3 imply that $\lim_{j\to\infty} H_n(\underline{x}|\theta_j) = H_n(\underline{x}|\theta)$ for (almost) all $x \epsilon X^{(n)}$, so that $H_n(x|\theta)$ is the limit of $\beta^{(n)}$-measurable functions and therefore is $\beta^{(n)}$-measurable. Consequently, Conditions 2.1, 2.2 and 2.3 imply that $H_n(\underline{x}|\theta)$ is $\beta^{(n)}$-measurable for all $\theta \epsilon \bar{\Theta}$.

Conditions 2.2 and 2.3 also imply that for (almost) all $\underline{x} \in X^{(n)}$, $H_n(\underline{x}|\theta)$ is continuous in θ for $\theta \in \bar{\Theta}$.

Lemma 2.1. Let S be a sample space with a sigma field Δ of subsets, and let Y be a compact subset of Euclidean k-dimensional space (a metric space under the usual Euclidean distance $|y-y'|$). Let Γ be the appropriate Borel sigma field of subsets of Y. Let $U(s,y)$ map $S \times Y$ into $[0,\infty)$. If for (almost) every $s \in S$, $U(s,.)$ is a continuous function on Y, and if for every $y \in Y$, $U(.,y)$ is a Δ-measurable function (measurable with respect to (S,Δ)), then

(i) $U(.,.)$ is a $\Delta \times \Gamma$-measurable function,

(ii) $\sup_{y \in Y} U(s.y)$ exists and is a Δ-measurable function,

(iii) there exists a Δ-measurable function $T: S \to Y$ satisfying $U(s, T(s)) = \sup_{y \in Y} U(s,y)$ for (almost) all $s \in S$.

Proof. Since Y is a compact subset of k-dimensional space, Y is bounded. Thus, there exists a k-dimensional cube C containing Y. We may partition C into m^k

subcubes, each subcube having volume m^{-k} times the volume of C. Let $C(p,m)$ be the p-th such subcube formed in this fashion, and let $z(p,m)$ be any fixed point in $C(p,m) \cap Y$ (if $C(p,m) \cap Y = \emptyset$, this does not effect the proof), $p = 1,2,\ldots,m^k$; $m = 1,2,\ldots$. Define $U_m(s,y)$ to be equal to $U\big(s, z(p,m)\big)$ whenever $y \in C(p,m)$, $p = 1,2\ldots,m^k$; $m = 1,2,\ldots$. Then clearly $U_m(s,y)$ is $\Delta \times \Gamma$-measurable. Since $U(s,.)$ is a continuous function on Y for (almost) every $s \in S$, it follows that $\lim_{m\to\infty} U_m(s,y) = U(s,y)$ for (almost) all $s \in S$ and for all $y \in Y$. Hence, $U(s,y)$ is the pointwise limit of $\Delta \times \Gamma$-measurable functions, and therefore $U(.,.)$ is $\Delta \times \Gamma$-measurable, proving (i).

Define t_m: $S \to [0,\infty)$ by

$$t_m(s) = \Big[\int_Y \big(U(s,y)\big)^m dy\Big]^{1/m} \equiv \|U(s,.)\|_m \quad ,$$

for $m = 1,2,\ldots$. Since Y is compact and $U(s,.)$ is continuous on Y for almost all s, it follows that $\|U(s,.)\|_m < \infty$ for almost all $s \in S$. Let

$$t(s) = \sup_{y \in Y} U(s,y) \equiv ||U(s,.)||_{\infty} \ .$$

For each m, $t_m(s)$ is Δ-measurable (Fubini's Theorem), and since $\lim_{m \to \infty} ||U(s,.)||_m = ||U(s,.)||_{\infty}$ for almost all $s \in S$, it follows that $t(s)$ is a pointwise limit of Δ-measurable functions and consequently is Δ-measurable. This verifies (ii). The existence of a Δ-measurable function T: $S \to Y$ which is Δ-measurable and satisfies $U\big(s, T(s)\big) = \sup_{y \in Y} U(s,y)$ for almost all $s \in S$ now follows as a direct consequence of Theorem 2 of Olech [1965]. QED

Lemma 2.2. Let S be a sample space with a sigma field Δ of subsets, and let Y be a closed subset of Euclidean k-dimensional space with the appropriate Borel sigma field Γ of subsets. Let $U(s,y)$ map $S \times Y$ into $[0,\infty)$. If for (almost) every $s \in S$, $U(s,.)$ is a continuous function on Y, and if for every $y \in Y$, $U(.,y)$ is a Δ-measurable function, then for any σ-compact subset F of Y, the restriction of $U(s,y)$ to $S \times F$ is $\Delta \times \Gamma_F$-measurable (where Γ_F is the sub-Borel-field of Γ consisting of subsets of F which belong to Γ) and

$U(s,F) \equiv \sup_{y \in F} U(s,y)$ is Δ-measurable. Further, if there exists an increasing sequence $\{K_t, t=1,2,\ldots\}$ of compact subsets of Y such that $Y = \bigcup_{t=1}^{\infty} K_t$ and

$$\lim_{t\to\infty} \sup_{y \in K_t^c \cap Y} U(s,y) = 0$$

for all $s \in S$, then there exists a Δ-measurable function $T: S \to Y$ for which $\sup_{y \in Y} U(s,y) = U\big(s,T(s)\big)$, all $s \in S$.

Proof. Since F is σ-compact, there exists a sequence $\{K_t: t=1,2,\ldots\}$ of compact subsets of F (and hence of Y) such that $F = \bigcup_{t=1}^{\infty} K_t$. Let $U_t(s,y)$ be equal to $U(s,y)$ on $S \times K_t$ and be equal to 0 otherwise. Since $K_t \subseteq K_{t+1}$, all t, and $F = \bigcup_{t=1}^{\infty} K_t$, it follows that $U(s,y)$ restricted to $S \times F$ is equal to $\lim_{t\to\infty} U_t(s,y)$ for all $(s,y) \in S \times F$. Since $U(s,y)$ restricted to $S \times K_t$ is $\Delta \times \Gamma_{K_t}$-measurable for all $t = 1,2,\ldots$, by Lemma 2.1, it follows that $U_t(s,y)$ is $\Delta \times \Gamma_F$-measurable for each $t = 1,2,\ldots$. Hence, $U(s,y)$ is the limit of $\Delta \times \Gamma_F$ measurable functions, therefore is $\Delta \times \Gamma_F$-measurable. Also, since by Lemma

2.1, $U(s,K_t) \equiv \sup_{y \in K_t} U(s,y)$ exists and is Δ-measurable for all $t = 1,2,\ldots$, and since $F = \bigcup_{t=1}^{\infty} K_t$, then $U(s,F) = \sup_{1 \leq t \leq \infty} U(s,K_t)$ is the supremum of a countable number of Δ-measurable functions, and thus is Δ-measurable.

For future reference, we remark that since Euclidean k-dimensional space $\mathcal{E}^{(k)}$ is itself σ-compact, every open set and every closed set in $\mathcal{E}^{(k)}$ is σ-compact. Hence Y is σ-compact, and $U(s,y) = \sup_{y \in Y} U(s,y)$ exists and is Δ-measurable.

Now assume that $Y = \bigcup_{t=1}^{\infty} K_t$, where $\{K_t, t=1,2,\ldots\}$ is an increasing sequence of compact subsets of Y and

$$\lim_{t \to \infty} \sup_{y \in K_t^c \cap Y} U(s,y) = 0 \quad , \quad \text{all } s \in S. \tag{2.3}$$

By Lemma 2.1, for each $t = 1,2,\ldots$, $U(s,K_t) \equiv \sup_{y \in K_t} U(s,y)$ exists and is Δ-measurable. Also by Lemma 2.1, for each $t = 1,2,\ldots$, there exists a Δ-measurable function $T_t\colon\ S \to K_t$ satisfying

$$U(s,K_t) = U\Big(s,T_t(s)\Big)\,, \quad \text{all} \quad s \in S. \tag{2.4}$$

Let $T_1^*(s) = T_1(s)$, all $s \in S$, and let

$$T_t^*(s) = \begin{cases} T_{t-1}(s)\,, & \text{if} \quad U(s,Y) = U(s,K_t) = U(s,K_{t-1})\,, \\ T_t(s)\,, & \text{if} \quad U(s,Y) = U(s,K_t) > U(s,K_{t-1})\,, \\ y_o\,, & \text{if} \quad U(s,Y) > U(s,K_t)\,, \end{cases}$$

where y_o is some fixed point in Y. Since $U(s,K_t)$ is non-decreasing in t for each fixed $s \in S$, $T_t^*(s)$ is well-defined for all $s \in S$. By the Δ-measurability of $T_t(s)$ and $U(s,K_t)$, all $t = 1,2,\ldots$, by the Δ-measurability of $U(s,y)$, and from the definition of $T_t^*(s)$, it follows that $T_t^*(s)$ is Δ-measurable for all $t = 1,2,\ldots$. Further, from the fact that $T_t(s) \in K_t \subseteq Y$, all $s \in S$, all $t = 1,2,\ldots$, it follows that $T_t^*(s) \in Y$, all $s \in S$, all $t = 1,2,\ldots$. Finally, for every $s \in S$, it follows from (2.3), and the fact that for all $s \in S$, $t = 1,2,\ldots$,

$$U(s|Y) = \max\{U(s|K_t), U(s,K_t^c \cap Y)\}\,,$$

that for every $s \in S$, there exists integer $t_o = t_o(s) \geq 1$ such that for all $t \geq t_o$,

$$T_t^*(s) = T_{t_o}^*(s) \ . \tag{2.5}$$

Hence $\lim_{t\to\infty} T_t^*(s)$ exists in Y for each $s \in S$. Let $T(s) = \lim_{t\to\infty} T_t^*(s)$. Then $T\colon\ S \to Y$ is the pointwise limit of a sequence of Δ-measurable functions and hence is Δ-measurable. From (2.5) it follows that for each $s \in S$, there exists $t_o(s) \geq 1$ such that for all $t \geq t_o(s)$,

$$U(s,Y) = U(s,K_t) = U(s,K_{t_o(s)}) = U(s,T_t^*(s)).$$

Hence, since $U(s,y)$ is continuous in y for all $y \in Y$,

$$U(s,Y) = \lim_{t\to\infty} U\Big(s,T_t^*(s)\Big) = U\Big(s,T(s)\Big) \ ,$$

for all $s \in S$. This completes the proof. Q.E.D.

Let

$$\begin{aligned} H_n(\underline{x}|\overline{\Theta}) &= \sup_{\theta\in\overline{\Theta}} H_n(\underline{x}|\theta) \ , \\ H_n(\underline{x}|\Theta) &= \sup_{\theta\in\Theta} H_n(\underline{x}|\theta). \end{aligned} \tag{2.6}$$

By Lemma 2.2, $H_n(\underline{x}|\overline{\Theta})$ is $\beta^{(n)}$-measurable. Further, since $H_n(\underline{x}|\theta)$ is, for each $\underline{x} \in X^{(n)}$, continuous in θ for all $\theta \in \overline{\Theta}$,

$$H_n(\underline{x}|\Theta) = H_n(\underline{x}|\overline{\Theta}) \qquad \text{all} \quad \underline{x} \in X^{(n)},$$

so that $H_n(\underline{x}|\Theta)$ is also $\beta^{(n)}$-measurable.

<u>Theorem 2.1</u>. If for a given integer $n \geq 1$, Conditions 2.1 through 2.4 hold, there exists a $\beta^{(n)}$-measurable function $\theta_n^*(\underline{x})$: $X^{(n)} \to \overline{\Theta}$ satisfying

$$H_n(\underline{x}|\Theta) = H_n(\underline{x}|\overline{\Theta}) = H_n(\underline{x}|\theta_n^*(\underline{x}))$$

for (almost) all $\underline{x} \in X^{(n)}$. That is, for that n, the M.P.E. $\theta_n^*(\xi)$ exists and is $\beta^{(n)}$-measurable.

<u>Proof</u>. Since by Conditions 2.1 through 2.3, $H_n(\underline{x}|\theta)$ is $\beta^{(n)}$-measurable in $\underline{x}$ for each fixed θ, and continuous in θ for each fixed $\underline{x} \in X^{(n)}$, and since by Condition 2.4, there exists an increasing sequence $\{K_t:\ t=1,2,\ldots\}$ of compact subsets of $\overline{\Theta}$ such that $\overline{\Theta} = \bigcup_{t=1}^{\infty} K_t$ and

$\lim_{t\to\infty} \sup_{\theta \in K_t^c \cap \overline{\Theta}} H_n(\underline{x}|\theta| = 0$ for (almost) all $\underline{x} \in X^{(n)}$, the asserted result is a direct consequence of Lemma 2.2. Q.E.D.

Before turning to a proof of strong consistency, we note that if for some $n_o \geq 1$, Conditions 2.1, 2.2, and 2.4 are satisfied, then since

$$\prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \Big(\prod_{j \neq i} f(x_j|\theta)\Big)^{1/n-1} ,$$

it follows by induction that Conditions 2.1, 2.2, and 2.4 hold for all $n \geq n_o$, and thus that $\theta_n^*(\underline{x})$ exists and is $\beta^{(n)}$-measurable for all $n \geq n_o$.

For a proof of the strong consistency of θ_n^*, we add four additional conditions. In what follows, N equals a fixed integer n for which Conditions 2.1, 2.2, and 2.4 hold simultaneously.

Condition 2.5. For all $\theta \in \Theta$, $g(\theta) > 0$.

Condition 2.6. If θ and θ', $\theta \neq \theta'$, are two distinct points in $\bar{\Theta}$, we have

$$\mu^{(N)}\Big(\{\underline{x}: \ \underline{x} \in X^{(N)}, \ L_n(\underline{x}|\theta) \neq L_N(\underline{x}|\theta')\}\Big) > 0 .$$

Condition 2.7. For every $\theta \in \bar{\Theta} \sim \Theta$,

$$\int_{\mathfrak{X}^{(N)}} L_N(\underline{x}|\theta)d\mu(\underline{x}) \leq 1 .$$

Condition 2.8. For every $\theta \in \Theta$,

$$E_\theta \log \left[\frac{L_N\big((X_1,X_2,\ldots,X_N)|\ \Theta\big)}{L_N\big((X_1,X_2,\ldots,X_N)|\ \theta\big)} \right] < \infty .$$

It is important to note that we do not require that Conditions 2.6, 2.7 and 2.8 hold for the smallest $n = n_o$ such that Conditions 2.1, 2.2, and 2.4 hold simultaneously, but only that these conditions hold for some N greater than or equal to that smallest n.

For $n \geq N$ and for fixed $\underline{x} = (x_1,x_2,\ldots,x_n) \in \mathfrak{X}^{(n)}$, define the sets

$$\begin{aligned} \Lambda_n(\underline{x}) &= \{\theta:\ \theta \in \Theta,\ H_n(\underline{x}|\theta) = H_n(\underline{x}|\Theta)\} , \\ \Lambda_n^*(\underline{x}) &= \{\theta:\ \theta \in \Theta,\ H_n(\underline{x}|\theta) \geq \tfrac{1}{2} H_n(\underline{x}|\ \Theta)\} . \end{aligned} \tag{2.7}$$

Note that for all $n \geq N$, and all $\underline{x} \in X^{(n)}$,

$$\Lambda_n(\underline{x}) \subset \Lambda_n^*(\underline{x}) \ . \tag{2.8}$$

Finally for any $\theta' \in \overline{\Theta}$, any $\epsilon > 0$, let $N(\theta', \epsilon)$ denote the set

$$N(\theta', \epsilon) = \{\theta'' : \ \theta'' \in \overline{\Theta}, \ |\theta' - \theta''| < \epsilon\} \ . \tag{2.9}$$

Lemma 2.3. Assume that Conditions 2.1 through 2.8 hold. Then for any $\epsilon > 0$ and for each given $\theta \in \Theta$,

$$P_\theta^{(\infty)}\{\Lambda_n^*(\xi) = \Lambda_n^*\big((X_1, X_2, \ldots, X_n)\big) \subset N(\theta, \epsilon)$$
$$\text{for all sufficiently large } n\} = 1 \ . \tag{2.10}$$

Proof. Let $\theta \in \Theta$ be given. Let $\epsilon > 0$ be fixed. Let $\{K_t, t=1,2,\ldots\}$ be the increasing sequence of compact sets guaranteed by Condition 2.4. There exists a large enough integer t_o so that $N(\theta, \epsilon) \subset K_{t_o}$. Since for any $t \geq 1$, $K_t(\theta, \epsilon) = K_t \sim N(\theta, \epsilon)$ is the intersection of the compact set K_t and the closed set

$N^c(\theta,\epsilon)$, $K_t(\theta,\epsilon)$ is compact.

Consider any $\theta' \epsilon K_t(\theta,\epsilon)$, for fixed $t \geq t_o$. For all $\delta > 0$,

$$\sup_{\theta''\epsilon N(\theta',\delta)} L_N(\underline{x}|\theta'') = L_N\left(\underline{x}|N(\theta',\delta)\right)$$

is an $\beta^{(N)}$-measurable function. This assertion follows since $L_N(\underline{x}|\theta)$ is $\beta^{(N)}$-measurable for all $\theta \epsilon \bar{\Theta}$ (follows from Conditions 2.1 and 2.2), since $L_N(\underline{x}|\theta)$ is for every fixed $\underline{x} \epsilon X^{(N)}$ continuous in θ for all $\theta \epsilon \bar{\Theta}$ (Condition 2.2), and from Lemma 2.2. Since $L_N(\underline{x}|\theta)$ is continuous in θ,

$$\lim_{\delta\to 0} L_N\left(\underline{x}|N(\theta',\delta)\right) = L_N(\underline{x}|\theta')$$

for all $\underline{x} \epsilon X^{(N)}$. Hence.

$$\lim_{\delta\to 0} \log\left[\frac{L_N\left(\underline{x}|N(\theta',\delta)\right)}{L_N(\underline{x}|\theta)}\right] = \log\left[\frac{L_N(\underline{x}|\theta')}{L_N(\underline{x}|\theta)}\right]. \qquad (2.11)$$

Since for all δ,δ', $0 < \delta' < \delta$, and all $\underline{x} \in X^{(N)}$,

$$\log\left[\frac{L_N\left(\underline{x}\,|\,N(\theta',\delta')\right)}{L_N(\underline{x}\,|\,\theta)}\right] \leq \log\left[\frac{L_N\left(\underline{x}\,|\,N(\theta',\delta)\right)}{L_N(\underline{x}\,|\,\theta)}\right]$$

$$\leq \log\left[\frac{L_N(\underline{x}\,|\,\Theta)}{L_N(\underline{x}\,|\,\theta)}\right] . \tag{2.12}$$

and since $\log[L_N(\underline{x}\,|\,\Theta) \,/\, L_N(\underline{x}\,|\,\theta)] \geq \log 1 = 0$ for all $\underline{x} \in X^{(N)}$, it follows from Condition 2.8, (2.11), and from the Lebesgue Monotone Convergence Theorem that

$$\lim_{\delta \to 0} E_\theta \log\left[\frac{L_N\left(\underline{X}\,|\,N(\theta',\delta)\right)}{L_N(\underline{X}\,|\,\theta)}\right] = E_\theta \log\left[\frac{L_N(\underline{X}\,|\,\theta')}{L_N(\underline{X}\,|\,\theta)}\right] . \tag{2.13}$$

But by Jensen's inequality,

$$E_\theta \log\left[\frac{L_N(\underline{X}\,|\,\theta')}{L_N(\underline{X}\,|\,\theta)}\right] < \log E_\theta \left[\frac{L_N(\underline{X}\,|\,\theta')}{L_N(\underline{X}\,|\,\theta)}\right]$$

$$= \log \int_{X^{(N)}} L_N(\underline{x}\,|\,\theta')d\mu^{(N)}(x) \geq \log 1 = 0, \tag{2.14}$$

where the strictness of the first inequality in (2.14) follows from the strict concavity of the logarithm and

Condition 2.6, and where the last inequality follows from the fact that $L_N(\underline{x}|\theta')$ is a density function on $\mathcal{X}^{(N)}$ for $\theta' \in \Theta$ and from Condition 2.7. Hence, we conclude from (2.14) that given $\theta' \in K_t\ (\theta, \epsilon)$, there exists $\delta(\theta')$, $\eta(\theta') > 0$ such that

$$E_\theta \log \left[\frac{L_N(\underline{X}|N(\theta', \delta(\theta', \eta)))}{L_N(\underline{X}|\theta)} \right] < -\ \eta(\theta')\ . \tag{2.15}$$

Let t be any integer $\geq t_o$. Since $K_t(\theta,\ \epsilon)$ is compact, there exist $\theta_1, \theta_2, \ldots, \theta_s \in K_t(\theta, \epsilon)$ such that $K_t(\theta, \epsilon) \subset \bigcup_{i=1}^{s} N\left(\theta_i,\ \delta(\theta_i)\right)$. Hence since

$$\begin{aligned} L_n\left(\underline{x}|K_t(\theta, \epsilon)\right) &\equiv \sup_{\theta' \epsilon K_t(\theta, \epsilon)} L_N(\underline{x}|\theta') \\ &\leq \max_{1 \leq i \leq k} L_N\left(\underline{x}\,|N(\theta_i,\ \delta(\theta_i))\right)\ , \end{aligned}$$

it follows that

$$E_\theta \log \left[\frac{L_N\left(\underline{X}|K_t(\theta, \epsilon)\right)}{L_N(\underline{X}|\theta)} \right] < \max_{1 \leq i \leq s} -\eta(\theta_i) \equiv -\eta < 0\ . \tag{2.16}$$

Consider

$$L_N(\underline{x}|K_t^c) = \sup_{\theta' \epsilon K_t^c \cap \bar{\Theta}} L_N(\underline{x}|\theta'). \tag{2.17}$$

Since $K_t^c \cap \bar{\Theta}$ is σ-compact, it follows from Lemma 2.2 that $L_N(\underline{x}|K_t^c)$ is $\beta^{(N)}$-measurable for each $t = 1,2,\ldots$. By Condition 2.4, for all $\underline{x} \epsilon X^{(N)}$

$$\lim_{t\to\infty} L_N(\underline{x}|K_t^c) = 0 . \tag{2.18}$$

Further for each $\underline{x} \epsilon K_t^c$, $L_N(\underline{x}|K_t^c)$ is decreasing in t and $L_N(\underline{x}|K_t^c) \leq L_N(\underline{x}|\Theta)$. From the Lebesgue Monotone Convergence Theorem and Condition 2.8, we can thus conclude that

$$\lim_{t\to\infty} E_\theta \log \left[\frac{L_N(\underline{X}|K_t^c)}{L_N(\underline{X}|\theta)} \right] = -\infty .$$

Hence, we can certainly choose $t \geq t_o$ large enough so that for the fixed $\eta > 0$ given in (2.16),

$$E_\theta \log \left[\frac{L_N(\underline{X}|K_t^c)}{L_N(\underline{X}|\theta)} \right] < -\eta . \tag{2.19}$$

Let

$$L_N\left(\underline{x}\,|\,N^c(\theta,\epsilon)\right) = \sup_{\theta' \in N^c(\theta,\epsilon)\cap\bar{\Theta}} L_N(\underline{x}\,|\,\theta') \ . \tag{2.20}$$

Then

$$L_N\left(\underline{x}\,|\,N^c(\theta,\epsilon)\right) = \max\left\{L_N(\underline{x}\,|\,K_t^c)\ ,\ L_N\left(\underline{x}\,|\,K_t(\theta,\epsilon)\right)\right\}$$

is $\beta^{(N)}$-measurable; and, given $\eta > 0$, it follows from (2.16) and (2.19) that

$$E_\theta \log\left[\frac{L_N\left(\underline{X}\,|\,N^c(\theta,\epsilon)\right)}{L_N(\underline{X}\,|\,\theta)}\right] < -\eta \ . \tag{2.21}$$

Let n be any integer $\geq N$. We note that

$$\log H_n\left((x_1,x_2,\ldots,x_n)\,|\,\theta'\right) = \frac{n}{\binom{n}{N}} \sum \log L_N\left((x_{i_1},x_{i_2}\ldots,x_{i_N})\,|\,\theta'\right) + \log g(\theta') \ , \tag{2.22}$$

where the summation is taken over the $\binom{n}{N}$ possible choices of N indices $i_1,i_2,\ldots,i_N$ from among the n indices $1,2,3,\ldots,n$. Thus by the superadditivity of the supremum function and the fact that the logarithm is a monotonic non-decreasing function, it follows that

$$\frac{1}{n} \log \left[\frac{H_n\left((x_1, x_2, \ldots x_n) \,|\, N^c(\theta, \epsilon)\right)}{H_n\left((x_1, x_2, \ldots x_n) \,|\, \theta\right)} \right.$$

$$\leq \frac{1}{\binom{n}{N}} \sum \log \left[\frac{L_N\left(x_{i_1}, x_{i_2}, \ldots, x_{i_N} \,|\, N^c(\theta, \epsilon)\right)}{L_N\left((x_{i_1}, x_{i_2}, \ldots, x_{i_N}) \,|\, \theta\right)} \right. + \frac{1}{n} \log \frac{g(\Theta)}{g(\theta)} \quad . \tag{2.23}$$

We recognize the first term on the right-hand side of inequality (2.23) as a U-statistic. Since the sequence of such statistics (indexed by n) is a reverse Martingale, it follows from the Martingale Convergence Theorem (see Berk (1966)) that

$$\frac{1}{\binom{n}{N}} \sum \log \left[\frac{L_N\left((X_{i_1}, X_{i_2}, \ldots, X_{i_N}) \,|\, N^c(\theta, \epsilon)\right)}{L_N\left((X_{i_1}, X_{i_2}, X_{i_N}) \,|\, \theta\right)} \right]$$

$$\xrightarrow{\text{a.s.}} E_\theta \log \left[\frac{L_N\left((X_1, X_2, \ldots X_N) \,|\, N^c(\theta, \epsilon)\right)}{L_N\left((X_1, X_2, \ldots X_N) \,|\, \theta\right)} \right] - \eta < 0 \quad . \tag{2.24}$$

Since by Conditions 2.3 and 2.5, $n^{-1} \log [g(\Theta)/g(\theta)] \to 0$ as $n \to \infty$, it follows that

$$P_\theta \left\{ \frac{1}{n} \log \left[\frac{H_n\big((X_1, X_2, \dots X_n) \,|\, N^c(\theta,\epsilon)\big)}{H_n((X_1, X_2, \dots X_n) \,|\, \theta)} \right] < -\frac{\eta}{2} \right.$$

$$\left. \text{for all sufficiently large } n \right\} = 1 . \tag{2.25}$$

Hence, for all $\theta' \in N^c(\theta,\epsilon)$

$$H_n(\xi|\theta') < e^{-\frac{1}{2}n\eta} H_n(\xi|\theta) \leq e^{-\frac{1}{2}n\eta} H_n(\xi|\Theta) \leq \tfrac{1}{2} H_n(\xi|\Theta)$$

for all sufficiently large n with probability one under $P_\theta^{(\infty)}$, all $\theta \in \Theta$. This establishes (2.10). Q.E.D.

Theorem 2.2. Under Conditions 2. through 2.8, there exists an integer $N \geq 1$ such that $\theta_n^*(s) = \theta_n^*((x_1, x_2, \dots, x_n))$ exists and is $X^{(n)}$-measurable for all $n \geq N$ and such that

$$\theta_n^*(\xi) \to \theta \quad \text{a.s.} \quad \text{as } n \to \infty$$

for all $\theta \in \Theta$.

Proof. The first part of the assertion is a restatement of Theorem 2.1. Since θ_n^* exists and is measurable for $n \geq N$, $\theta_n^*(s) \in \Lambda_n(s) \subset \Lambda_n^*(s)$, for all $s \in X^{(\infty)}$, so that

$\Lambda_n^*(s)$ is non-empty for $n \geq N$. The asserted strong convergence of $\theta_n^*(\xi)$ now follows as a direct consequence of Lemma 2.3.

The regularity conditions which we have used in this section closely resemble Kiefer and Wolfowitz's (1962) and Bahadur's (1967) modifications of the regularity conditions originally adopted by Wald (1949). Of course, to cover the greater generality of our estimator $\theta_n^*(\xi)$, we have added Conditions 2.3 and 2.5, which are regularity conditions on the prior density $g(\theta)$. Since we nowhere used (or assumed) the requirement that $g(\theta)$ be a probability density on Θ, the density $g(\theta) \equiv 1$, all $\theta \in \Theta$, satisfies our conditions. Thus, our conditions cover the special case of the M.L.E. $\hat{\theta}_n$.

Our assumptions do differ in some respe t from the regularity conditions of Kiefer and Wolfowitz (1962). The major differences between Conditions 2.1-2.8 and the conditions of Kiefer and Wolfowitz (1962) are:

(a) Our Condition 2.2 requires continuity in θ of $L_N(\underline{x}|\theta)$. In contrast, Kiefer and Wolfowitz (1962) require that $\sup_{\theta' \in N(\theta,\delta)} f(x|\theta') \to f(x|\theta)$ as $\delta \to 0$.

(b) A comment by Kiefer and Wolfowitz (1962) that if their regularity conditions do not hold for $L_1(x|\theta) = f(x|\theta)$, everything still goes through if the regularity conditions hold for $L_N(x|\theta)$, some $N \geq 1$, has been directly incorporated into our regularity conditions (see also Berk (1966)). This enables us, for example, to verify existence and strong consistency of the M.L.E. $\hat{\theta}_n$ and M.P.E. θ_n^* for the case of the normal distribution with unknown mean and variance.

(c) We have confined ourselves to parameters which are points in k-dimensional Euclidean space under the usual metric. Further, we have not tried to introduce special measures of distance or to transform our regularity conditions on the densities $f(x|\theta)$ to conditions on transformations of the density functions $f(x|\theta)$ — devices mentioned by Kiefer and Wolfowitz (1962), and used more explicitly by Huber (1967).

The restriction in our Condition 2.2 pointed out in remark (a) above can probably be removed by modifying Theorem 2 of Olech (1965) used in proving Lemma 2.1 in an appropriate manner. However, until we can remove this restriction, our conditions do not cover the case of estimation of the parameters for the uniform distribution and other similar distributions. The conditions of Kiefer and Wolfowitz, however, do cover these distributions. The restrictions noted in point (c) above may also limit the applicability of our results somewhat. These restrictions were needed to prove Lemma 2.1. On the other hand, as noted above, the incorporation of Kiefer and Wolfowitz's remark directly into our regularity conditions provides added flexibility and applicability to our results. Actually, Kiefer and Wolfowitz only mentioned looking at certain integrability results (Condition 2.8) for large enough n; we have extended their suggestion to continuity and measurability conditions. Our proof of strong consistency of the M.L.E. also differs somewhat from that indicated by Kiefer and Wolfowitz (1962).

Perhaps the major novelty of our results is the rigorous proof (Theorem 2.1) we give for the existence and measurability of the M.P.E. θ_n^* for fixed sample size n. The existence of $\theta_n^*(s)$ for fixed $s = (x_1, x_2, \ldots)$ and large enough n has been argued by many authors (Bahadur (1967), LeCan (1953), etc.), but this does not prove that $\theta_n^*\big((x_1, x_2, \ldots, x_n)\big)$ exists and is $\beta^{(n)}$-measurable for each n. The truth of this missing fact has been asserted by the authors mentioned above, but (probably due to considerations of space) they have ommitted mentioning sufficient conditions or providing a proof for such a result. Without a proof of the existence and $\beta^{(n)}$-measurability of $\theta_n^*(s)$ for all large enough n, the usual proof of the consistency of such estimators (as given in our Lemma 2.3 - although we have modified the usual proof slightly) is not completely rigorous (although, as remarked by Huber (1967), it can probably be done by judicious use of inner and outer probabilities).

3. Asymptotic Normality

In the previous section we established conditions sufficient to prove that the M.P.E. θ_n^* with respect to a prior density $g(\theta)$ is a strongly consistent estimator of θ. In the present section, we show that under certain additional regularity conditions when θ is a one-dimensional parameter, the M.P.E. is asymptotically normally distributed. We also derive a useful representation for the sequence $\{\theta_n^*\}$, and indicate a rate for the almost sure convergence of θ_n^* to θ.

Denote $\log f(x|\theta)$ by $\ell(x|\theta)$, and let

$$\ell^{(i)}(x|\theta') = \left(\frac{\partial}{\partial\theta}\right)^i \log f(x|\theta)\Big|_{\theta=\theta'} = \left(\frac{\partial}{\partial\theta}\right)^i \ell(x|\theta)\Big|_{\theta=\theta'}$$

for $i = 1,2,\ldots$. Let

$$\ell_n^{(i)}(\xi|\theta) = \sum_{i=1}^{n} \ell^{(i)}(X_j|\theta),$$

where $\xi = (X_1,X_2,\ldots)$. In what follows we assume that the M.L.E. $\hat{\theta}_n(\xi)$ and the M.P.E. $\theta_n^*(\xi)$ with respect to a given prior density $g(\theta)$ exist for large enough n and are both strongly consistent estimators of θ. Further, we assume that for all $s = (x_1, x_2, \dots) \in X^{(\infty)}$, $\hat{\theta}_n(s) = \hat{\theta}_n\big((x_1, x_2, \dots, x_n)\big)$ is a solution of the equation

$$\ell_n^{(1)}(s|\theta) = 0, \tag{3.1}$$

and that $\theta_n^*(s) = \theta_n^*\big((x_1, x_2, \dots, x_n)\big)$ is a solution to the equation

$$\ell_n^{(1)}(s|\theta) + \frac{\partial}{\partial\theta} \log g(\theta) = 0, \tag{3.2}$$

for all large enough n. This will be the case if $\ell^{(1)}(x|\theta)$ exists and $\partial/\partial\theta \log g(\theta)$ exists for all $\theta \in \Theta$, all $x \in X$, and if the maxima of $\prod_{i=1}^{n} f(x_i|\theta)$ and of $g(\theta) \prod_{i=1}^{n} f(x_i|\theta)$ exist within the interior of Θ for all (large enough) n. To obtain the main results of this section, we need the following added conditions:

<u>Condition 3.1</u>. For each $x \in X$, $\ell^{(2)}(x|\theta)$ exists and is continuous in θ for all $\theta \in \Theta$.

<u>Condition 3.2</u>. For each $\theta \in \Theta$, we have $E_\theta \ell^{(1)}(x|\theta) = 0$ and

$$0 < E_\theta(\ell^{(1)}(X|\theta))^2 = -E_\theta \ell^{(2)}(X|\theta) \equiv I(\theta) < \infty .$$

Condition 3.3. For each $\theta \in \Theta$, there exists a δ-neighborhood, say $N(\theta,\delta) = \{\theta' : |\theta-\theta'| < \delta\} \subset \Theta$, of θ and a measurable function $M(x)$ on $(\mathcal{X},\beta)$ such that

$$|\ell^{(2)}(x|\theta') - \ell^{(2)}(x|\theta)| < M(x)$$

for all $x \in \mathcal{X}$ and all $\theta' \in N(\theta,\delta)$, and such that $E_\theta\, M(X) < \infty$.

Condition 3.4. $\partial/\partial\theta \log g(\theta)$ exists for all $\theta \in \Theta$. Conditions 3.1 through 3.3 have previously been adopted by Bahadur (1964).

Before stating the main result of this section, we need to introduce some new notation. By the symbol $o_a(1/b_n)$ we represent any sequence of random variables $\{Z_n\}$ defined on $(\mathcal{X}^{(\infty)},\beta^{(\infty)})$ which has the property that for the sequence $\{b_n\}$ of constants, $b_n \to \infty$ as $n \to \infty$,

$$b_n Z_n \to 0 \text{ almost surely } \left(P_\theta^{(\infty)}\right), \quad n \to \infty .$$

The symbol $o_a(1)$ represents any sequence of random variables $\{Z_n\}$ defined on $\left(\mathfrak{X}^{(\infty)},\beta^{(\infty)}\right)$ which converges almost surely $\left(P_\theta^{(\infty)}\right)$ to zero as $n \to \infty$. Addition, subtraction, multiplication, or division of the symbols $o_a(1/b_n)$ refer to the same operations performed on the corresponding sequences of random variables. These operations almost surely obey the calculus of $o(1/b_n)$. Also, in what follows, the symbols $o_p(1)$, $o_p(1/b_n)$, $O_p(1)$, $o_p(1/b_n)$ have their usual meanings (see Mann and Wald (1943)).

<u>Theorem 3.1</u>. Under Conditions 3.1 through 3.4 and under the assumption that the M.L.E. $\hat{\theta}_n(\xi)$ and the M.P.E. $\theta_n^*(\xi)$ exist (for all sufficiently large n) and are strongly consistent estimators of θ satisfying Equation (3.1) and Equation (3.2) respectively, it follows that for each $\theta \in \Theta$,

$$\hat{\theta}_n = \theta + \frac{1}{nI(\theta)} \ell_n^{(1)}(\xi|\theta) \left(1+o_a(1)\right) , \tag{3.3}$$

$$\theta_n^* = \theta + \frac{1}{nI(\theta)} \left[\ell_n^{(1)}(\xi|\theta) + \frac{\partial}{\partial\theta}\log g(\theta)\Big|_{\theta=\theta_n^*}\right]\left[1+o_a(1)\right], \tag{3.4}$$

as $n \to \infty$.

<u>Proof</u>. Since $\hat{\theta}_n(\xi)$ and $\theta_n^*(\xi)$ both are strongly consistent estimators of θ, we can assume that

$$\hat{\theta}_n(\xi) = \theta + k_n(\xi,\theta), \quad \theta_n^*(\xi) = \theta + h_n(\xi,\theta) \ , \tag{3.5}$$

where $k_n(\xi,\theta) = o_a(1)$, $h_n(\xi,\theta) = o_a(1)$ as $n \to \infty$. Since $\hat{\theta}_n(\xi)$ satisfies (3.1) and $\theta_n^*(\xi)$ satisfies (3.2), we have

$$\ell_n^{(1)}\left(\xi|\hat{\theta}_n(\xi)\right) = 0, \tag{3.6}$$

and

$$\ell_n^{(1)}\left(\xi|\theta_n^*(\xi)\right) + \frac{\partial}{\partial\theta}\log g(\theta)\Big|_{\theta=\theta_n^*(\xi)} = 0 \ , \tag{3.7}$$

for all large enough n. By Condition 3.1, we can expand $\ell^{(1)}\left(\xi|\hat{\theta}_n(\xi)\right)$ and $\ell^{(1)}\left(\xi|\theta_n^*(\xi)\right)$ each in a Taylor's expansion around the true θ, obtaining

$$0 = \ell_n^{(1)}\left(\xi|\hat{\theta}(\xi)\right) = \ell_n^{(1)}(\xi|\theta) + k_n(\xi,\theta)\ell_n^{(2)}(\xi|\eta_n), \tag{3.8}$$

and

$$0 = \ell_n^{(1)}\left(\xi|\theta_n^*(\xi)\right) + \frac{\partial}{\partial\theta} \log g(\theta)\Big|_{\theta=\theta_n^*(\xi)} = \ell_n^{(1)}(\xi|\theta) +$$

$$+ h_n(\xi,\theta)\ell_n^{(2)}(\xi|\alpha_n) + \frac{\partial}{\partial\theta} \log(\theta)\Big|_{\theta=\theta_n^*(\xi)},$$

(3.9)

where $\eta_n = \eta_n(\xi)$ lies between $\hat{\theta}_n(\xi)$ and θ, and where $\alpha_n = \alpha_n(\xi)$ lies between $\theta_n^*(\xi)$ and θ. Divide both sides of Equations (3.8) and (3.9) by $nI(\theta)$, where $I(\theta) > 0$ is defined in Condition 3.2. We obtain

$$\frac{\ell_n^{(1)}(\xi|\theta)}{nI(\theta)} = \left(k_n(\xi,\theta)\right)\left(\frac{-\ell_n^{(2)}(\xi|\eta_n)}{nI(\theta)}\right), \qquad (3.10)$$

and

$$\frac{1}{nI(\theta)}\left[\ell_n^{(1)}(\xi|\theta) + \frac{\partial}{\partial\theta} \log g(\theta)\Big|_{\theta=\theta_n^*}\right]$$

$$= \left(h_n(\xi,\theta)\right)\left(- \frac{\ell_n^{(2)}(\xi|\alpha_n)}{nI(\theta)}\right) . \qquad (3.11)$$

Since $\eta_n(\xi)$ is between $\hat{\theta}_n(\xi)$ and θ, and since $\alpha_n(\xi)$ is between $\theta_n^*(\xi)$ and θ, it follows from the fact that $\hat{\theta}_n(\xi)$ and $\theta_n^*(\xi)$ are strongly consistent estimators of θ that

$$\eta_n(\xi) - \theta = .o_a(1) \quad , \quad \alpha_n(\xi) - \theta = o_a(1) \quad .$$

Comparing (3.3) with (3.10), (3.4) with (3.11), and taking note of the definitions (3.5) of $k_n(\xi,\theta)$ and $h_n(\xi,\theta)$, we see that (3.3) and (3.4) are certainly verified if we can show that for any strongly consistent estimator $\{T_n(\xi)\}$ of θ, where $T_n(\xi) \equiv T_n(X_1,X_2,\ldots,X_n)$,

$$\frac{\ell_n^{(2)}(\xi|T_n)}{nI(\theta)} = -1 + o_a(1) \quad . \tag{3.12}$$

Since

$$\left|\frac{\ell_n^{(2)}(\xi|T_n)}{n} + I(\theta)\right| \leq \left|\frac{\ell_n^{(2)}(\xi|\theta)}{n} + I(\theta)\right| + \left|\frac{\ell_n^{(2)}(\xi|T_n) - \ell_n^{(2)}(\xi|\theta)}{n}\right|$$

and since by Condition 3.2 and the strong law of large numbers

$$\left|\frac{\ell_n^{(2)}(\xi|\theta)}{n} + I(\theta)\right| = o_a(1) ,$$

to prove (3.12) we need only demonstrate that
$n^{-1}|\ell_n^{(2)}(\xi|T_n) - \ell_n^{(2)}(\xi|\theta)| = o_a(1)$. But

$$n^{-1}|\ell_n^{(2)}(\xi|T_n) - \ell_n^{(2)}(\xi|\theta)| \leq \frac{1}{n}\sum_{i=1}^{n}|\ell^{(2)}(X_i|T_n) - \ell^{(2)}(X_i|\theta)| , \tag{3.13}$$

so that if the right-hand side of (3.13) tends almost surely to 0 as $n \to \infty$, we are done.

Let

$$A(x,\delta) = \sup_{\theta' \in N(\theta,\delta)} \{|\ell^{(2)}(x|\theta') - \ell^{(2)}(x|\theta)|\}$$

and let

$$A(\delta) = E_\theta\, A(X,\delta) .$$

(Note. It can be shown using Condition 3.1 and Lemma 2.2 of Section 2 that $A(x,\delta)$ is measurable).

It follows from Condition 3.1 that for every x,

$A(x,\delta) \to 0$ as $\delta \to 0$. Hence, by Condition 3.3 and the Lebesgue Dominated Convergence Theorem, $A(\delta) \to 0$ as $\delta \to 0$. Given any $\epsilon > 0$, chose δ so small that $A(\delta) < \epsilon$. Then since $P_\theta\{T_n \epsilon N(\theta,\delta)$ for all large enough $n\} = 1$, it follows that

$$P_\theta\left\{\frac{1}{n}\sum_{i=1}^{n} |\ell^{(2)}(X_i|T_n) - \ell^{(2)}(X_i|\theta)| \le \frac{1}{n}\sum_{i=1}^{n} A(X_i,\delta),\right.$$

$$\left.\text{for all large enough } n\right\} = 1.$$

Since by the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} A(X_i,\delta) \xrightarrow{a.s.} A(\delta) < \epsilon \quad , \text{ as } n \to \infty ,$$

and since $\epsilon > 0$ can be chosen arbitrarily small, it follows that

$$\frac{1}{n}\sum_{i=1}^{n} |\ell^{(2)}(X_i|T_n) - \ell^{(2)}(X_i|\theta)| = o_a(1) \ .$$

This establishes (3.3) and (3.4). Q.E.D.

Corollary 3.1. Under the conditions of Theorem 3.1, and the additional condition that

Condition 3.4'. Given $\theta \in \Theta$, there exists $\epsilon > 0$ such that

$$\sup_{\theta' \in N(\theta_o, \epsilon)} \left| \frac{\partial}{\partial\theta} \log g(\theta) \Big|_{\theta=\theta'} \right| < \infty \; ;$$

then the following relationships hold for any ρ , $0 < \rho \leq \frac{1}{2}$:

$$n^{\frac{1}{2}-\rho}(\hat{\theta}_n - \theta) = o_a(1) \quad , \quad n^{\frac{1}{2}-\rho}(\theta_n^* - \theta) = o_a(1), \tag{3.14}$$

$$n^{\frac{1}{2}-\rho}(\hat{\theta}_n - \theta_n^*) = o_a(1) \; . \tag{3.15}$$

Further,

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta_n^*) = o_p(1) \quad , \tag{3.16}$$

and

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \frac{1}{I(\theta)}) \; , \quad n^{\frac{1}{2}}(\theta_n^* - \theta) \xrightarrow{\mathcal{L}} N(0, \frac{1}{I(\theta)}). \tag{3.17}$$

It follows that under Condition 3.4' and the conditions of Theorem 3.1, the M.P.E. θ_n^* with respect to the prior density $g(\theta)$ is B.A.N.

<u>Proof</u>. By the law of the iterated logarithm for i.i.d. variables having zero mean and finite variance, and by Condition 3.2, for all ρ, $0 < \rho \leq \frac{1}{2}$,

$$n^{\frac{1}{2}-\rho} \left(\frac{\ell_n^{(1)}(\xi|\theta)}{n \cdot I(\theta)}\right) = o_a(1) \; . \tag{3.18}$$

From Condition 3.4' we know that there exists $\epsilon > 0$ for which

$$\sup_{\theta' \in N(\theta, \epsilon)} \left| \frac{\partial}{\partial\theta} \log g(\theta) \Big|_{\theta=\theta'} \right| < \infty \; .$$

Since $\theta_n^* - \theta = o_a(1)$, it follows that $P_\theta\{\theta_n^* \in N(\theta, \epsilon)$ for all sufficiently large $n\} = 1$ and thus that for all ρ, $0 < \rho \leq \frac{1}{2}$,

$$\frac{n^{\frac{1}{2}-\rho}}{nI(\theta)} \frac{\partial}{\partial\theta} \log g(\theta)\Big|_{\theta=\theta_n^*} = o_a(1) \; . \tag{3.19}$$

Assertion (3.14) now follows from (3.18) and (3.19). Assertion (3.15), of course, is an immediate consequence of (3.14).

Since almost sure convergence to zero implies convergence in probability to zero, and since by Condition

3.2 and the Central Limit Theorem

$$\frac{\sqrt{n}}{nI(\theta)} \ell_n^{(1)}(\xi|\theta) \xrightarrow{\mathcal{L}} N(0,\frac{1}{I(\theta)}) , \qquad (3.20)$$

it follows from (3.3), (3.4), (3.19), and the calculus of $o_p(1)$ and $O_p(1)$, that (3.16) and (3.17) hold. This completes the proof. Q.E.D.

Remark I. Results similar to Assertion (3.15) of Corollary 3.1 have been obtained by Bickel and Yahav (1966). These authors have shown that under certain regularity conditions, any Bayes estimator B_n under a convex loss function converges almost surely to the M.L.E. $\hat{\theta}_n$ at rate $1/\sqrt{n}$; i.e., that $\sqrt{n}(B_n-\hat{\theta}_n) = o_a(1)$. Their results have been generalized to the multiparameter case by Chao (1970). Although the results of Bickel and Yahav (1966) and of Chao (1970) are slightly stronger than our result (3.15), the regularity conditions assumed by these authors seem (as best we can compare them) to be somewhat more restrictive than ours. In any case, the results of these authors do not apply to our case since the M.P.E. θ_n^* is not necessarily a Bayes estimator for any convex loss function.

Remark II. The representation (3.3) for $\hat{\theta}_n$ has previously been used without proof by Wolfowitz (1966).

Remark III. The results in Theorem 3.1 and Corollary 3.1 can be rather straightforwardly generalized to the k-parameter case. Appropriate conditions for such an extension (except for conditions on the partial derivatives of $\log g(\theta)$) are given by Bahadur (1964). Essentially, what is needed is that the information matrix exists and is positive definite, that the second partials of $\ell(x|\theta)$ exist and are continuous in θ, and that we can apply the Lebesgue Dominated Convergence Theorem as in the proof of Theorem 3.1.

The results of this section show that under fairly weak regularity conditions on $g(\theta)$, the M.P.E. θ_n^* is asymptotically efficient in the classical Fisherian sense. In Section 4, we show that θ_n^* also has the maximum asymptotic probability concentration property defined by Bahadur (1960, 1967).

4. Asymptotic Efficiency in the Sense of Probability Concentration.

We have shown that under certain regularity conditions the maximum probability estimator $\theta_n^*(\xi)$ is a strongly consistent estimator of θ and that $\sqrt{n}(\theta_n^* - \theta)$ is asymptotically normally distributed with variance $1/I(\theta)$. This implies that the estimator θ_n^* is asymptotically efficient in the classical sense. Since it is important to our further work, it is worthwhile to recall some historical examples relating to the classical definition of asymptotic efficiency of an estimator.

It was believed (Fisher's program) that for any consistent, asymptotically normal estimator $T_n(\xi)$ such that

$$\sqrt{n}\ (T_n(\xi) - \theta) \xrightarrow{\mathcal{L}} Z\ ,$$

where Z has a normal distribution with mean zero and variance $V(\theta)$, the asymptotic variance $V(\theta)$ satisfies the inequality

$$V(\theta) \geq 1/I(\theta)\ , \qquad (4.1)$$

and that an estimator with smallest asymptotic variance has maximum probability concentration about the true value θ in sufficiently large samples. Unfortunately, both of these results are not true without further restrictions. Hodges (1953) showed by an example that the first part of Fisher's program is not in general true. Let the sequence $\{X_i\}$ be a sequence of i.i.d. normally distributed random variables with mean θ and variance $0 < I(\theta) < \infty$, and let

$$T_n^*(X_1,\ldots X_n) = \begin{cases} \bar{X}_n , & \text{if } |\bar{X}_n| > n^{-\frac{1}{4}} , \\ 0 , & |\bar{X}_n| \leq n^{-\frac{1}{4}} , \end{cases} \qquad (4.2)$$

be an estimator of θ. Clearly, the estimator T_n^* is a consistent asymptotically normal estimator with asymptotic variance

$$V_{T^*}(\theta) = \begin{cases} \frac{1}{I(\theta)} & \text{if } \theta \neq 0 , \\ 0 & \text{if } \theta = 0 . \end{cases} \qquad (4.3)$$

Hence, the inequality (4.1) does not hold for all θ. The estimator T_n^* is said to be a superefficient estimator. LeCam (1953, 1955) and Wolfowitz (1953) have shown that any superefficient estimator violates the inequality (4.1) only for a null subset of R^1 (with respect to Lebesgue measure).

A simple and very ingenious proof of this fact has been devised by Bahadur (1964). It has been shown by Rao (1963) and Wolfowitz (1965) that if $f(x|\theta)$ is subject to certain so-called "regularity conditions," and if T_n is an estimator satisfying

$$\sqrt{n}\ (T_n(\xi) - \theta) \xrightarrow{U\mathcal{L}} Z \quad , \qquad (4.4)$$

where Z is normally distributed with mean zero and variance $V(\theta)$, and where $\xrightarrow{U\mathcal{L}}$ means uniform convergence in law with respect to θ, then the inequality (4.1) holds for all θ. The intuitive reason for this result is that uniformity in θ of the convergence in (4.4) implies that the function $V(\theta)$ is a continuous function of θ. If both $V(\theta)$ and $I(\theta)$ are continuous in θ, then the set of θ for which the inequality (4.1) does not hold is, of course, the empty set.

Basu(1956) brings out the fact that an estimator, though asymptotically much less efficient (in the classical sense) than another estimator, may yet have much greater probability concentration than the latter one. Let $\{X_i\}$ be i.i.d. normal random variables with mean θ and variance one.

Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $S_n = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$. The random variables $\bar{X}_n$ and S_n are stochastically independent. Consider the estimator $T_n^{(1)}$,

$$T_n^{(1)}(X_1, \ldots X_n) = (1-\varphi_n)\bar{X}_n + n\,\varphi_n \quad , \tag{4.5}$$

where

$$\varphi_n(X_1, \ldots X_n) = \begin{cases} 1 & \text{if } S_n > a_n \, , \\ 0 & \text{if } S_n \leq a_n \, , \end{cases} \tag{4.6}$$

and where a_n is the upper $100 \frac{1}{n}$ % point of the chi-square distribution with n-1 degrees of freedom. Let

$$T_n^{(2)}(X_1, \ldots X_n) = \bar{X}_{(\sqrt{n}\,)} = \frac{X_1 + \ldots + X_{[\sqrt{n}\,]}}{[\sqrt{n}\,]} \quad , \tag{4.7}$$

where $[\sqrt{n}\,]$ is the integer part of $\sqrt{n}$. Since $P_\theta\{\varphi_n = 0\} = 1 - \frac{1}{n} \to 1$, it follows that

$$\sqrt{n}\;(T_n^{(1)} - \theta) = \sqrt{n}\;(\bar{X}_n - \theta) + \sqrt{n}\;(n - \bar{X}_n)\varphi_n \xrightarrow{\mathcal{L}} Z \, , \tag{4.8}$$

where Z is a standard normal random variable. The estimator $T_n^{(1)}$ has asymptotic variance $\sigma_n^2(1) = \frac{1}{n}$, and $T_n^{(2)}$ has asymptotic variance $\sigma_n^2(2) = \frac{1}{\sqrt{n}}$. Hence

$$\lim_{n\to\infty} \frac{\sigma_n^2(1)}{\sigma_n^2(2)} = 0 . \tag{4.9}$$

Since, for large n

$$P_\theta\{|T_n^{(1)}-\theta| \geq \epsilon\} = P_\theta\{\varphi_n = 0\} P_\theta\{|\bar{X}_n-\theta| \geq \epsilon\} + P\{\varphi_n = 1\}$$

$$= o(\tfrac{1}{n}) + \tfrac{1}{n} \quad , \tag{4.10}$$

and

$$P_\theta\{|T_n^{(2)}-\theta| \geq \epsilon\} = P_\theta\{|\bar{X}_{[\sqrt{n}]} - \theta| \geq \epsilon\} = o(\tfrac{1}{n}) \; , \tag{4.11}$$

it follows that

$$\lim_{n\to\infty} \frac{P_\theta\{|T_n^{(1)}(\xi) - \theta| \geq \epsilon\}}{P_\theta\{|T_n^{(2)}(\xi) - \theta| \geq \epsilon\}} = \infty . \tag{4.12}$$

This example shows that the second part of Fisher's program (that smaller asymptotic variance implies larger probability concentration about θ in large samples) is also not true. In the above-mentioned paper, Basu suggests that the rapidity with which $P_\theta\{|T_n-\theta| \geq \epsilon\}$ converges to zero may be considered to be a measure of the asymptotic concentration of T_n. Based on this criterion, Bahadur (1960, 1967) has shown that if T_n is any consistent estimator, the logarithm of $P_\theta\{|T_n-\theta| \geq \epsilon\}$ cannot tend to $-\infty$ faster than the rate given by $-\frac{1}{2}\epsilon^2 n\, I(\theta)$; or, in looser terminology, that $P_\theta\{|T_n-\theta| \geq \epsilon\}$ cannot tend to zero faster than the exponential rate given by $\exp\{-\frac{1}{2}n\epsilon^2 I(\theta)\}$. In terms of the loose terminology used above, Bahadur (1960, 1967) also shows that the probability that the M.L.E. $\hat{\theta}_n(\xi)$ is located outside of an ϵ-neighborhood of θ tends to zero nearly at the optimal exponential rate $\exp\{-\frac{1}{2}n\epsilon^2 I(\theta)\}$. Stated formally, the above assertions mean that

$$\underline{\lim}_{\epsilon\to 0}\ \underline{\lim}_{n\to\infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} \geq -\tfrac{1}{2} I(\theta) , \qquad (4.13)$$

and

$$\lim_{\epsilon\to 0}\ \lim_{n\to\infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|\hat{\theta}_n-\theta| \geq \epsilon\} = -\tfrac{1}{2} I(\theta). \qquad (4.14)$$

In this section, we show that $P_\theta\{|\theta_n^*-\theta| \geq \epsilon\}$, the probability that the M.P.E. θ_n^* with respect to the prior density $g(\theta)$ is located outside of an ϵ-neighborhood of θ, tends to zero with the optimal rate $\exp\{-\frac{1}{2}\epsilon^2 n I(\theta)\}$; i.e. we show that

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|\theta_n^*-\theta| \geq \epsilon\} = -\tfrac{1}{2} I(\theta) . \qquad (4.15)$$

This result can be shown to hold under basically the same regularity conditions on the density $f(x|\theta)$ as were assumed by Bahadur (1960,1967) to prove (4.14), plus added assumptions to provide regularity conditions on $g(\theta)$. However, we shall take a slightly different approach.

We begin by proving some needed lemmas. Let Z be a random variable, and let

$$\varphi_Z(t) = E[e^{tZ}] \qquad , \quad -\infty < t < \infty ,$$

be the moment generating function of Z. For any constant a, $-\infty < a < \infty$, define

$$\rho(a) = \inf_{t\geq 0} e^{-at} \varphi_Z(t) . \qquad (4.16)$$

Lemma 4.1. If $EZ < 0$, $P(Z > 0) > 0$ and $\varphi_Z(t)$ is finite for all $t > 0$, then:

(a) $\varphi_Z(t)$ is a continuous and convex function of t,

(b) $0 < \rho(0) < 1$,

(c) there exists a unique τ, $0 < \tau < \infty$, such that $\rho(0) = \varphi_Z(\tau)$,

(d) $\varphi_Z^{(1)}(t) \equiv \frac{d}{dt}\varphi_Z(t)$ exists on $(0, \infty)$ and τ is the unique solution of $\varphi_Z^{(1)}(t) = 0$.

Proof: See Bahadur (1968) or Johns (1968).

Lemma 4.2. If $EZ < 0$, $P(Z > 0) > 0$ and $\varphi_Z^{(1)}(t)$ exists and is continuous for $t > 0$, then $\rho(a)$ is continuous at $a = 0$.

Proof: Since $EZ < 0$ and $P\{Z > 0\} > 0$, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $EZ < -\epsilon_1 < 0$ and $P\{Z > \epsilon_2\} > 0$. Let $\epsilon = \min(\epsilon_1, \epsilon_2)$ and let a be any number in $(-\epsilon, \epsilon)$. Let $Z_a = Z - a$. Clearly

$$EZ_a < 0 \text{ , all } a \in (-\epsilon, \epsilon) \text{ ,} \tag{4.17}$$

and

$$P\{Z_a > 0\} > 0, \text{ all } a \in (-\epsilon, \epsilon) \text{ .} \tag{4.18}$$

Hence we may apply Lemma 4.1, so that there exists a unique solution of $\varphi_{Z_a}^{(1)}(t) = 0$, say $t = \tau(a)$. Since

$\varphi_{Z_a}(t) = e^{-a\,t}\varphi_Z(t)$ and $\varphi_Z(t) > 0$ for all $0 \leq t < \infty$,

it follows that $\tau(a)$ is also the unique solution of the equation

$$\frac{\varphi_Z^{(1)}(t)}{\varphi_Z(t)} - a = 0 \quad . \tag{4.19}$$

Since $\varphi_Z^{(1)}(t)$ and $\varphi_Z(t)$ are continuous (by assumption and Lemma 4.1, respectively), and since $\tau(a)$ is the unique solution of (4.19), $\tau(a)$ is continuous for all $a \in (-\epsilon, \epsilon)$. But

$$\rho(a) = e^{-a\,\tau(a)}\varphi_Z(\tau(a)),$$

so that $\rho(a)$ is continuous for all $a \in (-\epsilon,\ \epsilon)$. Q.E.D.

Let the sequence $\{Z_n\}$ be a sequence of i.i.d. random variables with finite means. Let $\{a_n\}$ be a sequence of real numbers such that $\lim_{n\to\infty} a_n = a$, where $a > EZ_i$, all i. Define $p_n(a_n) = P\{n^{-1}\sum_{i=1}^{n} Z_i \geq a_n\}$. Parts of the following basic lemma have been proved by Bahadur (1960, 1968), by Chernoff (1952) and by Bernstein.

Lemma 4.3. Let $Z_1, Z_2, \ldots$ be i.i.d., and let each Z_i have the distribution of Z, where Z has moment generating function $\varphi_Z(t)$. Let $\rho(a)$ be defined by (4.16). Then if the random variable Z satisfies the conditions of Lemma 4.2,

$$\text{(a)} \quad p_n(a_n) \le (\rho(a_n))^n \quad , \quad \text{for all } n \, , \tag{4.20}$$

$$\text{(b)} \quad \frac{1}{n} \log p_n(a_n) \to \log \rho(a) \, , \quad \text{as } n \to \infty. \tag{4.21}$$

Proof: See Bahadur (1960,1968).

The following leamma, due to Daniels (1961), gives expansions for $E_\theta \ell^{(1)}(X|\theta')$ and $E_\theta[\ell^{(1)}(X|\theta')]^2$ for θ' in a small neighborhood of θ.

Lemma 4.4. Let the following conditions hold:

Condition 4.1. For all x, $\ell(x|\theta)$ is a continuous function of θ. Further for every $\theta \in \Theta$, there exists a δ-neighborhood of θ, say $N(\theta, \delta)$, and a measurable function $A(x,\theta)$, $E_\theta A^2(X,\theta) < \infty$, such that

$$|\ell(x|\theta') - \ell(x|\theta'')| < A(x,\theta)|\theta''-\theta'|, \tag{4.22}$$

for all $\theta', \theta'' \in N(\theta, \delta)$.

Condition 4.2. $\ell^{(1)}(x|\theta)$ exists, is not almost everywhere zero, and is continuous in θ for (almost) all $x \in X$.

then if θ' is sufficiently close to θ, the following statements are true:

(a) $E_\theta \ell^{(1)}(X|\theta) = 0$,

(b) $0 < E_\theta[\ell^{(1)}(X|\theta)]^2 = I(\theta) < \infty$,

(c) $E_\theta \ell^{(1)}(X|\theta') = -(\theta'-\theta)\, I(\theta) + o\,(\theta'-\theta)$,

(d) $E_\theta[\ell^{(1)}(X|\theta')]^2 = I(\theta) + o(1)$.

Proof: See Daniels (1961).

Let $\theta' = \theta + \gamma$, where γ is a small (positive or negative) constant. Under Conditions 4.1 and 4.2, we have (Lemma 4.4) $E_\theta \ell^{(1)}(X|\theta + \gamma) = -\gamma\, I(\theta) + o(\gamma)$, and we also know (Lemma 4.1) that if for small enough γ, $P_\theta\{\ell^{(1)}(X|\theta + \gamma) > 0\} > 0$, then for such γ there exists a unique solution, say $t = \tau(\gamma)$, of the equation

$$\frac{d}{dt} E_\theta\left(e^{t\ell^{(1)}(X|\theta+\gamma)}\right) \equiv \varphi^{(1)}_{\ell^{(1)}(X|\theta+\gamma)}(t) = 0 \ .$$

The following lemma gives a basic property of $\tau(\gamma)$.

Lemma 4.5. Assume that Conditions 4.1 and 4.2 hold, and further assume that the following additional conditions hold:

Condition 4.3. For each θ, there exists $G_\theta > 0$, $T_\theta > 0$ such that

$$P_\theta\{\ell^{(1)}(X|\theta+\gamma) > 0\} > 0, \quad \text{all } \gamma,\ -G_\theta \leq \gamma \leq G_\theta\ ,$$

and such that the moment generating function

$$\varphi^{(1)}_{\ell^{(1)}(X|\theta+\gamma)}(t) = E_\theta\left[e^{t\ell^{(1)}(X|\theta+\gamma)}\right]$$

exists for all (t,γ), $-T_\theta \leq t \leq T_\theta$, $-G_\theta \leq \gamma \leq G_\theta$.

Condition 4.4. The second partial $\left(\frac{\partial}{\partial t}\right)^2 \varphi_{\ell^{(1)}(X|\theta+\gamma)}(t)$ is jointly continuous in t and γ for $-T_\theta < t < T_\theta$, $-G_\theta < \gamma < G_\theta$.

[Note: The existence of all partial derivatives $\left(\frac{\partial}{\partial t}\right)^i \varphi_{\ell^{(1)}(X|\theta+\gamma)}(t)$, $i = 1,2,\ldots$, of $\varphi_{\ell^{(1)}(X|\theta+\gamma)}(t)$ for $-T_\theta < t < T_\theta$, $-G_\theta < \gamma < G_\theta$, follows from Condition 4.3. This is not enough, however, to show that $\left(\frac{\partial}{\partial t}\right)^2 \varphi_{\ell^{(1)}(X|\theta+\gamma)}(t)$, is <u>jointly</u> continuous

in t and γ.]

<u>Condition</u> 4.5. The second partial $\frac{\partial}{\partial\gamma}\frac{\partial}{\partial t}\varphi_{\ell^{(1)}(X|\theta+\gamma)}(t)$,

exists and is continuous in (t,γ) for $-T_\theta < t < T_\theta$, $-G_\theta < \gamma < G_\theta$. Futher, $\ell^{(2)}(x|\theta) = \left(\frac{d}{d\theta}\right)^2 \ell(x|\theta)$ exists for all $x \in X$ and all θ , and

$$\frac{\partial}{\partial\gamma}\frac{\partial}{\partial t}\varphi_{\ell^{(1)}(x|\theta+\gamma)}(t) = \int_X \frac{\partial}{\partial\gamma}\ell^{(1)}(x|\theta+\gamma)\exp\{t\ell^{(1)}(x|\theta+\gamma)\}dP_\theta(x),$$

for all $-T_\theta < t < T_\theta$, $-G_\theta < \gamma < G_\theta$.

Under these conditions, there exists a unique single-valued function $\tau(\gamma)$ defined on $-G_\theta < \gamma < G_\theta$ such that

$$\frac{d}{dt}\varphi_{\ell^{(1)}(X|\theta+\gamma)}(t)\Big|_{t=\tau(\gamma)} \equiv \varphi^{(1)}_{\ell^{(1)}(X|\theta+\gamma)}(\tau(\gamma)) = 0 ,$$

and $\tau(\gamma) = \gamma + o(\gamma)$ as $\gamma \to 0$.

<u>Proof</u>: Let $u(t,\gamma) \equiv \varphi^{(1)}_{\ell^{(1)}(x|\theta+\gamma)}(t) = \partial/\partial t\ \varphi_{\ell^{(1)}(x|\theta+\gamma)}(t)$.

From Condition 4.3, it follows (see Parzen (1960,p.216)) that

$$u(t,\gamma) = \int_{\mathcal{X}} \ell^{(1)}(x|\theta+\gamma) \exp\{t\ell^{(1)}(x|\theta+\gamma)\}dP_\theta(x) \ , \qquad (4.23)$$

for $-T_\theta < t < T_\theta$, $-G_\theta < \gamma < G_\theta$. Conditions 4.4 and 4.5 imply that $u(t,\gamma)$ has continuous partial derivatives in t and γ in the region $|t| < T_\theta$, $|\gamma| < G_\theta$. From (4.23), Conditions 4.1 and 4.2, and Lemma 4.4, $u(0,0) = 0$. Further, from Conditions 4.1, 4.2, 4.3, and 4.5, and Lemma 4.4.,

$$\frac{\partial}{\partial t} u(t,\gamma) \Big|_{t=0,\gamma=0} = I(\theta) \neq 0 \quad .$$

Hence, applying the fundamental theorem of implicit functions (see e.g., Taylor (1955)) to $u(t,\gamma)$, we have:

(a) There exists a unique single-valued function $\tau(\gamma)$ defined on $-G_\theta < \gamma < G_\theta$, such that

$$u\Big(\tau(\gamma),\ \gamma\Big) = 0 \qquad (4.24)$$

and

$$\tau(0) = 0 \ . \qquad (4.25)$$

(b) The function $\tau(\gamma)$ has a continuous first derivative $\tau'(\gamma) = (d/d\gamma)\, \tau(\gamma)$ on $-G_\theta < \gamma < G_\theta$ given by

$$\tau'(\gamma) = -\frac{\left.\frac{\partial}{\partial\gamma} u(t,\gamma)\right|_{t=\tau(\gamma)}}{\left.\frac{\partial}{\partial t} u(t,\gamma)\right|_{t=\tau(\gamma)}}$$

$$= -\frac{\left.\frac{\partial}{\partial\gamma} \varphi^{(1)}_{\ell^{(1)}(X|\theta+\gamma)}(t)\right|_{t=\tau(\gamma)}}{\left.\frac{\partial}{\partial t} \varphi^{(1)}_{\ell^{(1)}(X|\theta+\gamma)}(t)\right|_{t=\tau(\gamma)}} \quad . \tag{4.26}$$

Set $\gamma = 0$. Using Lemma 4.4, Conditions 4.4 and 4.5, and (4.26), we find that

$$\tau'(0) = \frac{I(\theta)}{I(\theta)} = 1 \; . \tag{4.27}$$

Since $\tau'(\gamma)$ is continuous in a neighborhood of $\gamma = 0$ and since $\tau(0) = 0$ (Equation 4.25), the result $\tau'(\gamma) = \gamma + o(\gamma)$ follows from (4.27) by the Mean Value Theorem (expanding $\tau(\gamma)$ around $\gamma = 0$). Q.E.D.

<u>Theorem 4.1</u>. Assume that Conditions 4.1 through 4.5 are satisfied, and further that

Condition 4.6. $(d/d\theta) \log g(\theta)$ exists all $\theta \in \Theta$.

Condition 4.7. $\tilde{\ell}_n(s|\theta) = \ell_n(s|\theta) + \log g(\theta)$ is concave in θ for every $n \geq n_o$, n_o some fixed positive integer, and for all $s = (x_1, x_2, \ldots) \in X^{(\infty)}$,

both hold. Then

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|\theta_n^*(\xi) - \theta| \geq \epsilon\} = -\frac{I(\theta)}{2}, \qquad (4.28)$$

for all $\theta \in \Theta$.

Proof: Let $\tilde{\ell}_n^{(1)}(s|\theta) = d/d\theta\, \tilde{\ell}_n(s|\theta)$ for $s \in X^{(\infty)}$.

Conditions 4.2 and 4.6 imply that $\tilde{\ell}^{(1)}(s|\theta)$ exists for all $\theta \in \Theta$. From Condition 4.7 we know that

$$\tilde{\ell}_n^{(1)}(s|\theta') = \begin{cases} \geq 0, \text{ when } \theta' \leq \theta_n^*(s), \\ \leq 0, \text{ when } \theta' \geq \theta_n^*(s), \end{cases} \qquad (4.29)$$

for all $s \in X^{(\infty)}$, $\theta' \in \Theta$. Hence we have

$$P_\theta\{\tilde{\ell}_n^{(1)}(\xi|\theta') > 0\} \leq P_\theta\{\theta^* \geq \theta'\} \leq P_\theta\{\tilde{\ell}_n^{(1)}(\xi|\theta') \geq 0\}. \qquad (4.30)$$

Taking $\theta' = \theta + \epsilon$ and $a_n = -\frac{1}{n}\frac{\partial}{\partial\theta} \log g(\theta)\Big|_{\theta=\theta'}$, $\epsilon > 0$, we have

$$P_\theta\{\frac{1}{n}\sum_{i=1}^{n} \ell^{(1)}(X_i|\theta+\epsilon) > a_n\} \leq P_\theta\{\theta_n^* - \theta \geq \epsilon\}$$

$$\leq P_\theta\{\frac{1}{n}\sum_{i=1}^{n} \ell^{(1)}(X_i|\theta+\epsilon) \geq a_n\} .$$

Since $\lim_{n\to\infty} a_n \to 0$, it follows from the continuity of $\rho(a)$ at $a = 0$ and from the Bernstein-Chernoff-Bahadur theorem (Lemma 4.3), that

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta\{\theta_n^*(\xi) - \theta \geq \epsilon\} = \log \rho_1(\theta,\epsilon,0), \tag{4.31}$$

where

$$\rho_1(\theta,\epsilon,0) = \inf_{t\geq 0} \varphi_{\ell^{(1)}(X|\theta+\epsilon)}(t) \quad .$$

In a similar way, we have

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta\{\theta_n^*(\xi) - \theta \leq -\epsilon\} = \log \rho_2(\theta,\epsilon,0), \tag{4.32}$$

where

$$\rho_2(\theta,\epsilon,0) = \inf_{t\leq 0} \varphi_{\ell^{(1)}(X|\theta-\epsilon)}(t) \quad .$$

By Condition 4.4, Lemma 4.4 and Lemma 4.5, we have

$$\log \rho_1(\theta,\epsilon,0) = \log \inf_{t\geq 0} \varphi_{\ell^{(1)}(X|\theta+\epsilon)}(t)$$

$$= \log \varphi_{\ell^{(1)}(X|\theta+\epsilon)}(\tau(\epsilon))$$

$$= k_1(\theta,\epsilon)\tau(\epsilon)+k_2(\theta,\epsilon)\tau^2(\epsilon)+o(\tau^2(\epsilon))$$

$$= -\frac{I(\theta)}{2}\epsilon^2 + o(\epsilon^2)\ , \tag{4.33}$$

where $k_1(\theta,\epsilon)$ and $k_2(\theta,\epsilon)$ are the first two cumulants of the random variable $\ell^{(1)}(X|\theta+\epsilon)$. Hence

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{\epsilon^2 n}\log P_\theta\{\theta_n^* - \theta \geq \epsilon\} = -\frac{I(\theta)}{2} \tag{4.34}$$

follows immediately from (4.33). By a symmetric argument, we have

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{\epsilon^2 n}\log P_\theta\{\theta_n^* - \theta \leq -\epsilon\} = -\frac{I(\theta)}{2}\ . \tag{4.35}$$

Since $\lim_{n\to\infty}\frac{1}{n}\log P_\theta\{|\theta_n^* - \theta| \geq \epsilon\} = \log \rho(\theta,\epsilon,0)$, where $\rho(\theta,\epsilon,0) = \max(\rho_1(\theta,\epsilon,0),\ \rho_2(\theta,\epsilon,0))$, it follows from (4.33) and (4.35) that

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{\epsilon^2 n}\log P_\theta\{|\theta_n^* - \theta| \geq \epsilon\} = \frac{-I(\theta)}{2} \quad .$$

This completes our proof.

<u>Remark</u>. The part of the above proof that shows that

$\lim_{n\to\infty} n^{-1}\log P_\theta\{|\theta_n^*(\xi) - \theta| \geq \epsilon\} = \log \rho(\theta,\epsilon,0)$ depends greatly on Condition 4.7. This condition is rather restrictive, but when $g(\theta)$ is log-concave, it is satisfied for some distributions (such as the normal distribution with known variance, and other members of the Koopmans-Darmois class of probability distributions) of interest in both statistical theory and statistical practice. A somewhat different set of strong regularity conditions has been given by Bahadur (1960). Under his conditions, Bahadur proved that for any given small $\gamma > 0$ there exists β, $0 < \beta < 1$, such that for all sufficiently large n, the probability of the event that the function $\ell_n(s|\theta)$ is a concave function of θ and that the M.L.E. $\hat{\theta}_n(s)$ exists uniquely in the neighborhood $\{\theta' :\ |\theta'-\theta| < \gamma\}$ of θ is a number larger than $1-\beta^n$. Using this fact, Bahadur (1967) proved that

$$\lim_{n\to\infty} \frac{1}{n} P_\theta\{|\hat{\theta}_n(\xi) - \theta| \geq \epsilon\} = \log \rho(\theta, \epsilon, 0). \tag{4.36}$$

If we add to Bahadur's (1960) conditions some extra regularity conditions on $g(\theta)$ - such as the conditions that $(d/d\theta)^i \log g(\theta)$, $i = 1,2,3$, exist and $\sup_{\theta\in\Theta} |(d/d\theta)^i \log g(\theta)| < \infty$, $i = 1,2,3$ - then a modification of Bahadur's (1960, 1967) proof shows that

$$\lim_{n\to\infty} \frac{1}{n} \log P_\theta\{|\theta_n^*(\xi) - \theta| \geq \epsilon\} = \log \rho(\theta, \epsilon, 0). \tag{4.37}$$

Since Bahadur's regularity conditions are quite extensive (they involve the finiteness of three moment-generating functions in a neighborhood of the origin, the usual regularity conditions needed to prove the Cramer-Rao inequality, etc.) and a proof of (4.37) differs from Bahadur's proof of (4.36) in only minor details, the conditions and the proof will not be given here.

3. A MEASURE OF ASYMPTOTIC RELATIVE EFFICIENCY FOR CONSISTENT ESTIMATORS

1. Introduction

Let $\mathfrak{X}$ be a sample space with a sigma-field $\mathcal{B}$ of subsets, and let μ be a σ-finite measure defined on $\mathcal{B}$. Let $\{P_\theta: \theta \in \Theta\}$ be a collection of probability measures on $\mathcal{B}$ and dominated by μ, where Θ is an open subinterval of the real line R^1. Let $\xi = (X_1, X_2, \ldots.)$ be a sequence of i.i.d. random variables with common probability measure P_θ, $\theta \in \Theta$. Since P_θ is dominated by μ, there exists a density $f(x|\theta) = dP_\theta/d\mu$ for P_θ with respect to μ; that is, for any measurable subset $B \in \mathcal{B}$,

$$P_\theta(B) = \int_B f(x|\theta)d\mu(x) \quad .$$

In what follows, we denote the Cartesian product space $(\mathfrak{X}x\mathfrak{X}x\ldots, \mathcal{B}x\mathcal{B}x\ldots)$ by $(\mathfrak{X}^{(\infty)}, \mathcal{B}^{(\infty)})$ and the product probability measure $P_\theta x P_\theta x \ldots$ by $P_\theta^{(\infty)}$. Thus ξ is distributed according to $P_\theta^{(\infty)}$ on $(\mathfrak{X}^{(\infty)}, \mathcal{B}^{(\infty)})$. Whenever our intention is clear from the context, we shall leave off the superscripts. Thus, $P_\theta^{(\infty)}$ will often be written P_θ.

Let $\{T_n\}$, $T_n(\xi) = T_n(X_1, X_2, \ldots X_n)$, be a consistent sequence of estimators for θ, and let $\epsilon > 0$ be any fixed, positive constant. Basu (1956) suggests using the rate at which the probability $P_\theta\{|T_n - \theta| \geq \epsilon\}$ tends to zero as $n \to \infty$ as a criterion of asymptotic performance (efficiency) for $\{T_n\}$. Assuming that the probability distribution of T_n has a density function with respect to a σ-finite measure (e.g., Lebesgue measure) on R^1, we can distinguish three levels of difficulty in evaluating Basu's criterion. In Level 1, we do not have a closed mathematical form either for the density function or for the cumulative distribution function (c.d.f.) of T_n (possibly due to the complexity of the mathematical form of T_n itself, or due to some other reason). At Level 2, we have a closed form for the density function of T_n, but find the c.d.f. too difficult to evaluate exactly. Finally in Level 3, we have a closed form for the c.d.f. of T_n. If we are in Level 3, we can (at least in theory) directly evaluate

$$P_\theta\{|T_n - \theta| \geq \epsilon\} = 1 - P_\theta\{T_n < \theta + \epsilon\} + P_\theta\{T_n \leq \theta - \epsilon\}$$

and obtain the exact rate at which this quantity tends to zero as $n \to \infty$. Hence, in this case, Basu's criterion can be applied directly. However, in most statistical problems, we are either at Level 1 or Level 2. In such cases, the exact rate at which $P_\theta\{|T_n-\theta| \geq \epsilon\}$ tends to zero is usually difficult to evaluate. If this is the case, we might at least hope to obtain bounds for the rate (or approximate rate) at which $P_\theta\{|T_n-\theta| \geq \epsilon\}$ tends to zero as $n \to \infty$.

In Section 2, we show under very weak regularity conditions on the probability structure $\{P_\theta: \theta \in \Theta\}$ that for small enough $\epsilon > 0$,

$$\varliminf_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} \geq - \inf_{\theta'} \{K(\theta',\theta): |\theta'-\theta| > \epsilon\}$$

holds for all consistent sequences $\{T_n\}$ of estimators, where $K(\theta',\theta)$ is the Kullback-Leibler information obtained from P_θ and $P_{\theta'}$. We also give another lower bound for $n^{-1} \log P_\theta\{|T_n-\theta| \geq \epsilon\}$ for use in cases when we have a closed form for the density function of T_n, but not necessarily a closed form for the c.d.f. of T_n.

In Section 3 we discuss a measure of asymptotic relative efficiency (A.R.E.) for two sequences $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ of consistent estimators. This measure is based on the rates at which $n^{-1} \log P_\theta\{|T_n^{(i)} - \theta| \geq \epsilon\}$ tend to zero as $n \to \infty$ (and as $\epsilon \to 0$), $i = 1,2$. In Section 4, we use the measure of A.R.E. discussed in Section 3 to compare the sample mean and sample median as estimators of a location parameter.

2. A Lower Bound for the Rate of Convergence.

We are given a sequence $\{X_n\}$ of i.i.d. random variables, each with probability measure P_θ, $\theta \in \Theta$. Recalling that each P_θ has a density $f(x|\theta)$ with respect to μ, define

$$K(\theta',\theta) = \int_{\mathcal{X}} f(x|\theta') \log \frac{f(x|\theta')}{f(x|\theta)} \, d\mu(x), \tag{2.1}$$

when P_θ and $P_{\theta'}$ are absolutely continuous with respect to one another, and let $K(\theta',\theta) = \infty$ otherwise. It can be shown that $K(\theta',\theta) \geq 0$ for all $\theta',\theta \in \Theta$. If for $\theta,\theta' \in \Theta$, $\theta \neq \theta'$, we have

$$\mu\left(\{x: \; f(x|\theta) \neq f(x|\theta')\}\right) > 0 ,$$

then $K(\theta',\theta) = 0$ if and only if $\theta = \theta'$. The quantity $K(\theta',\theta)$ is called the Kullback-Liebler information.

For each $\theta,\theta' \in \Theta$ and for any small number $\delta > 0$, define the ($\beta^{(\infty)}$-measurable) set

$$B_n(\theta',\theta,\delta) = \left\{(x_1,x_2,\ldots.): \prod_{i=1}^{n} f(x_i|\theta) \geq \gamma^n(\theta',\theta,\delta) \prod_{i=1}^{n} f(x_i|\theta')\right\},$$

(2.2)

where

$$\gamma(\theta',\theta,\delta) = \exp\{-K(\theta',\theta) - \delta\} . \tag{2.3}$$

Also let

$$U_n(T_n,\theta,\epsilon) = \{(x_1,x_2,\ldots): \ |T_n(x_1,x_2,\ldots,x_n) - \theta| \geq \epsilon\}.$$

Before we prove our main result, we need to prove the following lemma.

Lemma 2.1. If $\{T_n\}$ is a sequence of consistent estimators and if for $|\theta'-\theta| > \epsilon$ we have $K(\theta',\theta) < \infty$, then

$$\lim_{n\to\infty} \left[P_{\theta'}\{U_n(T_n,\theta,\epsilon)\} - P_{\theta'}\{B_n^c(\theta',\theta,\delta)\}\right] = 1. \tag{2.4}$$

Proof. Since $|\theta'-\theta| > \epsilon$ and since $\{T_n\}$ is a consistent sequence of estimators for θ,

$$\lim_{n\to\infty} P_{\theta'}\ \{U_n(T_n,\theta,\epsilon)\} = 1 . \tag{2.5}$$

Further, since $K(\theta',\theta) = E_{\theta'}\left(\log\ [f(x|\theta')/f(x|\theta)]\right) < \infty$, and since from (2.2),

$$B_n(\theta',\theta,\delta) = \left\{(x_1,x_2,\ldots): \ \frac{1}{n}\sum_{i=1}^{n}\log\ [f(x_i|\theta')/f(x_i|\theta)] \right.$$

$$\left. \leq K(\theta',\theta) + \delta \right\},$$

it follows from the strong law of large numbers that

$$\lim_{n\to\infty} P_{\theta'}\ \{B_n^c(\theta',\theta,\delta)\} = 0\ . \tag{2.6}$$

Equation (2.4) now follows from (2.5) and (2.6). Q.E.D.

We are now ready to prove the following theorem.

Theorem 2.1. If $\{T_n\}$ is a consistent sequence of estimators, then for each $\theta \in \Theta$ and each small $\epsilon > 0$,

$$\varliminf_{n\to\infty} \frac{1}{n}\log\ P_\theta\{|T_n-\theta| \geq \epsilon\} \geq -\inf_{\theta'}\ \{K(\theta',\theta): \ |\theta'-\theta| > \epsilon\}. \tag{2.7}$$

Proof. If $\inf_{\theta'} \{K(\theta',\theta): \ |\theta'-\theta| > \epsilon\} = \infty$, then (2.7) holds trivially. Otherwise, there exists at least one θ', $|\theta'-\theta| > \epsilon$, for which $K(\theta',\theta) < \infty$. Without loss of generality, we can restrict attention only to those θ' for which $|\theta'-\theta| > \epsilon$ and $K(\theta',\theta) < \infty$.

Now

$$P_\theta\{U_n(T_n,\theta,\epsilon)\} \geq P_\theta\{U_n \cap B_n\} = \int_{U_n \cap B_n} \prod_{i=1}^{n} f(x_i|\theta)\, d\mu(x_i), \qquad (2.8)$$

where $U_n \equiv U_n(T_n,\theta,\epsilon)$, $B_n \equiv B_n(\theta',\theta,\delta)$. From the definition of B_n,

$$\int_{U_n \cap B_n} \prod_{i=1}^{n} f(x_i|\theta) d\mu(x_i) \geq \gamma^n(\theta',\theta,\delta) \int_{U_n \cap B_n} \prod_{i=1}^{n} f(x_i|\theta') d\mu(x_i)$$

$$= \gamma^n(\theta',\theta,\delta)\, P_{\theta'}\{U_n \cap B_n\}\ , \qquad (2.9)$$

and since $U_n \cap B_n = U_n \sim (U_n \cap B_n^c)$, $P_{\theta'}\{U_n \cap B_n\} \geq P_{\theta'}\{U_n\} - P_{\theta'}(B_n^c)$.

Hence, from (2.8), (2.9) and Lemma 2.1, we conclude that for all sufficiently large n, and any d, $0 < d < 1$,

$$P_\theta\{U_n(T_n,\theta,\epsilon)\} \geq \left[\gamma^n(\theta',\theta,\delta)\right] d. \qquad (2.10)$$

Hence for all sufficiently large n,

$$\frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} \geq - K(\theta',\theta) - \delta$$

holds for all θ', $|\theta'-\theta| > \epsilon$. Since $\delta > 0$ is an arbitrary constant,

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} \geq - \inf_{\theta'} \{K(\theta',\theta): \ |\theta'-\theta| > \epsilon\}.$$

This completes the proof.

A result similar to Theorem 2.1 has been proved by Bahadur (1960) under somewhat stronger conditions. Similar conditions and arguments are used by Bahadur (1967) in a different context (hypothesis-testing).

If we assume that $K(\theta',\theta)$ is a continuous function of θ' in a neighborhood of $\theta + \epsilon$ and in a neighborhood of $\theta-\epsilon$ (this assumption holds under rather weak conditions on the density functions $f(x|\theta)$, $\theta \in \Theta$ - in particular, this assumption holds under Kullback's "regularity conditions" mentioned later in this section), then

$$\inf_{\theta'} \{K(\theta',\theta): \ |\theta'-\theta| > \epsilon\} = \inf_{\theta'} \{K(\theta',\theta): \ |\theta'-\theta| \geq \epsilon\} .$$

Since $\inf_{\theta'} \{K(\theta',\theta): \ |\theta'-\theta \geq \epsilon\} \leq \min \{K(\theta+\epsilon,\theta), K(\theta-\epsilon,\theta)\}$, it follows from Theorem 2.1 that under the above-mentioned continuity condition on $K(\theta',\theta)$,

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} \geq - \min \{K(\theta+\epsilon,\theta),\ K(\theta-\epsilon,\theta)\}. \tag{2.11}$$

Under certain "regularity conditions" (involving the first three partial derivatives of $\log f(x|\theta)$ with respect to θ), Kullback (1958, p.26) proves that

$$\begin{aligned} K(\theta + \epsilon,\ \theta) &= \tfrac{1}{2}\, I(\theta)\ \epsilon^2 + o(\epsilon^2)\ , \\ K(\theta - \epsilon,\ \theta) &= \tfrac{1}{2}\, I(\theta)\ \epsilon^2 + o(\epsilon^2)\ , \end{aligned} \tag{2.12}$$

as $\epsilon \to 0$, where $I(\theta)$ is Fisher's information. From (2.11) and (2.12), we obtain the following result.

<u>Theorem 2.2</u>. If $\{T_n\}$ is any consistent sequence of estimators, and if the "regularity conditions" of Kullback (1968, p.26) hold, then

$$\underline{\lim}_{\epsilon\to 0}\ \underline{\lim}_{n\to\infty} \frac{1}{n\epsilon^2} \log P_\theta\{|T_n-\theta| > \epsilon\} \geq - \frac{I(\theta)}{2}\ . \tag{2.13}$$

The result (2.13) has also been obtained by Bahadur (1960,1967). The proof given here, however, differs somewhat from those given earlier by Bahadur.

Theorems 2.1 and 2.2 are useful when we are at level

1 (as discussed in Section 1) and do not know the exact forms of the density or c.d.f. of T_n, $n = 1,2,\ldots$. Suppose, instead, we are at Level 2 and know the exact form of the density function $f_n(t|\theta)$ of $T_n(\xi)$, $n = 1,2,\ldots$, but not the exact form of the c.d.f. of T_n. Assume that for each θ' $\theta \in \Theta$, there exists a constant $R(\theta',\theta|\theta')$ such that

$$P_{\theta'}\left[\lim_{n\to\infty} \frac{1}{n} \log \left[\frac{f_n(T_n(\xi)|\theta')}{f_n(T_n(\xi)|\theta)}\right] = R(\theta',\theta|\theta')\right] = 1 \ . \qquad (2.14)$$

It can be shown that $R(\theta',\theta|\theta') \geq 0$ all θ', $\theta \in \Theta$, and that $R(\theta',\theta|\theta') = 0$ if and only if $\theta' = 0$. Consider the measurable set

$$A_n(\theta',\theta,\delta) = \{(x_1,x_2,\ldots): \ f_n(T_n(x_1,x_2,\ldots)|\theta) \geq e^{-n(R(\theta',\theta|\theta')+\delta)}$$

$$f_n(T_n(x_1,x_2,\ldots)|\theta')\},$$

where $\delta > 0$ is any positive constant. Stepping through the proofs of Lemma 2.1 and Theorem 2.1 with $B_n(\theta',\theta,\delta)$ replaced by $A_n(\theta',\theta,\delta)$, we can establish the following theorem.

<u>Theorem 2.3</u>. If $\{T_n\}$ is a consistent sequence of estimators,

if T_n has density $f_n(t|\theta)$ with respect to some σ-finite measure on R^1 for all $\theta \in \Theta$, $n = 1,2,\ldots$, and if (2.14) holds for all $\theta, \theta' \in \Theta$, then

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n(\xi)-\theta| \geq \epsilon\} \geq - \inf_{\theta'} \{R(\theta',\theta|\theta'): |\theta'-\theta| > \epsilon\}. \tag{2.15}$$

It is of obvious interest to ask whether (2.15) is actually an equality when (2.14) holds; that is, whether (2.14) implies that

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n-\theta| \geq \epsilon\} = - \inf_{\theta'} \{R(\theta',\theta|\theta'): |\theta'-\theta| > \epsilon\}. \tag{2.16}$$

For a variety of reasons, we feel that (2.14) should imply (2.16), and in special cases we have been able to establish this result. A general assertion of this implication, however, remains only a conjecture.

3. A Measure of Asymptotic Relative Efficiency For Consistent Estimators

Let $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ be two consistent sequences of estimators for θ. For given $\epsilon > 0$, we denote

$$\lambda_i(\epsilon,\theta) = \varliminf_{n\to\infty} \frac{1}{n} \log P_\theta\{|T_n^{(i)}(\xi) - \theta| \geq \epsilon\}, \tag{3.1}$$

$i = 1,2$. For cases in which one of the quantities $\lambda_1(\epsilon,\theta)$ or $\lambda_2(\epsilon,\theta)$ is non-zero (i.e., at least one of the estimators $\{T_n^{(1)}\}$ or $\{T_n^{(2)}\}$ has probability of falling outside of the interval $(\theta-\epsilon, \theta+\epsilon)$ which converges exponentially fast to zero as $n \to \infty$), the ratio

$$\text{eff}_{1,2}(\epsilon,\theta) = \frac{\lambda_1(\epsilon,\theta)}{\lambda_2(\epsilon,\theta)}, \tag{3.2}$$

serves as a meaningful measure of asymptotic relative efficiency of $\{T_n^{(1)}\}$ relative to $\{T_n^{(2)}\}$. As we have remarked in Section 1, this measure of efficiency is frequently difficult to compute (if we do not know the c.d.f.'s of $T_n^{(1)}$ and $T_n^{(2)}$, $n = 1,2,\ldots$). One additional difficulty lies in the fact that $\text{eff}_{1,2}(\epsilon,\theta)$ depends on $\epsilon > 0$. Besides complicating the analysis, this dependence of the efficiency on a quantity other than the true parameter is somewhat non-intuitive, and leads to the question of how to choose $\epsilon > 0$ so as to give a valid comparison between the two sequences of estimators. Except when ϵ is given as part of the structure of a decision-theoretic model for the problem of estimating θ (e.g., we have a loss function $L(\theta,a)$ which equals 0 if $|\theta-a| < \epsilon$ and 1 otherwise), it is not easy to see how to choose ϵ.

From a commonly stated large-sample point of view, however, we would not be in a large-sample situation unless we demanded very high accuracy (that is, ϵ is very small). Thus, it is reasonable to consider the quantities

$$\begin{aligned} \lambda_1(\theta) &= \lim_{\epsilon \to 0} \lambda_1(\epsilon,\theta) , \\ \lambda_2(\theta) &= \lim_{\epsilon \to 0} \lambda_2(\epsilon,\theta) , \end{aligned} \tag{3.3}$$

and to define a measure of asymptotic relative efficiency for $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ to be

$$e_{1,2}(\theta) = \frac{\lambda^{1}(\theta)}{\lambda_2(\theta)} . \tag{3.4}$$

Strictly speaking, this measure is not an exact measure of asymptotic relative efficiency - although if $\lim_{\epsilon\to 0} \lambda_1(\epsilon,\theta)$ and $\lim_{\epsilon\to 0} \lambda_2(\epsilon,\theta)$ exist, $e_{1,2}(\theta)$ has almost all of the properties of a legitimate measure of asymptotic efficiency since then $e_{1,2}(\theta) = \lim_{\epsilon\to 0} \mathrm{eff}_{1,2}(\epsilon,\theta)$. Neither the measure (3.2) nor the measure (3.4) coincide in general with the classical measure of A.R.E.(asymptotic relative efficiency) based on the asymptotic variances $V_i(\theta)$ of $\{T_n^{(i)}\}$, $i = 1,2$, even when $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are both consistent asymptotically normal estimators.

Even after the reduction in complexity achieved by taking the limit of (3.2) as $\epsilon \to 0$, we are still left with a measure of A.R.E. that is (in most cases) exceedingly difficult to obtain. Research is in progress towards developing a general theory that will allow us to calculate the quantities $\lambda_i(\theta)$. At present, results are known only for maximum likelihood estimators (Bahadur (1960,1967) and, more generally, for maximum probability estimators (see Fu (1971)). In the next section, we show how to apply the measure of A.R.E. $e_{1,2}(\theta)$ to the comparison of the performance of the sample mean and sample median when estimating the location parameter of a symmetric distribution.

4. Comparison of A.R.E. of Sample Mean and Sample Median for the Location Parameter of a Symmetric Distribution.

Let Θ and $\mathcal{X}$ be the real line R^1. Let $X_1, X_2, \ldots$ be i.i.d. random variables having probability distribution P_θ which is absolutely continuous with respect to Lebesgue measure on R^1, $\theta \in \Theta$. Assume further that θ is a location parameter - that is, the density function $f(x|\theta)$ has the form

$$f(x|\theta) = g(x-\theta), \qquad \text{all } \theta \in \Theta,\ x \in \mathcal{X} .$$

Assume that $g(u)$ is symmetric about $u = 0$; i.e., $g(u) = g(-u)$, all $u \in R^1$. For convenience, assume that we can take only odd sample sizes $n = 2m - 1$, $m = 1,2,\ldots$. Let $\bar{X}_n = n^{-1} \sum_{i=1} X_i$ be the sample mean, and let $X_{(m)}$, $m = (n+1)/2$, be the sample median. Both $\bar{X}_n$ and $X_{(m)}$ are consistent estimators of θ when $E_\theta|X| < \infty$, all $\theta \in \Theta$, and when $g(0) > 0$.

If we compare $\bar{X}_n$ and $X_{(m)}$ by the classic measure of A.R.E. based on asymptotic variances, it is known that in some cases (e.g., when $X_1, X_2, \ldots$ are normally distributed) $\bar{X}_n$ is more efficient than $X_{(m)}$, while in other cases $X_{(m)}$ is the more efficient (e.g., when the X_i's are Cauchy distributed). Here, we compare $\bar{X}_{(n)}$ and $X_{(m)}$ through use of the measure of A.R.E. defined in Section 3.

Let

$$\lambda_1(\theta) = \underline{\lim_{\epsilon\to 0}}\ \underline{\lim_{n\to\infty}} \frac{1}{\epsilon^2 n} \log P_\theta\{|\bar{X}_n - \theta| \geq \epsilon\} ,$$

$$\lambda_2(\theta) = \underline{\lim_{\epsilon\to 0}}\ \underline{\lim_{n\to\infty}} \frac{1}{\epsilon^2 n} \log P_\theta\{|X_{(m)} - \theta| \geq \epsilon\} , \tag{4.1}$$

where $m = (n+1)/2$.

Theorem 4.1. Assume that

Condition 4.1. There exists $T \geq 0$ such that

$$\varphi(t) = \int_{-\infty}^{\infty} e^{tx} g(x)dx \tag{4.2}$$

exists (is finite) for all t, $|t| \leq T$.

Then,

$$\lambda_1(\theta) = \lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{\epsilon^2 n} \log P_\theta\{|\bar{X}_n - \theta| \geq \epsilon\} = -\frac{1}{2\sigma^2} , \tag{4.3}$$

where

$$\sigma^2 = \int_{-\infty}^{\infty} x^2\, g(x)dx .$$

Proof. Since θ is a location parameter,

$$P_\theta\{|\overline{X}_n-\theta| \geq \epsilon\} = P_{\theta=0}\{|\overline{X}_n-\theta| \geq \epsilon\},$$

so that without loss of generality we can assume that $\theta = 0$ and that $X_1, X_2, \ldots$ are i.i.d. with density $g(x)$. Condition 4.1 and the symmetry of $g(u)$ about $u = 0$ implies by the theorem of Bernstein, Chernoff, and Bahadur (Chapter 2, Lemma 4.3) that

$$\lim_{n\to\infty} \frac{1}{n} P_0\{|\overline{X}_n| \geq \epsilon\} = -\log \rho_1(\epsilon) , \tag{4.4}$$

where

$$\rho_1(\epsilon) = \inf_{t\geq 0} e^{-t\epsilon}\varphi(t) = \inf_{t\geq 0} \varphi_{X-\epsilon}(t)$$

and

$$\varphi_{X-\epsilon}(t) = \int_{-\infty}^{\infty} e^{tu}g(u+\epsilon)du .$$

Since $\varphi(t)$ exists for $|t| \leq T$, all moments of the distribution $g(x)$ exist. Since $g(u) = g(-u)$, all u, all odd moments of $g(.)$ are zero; hence,

$$\int_{-\infty}^{\infty} x\, g(x)dx = 0.$$

Since $g(u)$ is symmetric about $u = 0$, we know that there exists an $\eta > 0$ such that $P_0\{X > \eta\} > 0$, $P_0\{X < -\eta\} < 0$. Hence, by Lemma 4.2 of Chapter 2, for small enough $\epsilon > 0$, there exists a unique $\tau(\epsilon) > 0$ such that

$$\frac{d}{dt}\left[\varphi(t)\, e^{-\epsilon t}\right]_{t=\tau(\epsilon)} = 0 ,$$

$\tau(\epsilon)$ is a continuous function of ϵ in a small enough neighborhood of $\epsilon = 0$, and $\tau(\epsilon) \to 0$ as $\epsilon \to 0$. Since it is easily shown that $(d/dt)^2\varphi_{X-\epsilon}(t)$ exists and is continuous in (t, ϵ) for all $|t| < T$, small enough ϵ, it follows by a use of Taylor's expansion and the Mean Value Theorem that

$$\tau(\epsilon) = \frac{\epsilon}{\left(\frac{d}{dt}\right)^2 \varphi_{X-\epsilon}\left(h\tau(\epsilon)\right)}$$

where $0 < h < 1$. Since $\tau(\epsilon) \to 0$ as $\epsilon \to 0$, it follows that $(d/dt)^2 \varphi_{X-\epsilon}(h\tau(\epsilon)) \to (d/dt)^2\varphi(0)$ and hence

$$\tau(\epsilon) = \frac{\epsilon}{\left(\frac{d}{dt}\right)^2 \varphi(t)\Big|_{t=0}}(1+o(1)) = \frac{\epsilon}{\sigma^2}(1+o(1))$$

as $\epsilon \to 0$. But expanding $\log \varphi_{X-\epsilon}(t)$ in terms of the cumulants of $X-\epsilon$, we obtain

$$\log \rho_1(\epsilon) = \log \varphi_{X-\epsilon}(\tau(\epsilon)) = K_1(\epsilon)\tau(\epsilon) + K_2(\epsilon)\frac{\tau^2(\epsilon)}{2} + o(\tau^2(\epsilon)) \quad ,$$

where $K_1(\epsilon) = -\epsilon$ and $K_2(\epsilon) = \sigma^2+\epsilon^2$ are the first two cumulants of $X-\epsilon$. Thus,

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2} \log \rho_1(\epsilon) = -\frac{1}{2\sigma^2} \quad . \qquad (4.5)$$

Combining (4.4) with (4.5) gives us (4.3). Q.E.D.

Theorem 4.2. Assume that $g(0) > 0$ and that

Condition 4.2 $\frac{d}{dx} g(x)$ exists and is continuous in x for all $x \epsilon X$.

Then

$$\lambda_2(\theta) \geq -2g^2(0) \quad . \qquad (4.6)$$

<u>Proof</u> The density of $X_{(m)}$ is given by

$$h(x|\theta) = \frac{(2m-1)!}{\left[(m-1)!\right]^2} F^{m-1}(x|\theta)\, f(x|\theta)\left(1-F(x|\theta)\right)^{m-1} .$$

Since $X_{(m)} \xrightarrow{a.s.} \theta$, all θ, it is not hard to show that

$$P_\theta\{\lim_{n\to\infty} \frac{1}{n} \log \left[h(X_{(m)}|\theta')/h(X_{(m)}|\theta)\right] = R(\theta',\theta|\theta')\} = 1 ,$$

where

$$\begin{aligned} R(\theta',\theta|\theta') &= \tfrac{1}{2} \log \frac{1}{4F(\theta'|\theta)\,(1-F(\theta'|\theta))} \\ &= \tfrac{1}{2} \log \frac{1}{4F(\theta'-\theta|0)\,(1-F(\theta'-\theta|0))} . \end{aligned} \tag{4.7}$$

Hence, Theorem 2.3 of Section 2 shows that

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P_\theta\{|X_{(m)}-\theta| \geq \epsilon\} \geq - \inf_{\theta'} \{R(\theta',\theta|\theta') : |\theta'-\theta| \geq \epsilon\}$$

$$\begin{aligned} &= - \inf_{|\theta'| \geq \epsilon} \left[\tfrac{1}{2} \log \frac{1}{4F(\theta'|0)\,(1-F(\theta'|0))}\right] \\ &\geq \tfrac{1}{2} \log \left[4\, F(\epsilon|0)\, (1-F(\epsilon|0))\right] . \end{aligned} \tag{4.8}$$

Since $d/dx\ g(x)$ exists and is continuous, it follows that

$$F(x|0) = \int_{-\infty}^{x} g(u)du$$

is twice differentiable and that $(d/dx)^2 F(x|0)$ is continuous in x. We conclude that $\log[F(x|0)(1-F(x|0))]$ is twice diffentiable in x and that the second derivative of $\log\ [F(x|0)\ (1-F(x|0))]$ is continuous in a neighborhood of $x = 0$. Further since $F(0|0) = \frac{1}{2}$, it follows that $\log\ [4F(0|0)\ (1-F(0,0))] = 0$,

$$\frac{d}{dx} \log F(x|0)\ (1-F(x|0))\Big|_{x=0} = 0 ,$$

$$\left(\frac{d}{dx}\right)^2 \log F(x|0)\ (1-F(x|0))\Big|_{x=0} = -\ 8\ g^2(0) ,$$

and hence (expanding $\log\ [4F(x|0)\ (1-F(x|0))]$ in a Taylor Series about $x = 0$),

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2} \log\ [4F(\epsilon|0)\ (1-F(\epsilon|0))] = -4\, g^2(0) . \qquad (4.9)$$

From (4.8) and (4.9), (4.6) follows. Q.E.D.

Theorem 4.3. Under Conditions 4.1 and 4.2, and the further condition that $g(0) > 0$,

$$e_{1,2}(\theta) = \frac{\lambda_1(\theta)}{\lambda_2(\theta)} \leq \frac{1}{4\sigma^2 g^2(0)} \quad . \qquad (4.10)$$

Proof. This result is a direct consequence of Theorems 4.1 and 4.2. Q.E.D.

We recognize the upper bound to $e_{1,2}(\theta)$ in (4.10) as being the classical measure of A.R.E. based on asymptotic variances (the so-called Pitman efficiency) for $\bar{X}_n$ and $X_{(m)}$. Although we have not been able to verify our conjecture in all cases, we believe that under Conditions 4.1 and 4.2, the inequality (4.6) becomes an equality. If this is the case, the classical Pitman measure of A.R.E. and the measure A.R.E. introduced in Section 3 coincide. This assertion is certainly true when $X_1, X_2, \ldots$ are i.i.d. normally distributed with mean θ and variance 1 (where $\lambda_1(\theta) = -\frac{1}{2}$, $\lambda_2(\theta) = -1/\pi$, and $e_{1,2}(\theta) = \pi/2$).

BIBLIOGRAPHY

Bahadur, R. R. (1960). Asymptotic efficiency of test and estimates. Sankhyā 22 229-252.

Bahadur, R. R. (1964). On Fisher's bound for asymptotic variances. Ann. Math. Statist. 35 1545-1552.

Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. Ann. Math. Statist. 38 303-324.

Bahadur, R. R. (1968). Lecture Notes at Department of Statistics, Florida State University.

Basu, D. (1956). On the concept of asymptotic efficiency. Sankhyā 17 193-196.

Berk, R. H. (1966). Limiting behaviour of posterior distributions when the model is incorrect. Ann. Math. Statist. 37 51-58.

Bickel, P. and Yahav, J. A. (1966). The asymptotic theory of Bayes' solutions. Technical Report, Department of Statistics, Stanford University.

Bickel, P. and Yahav, J. A. (1968). Asymptotically optimal Bayes; and minimax Procedures in sequential estimation. Ann. Math. Statist. 39 442-456.

Chao, M. T. (1970). The asymptotic behavior of Bayes' estimators. _Ann. Math. Statist._ _41_ 601-608.

Chernoff, H. (1952) A measure of asymptotic efficiency for tests of an hypothesis based on sum of observations. _Ann. Math. Statist._ _23_ 493-502.

Cramer, H. (1946). _Mathematical methods of statistics._ Princeton University Press, Princeton, N.J.

Daniels, H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. _Proc. of the Fourth Berkeley Symposium on Math. Statist. and Prob._ _1_ 151-164.

Fu, James C. (1971). Asymptotic properties of a certain estimator related to the maximum likelihood estimator. _Ann. Math. Statist._ (Abstract). To appear.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. _Proc. Fifth Berkeley Symposium on Math. Statist. and Prob._ _1_ 221-233.

Johns, M. V. (1968). _Lecture Notes at Department of Statistics, The Johns Hopkins University._

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist. 27 887-906.

Kullback, K. (1968). Information Theory and Statistics, Dover Publication, New York

LeCam, L. (1953). On some asymptotic properties of maximum likelihood and related Bayes estimates. University of California Publications in Statistics.

LeCam, L. (1955). On the asymptotic theory of estimation and testing hypotheses. Proc. Third Berkeley Symp. Math. Statist. and Prob. 1 129-156.

LeCam, L. (1958). Les propriétés asymptotiques de solutions des Bayes. Publ. Inst. Statist. Univ. Paris 7 17-35.

LeCam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. Ann. Math. Statist. 41 801-828.

Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships. Ann. Math. Statist. 3 217-226.

Olech, C. (1965). A note concerning set-valued measurable functions. Bulletin De L'académie Polonaise Des Sciences Serie des Sciences Math., Astr. et Phys. No. 4 317-322.

Parzen, E. (1953). On uniform convergence of families of sequences of random variables. University of California Publication in Statistics 1 23-53.

Parzen, E. (1960). Modern Probability Theory and Its Applications. John Wiley and Sons, Inc., New York.

Rao, C. R. (1963). Criteria of estimation in large samples. Sankhyā Ser. A25 189-206.

Taylor, A. E. (1955). Advanced Calculus. Blaisdell Publishing, Waltham, Mass.

Wald, A. (1949). Note on the consistency of maximum likelihood estimates. Ann. Math. Statist. 20 595-601.

Weiss, L. and Wolfowitz, J. (1966). Generalized maximum likelihood estimators. Teoriya Vyerogatnostey. 11 68-93.

Weiss, L. and Wolfowitz, J. (1967). Maximum probability estimators. Ann. Inst. Stat. Math. 19 193-206.

Weiss, L. and Wolfowitz, J. (1968). Maximum probability estimators with a general loss function. _Proceedings of the International Symposium on Probability and Information Theory held at McMaster University_. Springer Verlag, Heidelberg and New York.

Weiss, L. and Wolfowtiz, J. (1970). Maximum probability estimators and asymptotic sufficiency. _Inst. of Statistical Mathematics_ 22 225-244.

Wolfowitz, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate. _Ann. Math. Statist._ _20_ 601-602.

Wolfowitz, J. (1953). The method of maximum likelihood and the Wald theory of decision functions. _Indag. Math._ 15 114-119.

Wolfowitz, J. (1965). Asymptotic efficiency of maximum likelihood estimator. _Theor. Prob. Appl._ _10_ 247-260.

Zacks, S. (1970). _The Theory of Statistical Inference_. John Wiley and Sons, Inc., New York.

VITA

James Chuan Fu was born in Nanking, China on July 11, 1937; and all his early education was in Kangshan, Taiwan, China. In 1960 the Provincial Cheng - Kung University awarded him a B.A. degree in mathematics. From 1961 to 1963 he worked in the Biostatistical Laboratory, Institute of Botany, Academia Sinica, China, as a research assistant. In 1964 he came to the United States. He received his M.S. degree in biometry in 1966 and M.A. degree in mathematics in 1968 both from Cornell University, Ithaca, N.Y. Since June 1968 he has been a graduate student working for his Ph.D. degree in the Department of Statistics, The Johns Hopkins University.