APPENDIX

The purpose of this appendix is to illustrate the application of the theory to various types of cases. The first part of the appendix gives sufficient regularity conditions for the theory of m.ℓ. estimators described in Section 6(1) to hold in certain cases where the components of $X(n)$ are not necessarily independent and identically distributed. The second part of the appendix gives some examples where the regularity conditions are violated, but m.p. theory can still be applied.

In several of the examples to be discussed, the dimension m of Θ is greater than one, and each component of the vector of parameters requires its own normalizing factor. $k_i(n)$ will be the symbol used for the normalizing factor for the parameter θ_i $(i = 1,\ldots,m)$. This differs from the notation used in the rest of the monograph, where the discussion was carried out in detail only for the case $m = 1$.

PART I

In this first part of the appendix, we assume the following:

A(1). There exist m sequences of nonrandom positive quantities, $\{k_1(n)\}, \ldots, \{k_m(n)\}$, with $\lim_{n\to\infty} k_i(n) = \infty$ for $i = 1,\ldots,m$, such that for any $\theta^0 = (\theta_1^0,\ldots,\theta_m^0)$ in Θ, $-\frac{1}{k_i(n)k_j(n)} \frac{\partial^2}{\partial\theta_i\partial\theta_j} \log K_n(X(n)|\theta_1,\ldots,\theta_m)]_{\theta^0}$ converges stochastically as n increases to a nonrandom quantity, say $B_{ij}(\theta^0)$, when θ^0 is the true parameter value, for $i,j = 1,\ldots,m$. $B_{ij}(\theta^0)$ is assumed to be a continuous function of θ^0. Let $B(\theta^0)$ denote the m by m matrix with $B_{ij}(\theta^0)$ in row i and column j. We assume $[B(\theta^0)]^{-1}$ exists, and denote it by $I(\theta^0)$.

A(2). For each θ^0 in Θ, we assume there exist m sequences of nonrandom positive quantities $\{M_1^*(n,\theta^0)\}$, ..., $\{M_i^*(n,\theta^0)\}$, ..., $\{M_m^*(n,\theta^0)\}$, satisfying the following conditions:

(a) $\lim_{n\to\infty} M_i^*(n,\theta^0) = \infty$, $i = 1,\ldots,m$.

(b) $\lim_{n\to\infty} \frac{M_i^*(n,\theta^0)}{k_i(n)} = 0$, $i = 1,\ldots,m$.

(c) Let $N_n(\theta^0)$ denote the set of all vectors $\theta = (\theta_1,\ldots,\theta_k)$ such that $|\theta_i - \theta_i^0| \leq \frac{M_i^*(n,\theta^0)}{k_i(n)}$ for $i = 1,\ldots,m$. (Note that for all sufficiently large n, $N_n(\theta^0)$ is contained in Θ.) We denote

$$- \frac{1}{k_i(n)k_j(n)} \frac{\partial^2}{\partial\theta_i\partial\theta_j} \log K_n(X(n)|\theta) - B_{ij}(\theta^0)$$

by $\varepsilon_{ij}(\theta,\theta^0,n)$. For any $\gamma > 0$, let $S_n(\theta^0,\gamma)$ denote the region in $X(n)$-space where

$\sum_{i=1}^{m} \sum_{j=1}^{m} M_i^*(n,\theta^0)M_j^*(n,\theta^0) \sup_{\theta \text{ in } N_n(\theta^0)} |\varepsilon_{ij}(\theta,\theta^0,n)| < \gamma$. We assume that there exist two sequences of nonrandom positive quantities $\{\gamma(n,\theta^0)\}$, $\{\delta(n,\theta^0)\}$, with $\lim_{n\to\infty} \gamma(n,\theta^0) = 0$ and $\lim_{n\to\infty} \delta(n,\theta^0) = 0$, such that for each n and each θ in $N_n(\theta^0)$, $P_\theta[X(n)$ in $S_n(\theta;\gamma(n,\theta^0))] > 1 - \delta(n,\theta^0)$.

The list of assumptions is now complete. Before motivating these assumptions, we show that in the special case $m = 1$ and $X(n) = (X_1,\ldots,X_n)$ with X_1, ..., X_n independent with common marginal density $f(x|\theta)$, so that $K_n(X(n)|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$, these assumptions are much less restrictive than those given in 6(1), which are typical of the standard literature. Since $m = 1$, we drop all subscripts i,j. In our discussion, $\{\Delta_i(n,\theta^0)\}$ is, for each i, some sequence of nonrandom positive quantities depending only on n and θ^0, with $\lim_{n\to\infty} \Delta_i(n,\theta^0) = 0$. Setting $k(n) = \sqrt{n}$ and $B(\theta) = E_\theta\{\frac{\partial^2}{\partial\theta^2} \log f(X_1|\theta)\}$, our assumption A(1) is seen to hold, with the positivity of $B(\theta^0)$ following from (6.5)(c) and the continuity of $B(\theta^0)$ following from (6.8). From (6.6)(b) we get

(A.1.1) For each n and each θ in $N_n(\theta^0)$,

$$P\,[\,|\,-\frac{1}{n}\,\frac{\partial^2}{\partial\theta^2}\log K_n(X(n)|\theta) - B(\theta)\,| < \Delta_1(n,\theta^0)]$$

$$> 1 - \Delta_2(n,\theta^0)\ .$$

From the continuity of $B(\theta)$, we get

(A.1.2) For each n and each θ in $N_n(\theta^0)$, $|B(\theta) - B(\theta^0)\,|$

$$< \Delta_3(n,\theta^0)\ .$$

From (6.8) we get that for all sufficiently large n,

(A.1.3)

$$\sup_{\theta^{(1)},\theta^{(2)} \text{ in } N_n(\theta^0)} \{\,|-\frac{1}{n}\,\frac{\partial^2}{\partial\theta^2}\log K_n(X(n)|\theta)]_{\theta^{(1)}} + \frac{1}{n}\,\frac{\partial^2}{\partial\theta^2}\log K_n(X(n)|\theta)]_{\theta^{(2)}}\,|\}$$

$$\leqq k_{\theta^0}\frac{2M^*(n,\theta^0)}{\sqrt{n}}$$

From (A.1.2) and (A.1.3), for all sufficiently large n and all θ^*, θ in $N_n(\theta^0)$, we get

$$(A.1.4) \qquad |\varepsilon(\theta^*, \theta^0, n)| \leq \left|-\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log K_n(X(n)|\theta) - B(\theta)\right| + \Delta_3(n,\theta^0) + k_{\theta^0}\frac{2M^*(n,\theta^0)}{\sqrt{n}} .$$

From (A.1.1) and (A.1.4), we get that for all sufficiently large n, and all θ in $N_n(\theta^0)$,

$$P_\theta\left[(M^*(n,\theta^0))^2 \sup_{\theta^* \text{ in } N_n(\theta^0)} |\varepsilon(\theta^*,\theta^0,n)| \leq (M^*(n,\theta^0))^2\left[\Delta_1(n,\theta^0) + \Delta_3(n,\theta^0) + \frac{2k_{\theta^0}M^*(n,\theta^0)}{\sqrt{n}}\right]\right] > 1 - \Delta_2(n,\theta^0) .$$

Now if we define $M^*(n,\theta^0)$ as

$$\min\left[\left[\frac{1}{\Delta_1(n,\theta^0)}\right]^{1/4}, \left[\frac{1}{\Delta_3(n,\theta^0)}\right]^{1/4}, n^{1/12}\right]$$

it is easy to verify that assumption A(2) is satisfied.

The motivation for assumptions A(1) and A(2) is fairly obvious. The purpose of A(1) is to guarantee (asymptotically) that $K_n(X(n)|\theta)$ will have a peak near the true value θ^0. The assumption A(2) guarantees that a small change in θ does not lead to a large change in the asymptotic behavior of $K_n(X(n)|\theta)$.

What conclusions follow from assumptions A(1) and A(2)? It was shown in [10] and [11] that if A(1) and A(2) hold, the following hold:

(1) Under any θ^0 in Θ, and for each n, there is a neighborhood $C_n(\theta^0)$ of θ^0, with diameter approaching zero as n increases, such that $\lim_{n\to\infty} P_{\theta^0}[K_n(X(n)|\theta)$ has a relative maximum w.r.t. θ in $C_n(\theta^0)] = 1$. If for each n, $\hat{\theta}(n) = (\hat{\theta}_1(n),\ldots,\hat{\theta}_m(n))$ is a point at which such a relative maximum occurs, then $k_1(n)(\hat{\theta}_1(n)-\hat{\theta}_1^0)$, ..., $k_m(n)(\hat{\theta}_m(n)-\theta_m^0)$ have asymptotically an m-variate normal distribution with zero means and covariance matrix $I(\theta^0)$.

(2) $\hat{\theta}(n)$ is an m.p. estimator with respect to any measurable convex R which is symmetric about the origin.

(3) Suppose for each n we have available a vector $\bar{\theta}(n) = (\bar{\theta}_1(n),\ldots,\bar{\theta}_m(n))$ such that for any sequence $\{L(n)\}$ of positive nonrandom quantities with $\lim_{n\to\infty} L(n) = \infty$, $\lim_{n\to\infty} P_{\theta^0}[k_i(n)|\bar{\theta}_i(n) - \theta_i^0| < L(n);\ i = 1,\ldots,m] = 1$, for each θ^0 in Θ. Define the vector $A(n;\bar{\theta}(n))$ as the row vector with i^{th} element given by $\frac{1}{k_i(n)} \frac{\partial}{\partial\theta_i} \log K_n(X(n)|\theta]_{\bar{\theta}(n)}$. Define the vector $\hat{\theta}^*(n)$ by the matrix equation $(k_1(n)(\hat{\theta}_1^*(n) - \bar{\theta}_1(n)), \ldots, k_m(n)(\hat{\theta}_m^*(n) - \bar{\theta}_m(n))) = A(n;\bar{\theta}(n))I(\bar{\theta}(n))$. Then $\hat{\theta}^*(n)$ has the same asymptotic distribution as $\hat{\theta}(n)$, and is therefore an m.p. estimator. Note that here $\bar{\theta}(n)$ is a "preliminary estimator" which does not have to be asymptotically efficient.

We illustrate with five examples.

<u>Example 1</u>. For each n, X_1', ..., X_n' are independent and identically distributed, with common density function $f(x-\theta)$, common distribution function $F(x-\theta)$, where f is a known

function and θ is an unknown location parameter. Thus $m = 1$ in this example, and we dispense with subscripts. We assume f is positive everywhere, and has a continuous second derivative everywhere. p, q are given values with $0 < p < q < 1$. $X_1 < \cdots < X_n$ are the ordered values of X_1', ..., X_n'. The estimation of θ is to be based only on $X_{[np]}$, ..., $X_{[nq]}$. Thus the vector $X(n)$ is $(X_{[np]}, \ldots, X_{[nq]})$, and $K_n(X(n)|\theta)$ is given by

$$\frac{n!}{([np]-1)!(n-[nq])!}(F(X_{[np]}-\theta))^{[np]-1}(1-F(X_{[nq]}-\theta))^{n-[nq]}\prod_{i=[np]}^{[nq]} f(X_i-\theta)$$

for $X_{[np]} < \cdots < X_{[nq]}$, and $K_n(X(n)|\theta)$ is zero otherwise. In this example, for any $\Delta > 0$, $n^{1/2-\Delta} \max_{[np]\leq i\leq [nq]} |X_i-\theta^0-F^{-1}(\frac{i}{n})|$ converges stochastically to zero as n increases. From this it follows easily that assumptions A(1) and A(2) are satisfied with $k(n) = \sqrt{n}$, and $B(\theta^0) = \frac{f^2(F^{-1}(p))}{p} + \frac{f^2(F^{-1}(q))}{1-q}$ $+ \int_{F^{-1}(p)!}^{F^{-1}(q)} \left[\frac{f'(y)}{f(y)}\right]^2 f(y)dy$. Since $\sqrt{n}\ (X_{[np]} - \theta^0 - F^{-1}(p))$ has asymptotically a normal distribution with mean zero and finite variance, we can use $X_{[np]} - F^{-1}(p)$ as $\overline{\theta}(n)$. We note that $F^{-1}(p)$ is known.

In the paper [16], an estimation problem where f is known to be symmetric and to have f'' satisfy a Lipschitz condition, but is otherwise unknown, was solved by estimating f and then using the estimate of f to construct the estimate of θ.

Example 2. This example modifies example 1 by introducing a scale parameter, so the common density is $\frac{1}{\theta_2} f\left(\frac{x-\theta_1}{\theta_2}\right)$, with f known, θ_1 and $\theta_2 > 0$ unknown. Here $m = 2$, and $K_n(X(n)|\theta) =$

$$\frac{n!}{([np]-1)!(n-[nq])!}\left[F\left(\frac{X_{[np]}-\theta_1}{\theta_2}\right)\right]^{[np]-1}$$

$$\left[1-F\left(\frac{X_{[nq]}-\theta_1}{\theta_2}\right)\right]^{n-[nq]} \prod_{i=[np]}^{[nq]}\left[\frac{1}{\theta_2} f\left(\frac{X_i-\theta_1}{\theta_2}\right)\right]$$

if $X_{[np]} < \cdots < X_{[nq]}$. Using the fact that for any $\Delta > 0$,

$n^{1/2-\Delta} \max\limits_{[np]\leq i\leq [nq]} \left|\frac{X_i-\theta_1}{\theta_2} - F^{-1}\left(\frac{i}{n}\right)\right|$ converges stochastically to zero as n increases, we can proceed as in example 1. We find $k_1(n) = k_2(n) = \sqrt{n}$, and

$$B_{11}(\theta^0) = \left[\frac{1}{\theta_2^0}\right]^2\left[\frac{f^2(F^{-1}(p))}{p} + \frac{f^2(F^{-1}(q))}{1-q} + \int_{F^{-1}(p)}^{F^{-1}(q)}\left(\frac{f'(y)}{f(y)}\right)^2 f(y)dy\right],$$

$$B_{12}(\theta^0) = \left[\frac{1}{\theta_2^0}\right]^2\left[\int_{F^{-1}(p)}^{F^{-1}(q)} y\left(\frac{f'(y)}{f(y)}\right)^2 f(y)dy + f(F^{-1}(q)) - f(F^{-1}(p)) + F^{-1}(p)\frac{f^2(F^{-1}(p))}{p} + F^{-1}(q)\frac{f^2(F^{-1}(q))}{1-q}\right]$$

$$B_{22}(\theta^0) = \left[\frac{1}{\theta_2^0}\right]^2\left[\frac{(F^{-1}(p))^2 f^2(F^{-1}(p))}{p} + \frac{(F^{-1}(q))^2 f^2(F^{-1}(q))}{1-q} - q+p + \int_{F^{-1}(p)}^{F^{-1}(q)}\left(y\frac{f'(y)}{f(y)}\right)^2 f(y)dy + 2F^{-1}(q)f(F^{-1}(q)) - 2F^{-1}(p))f(F^{-1}(p))\right]$$

In this example, $\sqrt{n}\left[\frac{X_{[np]}-\theta_1^0}{\theta_2^0} - F^{-1}(p)\right]$, $\sqrt{n}\left[\frac{X_{[nq]}-\theta_1^0}{\theta_2^0} - F^{-1}(q)\right]$ have asymptotically a joint normal distribution with zero means and finite variances. It follows that we can define $\overline{\theta}_2(n) = \frac{X_{[nq]}-X_{[np]}}{F^{-1}(q)-F^{-1}(p)}$, $\overline{\theta}_1(n) = X_{[nq]} - \overline{\theta}_2(n)F^{-1}(q)$.

Example 3. We have a 2-state stationary Markov chain, with transition matrix $\begin{bmatrix}\theta_1 & 1-\theta_1 \\ \theta_2 & 1-\theta_2\end{bmatrix}$, $0 < \theta_1 < 1$, $0 < \theta_2 < 1$. $X_0 = 1$, and $X_0, X_1, \ldots, X_n$ are the observed states. Define N_{ij} as the number of transitions from state i to state j in the sequence $X_0, X_1, \ldots, X_n$, for $i,j = 1,2$. Then $K_n(X(n)|\theta) = \theta_1^{N_{11}}(1-\theta_1)^{N_{12}}\,\theta_2^{N_{21}}(1-\theta_2)^{N_{22}}$. The stationary probabilities are

$\frac{\theta_2}{1-\theta_1+\theta_2}$, $\frac{1-\theta_1}{1-\theta_1+\theta_2}$, and it follows that as n increases, $\frac{N_{11}}{n}$ converges stochastically to $\frac{\theta_1^0\theta_2^0}{1-\theta_1^0+\theta_2^0}$, $\frac{N_{12}}{n}$ converges stochastically to $\frac{(1-\theta_1^0)\theta_2^0}{1-\theta_1^0+\theta_2^0}$, $\frac{N_{21}}{n}$ converges stochastically to $\frac{\theta_2^0(1-\theta_1^0)}{1-\theta_1^0+\theta_2^0}$, and $\frac{N_{22}}{n}$ converges stochastically to $\frac{(1-\theta_1^0)(1-\theta_2^0)}{1-\theta_1^0+\theta_2^0}$. From this, it follows that our assumptions are satisfied with $k_1(n) = k_2(n) = \sqrt{n}$, and $B_{11}(\theta^0) = \frac{\theta_2^0}{\theta_1^0(1-\theta_1^0)(1-\theta_1^0+\theta_2^0)}$, $B_{12}(\theta^0) = 0$,

$B_{22}(\theta^0) = \frac{1-\theta_1^0}{\theta_2^0(1-\theta_2^0)(1-\theta_1^0+\theta_2^0)}$. $\hat{\theta}_1(n) = \frac{N_{11}}{N_{11}+N_{12}}$, $\hat{\theta}_2(n) = \frac{N_{21}}{N_{21}+N_{22}}$.

It is clear that similar results hold for an ergodic chain with any finite number of states.

Example 4. We have a 2-state stationary Markov chain, with transition matrix $\begin{bmatrix} \theta & 1-\theta \\ \sin^2\frac{\pi}{2}\theta & \cos^2\frac{\pi}{2}\theta \end{bmatrix}$, $0 < \theta < 1$. $X_0 = 1$, and $X_0, X_1, \ldots, X_n$ are the observed states. Define N_{ij} as in example 3. Then $K_n(X(n)|\theta) = \theta^{N_{11}}(1-\theta)^{N_{12}}(\sin^2 \frac{\pi}{2}\theta)^{N_{21}}(\cos^2\frac{\pi}{2}\theta)^{N_{22}}$.

The stationary probabilities are $\frac{\sin^2 \frac{\pi}{2}\theta}{1-\theta+\sin^2 \frac{\pi}{2}\theta}$, $\frac{1-\theta}{1-\theta+\sin^2 \frac{\pi}{2}\theta}$, and it follows that as n increases, $\frac{N_{11}}{n}$ converges stochastically to $\frac{\theta^0 \sin^2 \frac{\pi}{2}\theta^0}{1-\theta^0+\sin^2 \frac{\pi}{2}\theta^0}$, $\frac{N_{12}}{n}$ converges stochastically to $\frac{(1-\theta^0)\sin^2 \frac{\pi}{2}\theta^0}{1-\theta^0+\sin^2 \frac{\pi}{2}\theta^0}$, $\frac{N_{21}}{n}$ converges stochastically to $\frac{(\sin^2 \frac{\pi}{2}\theta^0)(1-\theta^0)}{1-\theta^0+\sin^2 \frac{\pi}{2}\theta^0}$, and $\frac{N_{22}}{n}$ converges stochastically to $\frac{(1-\theta^0)\cos^2 \frac{\pi}{2}\theta^0}{1-\theta^0+\sin^2 \frac{\pi}{2}\theta^0}$. From this, it follows that A(1) and A(2) are satisfied with $k(n) = \sqrt{n}$, and $B(\theta^0) =$

$$\frac{1}{1-\theta^0+\sin^2 \frac{\pi}{2}\theta^0}\left[\frac{\sin^2 \frac{\pi}{2}\theta^0}{\theta^0(1-\theta^0)} + \frac{\pi^2(1-\theta^0)}{2(\sin^2 \frac{\pi}{2}\theta^0)(\cos^2 \frac{\pi}{2}\theta^0)}\right].$$

$\overline{\theta}(n)$ can be taken as $\frac{N_{11}}{N_{11}+N_{12}}$.

It is clear that similar results apply to an ergodic chain with any finite number of states, whose transition probabilities are any reasonable functions of a finite number of parameters.

Example 5. $X_1, \ldots, X_n$ are independent, each with density function $\frac{1}{\pi\theta_2}\left[\frac{1}{1+\left(\frac{x-\theta_1}{\theta_2}\right)^2}\right]$, $\theta_2 > 0$. This Cauchy distribution is a standard case, but the computation of $\hat{\theta}_1(n)$, $\hat{\theta}_2(n)$ in closed form is impossible. The purpose of this example is to compute $\hat{\theta}_1^*(n)$, $\hat{\theta}_2^*(n)$. Here of course $k_1(n) = k_2(n) = \sqrt{n}$. A simple computation gives $B_{11}(\theta^0) = \frac{1}{2(\theta_2^0)^2}$, $B_{12}(\theta^0) = 0$, $B_{22}(\theta^0) = \frac{2}{(\theta_2^0)^2}$. Let $Z_1(n)$, $Z_2(n)$, $Z_3(n)$ denote the first, second, and third sample quartiles, respectively. $\sqrt{n}\left[\frac{Z_2(n) - \theta_1^0}{\theta_2^0}\right]$, $\sqrt{n}\left[\frac{Z_3(n) - Z_1(n)}{\theta_2^0} - 2\right]$ have asymptotically a joint normal distribution with zero means and finite variances. It follows that we can define $\bar{\theta}_1(n) = Z_2(n)$, $\bar{\theta}_2(n) = \frac{Z_3(n) - Z_1(n)}{2}$. Thus $\hat{\theta}_1^*(n)$, $\hat{\theta}_2^*(n)$ can be written in closed form.

PART II

In various parts of Chapters 5 and 6, examples were given of the application of the theory to cases not satisfying the regularity conditions of 6(1) or Part I of this Appendix. We now give some additional examples of this type.

Example 6. $X_1, \ldots, X_n$ are independent each with density $f(x-\theta)$, where $f(y) = 0$ if $y > 0$, $f(y)$ is continuous on the left at $y = 0$, and $f(y)$ is nondecreasing in y for all $y < 0$. $R = (-r,r)$. (Note that this example is similar to some of the examples of Chapter 5, but we have replaced the assumptions made in Chapter 5 by a monotonicity assumption about f.) It is easily verified that $K_n(X(n)|\theta) = 0$ if $\theta < \max(X_1,\ldots,X_n)$, and $K_n(X(n)|\theta)$ is nonincreasing in θ if $\theta > \max(X_1,\ldots,X_n)$. Thus $\max(X_1,\ldots,X_n) + \frac{r}{k(n)}$ is an m.p. estimator. Since $\lim_{n\to\infty} P_\theta[n\{\max(X_1,\ldots,X_n) - \theta\} \leq y] = e^{yf(0-)}$ if $y \leq 0$, 1 if $y > 0$, $k(n)$ can be taken as n.

Example 7. This example is a two-dimensional analogue of example 6. $(X_{11},X_{21}), \ldots, (X_{1n},X_{2n})$ are n independent pairs, each with bivariate density $f(x_1-\theta_1, x_2-\theta_2)$, where $f(y_1,y_2) = 0$ if $y_1 > 0$ or $y_2 > 0$, $f(y_1,y_2)$ is nondecreasing in each of y_1, y_2 for all $y_1 < 0$ and $y_2 < 0$, and $\lim_{\substack{y_1\to 0- \\ y_2\to 0-}} f(y_1,y_2) = f(0,0)$, $\lim_{y_1\to 0-} f_1(y_1) = f_1(0)$, $\lim_{y_2\to 0-} f_2(y_2) = f_2(0)$, where $f_i(y)$ is the marginal density for the i^{th} component $(i = 1,2)$. Our region R is $\{(\theta_1,\theta_2) \mid |\theta_1| \leq r_1, |\theta_2| \leq r_2\}$. As in example 6, $K_n(X(n)|\theta_1,\theta_2) = 0$ if $\theta_1 < \max(X_{11},\ldots,X_{1n})$, or $\theta_2 < \max(X_{21},\ldots,X_{2n})$; and $K_n(X(n)|\theta_1,\theta_2)$ is nonincreasing in each θ if $\theta_1 > \max(X_{11},\ldots,X_{1n})$ and $\theta_2 > \max(X_{21},\ldots,X_{2n})$. Also, $\lim_{n\to\infty} P_{\theta_1,\theta_2}[n\{\max(X_{11},\ldots,X_{1n})-\theta_1\} \leq y_1$ and $n\{\max(X_{21},\ldots,X_{2n})-\theta_2\} \leq y_2] = \lim_{n\to\infty} P_{\theta_1,\theta_2}[n\{\max(X_{11},\ldots,X_{1n})-\theta_1\} \leq y_1]$ $\times \lim_{n\to\infty} P_{\theta_1,\theta_2}[n\{\max(X_{21},\ldots,X_{2n})-\theta_2\} \leq y_2] =$

$$= \begin{cases} 1 & \text{if } y_1, y_2 \text{ are both positive} \\ e^{y_1 f_1(0-)} & \text{if } y_1 < 0 \text{ and } y_2 > 0 \\ e^{y_2 f_2(0-)} & \text{if } y_1 > 0 \text{ and } y_2 < 0 \\ e^{y f_1(0-)+y_2 f_2(0-)} & \text{if } y_1 < 0 \text{ and } y_2 < 0. \end{cases}$$

It follows that $\left\{ \max(X_{11},\ldots,X_{1n}) + \frac{r_1}{n} , \max(X_{21},\ldots,X_{2n}) + \frac{r_2}{n} \right\}$ is an m.p. estimator.

<u>Example 8</u>. $(X_{11},X_{21}), \ldots, (X_{1n},X_{2n})$ are n independent pairs, each with bivariate density $f(x_1,x_2|\theta) = \frac{1}{\pi\theta^2}$ if $x_1^2 + x_2^2 \leqq \theta^2$, $f(x_1,x_2|\theta) = 0$ if $x_1^2 + x_2^2 > \theta^2$. Here $\theta > 0$, and $R(\theta) = (-r\theta, r\theta)$, $r > 0$. Here $K_n(X(n)|\theta) = 0$ if $\theta < \max\{ \sqrt{X_{11}^2+X_{21}^2}, \ldots, \sqrt{X_{1n}^2+X_{2n}^2}\}$, $K_n(X(n)|\theta) = \{\frac{1}{\pi\theta^2}\}^n$ if $\theta \geqq \max\{ \sqrt{X_{11}^2+X_{21}^2}, \ldots, \sqrt{X_{1n}^2+X_{2n}^2}\}$. $\lim_{n\to\infty} P_\theta[n\{\max_i \sqrt{X_{1i}^2+X_{2i}^2} - \theta\} \leqq -u$ $= e^{-2u/\theta}$ for $u > 0$. Thus $k(n) = n$, and the m.p. estimator is easily seen to be $\{1+\frac{r}{n}\} \max_i\{ \sqrt{X_{1i}^2+X_{2i}^2}\}$.

References

[1] Anderson, T.W.: The integral of a symmetric unimodal function. Proc. Amer. Math. Soc., 6 (1955), 170-176.

[2] Anderson, T.W.: Introduction to multivariate analysis. Wiley, 1958, New York, N.Y.

[3] Cramér, H.: Mathematical methods of statistics. Princeton University Press, 1946, Princeton, N.J.

[4] Kaufman, S.: Asymptotic efficiency of the maximum likelihood estimator. Ann. Inst. Stat. Math., 18 (1966), 155-178.

[5] Kiefer, J. and Wolfowitz, J.: Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Stat., 27, No. 4 (1956), 887-906.

[6] Michel, R. and Pfanzagl, J.: The accuracy of the normal approximation for minimum contrast estimates. Zeitschr. f. Wahrscheinlichkeitsrechnung verw. Gebiete, 18 (1971), 73-84.

[7] Neyman, J.: Existence of consistent estimates of the directional parameter in a linear structural relation between two variables. Ann. Math. Stat., 22 (1951), 497-512.

[8] Roussas, G.G.: Contiguity of probability measures: some applications in statistics. Cambridge University Press, 1972.

[9] Wald, A.: Note on the consistency of the maximum likelihood estimate. Ann. Math. Stat., 20 (1949), 595-601.

[10] Weiss, L.: Asymptotic properties of maximum likelihood estimators in some nonstandard cases. Jour. Amer. Stat. Assoc., 66 (1971), 345-350.

[11] Weiss, L.: Asymptotic properties of maximum likelihood estimators in some nonstandard cases, II. Jour. Amer. Stat. Assoc., 68 (1973), 428-430.

[12] Weiss, L. and Wolfowitz, J.: Generalized maximum likelihood estimators. Teoriya Vyeroyatnostey, 11 (1966), 68-93.

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

[14] Weiss, L. and Wolfowitz, J.: Asymptotically minimax tests of composite hypotheses. Z. Wahrscheinlichkeitstheorie verw. Geb., 14 (1969), 161-168.

[15] Weiss, L. and Wolfowitz, J.: Maximum probability estimators and asymptotic sufficiency. Ann. Inst. Stat. Math., 22 (1970), 225-244.

[16] Weiss, L. and Wolfowitz, J.: Asymptotically efficient non-parametric estimators of location and scale parameters. Z. Wahrscheinlichkeitstheorie verw. Geb., 16 (1970), 134-150.

[17] Weiss, L. and Wolfowitz, J.: Maximum likelihood estimation of a translation parameter of a truncated distribution. Ann. Stat., 1 (1973), 944-947.

[18] Wolfowitz, J.: Estimation by the minimum distance method. Ann. Inst. Stat. Math., 5 (1953), 9-23.

[19] Wolfowitz, J.: The method of maximum likelihood and the Wald theory of decision functions. Indagationes Mathematicae, 15 (1953).

[20] Wolfowitz, J.: The minimum distance method. Ann. Math. Stat., 28 (1957), 75-88.

[21] Wolfowitz, J.: Asymptotic efficiency of the maximum likelihood estimator. Teoriya Vyeroyatnostey, 10 (1965), 267-281.

[22] Woodroofe, M.: Maximum likelihood estimation of a translation parameter of a truncated distribution. Ann. Math. Stat., 43 (1972), 113-122.