Z. Wahrscheinlichkeitstheorie verw. Geb. 24, 123–133 (1972)

Contributions to Maximum Probability Estimators

Uwe Kuß

1. Introduction

In [4]–[6] L. Weiss and J. Wolfowitz developed the theory of "maximum probability" estimators (m.p.e.'s). This new estimation method as a generalization of R.A. Fisher's well-known maximum likelihood method, does not possess the inadequacies of the "classical" estimation method.

We will briefly mention the important theorems, upon which the maximum probability method is based, from [4], Sec. 2 and [5], Secs. 2–4 with some modifications in the formulation which would hold for any dimension of the parameter space.

2. Maximum Probability Estimators (m.p.e.'s)

For $n \in N$ let $(P_{\theta,n})_{\theta \in \Theta}$ be a family of probability distributions on the space of the observed outcomes X_n, where Θ is assumed to be a connected closed subset of R^m (Θ is called the parameter space). Let the space of possible decisions be also a connected closed subset $\bar{\Theta}$ of R^m, containing all points of Θ as inner points. (There is no compelling reason for giving the space A of decisions this special form; we do this merely for technical reasons.) $f_n(\cdot|\theta)$ denotes the density of $P_{\theta,n}$ with respect to a σ-finite (positive) measure μ_n. Consequently the family $(P_{\theta,n})_{\theta \in \Theta}$ is dominated by μ_n. Let $L_n(\theta, a)$ be the value of the loss function at the point $\theta \in \Theta$, $a \in \bar{\Theta}$. Suppose that $L_n(\theta, a) \geqq 0$.

For a given null sequence $h_2(n) > 0$ we define

$$s(n) := \sup \{L_n(\theta, a) \colon \; \|\theta - a\| \leqq h_2(n)\} \tag{2.1}$$

where, for $x, y \in R^m$,

$$\|x-y\| = \max_{i=1,\dots,m} |x_i - y_i| \quad \text{or} \quad \|x-y\| = \left\{\sum_{i=1}^{m} (x_i - y_i)^2\right\}^{\frac{1}{2}}. \tag{2.2}$$

We will assume $s(n)$ to be finite for all n. Let Y_n be an estimator (i.e. a Borel measurable function of X_n in $\bar{\Theta}$), whose values $Y_n(\xi_n)$, $\xi_n \in X_n$ all maximize the integral

$$I(d) = \int_{\{\|d-\theta\| \leqq h_2(n)\}} \big(s(n) - L_n(d, \theta)\big) f_n(\xi_n|\theta)\, d\theta \tag{2.3}$$

as a function of d, under the assumption, of course, that (the integral exists and is finite and) there is such a Y_n. As the authors of [4] remark, it is also sufficient to assume that

$$I(Y_n)/\sup_d I(d) \geqq 1 - l_n, \tag{2.3'}$$

where l_n is a null sequence with $l_n > 0$.

Let $\theta_0 \in \Theta$ be arbitrary, but fixed,

$$H_n = \{\theta \in \bar{\Theta}:\ \|\theta - \theta_0\| \leqq h_1(n)\},$$

where $h_1(n) > 0$ is a null sequence, for which

$$\lim h_2(n)/h_1(n) = 0$$

holds. Let us remark that the formulation "$F_n(\theta)$ converges uniformly in H_n to $F(\theta)$" means that

$$\lim_{n\to\infty} \left[\sup_{\theta \in H_n} |F_n(\theta) - F(\theta)|\right] = 0.$$

Weiss and Wolfowitz prove the following theorem (see [5]):

Theorem 2.1. *Let the estimator Y_n satisfy the following conditions:*

$$\lim_{n\to\infty} E\big(L_n(Y_n, \theta) | \theta\big) =: \beta \tag{2.4}$$

uniformly in H_n

$$\lim_{n\to\infty} s(n)\, P_n\big(\|Y_n - \theta\| > h_2(n) | \theta\big) = 0 \tag{2.5}$$

uniformly in H_n and

$$\lim_{n\to\infty} \int_{B_n(\theta)} L_n\big(Y_n(\xi_n), \theta\big) f_n(\xi_n | \theta)\, d\mu_n(\xi_n) = 0 \tag{2.6}$$

uniformly in H_n, where

$$B_n(\theta) = \{\xi_n \in X:\ \|Y_n(\xi_n) - \theta\| > h_2(n)\}. \tag{2.7}$$

If T_n is any estimator which satisfies the two conditions

$$\lim_{n\to\infty} [E\{L_n(T_n, \theta) | \theta\} - E\{L_n(T_n, \theta_0) | \theta_0\}] = 0 \tag{2.8}$$

uniformly in H_n and

$$\lim_{n\to\infty} s(n)\, P_n\big(\|T_n - \theta\| > h_2(n) | \theta\big) = 0 \tag{2.9}$$

uniformly in H_n, then we have

$$\beta \leqq \varliminf_{n\to\infty} E\big(L_n(T_n, \theta_0) | \theta_0\big), \tag{2.10}$$

the optimality property of Y_n.

We formulated the Theorem for one $\theta_0 \in \Theta$; we are, of course, interested in the case where a maximum probability estimator Y_n satisfies the hypotheses for all $\theta_0 \in \Theta$, and thus (2.10) is valid in general.

We make the following remark concerning the uniqueness of Y_n: The formulation (2.3′) itself shows that Y_n is not uniquely defined. Furthermore, it is clear that an estimator U_n with

$$\beta = \lim_{n\to\infty} E\big(L_n(U_n, \theta) | \theta\big) \tag{2.11}$$

uniformly in H_n has the same optimality property as Y_n, as has been stated above (2.10). That is the reason why every estimator possessing this optimality property will be regarded as a maximum probability estimator.

Theorem 2.2. *If for all* $n \geqq n_0$ $L_n(z, \theta)$ *is a monotonically increasing function of* $\|z-\theta\|$, *Theorem* 2.1 *holds even without the condition* (2.9).

Theorem 2.3. *If for all* $n \geqq n_0$ $L_n(z, \theta) = s(n)$ *for* $\|z-\theta\| \geqq h_2(n)$, *Theorem* 2.1 *holds even without* (2.9).

Detailed statements on the m.p.e.'s – including the idea and the conception underlying the new method – are to be found in [4]–[7], especially in [4], Sec. 1, [5], Secs. 1 and 5, and in [6], Sec. 1, which contain further references. In [5], Sec. 5 there is also a statistically operational justification of (2.8) and (2.9) of the conditions on competing estimators.

3. Sufficient Conditions for Equivalence, Especially for Condition (2.3′)

According to [4], Sec. 5 (and [6], Sec. 5), in the so-called regular case, the m.l.e. is also a m.p.e. for the important loss function

$$L_n(a, \theta) = \chi_{\{\|a-\theta\| > r_n\}} \quad \text{with } r_n > 0 \text{ and } \lim r_n = 0 \tag{3.1}$$

(χ_A denotes the indicator function of A), where we have to set $r_n = r\, n^{-\frac{1}{2}}$; for this loss function the optimality property (2.10) yields the result of R. A. Fisher on the asymptotic efficiency of the m.l.e. However, we now know that the m.l.e.'s (= m.p.e.'s) are not only better than asymptotically normally distributed (and unbiased) estimators but also than other estimators.

A sufficient condition – with arbitrary density function – for equivalence in the case (3.1) is contained in the following

Theorem 3.1. *We proceed from the same situation as in* 2. *and use the same notation, but we consider only the case* $m=1$, *i.e.* $\Theta, \overline{\Theta} \subseteq R$. *Let* $L_n(a, \theta)$ *– as we have announced – be of the form* (3.1).

Suppose that there exists a maximum likelihood estimator $\hat{\theta}_n$, *i.e.* $f_n(\xi_n|\theta)$ *is maximized at the point* $\hat{\theta}_n(\xi_n)$ *for each* ξ_n *as a function of* θ.

In addition, we assume that, from a certain n on, say, for $n \geqq n_0$, *the following holds for all* $\xi_n \in X_n$:

Let $h(n) = \min(h_2(n), r_n)$,

(a) $[\hat{\theta}_n - h(n), \hat{\theta}_n + h(n)] \subseteq \overline{\Theta}$,

(b) $\lim_{n\to\infty} P_n\{|\hat{\theta}_n - \theta| > h_2(n) | \theta\} = 0$

uniformly in H_n, *and*

$$\lim_{n\to\infty} P_n\{|\hat{\theta}_n - \theta| > r_n | \theta\}$$

exists uniformly in H_n (*for a null sequence* $h_2(n)$ *with* $\lim[h_2(n)/h_1(n)] = 0$),

(c)

$$f_n(\xi_n|\theta) \geqq f_n(\xi_n|\theta^*) \quad \textit{for all } \theta \in [\hat{\theta}_n, \hat{\theta}_n + h(n)], \quad \theta^* \geqq \hat{\theta}_n + h(n)$$

and

$$f_n(\xi_n|\theta') \geqq f_n(\xi_n|\bar{\theta}) \quad \textit{for all } \theta' \in [\hat{\theta}_n - h(n), \hat{\theta}_n], \quad \bar{\theta} \leqq \hat{\theta}_n - h(n).$$

(Hence, it is *not* assumed that $f_n(\xi_n|\theta) \geqq f_n(\xi_n|\theta^*)$ holds for all

$$\theta \in [\hat{\theta}_n - h(n), \hat{\theta}_n + h(n)], \quad \theta^* \notin [\hat{\theta}_n - h(n), \hat{\theta}_n + h(n)]\,!$$

(c) is satisfied, for example, for functions $f_n(\xi_n|\theta)$ which admit a single mode with respect to θ.)

(d) *For a null sequence* ε_n *there holds:*

$$|\log[f_n(\xi_n|\hat{\theta}_n+h(n))/f_n(\xi_n|\hat{\theta}_n-h(n))]|\leqq\varepsilon_n.$$

Then, $\hat{\theta}_n$ *is a maximum probability estimator (and fulfills the conditions of Theorem* 2.1).

Proof. The proof is quite simple. Write for short:

$$K(\tilde{\theta}):=[\tilde{\theta}-h(n),\tilde{\theta}+h(n)],\quad f(\theta):=f_n(\xi_n|\theta).$$

Let $\delta_n\in K(\hat{\theta}_n)-\{\hat{\theta}_n\}$. We define

$$\theta_1=\begin{cases}\hat{\theta}_n+h(n), & \text{if } \delta_n<\hat{\theta}_n\\ \hat{\theta}_n-h(n), & \text{if } \delta_n>\hat{\theta}_n\end{cases}$$

and θ_2 defined with the signs reversed.

On account of (c) and (d) we have for all $\theta\in K(\hat{\theta}_n)-K(\delta_n)$, $\theta^*\in K(\delta_n)-K(\hat{\theta}_n)$:

$$f(\theta)\geqq f(\theta_1)\geqq\exp(-\varepsilon_n)f(\theta_2)\geqq\exp(-\varepsilon_n)f(\theta^*).$$

Hence,

$$\begin{aligned}\int_{K(\hat{\theta}_n)} f(\theta)\,d\theta&=\int_{K(\hat{\theta}_n)-K(\delta_n)} f(\theta)\,d\theta+\int_{K(\hat{\theta}_n)\cap K(\delta_n)} f(\theta)\,d\theta=\int_{K(\delta_n)-K(\hat{\theta}_n)} f(\delta_n+\hat{\theta}_n-\theta')\,d\theta'\\&\quad+\int_{K(\hat{\theta}_n)\cap K(\delta_n)} f(\theta)\,d\theta\\&\geqq\exp(-\varepsilon_n)\int_{K(\delta_n)-K(\hat{\theta}_n)} f(\theta')\,d\theta'+\exp(-\varepsilon_n)\int_{K(\hat{\theta}_n)\cap K(\delta_n)} f(\theta')\,d\theta'\\&=\exp(-\varepsilon_n)\int_{K(\delta_n)} f(\theta')\,d\theta';\end{aligned}\tag{$*$}$$

Since $(*)$ is, of course, also valid for $\delta_n\notin K(\hat{\theta}_n)$, the condition (2.3′) of Theorem 2.1 is satisfied for $\hat{\theta}_n$ instead of Y_n. On account of $s(n)=1$, (2.4)–(2.6) also hold for $\hat{\theta}_n$ according to (b). Theorem 2.1 yields the assertion.

It is, of course, sufficient to require (a)–(d) only for all $\xi_n\in X_n-M_n$, where $\lim_{n\to\infty} P_n(M_n|\theta)=0$ uniformly in H_n, because for such M_n's the existence of the limit, uniformly in H_n,

$$\lim_{n\to\infty} P_n(\xi_n\in X_n\colon\ \|\hat{\theta}_n-\theta\|>r_n|\theta)$$

is tantamount to the existence of the limit (uniformly in H_n)

$$\lim_{n\to\infty} P_n(\xi_n\in X_n-M_n\colon\ \|\hat{\theta}_n-\theta\|>r_n|\theta),$$

and both limits are equal.

The transformation of (2.3′) in [4], 3.2 and [6], p. 198 et sq. can be used in (a)–(d), Theorem 3.1, to consider only the set

$$[\hat{\theta}_n-h_1(n),\hat{\theta}_n+h_1(n)]\subseteq\bar{\Theta}$$

instead of $\bar{\Theta}$.

Basically, Theorem 3.1 is only a sufficient condition for (2.3'). The following two examples of densities $f(\theta)=f_n(\xi_n|\theta)$ – as functions of θ – should clarify the significance of (c) and (d); that (a) is to hold is clear in any case. The loss function (3.1) – as in the Theorem – is taken as a starting point; let $\Theta=\bar{\Theta}=R$.

Example I (Fig. 1). Let f be as is shown in Fig. 1, $a(x)=\hat{\theta}_n+x\,h(n)$. We have $I'(\theta_n)/I'(\delta_n)=0.7\,h(n)/(2\,h(n))=0.35$, where

$$I'(\theta^*)=\int_{\{\|\theta^*-\theta\|\leqq h(n)\}} f_n(\xi_n|\theta)\,d\theta \quad \text{and} \quad \delta_n=\hat{\theta}_n\pm 2.5\,h(n); \tag{3.2}$$

i.e. (2.3') *is not satisfied* for $\hat{\theta}_n$. (a) and (d) hold, but (c) *does not!*

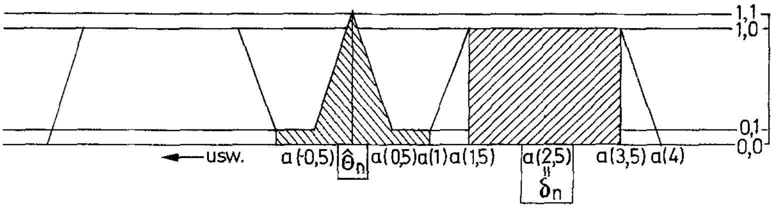

Fig. 1. Function f of Example I

Example II (Fig. 2). Let $f(\theta)$ be defined according to Fig. 2 (graph of f); the only important points are that $f(\theta)$ has the value $(4\,h(n))^{-1}(\theta-\hat{\theta}_n)+0.375$ for $\theta\in[\hat{\theta}_n-h(n),\hat{\theta}_n-0.5\,h(n)]$, its maximum value 0.75 when $\theta=\hat{\theta}_n$, and the value $(-4\,h(n))^{-1}(\theta-\hat{\theta}_n)+0.625$ for $\theta\in[\hat{\theta}_n+h(n),\hat{\theta}_n+1.5\,h(n)]$, and that condition (c) of Theorem 3.1 is satisfied. $a(x)$ is as in Example I, $\delta_n=\hat{\theta}_n+0.5\,h(n)$. We have $I'(\hat{\theta}_n)/I'(\delta_n)\leqq\frac{31}{32}$, hence (2.3') is *not fulfilled*. The reason for this is that although (a) and (c) are satisfied, (d) *is not*.

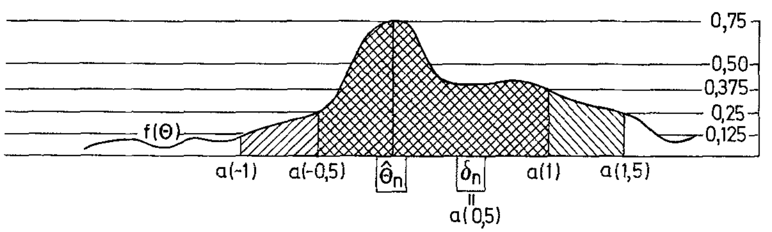

Fig. 2. Function f of Example II

Both examples can be modified to yield continuously differentiable functions; in the same way one can find $f_n(\xi_n|\theta)$ such that $f_n(\xi_n|\theta)$, as a function of θ, is of the given form for a set N_n of ξ_n's with $\lim_{n\to\infty} P_n(N_n|\theta)>0$.

A generalization of Theorem 3.1 for other "reasonable" loss functions and arbitrary dimension of the parameter space, which is now under question, requires a suitable modification of the conditions (c) and (d); viz., we will assume $L_n(a,\theta)$ to be a monotonically increasing function of $|a-\theta|$, because, without such an assumption, useful sufficient conditions for (2.3') can hardly be found. The following example shows that it is not sufficient to require (d) and to replace (c) (for $m=1$) by the condition "$f_n(\xi_n|\theta)$ increases monotonically for $\theta\leqq\hat{\theta}_n$ and decreases monotonically for $\theta\geqq\hat{\theta}_n$ for all $\xi_n\in X_n$".

Example III (Fig. 3). Choose

$$L_n(a,\theta)=\begin{cases}0; & |a-\theta|\leqq 0.5\,h_2(n)\\ \frac{9}{10}; & 0.5\,h_2(n)\leqq|a-\theta|\leqq h_2(n),\\ 1; & \text{otherwise}\end{cases}$$

hence, $s(n)=1$. Let $f(\theta):=f_n(\xi_n|\theta)$ be defined as in Fig. 3. The only relevant points are that $f(\theta)$, for $\theta\in[a(-1),a(-\frac{3}{4})]\cup[a(1),a(1.25)]$ is equal to 1 and assumes the value $(4/h_2(n)(\theta-\hat\theta_n)+3)$ for $\theta\in[a(-0.5),a(-0.25)]$ and the value

$$(4/h_2(n)(\hat\theta_n-\theta)+5)\qquad\text{for }\theta\in[a(0.5),a(0.75)].$$

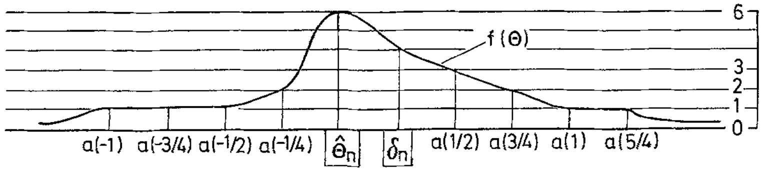

Fig. 3. Function f of Example III

Parallel to the above examples, let $a(x):=x\,h_2(n)+\hat\theta_n$. We have $I(\delta_n)-I(\hat\theta_n)=2.25\,h_2(n)$ (where I is as in 2., (2.3)), and since the first integral is definitely smaller than, or equal to, $12h_2(n)$, (2.3′) *does not hold* for $Y_n=\hat\theta_n$; our example for $f(\theta)$ has – as can immediately be seen – the above-mentioned properties.

That is why we will require a condition corresponding to (d) not only for $\hat\theta_n\pm h_2(n)$, but also for all $\hat\theta_n\pm a$, $|a|\leqq h_2(n)$, and instead of (c) require monotonicity for $\theta\geqq\hat\theta_n$ *or* $\theta\leqq\hat\theta_n$, or only supplement (c) by the requirement of monotonicity in $[\hat\theta_n,\hat\theta_n+h_2(n)]$ and $[\hat\theta_n-h_2(n),\hat\theta_n]$, respectively.

In addition, the function in the *example of* Fig. 4 (for $m=2$) shows why it is not sufficient to require the generalized conditions we have just described for $m=1$ for the functions $g_{\bar\theta}(t)=f_n(\xi_n|\hat\theta_n+t(-\hat\theta_n+\bar\theta))$, where $\bar\theta$ runs, say, through all points with $\|\ldots\|$-distance $h_2(n)$ from $\hat\theta_n$.

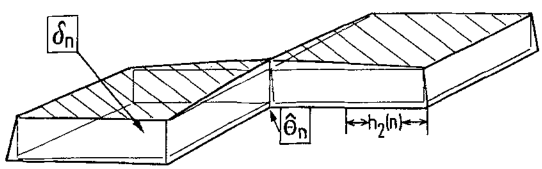

Fig. 4. Example of a function $f(\theta)$ (here $f(\theta)$ is a "surface")

We summarize these considerations in the following way:

Theorem 3.2. *Notation as in* 2. *We assume that for* $n\geqq n_0$

(a*) $L_n(a,\theta)$ *is a monotonically increasing function of* $\|a-\theta\|$, *where now* $\|a-\theta\|=|a-\theta|=\left(\sum_{i=1}^{m}(a_i-\theta_i)^2\right)^{\frac{1}{2}}$.

For all $\xi_n \in X_n$ *we have*

(b*)
$$K(\hat{\theta}_n) = \{\theta: \; |\hat{\theta}_n(\xi_n) - \theta| \leqq h_2(n)\} \subseteq \bar{\Theta},$$

where $\hat{\theta}_n$ *is again the maximum likelihood estimator whose value* $\hat{\theta}_n(\xi_n)$ *we have again shortened to* $\hat{\theta}_n$.

(c*) *There exists a null sequence* ε_n *with*

$$|\log\{f_n(\xi_n|\hat{\theta}_n + a)/f_n(\xi_n|\hat{\theta}_n + a')\}| \leqq \varepsilon_n$$

for all a, a' *with* $|a| = |a'| \leqq h_2(n)$.

(d*) α) *For any* $\bar{\theta} \in \text{boundary}\, K(\hat{\theta}_n)$ *there holds:* $f_n(\xi_n|\hat{\theta}_n + t(\bar{\theta} - \hat{\theta}_n))$ *is (as a function of t) monotonically decreasing in* $[0, 1]$;

β)

$f_n(\xi_n|\theta) \geqq f_n(\xi_n|\theta')\exp(-\varepsilon_n)$ (*or* $\log(f(\theta)/f(\theta')) \leqq \varepsilon_n$) *for all* $\theta \in K(\hat{\theta}_n), \theta' \notin K(\hat{\theta}_n)$.

(Remark: It is also sufficient to require the monotonicity in (d*) modulo a factor $\exp(-\varepsilon_n)$.)

Then, (2.3′) *is satisfied for* $\hat{\theta}_n$ *instead of* Y_n.

Proof. Next, we write following abbreviations:

$$f(\theta) := f_n(\xi_n|\theta); \quad L(\tilde{\theta}, \theta) := (s(n) - L_n(\tilde{\theta}, \theta));$$
$$K(\theta_0) = \{\theta \in \bar{\Theta}: \; |\theta - \theta_0| \leqq h_2(n)\}.$$

Let $\delta_n \in \bar{\Theta}$: We define

$$K = \{\theta \in K(\hat{\theta}_n) \cap K(\delta_n): \; |\theta - \hat{\theta}_n| \leqq |\theta - \delta_n|\}$$

and

$$\bar{K} = \{\theta \in K(\hat{\theta}_n) \cap K(\delta_n): \; |\theta - \hat{\theta}_n| \geqq |\theta - \delta_n|\}.$$

α) For arbitrary $\theta \in K$, if $t_1 = |\hat{\theta}_n - \theta|/h_2(n)$, $t_2 = |\delta_n - \theta|/h_2(n)$, there holds $t_1, t_2 \in [0, 1]$ and $t_1 \leqq t_2$, hence, according to (d*), α):

$$f(\hat{\theta}_n + t_1(\bar{\theta} - \hat{\theta}_n)) \geqq f(\hat{\theta}_n + t_2(\bar{\theta} - \hat{\theta}_n)); \quad \text{since} \quad f(\theta)/f(\hat{\theta}_n + t_1(\bar{\theta} - \hat{\theta})) \geqq \exp(-\varepsilon_n)$$

and

$$f(\hat{\theta}_n + t_2(\bar{\theta} - \hat{\theta}_n))/f(\hat{\theta}_n + \delta_n - \theta) \leqq \exp(-\varepsilon_n)$$

according to (c*) (because $\theta = \hat{\theta}_n + (\theta - \hat{\theta}_n)$ and $\bar{\theta} \in \text{boundary}\, K(\hat{\theta}_n)$), we obtain

$$f(\theta)/f(\hat{\theta}_n + \delta_n - \theta) \geqq \exp(-2\varepsilon_n). \tag{3.21}$$

β)
$$\int_{K(\hat{\theta}_n) - K(\delta_n)} L(\hat{\theta}_n, \theta) f(\theta)\, d\theta = \int_{K(\delta_n) - K(\hat{\theta}_n)} L(\hat{\theta}_n, \hat{\theta}_n + \delta_n - \theta'),$$
$$f(\hat{\theta}_n + \delta_n - \theta')\, d\theta' = (\text{according to (a*)})$$
$$= \int_{K(\delta_n) - K(\hat{\theta}_n)} L(\delta_n, \theta') f(\hat{\theta}_n + \delta_n - \theta')\, d\theta' \tag{3.22}$$
$$\geqq \exp(-\varepsilon_n) \int_{K(\delta_n) - K(\hat{\theta}_n)} L(\delta_n, \theta') f(\theta')\, d\theta';$$

the inequality holds according to (d*), β).

γ)

$$\begin{aligned}\int\limits_{K(\hat{\theta}_n)\cap K(\delta_n)} L(\hat{\theta}_n,\theta)f(\theta)\,d\theta &\\ = \int\limits_{\bar{K}}\cdots+\int\limits_{K}\cdots=\int\limits_{K} L(\hat{\theta}_n,\theta)f(\theta)\,d\theta+\int\limits_{K} L(\delta_n,\theta')f(\hat{\theta}_n+\delta_n-\theta')\,d\theta'.\end{aligned} \tag{3.23}$$

Likewise:

$$\int\limits_{K(\hat{\theta} n)\cap K(\delta_n)} L(\delta_n,\theta)f(\theta)\,d\theta=\int\limits_{K} L(\delta_n,\theta)f(\theta)\,d\theta+\int\limits_{K}\big(L(\hat{\theta}_n,\theta')f(\hat{\theta}_n+\delta_n-\theta')\,d\theta'. \tag{3.24}$$

From (3.23) and (3.24) there follows:

$$\begin{aligned}\int\limits_{K(\hat{\theta}_n)\cap K(\delta_n)} L(\hat{\theta}_n,\theta)f(\theta)\,d\theta-\int\limits_{K(\hat{\theta}_n)\cap K(\delta_n)} L(\delta_n,\theta)f(\theta)\,d\theta &\\ =\int\limits_{K}\big(L(\hat{\theta}_n,\theta)-L(\delta_n,\theta)\big)\big(f(\theta)-f(\delta_n+\hat{\theta}_n-\theta)\big)\,d\theta &\\ \geqq\big(\exp(-2\varepsilon_n)-1\big)\int\limits_{K}\big(L(\hat{\theta}_n,\theta)-L(\delta_n,\theta)\big)f(\delta_n+\hat{\theta}_n-\theta)\,d\theta &\\ =\big(\exp(-2\varepsilon_n)-1\big)\int\limits_{\bar{K}}\big(L(\delta_n,\theta')-L(\hat{\theta}_n,\theta')f(\theta')\,d\theta';\end{aligned} \tag{3.25}$$

the inequality holds, because $L(\hat{\theta}_n,\theta)\geqq L(\delta_n,\theta)$ for all $\theta\in K$ and $f(\theta)-f(\hat{\theta}_n+\delta_n-\theta)\geqq\big(\exp(-2\varepsilon_n)-1\big)f(\hat{\theta}_n+\delta_n-\theta)$ according to (3.21).

We set

$$A_1=\exp(-2\varepsilon_n)\int\limits_{\bar{K}}\big(L(\delta_n,\theta)-L(\hat{\theta}_n,\theta)\big)f(\theta)\,d\theta,$$

$$A_2=\int\limits_{\bar{K}} L(\hat{\theta}_n,\theta)f(\theta)\,d\theta \quad\text{and}\quad A_3=\int\limits_{K} L(\delta_n,\theta)f(\theta)\,d\theta,$$

and we transform (3.25) into

$$\begin{aligned}\int\limits_{K(\hat{\theta}_n)\cap K(\delta_n)} L(\hat{\theta}_n,\theta)f(\theta)\,d\theta &\geqq \exp(-2\varepsilon_n)\int\limits_{\bar{K}}\big(L(\delta_n,\theta)-L(\hat{\theta}_n,\theta)\big)f(\theta)\,d\theta\\ &\quad+\int\limits_{\bar{K}} L(\hat{\theta}_n,\theta)f(\theta)\,d\theta-\int\limits_{\bar{K}} L(\delta_n,\theta)f(\theta)\,d\theta\\ &\quad+\int\limits_{K(\hat{\theta}_n)\cap K(\delta_n)} L(\delta_n,\theta)f(\theta)\,d\theta\\ &=A_1+A_2+\int\limits_{K} L(\delta_n,\theta)f(\theta)\,d\theta\\ &\geqq\exp(-2\varepsilon_n)\int\limits_{\bar{K}} L(\delta_n,\theta)f(\theta)\,d\theta+\exp(-2\varepsilon_n)A_3\\ &=\exp(-2\varepsilon_n)\int\limits_{K(\delta_n)\cap K(\hat{\theta}_n)} L(\delta_n,\theta)f(\theta)\,d\theta;\end{aligned} \tag{3.26}$$

thus (3.22) and (3.26) yield

$$\int\limits_{K(\hat{\theta}_n)} L(\hat{\theta}_n,\theta)f(\theta)\,d\theta\geqq\exp(-2\varepsilon_n)\int\limits_{K(\delta_n)} L(\delta_n,\theta)f(\theta)\,d\theta, \qquad \text{qed.}$$

Corollary 3.3. (a′) *Let* $L_n(a, \theta)$ *be the loss function* (3.1).
(b′) (b*) *holds for* $h(n) = \min(r_n, h_2(n))$ *instead of* $h_2(n)$.
(c′) *There exists a null sequence* ε_n *with* (d*), β) (*where* $h_2(n)$ *is replaced by* $h(n)$).
Then (2.3′) *is also satisfied for* $\hat{\theta}_n$ *instead of* Y_n (*and the loss function* (3.1)).

Proof. The proof is clear, because (3.22) of Theorem 3.2 is likewise valid.

The remarks on the possible restriction of $\bar{\Theta}$ and X_n (p. 126) in the hypotheses hold correspondingly here. One chooses M_n so that

$$\lim_{n\to\infty} \int_{M_n} L_n(\hat{\theta}_n(\xi_n), \theta) f_n(\xi_n | \theta)\, \mu_n(d\,\xi_n) = 0$$

uniformly in H_n.

The theorems make it possible to confine oneself to the conditions of probability theory (2.4)–(2.6) in proving the equivalence of the two estimation methods. They thus apply to nonparametrical investigations as well. As in "regular" cases the equivalence of the two methods has been proved only for (3.1); in these cases the theorem can still be useful for other loss functions.

In most cases it would be more difficult to verify sufficient criteria for (2.4)–(2.6) (for (2.4) this is hardly imaginable) than to verify these conditions (2.4)–(2.6) themselves, which are in the most usable form possible, as is proved in [4]–[6]. For loss functions which are constant outside of $\{\|a-\theta\| \leqq h_2(n)\}$ and for which $s(n)$ is bounded (e.g. (3.1)), (2.5) and (2.6) mean that $\lim_{n\to\infty} P_n\{\|\hat{\theta}_n - \theta\| > h_2(n) | \theta\} = 0$ uniformly in H_n, a property similar to that of consistency, which can usually be proved easily. The reason why necessary conditions for (2.3′) are of no great importance is that maximum probability estimators (according to the definition) need not satisfy (2.3′).

Consider the classical problem of the estimation of the expected value of a sequence of independent, identically normally distributed random variables. As it is well-known that the normal distribution belongs to the regular cases, the maximum likelihood method always leads (see p. 125) to m.p.e.'s for (3.1) with $r_n = r\, n^{-\frac{1}{2}}$. Other loss functions are not considered in [4]–[6]

$$L_n(a, \theta) = n(a-\theta)^2 \tag{3.3}$$

is only mentioned. Taking (3.3) as a starting point we obtain, as an application of Theorem 3.2, that

$$\hat{\theta}_n(\xi) = \frac{1}{n}\sum_{i=1}^{n} \xi_i, \qquad \xi = (\xi_1, \ldots, \xi_n)$$

satisfies (2.3′), and even maximizes (2.3). We set, say,

$$h_2(n) = n^{-\frac{1}{4}}, \qquad h_1(n) = n^{-\frac{1}{8}};$$

then (2.4) and (2.5) are trivial; a proof of (2.6) requires, however, a longer calculation (see [9], pp. 21–27). Hence, we have

Theorem 3.4. $\hat{\theta}_n(\xi) = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ *is a* m.p.e. *for* (3.1) *and* (3.3).

An application of Theorem 3.1 will be given in 4.

4. $E(0, \theta)$ as an Application of Theorem 3.1

(See [1] and [3] for comparison.)

The density function of the exponential distribution $E(\alpha, \theta)$ is

$$f(\alpha, \theta) = \theta \exp\{-\theta(x-\alpha)\}, \quad x \geqq \alpha, \tag{4.1}$$

and otherwise equal to 0; $\theta > 0$.

We will estimate θ where α is known; hence, we can assume $\alpha = 0$ without loss of generality. Then we have

$$f_n(\xi|\theta) = \theta^n \exp\left\{-\theta \sum_{i=1}^{n} \xi_i\right\}, \quad \xi \in X_n = R^n, \ \xi_i > 0 \text{ all } i = 1, \ldots, n$$

and $f_n(\xi|\theta) = 0$ otherwise, and $\theta \in]0, \infty] = \Theta$.

The loss function (3.1) is taken as a starting point, $r_n = r n^{-\frac{1}{2}}$, which is the only reasonable possibility, and proceed in the following way. Let $\theta^* \in \Theta$: We choose n_0 such that $\theta^* - 1/n_0 > 0$, set

$$\Theta' = [\theta^* - 1/(n_0+1), \theta^* + 1], \qquad \bar{\Theta}' = [\theta^* - 1/n_0, \theta^* + 2]$$

and apply Theorem 3.1 to $\Theta', \bar{\Theta}'$ instead of $\Theta, \bar{\Theta}$.

The maximum likelihood estimator is

$$\hat{\theta}_n(\xi) = \left(\frac{1}{n} \sum_{i=1}^{n} \xi_i\right)^{-1}, \qquad \xi \in R^n = X_n.$$

Consider

$$M_n = \{\xi \in X_n : \hat{\theta}_n(\xi) \leqq (\theta^* - 1/n_0 + r_n)\} \cup \{\xi \in X_n : \hat{\theta}_n(\xi) \geqq (\theta^* + 2 - r_n)\}.$$

We have

$$\lim_{n \to \infty} P_n\{M_n|\theta\} = 0$$

uniformly in H_n. Then (a)-(c) are obvious; (d) is shown as follows: We have

$$Z_n := |\log\{f_n(\xi|\hat{\theta}_n + r_n)/f_n(\xi|\hat{\theta}_n - r_n)\}| = n|g_n(\hat{\theta}_n)|,$$

where

$$g_n(x) := \log(x + r_n) - \log(x - r_n) - 2r_n/x.$$

Let n be so large that $2r_n < \theta^* - 1/n_0$; $g_n(x)$ is monotonically decreasing and negative for $x \geqq 2r_n$. $(g_n(2r_n) = 2\log r_n - 1)$. Therefore, $|g_n(x)| \leqq |g_n(\theta^* + 2)|$ for all $x \in \bar{\Theta}'$; hence, $Z_n \leqq n|g_n(\theta^* + 2)|$. The fact that $n|g_n(\theta^* + 2)|$ is a null sequence, is shown by applying Taylor's theorem. Then (2.10) is true for all $\theta_0 \in \Theta'$, i.e. for θ^* in particular; θ^* was arbitrarily given from Θ, hence (2.10) holds for all $\theta^* \in \Theta$, and thus we have

Theorem 4.1. $\theta_n(\xi)$ *is a* m.p.e. *in the case of the exponential distribution* $E(0, \theta)$ *for* (3.1).

Acknowledgment. The author wishes to thank Professor K. Krickeberg for stimulating discussions and for his interest during the preparation of the paper.

References

1. El-Sayyad, G. M.: Estimation of the Parameter of an Exponential Distribution. J. roy. statist. Soc. **29**, 525–532 (1967).
2. Kaufman, S.: Asymptotic efficiency of the maximum likelihood estimator. Ann. Inst. statist. Math. **18**, 155–178 (1966).
3. Pfanzagl, J.: Consistent estimation of a location parameter in the presence of an incidental scale parameter. Ann. math. Statistics **40**, 1353–1357 (1969).
4. Weiss, L., and Wolfowitz, J.: Maximum probability estimators and asymptotic sufficiency. To appear in: Ann. Inst. statist. Math.
5. Weiss, L., and Wolfowitz, J.: Maximum probability estimators with a general loss function. Proc. McMaster Symposium, pp. 232–256. Lecture Notes in Mathematics. **89**. Berlin-Heidelberg-New York: Springer 1969.
6. Weiss, L., and Wolfowitz, J.: Maximum probability estimators. Ann. Inst. statist. Math. **19**, 193–206 (1967).
7. Weiss, L., and Wolfowitz, J.: Generalized Maximum Likelihood Estimators. Teor. Verojatn. Primen. **11**, 68–93 (1966).
8. Weiss, L., and Wolfowitz, J.: Asymptotically efficient nonparametric estimators of location and scale parameters. To appear.
9. Kuß, U.: Beitrag zur Maximum-Probability-Schätzung als einer Verallgemeinerung der Maximum Likelihood Methode. Diplomarbeit Heidelberg 1969 (unpublished).

Uwe Kuß
Institut für Angewandte Mathematik
der Universität
D-6900 Heidelberg 1
Im Neuenheimer Feld 5
Federal Republic of Germany

(Received October 2, 1971 / in revised form April 10, 1972)