KONINKL. NEDERL. AKADEMIE VAN WETENSCHAPPEN – AMSTERDAM
Reprinted from Proceedings, Series A, **60**, No. 1 and Indag. Math., **19**, No. 1, 1957

MATHEMATICS

MAXIMUM LIKELIHOOD ESTIMATION OF PARTIALLY OR COMPLETELY ORDERED PARAMETERS [1]. I

BY

CONSTANCE VAN EEDEN

(Communicated by Prof. D. VAN DANTZIG at the meeting of October 27, 1956)

1. *Introduction*

The problem treated in this paper concerns the maximum likelihood estimation of a finite partially or completely ordered set of parameters of probability distributions. A special case of this problem, the maximum likelihood estimation of a finite ordered set of probabilities, has been treated in [2].

The problem will be formulated in section 2; in section 4 and 5 methods will be given by means of which the estimates may be found. For the proofs of the theorems we need some lemma's which will be proved in section 3; in section 6 the consistency of the estimates will be investigated and in section 7 some examples will be given.

2. *The problem*

Consider k independent random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ [2] and n_i independent observations $x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}$ of $\mathbf{x}_i$ $(i=1, 2, \ldots, k)$. Assume that the distribution of $\mathbf{x}_i$ contains one unknown parameter θ_i $(i=1, 2, \ldots, k)$ and that its distribution function is

$$F_i(x_i \mid \theta_i) \overset{\text{def}}{=} \mathrm{P}\,[\mathbf{x}_i \leq x_i \mid \theta_i] \qquad (i = 1, 2, \ldots, k). \tag{2.1}$$

Two types of restrictions are imposed on the parameters $\theta_1, \theta_2, \ldots, \theta_k$. First let I_i be a closed interval such that $F_i(x_i \mid y)$ is a distribution function for each value of $y \in I_i$ $(i=1, 2, \ldots, k)$. By means of the choice of I_i restrictions of the type $c_i \leqq \theta_i \leqq d_i$ may be imposed. The second type of restrictions consists of a partial or complete ordering of the parameters, which may be described as follows. Let $\alpha_{i,j}$ $(i, j=1, 2, \ldots, k)$ be numbers satisfying the conditions

$$\text{(2.2)} \quad \begin{cases} 1. \quad \alpha_{i,j} = -\alpha_{j,i}, \\ 2. \quad \alpha_{i,j} = 0 \text{ if the intersection } I_i \cap I_j \text{ contains at most one point,} \\ 3. \quad \alpha_{i,j} = 0, +1 \text{ or } -1 \text{ in all other cases;} \end{cases}$$

[1]) Report SP 52 of the Statistical Department of the Mathematical Centre, Amsterdam.

[2]) Random variables will be distinguished from numbers (e.g. from the values they take in an experiment) by printing their symbols in bold type.

and

$$\alpha_{i,j} = 1 \text{ if } \alpha_{i,h} = \alpha_{h,j} = 1 \text{ for any } h. \tag{2.3}$$

The restrictions imposed on $\theta_1, \theta_2, \ldots, \theta_k$ are then

$$\begin{cases} 1. & \alpha_{i,j}(\theta_i - \theta_j) \leqq 0 \\ 2. & \theta_i \in I_i \end{cases} \qquad (i, j = 1, 2, \ldots, k) \tag{2.4}$$

and it will be supposed that the parameters $\theta_1, \theta_2, \ldots, \theta_k$ are numbered in such a way that

$$\alpha_{i,j} \geqq 0 \text{ for each pair of values } (i, j). \tag{2.5}$$

No other restrictions on $\theta_1, \theta_2, \ldots, \theta_k$ are admitted, such that all points $y_1, y_2, \ldots, y_k$ of the Cartesian product

$$G \stackrel{\text{def}}{=} \prod_{i=1}^{k} I_i, \tag{2.6}$$

satisfying

$$\alpha_{i,j}(y_i - y_j) \leqq 0 \qquad (i, j = 1, 2, \ldots, k) \tag{2.7}$$

belong to the parameterspace, which thus is a convex subdomain of G. This subdomain will be denoted by D.

Let

$$\begin{cases} 1. & \alpha_{i,j} = 0 \text{ for } r_0 \text{ pairs of values } (i, j) \text{ with } i < j, \\ 2. & \alpha_{i,j} = 1 \text{ for } r_1 \text{ pairs of values } (i, j) \text{ with } i < j, \end{cases} \tag{2.8}$$

then

$$r_0 + r_1 = \binom{k}{2}. \tag{2.9}$$

Let further $f_i(x_i \mid \theta_i)$ denote the density function of $\mathbf{x}_i$ if $\mathbf{x}_i$ possesses a continuous probability distribution and $\mathrm{P}[\mathbf{x}_i = x_i \mid \theta_i]$ if $\mathbf{x}_i$ possesses a discrete probability distribution and let

$$\begin{cases} 1. & L_i = L_i(y_i) \stackrel{\text{def}}{=} \sum_{\gamma=1}^{n_i} \ln f_i(x_{i,\gamma} \mid y_i) \qquad (i = 1, 2, \ldots, k), \\ 2. & L = L(y_1, y_2, \ldots, y_k) \stackrel{\text{def}}{=} \sum_{i=1}^{k} L_i(y_i). \end{cases} \tag{2.10}$$

Then the maximum likelihood estimates of $\theta_1, \theta_2, \ldots, \theta_k$ are the values of $y_1, y_2, \ldots, y_k$ which maximize L in the domain D. *Unless explicitely stated otherwise L will only be considered in this domain D*; the maximum likelihood estimates will throughout this paper be denoted by $t_1, t_2, \ldots, t_k$. Further the restrictions $\theta_i \leqq \theta_j$ (i.e. $\alpha_{i,j} = 1$) satisfying

$$\alpha_{i,h} \cdot \alpha_{h,j} = 0 \text{ for each } h \text{ between } i \text{ and } j \tag{2.11}$$

will be denoted by $R_1, R_2, \ldots, R_s$. Each R_λ thus corresponds with one pair (i, j); this pair will be denoted by (i_λ, j_λ).

Because of the transitivity relations (2.3) the system $R_1, R_2, \ldots, R_s$ is equivalent to (2.4.1) and uniquely determined by (2.4.1). The restrictions $R_1, R_2, \ldots, R_s$ will be called the *essential restrictions.*

Remark 1: H. D. BRUNK [1] described a method by means of which the estimates of $\theta_1, \theta_2, \ldots, \theta_k$ may be found if the distribution of $\mathbf{x}_i$ belongs to the "exponential family" $(i=1, 2, \ldots, k)$ and if moreover I_i is the set of *all* values of y for which $F_i(x_i \mid y)$ is a distribution function $(i=1, 2, \ldots, k)$. His method however leads to more complicated computations than ours.

3. *Lemma's*

Definition: *A function $\varphi(y)$ of a variable y will be called strictly unimodal in an interval J if there exists a value $y^* \in J$ such that*

$$\varphi(y) < \varphi(y') < \varphi(y^*) \tag{3.1}$$

for each pair of values $(y, y') \in J$ with

$$y < y' < y^* \tag{3.2}$$

and for each pair of values $(y, y') \in J$ with

$$y^* < y' < y. \tag{3.3}$$

It follows at once from this definition that a strictly unimodal function $\varphi(y)$ is bounded in every closed subdomain of J not containing y^*.

Now let $\varphi_\varkappa(y_\varkappa)$ be a strictly unimodal function of $y_\varkappa$ in the interval $J_\varkappa(\varkappa=1, 2, \ldots, K)$ and let further

$$\Phi(y_1, y_2, \ldots, y_K) \overset{\text{def}}{=} \sum_{\varkappa=1}^{K} \varphi_\varkappa(y_\varkappa), \tag{3.4}$$

then it will be clear that $\Phi(y_1, y_2, \ldots, y_K)$ possesses a unique maximum in

$$\Gamma \overset{\text{def}}{=} \prod_{\varkappa=1}^{K} J_\varkappa \tag{3.5}$$

in the point $(y_1^*, y_2^*, \ldots, y_K^*)$, where $\varphi_\varkappa(y_\varkappa^*)$ is the maximum of $\varphi_\varkappa$ in $J_\varkappa$ $(\varkappa=1, 2, \ldots, K)$.

We now define a function V as follows.
Let $y_1^0, y_2^0, \ldots, y_K^0$ be a given point in Γ with $y_\varkappa^0 \neq y_\varkappa^*$ for at least one value of $\varkappa$ and let

$$\begin{cases} Y_\varkappa(\beta) \overset{\text{def}}{=} (1-\beta)\, y_\varkappa^0 + \beta\, y_\varkappa^* \qquad (\varkappa = 1, 2, \ldots, K), \\ 0 \leqq \beta \leqq 1. \end{cases} \tag{3.6}$$

Then $\{Y_1(\beta), Y_2(\beta), \ldots, Y_K(\beta)\}$ is a point in Γ and V is defined by

$$V(\beta) \overset{\text{def}}{=} \Phi\{Y_1(\beta), Y_2(\beta), \ldots, Y_K(\beta)\}. \tag{3.7}$$

Lemma I: *$V(\beta)$ is a monotone increasing function of β in the interval $0 \leqq \beta \leqq 1$.*

Proof: Consider a value of $\varkappa$ with

$$y_\varkappa^0 = y_\varkappa^* \tag{3.8}$$

then

$$Y_\varkappa(\beta) = y_\varkappa^* \text{ for each } \beta \text{ with } 0 \leqq \beta \leqq 1. \tag{3.9}$$

Thus in this case we have

$$\varphi_\varkappa(y_\varkappa^0) = \varphi_\varkappa\{Y_\varkappa(\beta)\} = \varphi_\varkappa(y_\varkappa^*) \text{ for each } \beta \text{ with } 0 \leqq \beta \leqq 1. \tag{3.10}$$

Now consider a value of $\varkappa$ with

$$y_\varkappa^0 \neq y_\varkappa^*, \tag{3.11}$$

then it follows from the fact that $\varphi_\varkappa(y_\varkappa)$ is, in the interval $J_\varkappa$, a strictly unimodal function of $y_\varkappa$ and attains its maximum in $J_\varkappa$ for $y_\varkappa = y_\varkappa^*$, that

$$\varphi_\varkappa(y_\varkappa^0) < \varphi_\varkappa\{Y_\varkappa(\beta_1)\} < \varphi_\varkappa\{Y_\varkappa(\beta_2)\} < \varphi_\varkappa(y_\varkappa^*) \tag{3.12}$$

for each pair of values (β_1, β_2) with $0 < \beta_1 < \beta_2 < 1$.
From (3.4) and the fact that there exists at least one value of $\varkappa$ with (3.11) it follows then that

$$V(0) < V(\beta_1) < V(\beta_2) < V(1) \tag{3.13}$$

for each pair of values (β_1, β_2) with $0 < \beta_1 < \beta_2 < 1$.

Lemma II: *If C is a closed convex subdomain of Γ, not containing the point $(y_1^*, y_2^*, \ldots, y_K^*)$, then $\Phi(y_1, y_2, \ldots, y_K)$ attains its maximum in C only in one or more points on its boundary.*

Proof: Consider any inner point $y_1^0, y_2^0, \ldots, y_K^0$ of C and let $Y_\varkappa(\beta)$ be defined by (3.6) $(\varkappa = 1, 2, \ldots, K)$. Then, C being a closed convex domain not containing the point $(y_1^*, y_2^*, \ldots, y_K^*)$ there exists a value of β in the interval $0 < \beta < 1$, say β_0, such that $\{Y_1(\beta_0), Y_2(\beta_0), \ldots, Y_K(\beta_0)\}$ is a border point of C. Further it follows from Lemma I that

$$\Phi\{Y_1(\beta_0), Y_2(\beta_0), \ldots, Y_K(\beta_0)\} > \Phi(y_1^0, y_2^0, \ldots, y_K^0). \tag{3.14}$$

Thus for each inner point $(y_1^0, y_2^0, \ldots, y_K^0)$ of C there exists a border point $(Y_1, Y_2, \ldots, Y_K)$ of C with a larger value of Φ. Moreover Φ is bounded in C, because the point $(y_1^*, y_2^*, \ldots, y_K^*)$ is not contained in C. Thus Φ has a maximum in C, which can evidently only be attained in border points.

4. *The maximum likelihood estimates of $\theta_1, \theta_2, \ldots, \theta_k$*

Let M be a subset of the numbers $1, 2, \ldots, k$; let further

$$I_M \overset{\text{def}}{=} \bigcap_{i \in M} I_i \tag{4.1}$$

and if $I_M \neq 0$

(4.2) $$L_M(z) \stackrel{\text{def}}{=} \sum_{i \in M} L_i(z) \qquad z \in I_M.$$

Throughout this paper it will be supposed that the following condition is satisfied.

(4.3) Condition: *For each M with $I_M \neq 0$ the function $L_M(z)$ is strictly unimodal in the interval I_M.*

Now let M_ν $(\nu = 1, 2, \ldots, N)$ be subsets of the numbers $1, 2, \ldots, k$ with

(4.4) $$\begin{cases} 1. \quad \bigcup_{\nu=1}^{N} M_\nu = \{1, 2, \ldots, k\}, \\ 2. \quad M_{\nu_1} \cap M_{\nu_2} = 0 \text{ for each pair of values } \nu_1, \nu_2 = 1, 2, \ldots, N \text{ with } \nu_1 \neq \nu_2, \\ 3. \quad I_{M_\nu} \neq 0 \text{ for each } \nu = 1, 2, \ldots, N, \end{cases}$$

where

(4.5) $$I_{M_\nu} \stackrel{\text{def}}{=} \bigcap_{i \in M_\nu} I_i \qquad (\nu = 1, 2, \ldots, N).$$

Let further (cf. (2.6))

(4.6) $$G_N \stackrel{\text{def}}{=} \prod_{\nu=1}^{N} I_{M_\nu}$$

and

(4.7) $$L_{M_\nu}(z_\nu) \stackrel{\text{def}}{=} \sum_{i \in M_\nu} L_i(z_\nu) \qquad z_\nu \in I_{M_\nu} \ (\nu = 1, 2, \ldots, N).$$

Then for all points in G_N $L(y_1, y_2, \ldots, y_k)$ reduces to a function of N variables $z_1, z_2, \ldots, z_N$; we denote this function by $L'(z_1, z_2, \ldots, z_N)$ and thus have

(4.8) $$L'(z_1, z_2, \ldots, z_N) = \sum_{\nu=1}^{N} L_{M_\nu}(z_\nu),$$

which is according to (4.3), a sum of strictly unimodal functions.

Theorem I: *L possesses a unique maximum in D.*

Proof: This theorem will be proved by induction.
Let $M_1, M_2, \ldots, M_N$ be an arbitrary set of subsets of the numbers $1, 2, \ldots, k$ satisfying (4.4) and let

(4.9) $$D_{N,s} \stackrel{\text{def}}{=} D \cap G_N,$$

where s denotes the number of essential restrictions defining D and where G_N is defined by (4.6). Then $D_{N,s}$ is convex and:

for $N = k$ we have $I_{M_\nu} = I_\nu$ $(\nu = 1, 2, \ldots, N)$, thus $G_k = G$ and $D_{k,s} = D$
for $s = 0$ we have $D = G$ thus $D_{N,0} = G_N$.

We shall say that the function $L'(z_1, z_2, \ldots, z_N)$ can be monotonously traced to its maximum in $D_{N,s}$ if

(4.10)
1. $L'(z_1, z_2, \ldots, z_N)$ possesses a unique maximum in $D_{N,s}$,
2. every point of $D_{N,s}$ can be connected with the point in $D_{N,s}$ where L' assumes its maximum by means of a broken line, consisting of a finite number of segments, in $D_{N,s}$ such that L' increases monotonously along this line. (Such a line will be called a trace.)

For $s=0$ $L'(z_1, z_2, \ldots, z_N)$ has this property for every set $M_1, M_2, \ldots, M_N$ satisfying (4.4) and every N. This follows from the fact that L' is the sum of strictly unimodal functions and that $D_{N,0}$ is the Cartesian product of the intervals I_{M_ν} $(\nu=1, 2, \ldots, N)$ so that lemma I may be applied.

Let us now suppose that it has been proved that L' can be monotonously traced to its maximum for all values of $s \leqq s_0$ for every set $M_1, M_2, \ldots, M_N$ satisfying (4.4) and for every N. We then prove that the same holds for s_0+1 essential restrictions.

Consider, for a given set $M_1, M_2, \ldots, M_N$, satisfying (4.4), a domain D_{N,s_0+1} and the domain D_{N,s_0} which is obtained by omitting one of the essential restrictions defining D_{N,s_0+1}. Let this be the restriction R_λ: $\theta_{i_\lambda} \leqq \theta_{j_\lambda}$. Then clearly

$$D_{N,s_0+1} \subset D_{N,s_0}. \tag{4.11}$$

Now L' has a unique maximum in D_{N,s_0}, attained in (say) the point T. We first consider the case that T is outside D_{N,s_0+1}. Then an arbitrary point P of D_{N,s_0+1} with $z_{i_\lambda} < z_{j_\lambda}$ can be connected with T by means of a trace in D_{N,s_0} and this trace must contain at least one border point of D_{N,s_0+1} with $z_{i_\lambda} = z_{j_\lambda}$, because within D_{N,s_0+1} we have: $z_{i_\lambda} < z_{j_\lambda}$ and outside D_{N,s_0+1}: $z_{i_\lambda} > z_{j_\lambda}$. The first of these points when following the trace will be denoted by U; then L' assumes a larger value in U than in P. Now U lies in a domain $D_{N',s_0'}$, where $N'=N-1$ and $s_0' \leqq s_0$ and L' can thus monotonously be traced from U to its unique maximum in $D_{N',s_0'}$ by means of a trace within $D_{N',s_0'}$. The trace from P to U in D_{N,s_0+1} and from U to the maximum of L' in $D_{N',s_0'}$ together form a trace from P to the maximum of L' in D_{N,s_0+1}.

Consider next the case where T is a point of D_{N,s_0+1}. Then L' attains a unique maximum in D_{N,s_0+1} in T. If the maximum of L' in G_N is attained in this point T then, according to Lemma I, L' can be monotonously traced to its maximum from every point of D_{N,s_0+1} by means of a straight line, connecting this point with T. If T is not the point where L' assumes its maximum in G_N then it follows from Lemma II that T is a border point of D_{N,s_0+1} where at least two z_ν from $z_1, z_2, \ldots, z_N$ corresponding to an essential restriction for D_{N,s_0+1} are equal. Let this pair be

$$z_{i_\mu} = z_{j_\mu}, \tag{4.12}$$

then we consider the domain D'_{N,s_0} which is obtained from D_{N,s_0+1} by omitting the restriction R_μ: $\theta_{i_\mu} \leqq \theta_{j_\mu}$ from the essential restrictions defining D_{N,s_0+1}. The maximum of L' in D'_{N,s_0} then exists and the point where it is attained is a point of D'_{N,s_0} with $z_{i_\mu} \geqq z_{j_\mu}$. The rest of the proof for this case is then the same as for the first case considered.

Thus L' can be monotonously traced to its maximum in every $D_{N,s}$; thus, taking $N=k$, L can be monotonously traced to its maximum in D.

Remark 2: For $s=0$ and $N=k$ we have $D_{N,s}=G$. Thus L attains a unique maximum in G in a point which will be denoted by $v_1, v_2, \ldots, v_k$.

Theorem II: *If $t'_1, t'_2, \ldots, t'_k$ are the values of $y_1, y_2, \ldots, y_k$ which maximize L in G and under the restrictions $R_1, \ldots, R_{\lambda-1}, R_{\lambda+1}, \ldots, R_s$ then*

$$(4.13) \qquad \begin{cases} 1. & t_i = t'_i \quad (i = 1, 2, \ldots, k) \text{ if } t'_{i_\lambda} \leqq t'_{j_\lambda}, \\ 2. & t_{i_\lambda} = t_{j_\lambda} \text{ if } t'_{i_\lambda} > t'_{j_\lambda}. \end{cases}$$

Proof: The R_λ have not been arranged in a special order, thus we may take without any loss of generality $\lambda=s$. First consider the case that $t'_{i_s} \leqq t'_{j_s}$; then $t'_1, t'_2, \ldots, t'_k$ satisfy all restrictions $R_1, R_2, \ldots, R_s$; thus in this case we have

$$(4.14) \qquad t_i = t'_i \quad (i=1, 2, \ldots, k).$$

If $t'_{i_s} > t'_{j_s}$ then (4.13.2) may be proved as follows. The domain defined by the essential restrictions $R_1, R_2, \ldots, R_{s-1}$ will be denoted by D'. Then for each point $(y_1, y_2, \ldots, y_k)$ in D with $y_{i_s} < y_{j_s}$ there exists a trace in D' from the point $(y_1, y_2, \ldots, y_k)$ to the point $(t'_1, t'_2, \ldots, t'_k)$ and this trace contains a point $(y'_1, y'_2, \ldots, y'_k)$ with

$$(4.15) \qquad \begin{cases} 1. & y'_{i_s} = y'_{j_s}, \\ 2. & L(y'_1, y'_2, \ldots, y'_k) > L(y_1, y_2, \ldots, y_k). \end{cases}$$

Thus, if $t'_{i_s} > t'_{j_s}$, then $L(y_1, y_2, \ldots, y_k)$ attains its maximum in D for $y_{i_s} = y_{j_s}$; (4.13.2) then follows from the uniqueness of this maximum.

Remark 3:

If

$$(4.16) \qquad \mathrm{P}[\mathbf{x}_i = 1] = \theta_i, \ \mathrm{P}[\mathbf{x}_i = 0] = 1 - \theta_i \qquad (i = 1, \ldots, k)$$

and

$$(4.17) \qquad a_i \overset{\text{def}}{=} \sum_{\gamma=1}^{n_i} x_{i,\gamma}, \ b_i \overset{\text{def}}{=} n_i - a_i \qquad (i = 1, 2, \ldots, k)$$

then

$$(4.18) \qquad L(y_1, y_2, \ldots, y_k) = \sum_{i=1}^{k} \{a_i \ln y_i + b_i \ln(1 - y_i)\}.$$

In [2] it has been proved that, if I_i is the interval (0,1), this function L satisfies the following condition.

(4.19) Condition: *If* $(y_1, y_1, \ldots, y_k)$ *and* $(y'_1, y'_2, \ldots, y'_k)$ *are any two points in G with* $y_i \neq y'_i$ *for at least one value of i and if*

$$Y_i(\beta) = (1-\beta)\, y_i + \beta y'_i \qquad (i = 1, 2, \ldots, k),$$

then $L\{Y_1(\beta), Y_2(\beta), \ldots, Y_k(\beta)\}$ *is a strictly unimodal function of* β *in the interval* $0 \leqq \beta \leqq 1$.

This condition is stronger than condition (4.3) and the theorems I and II of this paper have been proved in [2] by using condition (4.19).

Further if condition (4.19) is satisfied then theorem I of this paper may be proved in a more simple way then we did in [2] as follows. Consider any two points $(y_1, y_2, \ldots, y_k)$ and $(y'_1, y'_2, \ldots, y'_k)$ in D with $y_i \neq y'_i$ for at least one value of i and

$$L(y_1, y_2, \ldots, y_k) = L(y'_1, y'_2, \ldots, y'_k). \tag{4.20}$$

Then it follows from condition (4.19) that there exists a point $(Y_1, Y_2, \ldots, Y_k)$ in D with

$$L(Y_1, Y_2, \ldots, Y_k) > L(y_1, y_2, \ldots, y_k). \tag{4.21}$$

Thus L possesses a unique maximum in D.

The maximum likelihood estimates of $\theta_1, \theta_2, \ldots, \theta_k$ may always be found by applying theorem II repeatedly. This follows from the fact that $L'(z_1, z_2, \ldots, z_N)$ is a sum of strictly unimodal functions and that $D_{N,s}$ is a convex subdomain of the Cartesian product of the intervals I_{M_ν} $(\nu = 1, 2, \ldots, N)$ for each set $M_1, M_2, \ldots, M_N$ and each N.

This leads however to a rather complicated procedure which may often be simplified by using one of the theorems of the following section.

5. *Some special theorems*

The theorems III–VI in this section may be proved in precisely the same way as the theorems II–V in [2].

Theorem III: *If* $\alpha_{i,j}(v_i - v_j) \leqq 0$ *for each pair of values* (i, j) *then*

$$t_i = v_i \quad (i = 1, 2, \ldots, k). \tag{5.1}$$

Theorem IV: *If* $l_1, l_2, \ldots, l_m$ *is a set of values satisfying*

$$\alpha_{i,l_1} = \alpha_{i,l_2} = \ldots = \alpha_{i,l_m} = 0 \quad \text{for each } i \neq l_1, l_2, \ldots, l_m \tag{5.2}$$

then the maximum likelihood estimates of $\theta_{l_1}, \theta_{l_2}, \ldots, \theta_{l_m}$ *are the values of* $y_{l_1}, y_{l_2}, \ldots, y_{l_m}$ *which maximize* $L_{l_1} + L_{l_2} + \ldots + L_{l_m}$ *in the domain*

$$D_1 \begin{cases} \alpha_{i,j}(y_i - y_j) \leqq 0 \\ y_i \in I_i \end{cases} \qquad (i, j = l_1, l_2, \ldots, l_m). \tag{5.3}$$

Theorem V: *If for some pair of values* (i, j) *with* $i < j$

$$\alpha_{i,j}(v_i - v_j) > 0 \tag{5.4}$$

and

$$(5.5)\quad \begin{cases} 1. & \alpha_{i,h} = \alpha_{h,j} = 0 \text{ for each } h \text{ between } i \text{ and } j, \\ 2. & \alpha_{h,i} = \alpha_{h,j} \quad \text{for each } h < i, \\ 3. & \alpha_{i,h} = \alpha_{j,h} \quad \text{for each } h > j, \end{cases}$$

then

$$(5.6)\qquad t_i = t_j.$$

Theorem VI: *If (i, j) is a pair of values satisfying*

$$(5.7)\qquad v_i \leqq v_j$$

and

$$(5.8)\quad \begin{cases} 1. & \alpha_{i,j} = 0, \\ 2. & \alpha_{h,i} \leqq \alpha_{h,j} \text{ for each } h < i, \\ 3. & \alpha_{i,h} \geqq \alpha_{j,h} \text{ for each } h > j, \end{cases}$$

then

$$(5.9)\qquad t_i \leqq t_j.$$

Theorem VII: *If (i, j) is a pair of values with*

$$(5.10)\qquad \alpha_{i,j} = 0,$$

if D' is the subdomain of D where $y_i \leqq y_j$ and if $(t_1', t_2', \ldots, t_k')$ is the point where L assume its maximum in D' then

$$(5.11)\quad \begin{cases} 1. & t_1 = t_1', t_2 = t_2', \ldots, t_k = t_k' \text{ if } t_i' < t_j', \\ 2. & t_i \geqq t_j \text{ if } t_i' = t_j'. \end{cases}$$

Proof: The proof of this theorem differs from the one given for theorem VI in [2] only in the form of the trace from a point in D' to the maximum of L in D. This trace which is a straight line in [2], need not be straight now (cf. the proof of theorem II of the present paper).

(To be continued)

REFERENCES

1. Brunk, H. D., Maximum likelihood estimates of monotone parameters, Ann. Math. Stat. **26**, 607–615 (1955).
2. Eeden, Constance van, Maximum likelihood estimation of ordered probabilities, Proc. Kon. Ned. Akad. v. Wet., **A 59**, 444–455 (1956), Indagationes Mathematicae **18**, 444–455 (1956).

KONINKL. NEDERL. AKADEMIE VAN WETENSCHAPPEN – AMSTERDAM
Reprinted from Proceedings, Series A, **60**, No. 2 and Indag. Math., **19**, No. 2, 1957

MATHEMATICS

MAXIMUM LIKELIHOOD ESTIMATION OF PARTIALLY OR COMPLETELY ORDERED PARAMETERS. II

BY

CONSTANCE VAN EEDEN

(Communicated by Prof. D. VAN DANTZIG at the meeting of November 24, 1956)

6. *The consistency of the estimates*

In this section the consistency of the estimates will be investigated. The method used stems from a paper by A. WALD [3], but is modified by condition (4.3), which does not occur in his paper.

Let, for $f_i(x_i|\theta_i)>0$,

$$g_i(x_i|y_i,\theta_i) \stackrel{\text{def}}{=} \ln \frac{f_i(x_i|y_i)}{f_i(x_i|\theta_i)} \qquad (i=1,2,\ldots,k), \tag{6.1}$$

then

$$L_i(y_i)=\sum_{\gamma=1}^{n_i} g_i(x_{i,\gamma}|y_i,\theta_i)+\sum_{\gamma=1}^{n_i}\ln f_i(x_{i,\gamma}|\theta_i) \qquad (i=1,2,\ldots,k) \tag{6.2}$$

and the maximum likelihood estimates of $\theta_1, \theta_2, \ldots, \theta_k$ are the values of $y_1, y_2, \ldots, y_k$ which maximize $\sum_{i=1}^{k}\sum_{\gamma=1}^{n_i} g_i(x_{i,\gamma}|y_i,\theta_i)$ in the domain D, the last term in (6.2) being constant.

Further

$$g_i(x_i\,|\,\theta_i,\theta_i)=0 \text{ for each } x_i \quad (i=1,2,\ldots,k) \tag{6.3}$$

and from condition (4.3) it follows that $g_i(x_i\,|\,y_i,\theta_i)$ is, for each x_i, a strictly unimodal function of y_i in the interval $I_i (i=1,2,\ldots,k)$.

Let I_i $(i=1,2,\ldots,k)$ be the interval $c_i \leqq y_i \leqq d_i$ (with $c_i<d_i$) and let $\eta_1,\eta_2,\ldots,\eta_k$ be k numbers satisfying

$$\begin{cases} 0<\eta_i \leqq \min(\theta_i-c_i, d_i-\theta_i) \text{ if } \theta_i \text{ is an innerpoint of } I_i, \\ 0<\eta_i\leqq d_i-c_i \text{ if } \theta_i \text{ is a borderpoint of } I_i. \end{cases} \tag{6.4}$$

Let further $I_i(\eta_i)$ denote the set of all values $y_i \in I_i$ satisfying

$$|y_i-\theta_i|\leqq \eta_i \qquad (i=1,2,\ldots,k). \tag{6.5}$$

In the following it will be supposed that the following condition is satisfied.

(6.6) *Condition: There exist k values $\eta_1,\eta_2,\ldots,\eta_k$ satisfying* (6.4) *such that*

$$\begin{cases} 1. \quad \mathscr{E}\{g_i(\mathbf{x}_i|y_i,\theta_i)\,|\,\theta_i\}<0, \quad \text{for each } y_i\in I_i(\eta_i) \text{ with } y_i\neq\theta_i \\ 2. \quad \dfrac{\sigma^2\{g_i(\mathbf{x}_i|y_i,\theta_i)\,|\,\theta_i\}}{[\mathscr{E}\{g_i(\mathbf{x}_i|y_i,\theta_i)\,|\,\theta_i\}]^2}<\infty \qquad (i=1,2,\ldots,k). \end{cases}$$

Some of the conditions mentioned in WALD's paper [3] are in our case sufficient for (6.6.1) and may therefore be useful for the application of our theorems. These conditions may be stated as follows.

Lemma III: *If condition* (4.3) *is satisfied, if* η_i *satisfies* (6.4) *and if*

$$(6.7)\quad \begin{cases} 1. & \mathscr{E}\{\ln f_i(\mathbf{x}_i\,|\,y_i)\,|\,\theta_i\} < \infty \text{ for each } y_i \in I_i(\eta_i) \text{ with } y_i \neq \theta_i, \\ 2. & -\infty < \mathscr{E}\{\ln f_i(\mathbf{x}_i\,|\,\theta_i)\,|\,\theta_i\} < \infty \end{cases}$$

then

$$(6.8)\qquad \mathscr{E}\{g_i(\mathbf{x}_i|y_i,\theta_i)|\theta_i\} < 0 \text{ for each } y_i \in I_i(\eta_i) \text{ with } y_i \neq \theta_i.$$

Proof: Consider any $y_i \in I_i(\eta_i)$, then $\mathscr{E}\{\ln f_i(\mathbf{x}_i|y_i)|\theta_i\} < \infty$. Clearly $\mathscr{E}\{g_i(\mathbf{x}_i|y_i,\theta_i)|\theta_i\} < 0$ if $\mathscr{E}\{\ln f_i(\mathbf{x}_i|y_i)|\theta_i\} = -\infty$.

Now consider the case that $\mathscr{E}\{\ln f_i(\mathbf{x}_i|y_i)|\theta_i\} > -\infty$; then

$$(6.9)\qquad -\infty < \mathscr{E}\{g_i(\mathbf{x}_i|y_i,\theta_i)|\theta_i\} < \infty$$

and from (6.9) it follows that

$$(6.10)\qquad \begin{cases} \mathscr{E}\{g_i(\mathbf{x}_i\,|\,y_i,\theta_i)\,|\,\theta_i\} \leqq \ln \mathscr{E}\{e^{g_i(\mathbf{x}_i|y_i,\theta_i)}\,|\,\theta_i\} = \\ = \ln \displaystyle\int_{f_i(x_i|\theta_i)>0} \frac{f_i(x_i|y_i)}{f_i(x_i|\theta_i)}\, dF_i(x_i\,|\,\theta_i) \leqq \ln 1 = 0. \end{cases}$$

Further

$$(6.11)\qquad \mathscr{E}\{g_i(\mathbf{x}_i\,|\,y_i,\theta_i)\,|\,\theta_i\} = \ln \mathscr{E}\{e^{g_i(\mathbf{x}_i|y_i,\theta_i)}\,|\,\theta_i\}$$

if and only if a value c exists such that

$$(6.12)\qquad \mathrm{P}\,[g_i(\mathbf{x}_i|y_i,\theta_i) = c|\theta_i] = 1.$$

Thus lemma III is proved if we show that such a value c does not exist.

This may be proved as follows. Suppose there exists a value c satisfying (6.12), then it follows from (6.9) that $|c| < \infty$ and further we have

$$(6.13)\qquad \mathrm{P}\,[f_i(\mathbf{x}_i|y_i) = e^c f_i(\mathbf{x}_i|\theta_i)|\theta_i] = 1.$$

From

$$(6.14)\qquad \int dF_i(x_i|y_i) = \int dF_i(x_i|\theta_i) = 1$$

it then follows that $c = 0$.
Further if

$$(6.15)\qquad \mathrm{P}\,[g_i(\mathbf{x}_i|y_i,\theta_i) = 0|\theta_i] = 1,$$

then it follows from (6.3) and the fact that $g_i(x_i|y_i,\theta_i)$ is, for each x_i, a strictly unimodal function of y_i in the interval I_i that

$$(6.16)\qquad \mathrm{P}\,[g_i(\mathbf{x}_i|y_i',\theta_i) > 0|\theta_i] = 1 \text{ for each } y_i' \text{ between } y_i \text{ and } \theta_i,$$

i.e.

$$(6.17)\qquad \mathrm{P}\,[f_i(\mathbf{x}_i|y_i') > f_i(\mathbf{x}_i|\theta_i)|\theta_i] = 1 \text{ for each } y_i' \text{ between } y_i \text{ and } \theta_i$$

and this is in contradiction with

$$\int dF_i(x_i|y_i') = \int dF_i(x_i|\theta_i) = 1. \tag{6.18}$$

Thus there does not exist a value c satisfying (6.12).

Now let (cf. section 4) M_ν ($\nu = 1, 2, \ldots, N$) be N subsets of the numbers $1, 2, \ldots, k$ with

$$\begin{cases} 1. & \bigcup_{\nu=1}^{N} M_\nu = \{1, 2, \ldots, k\}, \\ 2. & M_{\nu_1} \cap M_{\nu_2} \neq 0 \text{ for each pair } (\nu_1, \nu_2) \text{ with } \nu_1 \neq \nu_2, \\ 3. & \theta_i = \theta_j \text{ for each pair } i, j \in M_\nu, \text{ for any value of } \nu \end{cases} \tag{6.19}$$

and let I_{M_ν} be defined by (4.5); then $I_{M_\nu} \neq 0$ ($\nu = 1, 2, \ldots, N$). The value of θ_i for $i \in M_\nu$ will be denoted by θ_ν' ($\nu = 1, 2, \ldots, N$). From theorem I it then follows that

$$L'(z_1, z_2, \ldots, z_N) - L'(\theta_1', \theta_2', \ldots, \theta_N') = \sum_{\nu=1}^{N} \sum_{i \in M} \sum_{\gamma=1}^{n_i} g_i(x_{i,\gamma} | z_\nu, \theta_\nu') \tag{6.20}$$

possesses a unique maximum in (cf. (4.6))

$$G_N = \prod_{\nu=1}^{N} I_{M_\nu}, \tag{6.21}$$

say in the point $(z_1^*, z_2^*, \ldots, z_N^*)$. Let further

$$\eta_\nu' \stackrel{\text{def}}{=} \min_{i \in M_\nu} \eta_i \qquad (\nu = 1, 2, \ldots, N) \tag{6.22}$$

and

$$n \stackrel{\text{def}}{=} \sum_{i=1}^{k} n_i. \tag{6.23}$$

Then the following lemma holds

Lemma IV: *If*

$$\lim_{n \to \infty} n_i = \infty \quad \text{for each } i = 1, 2, \ldots, k, \tag{6.24}$$

then

$$\lim_{n \to \infty} \mathrm{P}\left[|\mathbf{z}_\nu^* - \theta_\nu'| \leq \varepsilon \text{ for each } \nu \,|\, \theta_1', \theta_2', \ldots, \theta_N'\right] = 1 \qquad \text{for each } \varepsilon > 0 \tag{6.25}$$

for each set $M_1, M_2, \ldots, M_N$ *satisfying* (6.19) *and each* N.

Proof: Let

$$\begin{cases} 1. & \beta_i(z_\nu) \stackrel{\text{def}}{=} \mathscr{E}\{g_i(\mathbf{x}_i | z_\nu, \theta_\nu') \,|\, \theta_\nu'\} \\ 2. & \delta_i(z_\nu) \stackrel{\text{def}}{=} \sigma^2\{g_i(\mathbf{x}_i | z_\nu, \theta_\nu') \,|\, \theta_\nu'\} \end{cases} \qquad i \in M_\nu \, (\nu = 1, 2, \ldots, N) \tag{6.26}$$

and let further ε_1 be a positive number satisfying

$$\varepsilon_1 \leq \min \eta_\nu'. \tag{6.27}$$

Then

$$(6.28)\quad \begin{cases} 1. \quad \theta'_\nu+\varepsilon_1 \in I_{M_\nu} \text{ and } \theta'_\nu-\varepsilon_1 \in I_{M_\nu} \text{ if } \theta'_\nu \text{ is an innerpoint of } I_{M_\nu}, \\ 2. \quad \theta'_\nu+\varepsilon_1 \in I_{M_\nu} \text{ or } \theta'_\nu-\varepsilon_1 \in I_{M_\nu} \text{ if } \theta'_\nu \text{ is a borderpoint of } I_{M_\nu}. \end{cases}$$

Now let S be a subset of the numbers $1, 2, \ldots, N$ such that

$$(6.29)\quad \begin{cases} 1. \quad \theta'_\nu+\varepsilon_1 \in I_{M_\nu} \text{ for } \nu \in S, \\ 2. \quad \theta'_\nu+\varepsilon_1 \notin I_{M_\nu} \text{ for } \nu \notin S, \end{cases}$$

then

$$(6.30)\quad z_\nu \leqq \theta'_\nu \text{ for } \nu \notin S.$$

Further it follows from (6.6.1), for $\nu \in S$, that

$$(6.31)\quad \mathscr{E}\{\sum_{i\in M_\nu}\sum_{\gamma=1}^{n_i} g_i(\mathbf{x}_{i,\gamma}\,|\,\theta'_\nu+\varepsilon_1, \theta'_\nu)\,|\,\theta'_\nu\} = \sum_{i\in M_\nu} n_i\,\beta_i(\theta'_\nu+\varepsilon_1) < 0$$

and from (6.31) and Bienaymé's inequality then follows

$$(6.32)\quad \begin{cases} \mathrm{P}\,[\sum_{i\in M_\nu}\sum_{\gamma=1}^{n_i} g_i(\mathbf{x}_{i,\gamma}\,|\,\theta'_\nu+\varepsilon_1, \theta'_\nu) \geqq 0\,|\,\theta'_\nu] \leqq \dfrac{\sum_{i\in M_\nu} n_i\,\delta_i(\theta'_\nu+\varepsilon_1)}{[\sum_{i\in M_\nu} n_i\beta_i(\theta'_\nu+\varepsilon_1)]^2} = \\ = \sum_{i\in M_\nu} \dfrac{n_i\,\delta_i(\theta'_\nu+\varepsilon_1)}{[\sum_{j\in M_\nu} n_j\,\beta_j(\theta'_\nu+\varepsilon_1)]^2} \leqq \sum_{i\in M_\nu} \dfrac{n_i\,\delta_i(\theta'_\nu+\varepsilon_1)}{n_i^2\,[\beta_i(\theta'_\nu+\varepsilon_1)]^2} = \sum_{i\in M_\nu} \dfrac{\delta_i(\theta'_\nu+\varepsilon_1)}{n_i\,[\beta_i(\theta'_\nu+\varepsilon_1)]^2}. \end{cases}$$

Thus

$$(6.33)\quad \begin{cases} \mathrm{P}\,[\sum_{i\in M_\nu}\sum_{\gamma=1}^{n_i} g_i(\mathbf{x}_{i,\gamma}\,|\,\theta'_\nu+\varepsilon_1, \theta'_\nu) < 0 \text{ for each } \nu \in S\,|\,\theta'_1, \theta'_2, \ldots, \theta'_N] \geqq \\ \qquad \geqq 1 - \sum_{\nu\in S}\sum_{i\in M_\nu} \dfrac{\delta_i(\theta'_\nu+\varepsilon_1)}{n_i\,[\beta_i(\theta'_\nu+\varepsilon_1)]^2}. \end{cases}$$

Further it follows from (6.3), (6.30) and the fact that

$$\sum_{i\in M_\nu}\sum_{\gamma=1}^{n_i} g_i(x_{i,\gamma}\,|\,z_\nu, \theta'_\nu)$$

is a strictly unimodal function of z_ν in the interval I_{M_ν} $(\nu=1, 2, \ldots, N)$ (cf. condition (4.3)) that

$$(6.34)\quad \begin{cases} z^*_\nu \leqq \theta'_\nu+\varepsilon_1\,(\nu=1, 2, \ldots, N) \text{ if} \\ \sum_{i\in M_\nu}\sum_{\gamma=1}^{n_i} g_i(x_{i,\gamma}\,|\,\theta'_\nu+\varepsilon_1, \theta'_\nu) < 0 \text{ for each } \nu \in S. \end{cases}$$

Thus (cf. 6.33))

$$(6.35)\quad \mathrm{P}\,[\mathbf{z}^*_\nu-\theta'_\nu \leqq \varepsilon_1 \text{ for each } \nu\,|\,\theta'_1, \theta'_2, \ldots, \theta'_N] \geqq 1 - \sum_{\nu\in S}\sum_{i\in M_\nu} \frac{\delta_i(\theta'_\nu+\varepsilon_1)}{n_i\,[\beta_i(\theta'_\nu+\varepsilon_1)]^2}.$$

From (6.6.2), (6.24) and (6.35) then follows

(6.36) $\lim_{n\to\infty} P[\mathbf{z}_\nu^* - \theta_\nu' \leqq \varepsilon \text{ for each } \nu \mid \theta_1', \theta_2', \ldots, \theta_N'] = 1$ for each $\varepsilon > 0$.

In an analogous way it may be proved that

(6.37) $\lim_{n\to\infty} P[\mathbf{z}_\nu^* - \theta_\nu' \geqq -\varepsilon \text{ for each } \nu \mid \theta_1', \theta_2', \ldots, \theta_N'] = 1$ for each $\varepsilon > 0$.

If we take $N = k$ in lemma IV then (6.25) reduces to (cf. remark 2 section 4)

(6.38) $\lim_{n\to\infty} P[|\mathbf{v}_i - \theta_i| \leqq \varepsilon \text{ for each } i \mid \theta_1, \theta_2, \ldots, \theta_k] = 1$ for each $\varepsilon > 0$.

Theorem VIII: *If t_i is the maximum likelihood estimate of θ_i ($i = 1, 2, \ldots, k$) under the restrictions $R_1, R_2, \ldots, R_s$ and if*

(6.39) $$\lim_{n\to\infty} n_i = \infty \text{ for each } i = 1, 2, \ldots, k,$$

then

(6.40) $\lim_{n\to\infty} P[|\mathbf{t}_i - \theta_i| \leqq \varepsilon \text{ for each } i \mid \theta_1, \theta_2, \ldots, \theta_k] = 1$ for each $\varepsilon > 0$.

Proof: This theorem will be proved by induction.

Consider the function $L'(z_1, z_2, \ldots, z_N) - L'(\theta_1', \theta_2', \ldots, \theta_N')$ (cf. (6.20)). From theorem I it follows that this function possesses a unique maximum in $D_{N,s}$ (cf. (4.9)), say in the point $(w_1^{(s)}, w_2^{(s)}, \ldots, w_N^{(s)})$.

From lemma IV then follows (for $s = 0$)

(6.41) $\lim_{n\to\infty} P[|\mathbf{w}_\nu^{(0)} - \theta_\nu'| \leqq \varepsilon \text{ for each } \nu \mid \theta_1', \theta_2', \ldots, \theta_N'] = 1$ for each $\varepsilon > 0$

for each set $M_1, M_2, \ldots, M_N$ satisfying (6.19) and each N.

Now suppose that it has been proved that

(6.42) $\lim_{n\to\infty} P[|\mathbf{w}_\nu^{(s)} - \theta_\nu'| \leqq \varepsilon \text{ for each } \nu \mid \theta_1', \theta_2', \ldots, \theta_N'] = 1$ for each $\varepsilon > 0$

for each $s \leqq s_0$, each set $M_1, M_2, \ldots, M_N$ satisfying (6.19) and each N. Then it will be proved that

(6.43) $\lim_{n\to\infty} P[|\mathbf{w}_\nu^{(s_0+1)} - \theta_\nu'| \leqq \varepsilon \text{ for each } \nu \mid \theta_1', \theta_2', \ldots, \theta_N'] = 1$ for each $\varepsilon > 0$

for each set $M_1, M_2, \ldots, M_N$, satisfying (6.19) and each N.

Consider, for a given set $M_1, M_2, \ldots, M_N$ satisfying (6.19), a domain D_{N,s_0+1} and the domain D_{N,s_0} which is obtained by omitting one of the essential restrictions defining D_{N,s_0+1}. Let this be the restriction: $\theta_{i_\lambda} \leqq \theta_{j_\lambda}$. Then the following two cases may be distinguished.

1. $\theta_{i_\lambda} < \theta_{j_\lambda}$; then a positive value ε_1 exists satisfying

(6.44) $$D_{N,s_0} \cap \prod_{\nu=1}^{N} I_{M_\nu}(\varepsilon_1) \subset D_{N,s_0+1}.$$

Further we have, for each ε_1 satisfying (6.44),

(6.45) $w_\nu^{(s_0+1)} = w_\nu^{(s_0)}$ $(\nu = 1, 2, \ldots, N)$ if $|w_\nu^{(s_0)} - \theta_\nu'| \leqq \varepsilon_1$ for each $\nu = 1, 2, \ldots, N$.

From (6.42) and (6.45) then follows

$$(6.46)\quad \begin{cases} \lim_{n\to\infty} \mathrm{P}\,[|\mathbf{w}_\nu^{(s_0+1)}-\theta'_\nu| \leqq \varepsilon_1 \text{ for each } \nu\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]=1 \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for each } \varepsilon_1 \text{ satisfying (6.44)} \end{cases}$$

and from (6.46) follows

$$(6.47)\quad \lim_{n\to\infty} \mathrm{P}\,[|\mathbf{w}_\nu^{(s_0+1)}-\theta'_\nu| \leqq \varepsilon \text{ for each } \nu\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]=1 \text{ for each } \varepsilon>0.$$

2. $\theta_{i_\lambda}=\theta_{j_\lambda}$; then we have for each $\varepsilon>0$

$$(6.48)\quad \begin{cases} \mathrm{P}\,[|\mathbf{w}_\nu^{(s_0+1)}-\theta'_\nu| \leqq \varepsilon \text{ for each } \nu\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]= \\ =\mathrm{P}\,[\mathbf{w}_{i_\lambda}^{(s_0)} < \mathbf{w}_{j_\lambda}^{(s_0)}\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]\cdot \\ \qquad\cdot\mathrm{P}\,[|\mathbf{w}_\nu^{(s_0)}-\theta'_\nu| \leqq \varepsilon \text{ for each } \nu\,|\,\mathbf{w}_{i_\lambda}^{(s_0)} < \mathbf{w}_{j_\lambda}^{(s_0)};\theta'_1,\theta'_2,\ldots,\theta'_N] \\ +\mathrm{P}\,[\mathbf{w}_{i_\lambda}^{(s_0)} \geqq \mathbf{w}_{j_\lambda}^{(s_0)}\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]\cdot \\ \qquad\cdot\mathrm{P}\,[|\mathbf{w}_\nu^{(s_0+1)}-\theta'_\nu| \leqq \varepsilon \text{ for each } \nu\,|\,\mathbf{w}_{i_\lambda}^{(s_0+1)}=\mathbf{w}_{j_\lambda}^{(s_0+1)};\theta'_1,\theta'_2,\ldots,\theta'_N], \end{cases}$$

because if $w_{i_\lambda}^{(s_0)}<w_{j_\lambda}^{(s_0)}$ then the maximum under s_0 restrictions coincides with the maximum under s_0+1 restrictions and if $w_{i_\lambda}^{(s_0)}\geqq w_{j_\lambda}^{(s_0)}$ then (according to theorem II) $w_{i_\lambda}^{(s_0+1)}=w_{j_\lambda}^{(s_0+1)}$.

Further $w_1^{(s_0+1)}, w_2^{(s_0+1)}, \ldots, w_N^{(s_0+1)}$ are, under the condition $w_{i_\lambda}^{(s_0+1)}=w_{j_\lambda}^{(s_0+1)}$, the values of $z_1, z_2, \ldots, z_N$ which maximize $L'(z_1, z_2, \ldots, z_N)-L'(\theta'_1, \theta'_2, \ldots, \theta'_N)$ in a domain $D_{N',s_0'}$ where $N'=N-1$ and $s'_0\leqq s_0-1$. Thus from (6.42) it follows that

$$(6.49)\quad \begin{cases} \lim_{n\to\infty} \mathrm{P}\,[|\mathbf{w}_\nu^{(s_0+1)}-\theta'_\nu| \leqq \varepsilon \text{ for each } \nu\,|\,\mathbf{w}_{i_\lambda}^{(s_0+1)}=\mathbf{w}_{j_\lambda}^{(s_0+1)};\theta'_1,\theta'_2,\ldots,\theta'_N]=1 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for each } \varepsilon>0. \end{cases}$$

Thus if

$$(6.50)\quad P_n \stackrel{\text{def}}{=\!=} \mathrm{P}\,[|\mathbf{w}_\nu^{(s_0+1)}-\theta'_\nu| \leqq \varepsilon \text{ for each } \nu\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]$$

and if A_n, B_n and $\bar{B}_n$ respectively denote the events

$$|w_\nu^{(s_0)}-\theta'_\nu| \leqq \varepsilon \text{ for each } \nu$$

$$w_{i_\lambda}^{(s_0)} < w_{j_\lambda}^{(s_0)}$$

and

$$w_{i_\lambda}^{(s_0)} \geqq w_{j_\lambda}^{(s_0)}$$

respectively then it follows from (6.42)

$$(6.51)\quad \lim_{n\to\infty} \mathrm{P}\,[A_n\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N] = 1$$

and from (6.48) and (6.49)

$$(6.52)\quad \begin{cases} 1 \geqq \overline{\lim_{n\to\infty}}\, P_n = \lim_{n\to\infty} \{\mathrm{P}\,[B_n\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]\cdot \\ \qquad\cdot\mathrm{P}\,[A_n\,|\,B_n;\theta'_1,\theta'_2,\ldots,\theta'_N]+\mathrm{P}\,[\bar{B}_n\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]\}= \\ =\lim_{n\to\infty} \{\mathrm{P}\,[A_n \text{ and } B_n\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]+\mathrm{P}\,[\bar{B}_n\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]\} \geqq \\ \geqq \lim_{n\to\infty} \mathrm{P}\,[A_n\,|\,\theta'_1,\theta'_2,\ldots,\theta'_N]=1. \end{cases}$$

Thus

$$\lim_{n\to\infty} P_n = 1. \tag{6.53}$$

7. *Examples*

In this section some examples will be given where the conditions (4.3) and (6.6) are satisfied.

Example 1

Let $\mathbf{x}_i$ possess a normal distribution with mean θ_i and known variance σ_i^2 $(i=1, 2, \ldots, k)$. Then

$$L_i(y_i) = -\tfrac{1}{2} n_i \ln 2\pi\sigma_i^2 - \tfrac{1}{2} \frac{\sum_{\gamma=1}^{n_i} (x_{i,\gamma} - y_i)^2}{\sigma_i^2} \qquad (i=1, 2, \ldots, k). \tag{7.1}$$

From (7.1) it follows that $L_i(y_i)$ is a strictly unimodal function of y_i in the interval $(-\infty, +\infty)$ and attains its maximum in this interval for

$$y_i = m_i \overset{\text{def}}{=} \frac{1}{n_i} \sum_{\gamma=1}^{n_i} x_{i,\gamma} \quad (i=1, 2, \ldots, k). \tag{7.2}$$

Thus $L_i(y_i)$ is a strictly unimodal function of y_i in each closed subinterval I_i of the interval $(-\infty, +\infty)$ and if I_i is the interval (c_i, d_i) then $L_i(y_i)$ attains its maximum in I_i for

$$y_i = \begin{cases} m_i \text{ if } c_i \leqq m_i \leqq d_i, \\ c_i \text{ if } m_i < c_i, \\ d_i \text{ if } m_i > d_i. \end{cases} \qquad (i=1, 2, \ldots, k) \tag{7.3}$$

Further if M is a subset of the numbers $1, 2, \ldots, k$ then (cf. (4.2))

$$L_M(z) = -\tfrac{1}{2} \sum_{i\in M} \left\{ n_i \ln 2\pi\sigma_i^2 + \frac{\sum_{\gamma=1}^{n_i} (x_{i,\gamma} - z)^2}{\sigma_i^2} \right\} \tag{7.4}$$

and from (7.4) it follows easily that $L_M(z)$ is a strictly unimodal function of z in the interval $(-\infty, +\infty)$. Thus L satisfies condition (4.3).

Further $L_M(z)$ attains its maximum in the interval $(-\infty, +\infty)$ for

$$z = m_M \overset{\text{def}}{=} \left(\sum_{i\in M} \frac{n_i}{\sigma_i^2} \right)^{-1} \sum_{i\in M} \frac{n_i m_i}{\sigma_i^2}. \tag{7.5}$$

Now let M consist of the numbers $h_1, h_2, \ldots, h_\mu$, then if $\sigma_i^2 = \sigma^2$ for each $i \in M$

$$L_M(z) = -\tfrac{1}{2} n_M \ln 2\pi\sigma^2 - \tfrac{1}{2} \frac{\sum_{\gamma=1}^{n_M} (x_{M,\gamma} - z)^2}{\sigma^2}, \tag{7.6}$$

where

$$n_M \overset{\text{def}}{=} \sum_{i\in M} n_i \tag{7.7}$$

and where $x_{M,\gamma}$ $(\gamma=1, 2, \ldots, n_M)$ denote the pooled samples of $\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \ldots, \mathbf{x}_{h_\mu}$. Thus if L attains its maximum for $y_{h_1}=y_{h_2}=\ldots=y_{h_\mu}$ then the samples of $\mathbf{x}_{h_1}, \mathbf{x}_{h_1}, \ldots, \mathbf{x}_{h_\mu}$ are to be pooled if $\sigma_i^2=\sigma^2$ for each $i \in M$.

Further

$$(7.8) \qquad g_i(x_i \mid y_i, \theta_i)=\frac{(y_i-\theta_i)(2x_i-y_i-\theta_i)}{2\sigma_i^2} \qquad (i=1,2,\ldots,k).$$

Thus

$$(7.9) \qquad \begin{cases} \mathscr{E}\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\}=-\dfrac{(y_i-\theta_i)^2}{2\sigma_i^2} \\ \sigma^2\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\}=\dfrac{(y_i-\theta_i)^2}{\sigma_i^2} \end{cases} \qquad (i=1,2,\ldots,k)$$

and

$$(7.10) \qquad \frac{\sigma^2\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\}}{[\mathscr{E}\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\}]^2}=\frac{4\sigma_i^2}{(y_i-\theta_i)^2} \qquad (i=1,2,\ldots,k).$$

From (7.9) and (7.10) it follows that condition (6.6) is satisfied if

$$(7.11) \qquad \sigma_i^2<\infty \qquad (i=1, 2, \ldots, k).$$

Remark 4: From (7.4) and (7.5) it follows that the estimates of $\theta_1, \theta_2, \ldots, \theta_k$ may also be found by means of the method described above if the σ_i^2 are unknown and σ_i^2/σ_j^2 is known for each pair of values $i, j=1, 2, \ldots, k$. Then if

$$(7.12) \qquad K_i \overset{\text{def}}{=} \frac{\sigma_i^2}{\sigma_1^2} \qquad (i=1, 2, \ldots, k)$$

the maximum likelihood estimate of σ_i^2 is

$$(7.13) \qquad s_i^2 \overset{\text{def}}{=} \frac{K_i}{n} \sum_{j=1}^{k} \sum_{\gamma=1}^{n_j} \frac{(x_{j,\gamma}-t_j)^2}{K_j} \qquad (i=1, 2, \ldots, k).$$

The procedure will now be illustrated by means of the following example.

Two preparations A and B, known to stimulate the growth of hogs, are added in two concentrations each to the food of four groups of hogs. Let these four additions be denoted by A_1, A_2, B_1 and B_2. It is known that B_1 is at least as good as A_1 (notation $A_1 \leq B_1$) and that in the same sense $A_1 \leq A_2$ and $B_1 \leq B_2$. No decisive knowledge however is available concerning the ordering of A_2 and B_2. The growths of the hogs during a certain period are then the four samples.

The fictitious numerical example given below concerns this partial ordering, but has been made a little more complicated by the introduction of unequal variances and of restrictions on the possible values of each θ_i separately:

Let

(7.14)

	A_1	A_2	B_1	B_2
i	1	2	3	4
$x_{i,\gamma}$	− 0,40 2,56 0,25 2,87	1,43 1,86 0,06 0,07 1,14 0,29 2,57 0,85 1,21	− 0,70 2,61 0,79 0,86 0,14	0,29 0 1,31 0,15 2,53 1,86
$\sum_{\gamma=1}^{n_i} x_{i,\gamma}$	5,28	9,48	3,70	6,14
n_i	4	9	5	6
m_i	1,32	1,05	0,74	1,02
σ_i^2	2	4	5	1
I_i	$(-\infty, 1)$	$(-\infty, +\infty)$	$(\frac{1}{2}, +\infty)$	$(-\infty, +\infty)$
v_i	1	1,05	0,74	1,02

and (cf.(2.8))

(7.15) $$\begin{cases} 1. \quad r_0=2,\ r_1=4, \\ 2. \quad \alpha_{1,2}=\alpha_{1,3}=\alpha_{3,4}=1. \end{cases}$$

From (7.14) and (7.15) it follows that the pairs $i=3$, $j=2$ and $i=4$, $j=2$ satisfy (5.7) and (5.8). Thus according to theorem VI L attains its maximum in D for

(7.16) $$y_1 \leqq y_3 \leqq y_4 \leqq y_2.$$

From (7.14), (7.16) and theorem V then follows

(7.17) $$t_1 = t_3,$$

i.e. L attains its maximum in D for

(7.18) $$y_1 = y_3 \leqq y_4 \leqq y_2.$$

From (7.14), (7.18) and (7.5) then follows

(7.19)

i	1	3	4	2
$\sum_{\gamma=1}^{n_i} x_{i,\gamma}$	5,28	3,70	6,14	9,48
n_i	4	5	6	9
m_{M_ν}	1,13	1,13	1,02	1,05
σ_i^2	2	5	1	4
I_{M_ν}	$(\frac{1}{2}, 1)$	$(\frac{1}{2}, 1)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
v_{M_ν}	1	1	1,02	1,05

From (7.19) and theorem III then follows

$$(7.20)\qquad t_1 = t_3 = 1,\ t_2 = 1{,}05,\ t_4 = 1{,}02.$$

Example 2. Let $\mathbf{x}_i$ possess a Poisson distribution with parameter θ_i ($0 < \theta_i < \infty$; $i = 1, 2, \ldots, k$).
Then

$$(7.21)\qquad L_i(y_i) = -n_i y_i + \sum_{\gamma=1}^{n_i} x_{i,\gamma} \ln y_i - \sum_{\gamma=1}^{n_i} \ln x_{i,\gamma}! \quad (i = 1, 2, \ldots, k);$$

thus

$$(7.22)\qquad \frac{dL_i(y_i)}{dy_i} \begin{cases} > 0 \text{ for } 0 \leqq y_i < m_i \overset{\text{def}}{=} \dfrac{1}{n_i} \displaystyle\sum_{\gamma=1}^{n_i} x_{i,\gamma}, \\ = 0 \text{ for } y_i = m_i, \\ < 0 \text{ for } y_i > m_i. \end{cases}$$

From (7.22) it follows that $L_i(y_i)$ is a strictly unimodal function of y_i in the interval $(0, \infty)$ ($i = 1, 2, \ldots, k$).

Further if M consists of the numbers $h_1, h_2, \ldots, h_\mu$ then

$$(7.23)\qquad L_M(z) = -n_M z + \sum_{\gamma=1}^{n_M} x_{M,\gamma} \ln z - \sum_{\gamma=1}^{n_M} \ln x_{M,\gamma}!,$$

where n_M is defined by (7.7) and where $x_{M,\gamma}$ ($\gamma = 1, 2, \ldots, n_M$) denote the pooled samples of $\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \ldots, \mathbf{x}_{h_\mu}$. Thus L satisfies condition (4.3) and if L attains its maximum for $y_{h_1} = y_{h_2} = \ldots = y_{h_\mu}$ then the samples of $\mathbf{x}_{h_1}, \mathbf{x}_{h_2}, \ldots, \mathbf{x}_{h_\mu}$ are to be pooled.

Further

$$(7.24)\qquad g_i(x_i \mid y_i, \theta_i) = \theta_i - y_i - x_i \ln \frac{\theta_i}{y_i} \quad (i = 1, 2, \ldots, k),$$

thus

$$(7.25)\qquad \begin{cases} \mathscr{E}\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\} = \theta_i - y_i - \theta_i \ln \dfrac{\theta_i}{y_i} \\ \sigma^2\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\} = \theta_i \left(\ln \dfrac{\theta_i}{y_i}\right)^2 \end{cases} \quad (i = 1, 2, \ldots, k)$$

and

$$(7.26)\qquad \frac{\sigma^2\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\}}{[\mathscr{E}\{g_i(\mathbf{x}_i \mid y_i, \theta_i) \mid \theta_i\}]^2} = \frac{\theta_i \left(\ln \dfrac{\theta_i}{y_i}\right)^2}{\left[\theta_i - y_i - \theta_i \ln \dfrac{\theta_i}{y_i}\right]^2} \quad (i = 1, 2, \ldots, k).$$

From (7.25) and (7.26) it may easily be proved that condition (6.6) is satisfied.

A practical situation of ordered parameters of Poisson distributions might occur if several toxicants are to be investigated as to their killing power for certain kinds of bacteria. If the toxicants are added in different concentrations to cultures of bacteria, knowledge may be available leading to a partial or complete ordering of the expected values of the number of survivors in the different experiments.

It may easily be verified that the conditions (4.3) and (6.6) are e.g. also satisfied if $\boldsymbol{x}_i$ possesses

1. a normal distribution with known mean μ_i and variance θ_i $(i=1, 2, \ldots, k)$,
2. an exponential distribution with parameter θ_i $(i=1, 2, \ldots, k)$,
3. a rectangular distribution between 0 and θ_i $(i=1, 2, \ldots, k)$,
4. a normal distribution with mean θ_i and known variance for $i=l_1, l_2, \ldots, l_g$ and a Poisson distribution with parameter θ_i for $i \neq l_1, l_2, \ldots, l_g$.

Acknowledgement

The author's thanks are due to Prof. Dr. J. HEMELRIJK for his stimulating help during the investigation and to Prof. Dr. D. VAN DANTZIG for reading the paper and for his constructive criticism.

(Mathematical Centre, Amsterdam)

REFERENCE

3. WALD, A., Note on the consistency of the maximum likelihood estimate, Ann. Math. Stat. 20, 595–601 (1949).

0216