MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Report S 188 (VP 5)

Prepublication

Maximum likelihood estimation of ordered probabilities

by

Constance van Eeden

January 1956

1. Introduction

The problem considered in this report concerns k ($k \geqq 2$) independent series of independent trials, each trial resulting in a success or a failure. The i-th series consists of n_i trials with $\underline{a}_i$ [1] successes and $\underline{b}_i = n_i - \underline{a}_i$ failures; π_i is the (unknown) probability of a success for each trial of the i-th series ($i = 1, 2, \ldots, k$) and $\pi_1, \pi_2, \ldots, \pi_k$ satisfy the inequalities

$$(1.1) \qquad \pi_1 \leqq \pi_2 \leqq \ldots \leqq \pi_k .$$

In section 2 a method will be described by means of which the maximum likelihood estimates may be found; in section 3 a generalization of the problem will be considered.

2. The maximum likelihood estimates of $\pi_1, \pi_2, \ldots, \pi_k$

2.1. The likelihood function

The maximum likelihood estimates of $\pi_1, \pi_2, \ldots, \pi_k$ are those values of $p_1, p_2, \ldots, p_k$ which maximize

$$(2.1.1) \quad L = L(p_1, p_2, \ldots p_k) \stackrel{\text{def}}{=} \sum_{i=1}^{k} \{a_i \lg p_i + (n_i - a_i) \lg q_i\} \qquad (q_i = 1 - p_i)$$

in the domain

$$(2.1.2) \qquad D : \begin{cases} p_1 \leqq p_2 \leqq \ldots \leqq p_k , \\ 0 \leqq p_i \leqq 1 \quad (i = 1, 2, \ldots, k). \end{cases}$$

<u>In this section L will, unless explicitely stated otherwise, only be considered in this domain D</u>; the maximum likelihood estimates will be denoted by $v_1, v_2, \ldots, v_k$ and

$$(2.1.3) \qquad L_i = L_i(p_i) \stackrel{\text{def}}{=} a_i \lg p_i + (n_i - a_i) \lg q_i \quad (i = 1, 2, \ldots, k).$$

2.2. The estimates for the case that $\frac{a_i}{n_i} \leqq \frac{a_{i+1}}{n_{i+1}}$ for each $i = 1, 2, \ldots, k-1$

<u>Theorem I</u>: <u>If</u> $\frac{a_i}{n_i} \leqq \frac{a_{i+1}}{n_{i+1}}$ <u>for each $i = 1, 2, \ldots, k-1$ then</u>

$$(2.2.1) \qquad v_i = \frac{a_i}{n_i} \quad (i = 1, 2, \ldots, k).$$

<u>Proof</u>: This follows immediately from the fact that the maximum of L in D coincides with the maximum of L in the domain: $0 \leqq p_i \leqq 1$ ($i = 1, 2, \ldots, k$) if $\frac{a_i}{n_i} \leqq \frac{a_{i+1}}{n_{i+1}}$ for each $i = 1, 2, \ldots, k-1$.

---

1) Random variables will be distinguished from numbers (e.g. from the value they take in an experiment) by underlining their symbols.

2.3. The estimates for the case that $\frac{a_i}{n_i} > \frac{a_{i+1}}{n_{i+1}}$ for at least one value of $i = 1, 2, \ldots, k-1$

In this section the following theorem will be proved.

Theorem II:

(2.3.1) $$v_i = v_{i+1} \quad \text{for each } i \text{ with } \frac{a_i}{n_i} > \frac{a_{i+1}}{n_{i+1}}.$$

Further a method will be described by means of which the estimates may be found.

For the proofs we need the following lemma and theorem.

Lemma I:

(2.3.2) $$L_i(p_i') > L_i(p_i)$$

if (p_i, p_i') is a pair of values satisfying

(2.3.3) $$0 \leqq p_i < p_i' \leqq \frac{a_i}{n_i} \quad \text{or} \quad \frac{a_i}{n_i} \leqq p_i' < p_i \leqq 1.$$

Proof:

From (2.1.3) follows

(2.3.4) $$\frac{dL_i}{dp_i} = \frac{a_i - n_i p_i}{p_i q_i}.$$

Therefore

(2.3.5) $$\frac{dL_i}{dp_i} \begin{cases} > 0 & \text{if} \quad p_i < \frac{a_i}{n_i}, \\ = 0 & \text{if} \quad p_i = \frac{a_i}{n_i}, \\ < 0 & \text{if} \quad p_i > \frac{a_i}{n_i} \end{cases}$$

and lemma I follows from (2.3.5).

Theorem III: If $\frac{a_i}{n_i} > \frac{a_{i+1}}{n_{i+1}}$ for any i and if $p_1, p_2, \ldots, p_k$ is any set in D with

(2.3.6) $$p_i < p_{i+1}$$

then a number p exists with

(2.3.7) $$p_i \leqq p \leqq p_{i+1}$$

which, substituted into $L(p_1, p_2, \ldots, p_k)$ for p_i and p_{i+1}, increases L.

Proof:

A number p which, substituted for p_i and p_{i+1} in L, increases L must satisfy the relation

(2.3.8) $$L_i(p) + L_{i+1}(p) > L_i(p_i) + L_{i+1}(p_{i+1}).$$

Further the following cases may be distinguished

1. $p_i < p_{i+1} \leqq \frac{a_i}{n_i}$; in that case we take $p = p_{i+1}$, satisfying (2.3.7).

According to lemma I we then have

$$(2.3.9) \qquad L_i(p) > L_i(p_i)$$

and p being equal to p_{i+1}

$$(2.3.10) \qquad L_{i+1}(p) = L_{i+1}(p_{i+1}).$$

(2.3.8) then follows from (2.3.9) and (2.3.10)

2. $\frac{a_i}{n_i} \leqq p_i < p_{i+1}$; in that case take $p = p_i$. In the same way as in case 1 it may be proved that this number p satisfies (2.3.7) and (2.3.8).

3. $p_i < \frac{a_i}{n_i} < p_{i+1}$; then if we take $p = \frac{a_i}{n_i}$, p satisfies (2.3.7) and

$$(2.3.11) \qquad p_i < p = \frac{a_i}{n_i}.$$

From lemma I and (2.3.11) then follows

$$(2.3.12) \qquad L_i(p) > L_i(p_i).$$

Further p satisfies

$$(2.3.13) \qquad p_{i+1} > p = \frac{a_i}{n_i} > \frac{a_{i+1}}{n_{i+1}}$$

and from lemma I and (2.3.13) follows

$$(2.3.14) \qquad L_{i+1}(p) > L_{i+1}(p_{i+1}).$$

(2.3.8) then follows from (2.3.12) and (2.3.14).

Further it will be clear that if $p_1, p_2, \ldots, p_k$ is a set in D and p a number satisfying (2.3.7) then $p_1, \ldots, p_{i-1}, p, p, p_{i+2}, \ldots p_k$ is also a set in D . Therefore from theorem III follows

<u>Theorem IV: If $\frac{a_i}{n_i} > \frac{a_{i+1}}{n_{i+1}}$ for $i = i_1$ then the maximum likelihood estimates of $\pi_1, \ldots \pi_{i_1}, \pi_{i_1+2}, \ldots, \pi_k$ are those values of $p_1, \ldots, p_{i_1}, p_{i_1+2}, \ldots, p_k$ which maximize</u>

$$(2.3.15) \qquad \sum_{i \neq i_1+1} \{ a_i' \lg p_i + (n_i' - a_i') \lg q_i \},$$

<u>where</u>

$$(2.3.16) \qquad \left.\begin{array}{l} a_i' = a_i \\ n_i' = n_i \end{array}\right\} i \neq i_1,\ i \neq i_1+1 \qquad \begin{array}{l} a_{i_1}' = a_{i_1} + a_{i_1+1} \\ n_{i_1}' = n_{i_1} + n_{i_1+1}, \end{array}$$

in the domain

$$(2.3.17) \qquad D' : \begin{cases} p_1 \leqq \dots \leqq p_{i_1} \leqq p_{i_1+2} \leqq \dots \leqq p_k, \\ 0 \leqq p_i \leqq 1 \quad (i = 1, \dots, i_1, i_1+2, \dots, k). \end{cases}$$

In this way the problem is reduced to the case of $k-1$ series of trials and may then be solved by means of theorem I or reduced to the case of $k-2$ series of trials by means of theorem IV. This procedure is necessarily finite, k being finite. Therefore it leads to a unique maximum for L.
Theorem II then follows from this uniqueness and the foregoing theorems.

2.4. Example

The procedure described in section 2.3 may be illustrated by means of the following example.

Suppose $k=4$ and

(2.4.1)

i	1	2	3	4
a_i	4	3	10	8
n_i	10	5	30	15
$\frac{a_i}{n_i}$	0,4	0,6	0,33	0,53.

From (2.4.1) and theorem II follows

$$(2.4.2) \qquad V_2 = V_3.$$

The problem is then reduced to the case of $k-1=3$ series of trials with (cf. theorem IV):

(2.4.3)

i	1	2	4
a'_i	4	13	8
n'_i	10	35	15
$\frac{a'_i}{n'_i}$	0,4	0,37	0,53.

From (2.4.3) and theorem II follows

$$(2.4.4) \qquad V_1 = V_2,$$

which reduces the problem to the case $k-2=2$ series of trials with

$$(2.4.5)\quad \begin{cases} i & 1 & 4 \\ a_i'' & 17 & 8 \\ n_i'' & 45 & 15 \\ \dfrac{a_i''}{n_i''} & 0,38 & 0,53. \end{cases}$$

Then from theorem I and (2.4.5) follows

$$(2.4.6)\qquad V_1 = 0,38 \ , \quad V_4 = 0,53$$

and from (2.4.2), (2.4.4) and (2.4.6)

$$(2.4.7)\qquad V_1 = V_2 = V_3 = 0,38 \ , \quad V_4 = 0,53.$$

3. A generalization of the problem

The problem treated in the foregoing sections may be generalized as follows:

Suppose the probabilities $\pi_1, \pi_2, \ldots, \pi_k$ satisfy the inequalities

$$(3.1)\qquad \alpha_{i,j}\,(\pi_i - \pi_j) \leqq 0 \qquad (i,j = 1,2,\ldots,k),$$

where

$$(3.2)\quad \begin{cases} \alpha_{i,j} = -\alpha_{j,i}\,, & \\ \alpha_{i,j} = 0 & \text{for } m_0 \text{ pairs of values } (i,j) \text{ with } i<j, \\ \alpha_{i,j} = 1 & \text{for } m_1 \text{ pairs of values } (i,j) \text{ with } i<j, \end{cases}$$

$$(3.3)\qquad m_0 + m_1 = \binom{k}{2}$$

and, if $i < l < j$ then

$$(3.4)\qquad \alpha_{i,j} = 1 \quad \text{if} \quad \alpha_{i,l} = \alpha_{l,j} = 1.$$

If $m_1 = 0$ then no restriction is imposed on $\pi_1, \pi_2, \ldots, \pi_k$ and it is well known that in this case the maximum likelihood estimate of π_i is : $\frac{a_i}{n_i}$ $(i = 1,2,\ldots k)$. Further, if $m_0 = 0$ then (3.1) is identical with: $\pi_1 \leqq \pi_2 \leqq \ldots \leqq \pi_k$ and this case has been considered in the foregoing sections. Therefore we suppose

$$(3.5)\quad \begin{cases} m_1 \geqq 1, \\ m_0 \geqq 1. \end{cases}$$

Then from (3.3) and (3.5) it follows that

(3.6) $$k \geqq 3.$$

In this report only the case $k=3$ will be considered; the maximum likelihood estimates will be denoted by v_1, v_2, v_3 and the domain

(3.7) $$\begin{cases} \alpha_{i,j}(p_i - p_j) \leqq 0 \\ 0 \leqq p_i \leqq 1 \end{cases}$$

will be denoted by D_1.

The following cases may be distinguished (cf. (3.3) and (3.5)).

(3.8) $$\begin{cases} 1. \quad m_1 = 1, \; m_0 = 2, \\ 2. \quad m_1 = 2, \; m_0 = 1. \end{cases}$$

In case (3.8.1) we may suppose, without any loss of generality

(3.9) $$\alpha_{1,2} = \alpha_{1,3} = 0, \quad \alpha_{2,3} = 1.$$

It will be clear that in this case

(3.10) $$v_1 = \frac{a_1}{n_1}$$

and that the estimates of π_2 and π_3 may be found by means of the procedure described in section 2.

In the case (3.8.2) we may suppose without any loss of generality

(3.11) $$\alpha_{1,2} = \alpha_{1,3} = 1, \quad \alpha_{2,3} = 0$$

and

(3.12) $$\frac{a_2}{n_2} \leqq \frac{a_3}{n_3}.$$

Theorem V: If $k=3$ and (3.11) and (3.12) are satisfied and if p_1, p_2, p_3 is a set in D_1 with

(3.13) $$p_2 > p_3$$

then a number p exists with

(3.14) $$\begin{cases} 1. \quad p_2 \geqq p \geqq p_3, \\ 2. \quad L_2(p) + L_3(p) > L_2(p_2) + L_3(p_3). \end{cases}$$

Proof: The proof is analogous to the proof of theorem IV. Here the following cases may be distinguished

1. $p_2 > p_3 \geqq \frac{a_2}{n_2}$; then take $p = p_3$,
2. $\frac{a_2}{n_2} \geqq p_2 > p_3$; then take $p = p_2$,
3. $p_2 > \frac{a_2}{n_2} > p_3$; then take $p = \frac{a_2}{n_2}$.

Further it will be clear that if p_1, p_2, p_3 is a set in D_1 with $p_2 > p_3$ then, for each number p satisfying (3.14.1), p_1, p, p is also a set in D_1. Therefore it follows from theorem V that

Theorem VI: If $k=3$ and (3.11) and (3.12) are satisfied then the maximum likelihood estimates of π_1, π_2, π_3 are the values of p_1, p_2, p_3 which maximize L in the domain

$$(3.15) \qquad p_1 \leqq p_2 \leqq p_3 .$$

In this way the problem may, for $k=3$, be reduced to the case treated in section 2.
This may be illustrated by means of the following example.

Suppose $k=3$,

(3.16)

i	1	2	3
a_i	13	12	6
n_i	20	25	15
$\frac{a_i}{n_i}$	0,65	0,48	0,4

and

$$(3.17) \qquad \alpha_{1,3} = \alpha_{2,3} = 1 \quad , \alpha_{1,2} = 0 .$$

If we define

$$(3.18) \qquad \begin{cases} \pi_1' \overset{\text{def}}{=} 1-\pi_3 , \\ \pi_2' \overset{\text{def}}{=} 1-\pi_1 , \\ \pi_3' \overset{\text{def}}{=} 1-\pi_2 , \end{cases}$$

then the problem is reduced to the case of 3 series of trials with

(3.19)

i	1	2	3
a_i'	9	7	13
n_i'	15	20	25
$\frac{a_i'}{n_i'}$	0,6	0,35	0,52

and

$$(3.20) \qquad \alpha_{1,2}' = \alpha_{1,3}' = 1 \quad , \quad \alpha_{2,3}' = 0 .$$

For these three series of trials (3.11) and (3.12) are satisfied and therefore the estimates of π_1', π_2', π_3' (denoted by v_1', v_2', v_3') may be found by means of theorem VI. This leads to

$$(3.21) \qquad v_1' = v_2' = 0{,}46 \quad , v_3' = 0{,}52$$

and from (3.18) and (3.21) follows

$$(3.22) \qquad v_1 = v_3 = 0{,}54 \quad , \quad v_2 = 0{,}48 \, .$$

The investigation of cases with $k > 3$ is in progress.