SA

stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 11/71 SEPTEMBER

W.N. VAN NOOTEN SA

# CONVERGENCE OF THE "ITERATIVE PROPORTIONAL FITTING PROCEDURE" FOR GENERALIZED FREQUENCY TABLES

Prepublication

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

## INTRODUCTION

This report deals with a class of models generalizing two- and more-dimensional frequency tables. Given such a model the concept of 'interaction' is defined. If certain quantities, comparable to marginal totals as well as the interaction terms are given, a uniquely defined probability vector exists satisfying the marginal conditions and having the prescribed interaction.

The aim of this report is to show how this probability vector can be found. An iterative procedure is described and it is shown that this procedure converges to the desired solution.

The solution vector p may satisfy parametric constraints, observational constraints or a mixture of these. The procedure is independent of these differences. This allows of a uniform treatment of situations, statistically different but computationally equivalent.

## 1. MULTINOMIAL DISTRIBUTIONS SUBJECT TO LINEAR RESTRICTIONS

In the statistical literature two- and more-dimensional frequency tables are usually treated as multinomial distributions subject to restrictions imposed on rows, columns, layers, row-column pairs and the like. The restrictions may be imposed by fixing marginal totals or by conditioning on them, or by fixing marginal probabilities.

It is essential, however, that a number of subsets of the k cells in the table play a special rôle. In particular a statement on the total number of observations will always be made.

For the present purpose it is irrelevant from what source the restrictions have arisen. The important thing is what subsets of cells are involved in the restrictions.

We identify the cells by labels, e.g. $1,2,\ldots,k$, and the set of all labels is indicated by $L = \{1,2,\ldots,k\}$. Probabilities $p_i$ are attached to the labels, summarized in a probability vector $p = (p_1,p_2,\ldots,p_k)'$. The prime denotes transposition. Subsets V of L will be used as labels as well. For instance, $p_V = \sum_{i\in V} p_i$ and $p_L = \sum_{i=1}^{k} p_i = 1$. The multinomial distribution with parameter vector p gives rise to the stochastic vector $\underline{n} = (\underline{n}_1,\underline{n}_2,\ldots,\underline{n}_k)'$ of possible outcomes on the probability space defined by L, p and $n_L = \sum_{i=1}^{k} \underline{n}_i$. By $\underline{n}_V$ we mean $\sum_{i\in V} \underline{n}_i$.

The restrictions may now be written in the form $p_V = \alpha_V$ or $\underline{n}_V = n_V$, where $\alpha_V$ is a fixed number, $n_V$ a fixed non-negative integer. In the latter case we switch over to $\underline{n}_V/n_L = \beta_V = n_V/n_L$ and we do not stress the difference any more.

The probability vectors p (and the set L) are chosen such that $p_i > 0$ for all $i \in L$. By $\mathcal{P}$ we indicate $\{p\colon p_i > 0 \text{ for } i \in L,\ p_L = 1\}$.

It is natural to summarize the sets $V_j$ $(j=1,2,\ldots,3)$, the $p_{V_j}$ or $n_{V_j}/n_L$ of which are fixed, in a matrix, the configuration matrix $H = (h_{ij})$, of the following form:

$$(1.1)\qquad \begin{cases} h_{ij} = 1 & \text{if } i \in V_j, \\ h_{ij} = 0 & \text{if } i \notin V_j. \end{cases}$$

Obviously, several configuration matrices belong to the same problem, since the order of the columns of H may be altered, columns may be repeated, and restrictions may be derived from others which leads to still other columns in H.

Two configuration matrices H and $H^*$ are said to be <u>equivalent</u> if the columns of H and those of $H^*$ span the same vector space $\mathcal{S}$. The collection of all equivalent configuration matrices to a given problem will be indicated by $\mathcal{H}$.

In the sequel we take configuration matrices of the following form:

(1.2) $H \in \mathcal{H}$ is a (k×s)-matrix. The s columns of H may be divided up into l groups, consisting of $s_1, s_2, \ldots, s_l$ columns, $s_\omega > 1$ for $\omega = 1,2,\ldots,l$ and having the following property: Let $t_0 = 0$, $t_\alpha = \sum_{\omega=1}^{\alpha} s_\omega$, then the α-th group consists of the columns $t_{\alpha-1}+1, t_{\alpha-1}+2, \ldots, t_\alpha$; now $\sum_{j=t_{\alpha-1}+1}^{t_\alpha} h_{ij} = 1$ must hold.

(1.2) implies that the subsets $V_j$ of L, characterized by the columns $t_{\alpha-1}+1, t_{\alpha-1}+2, \ldots, t_\alpha$ are disjoint and have L as their union. It is clear that H can always be chosen according to (1.2), for instance by taking the even-numbered columns to be the complement of the preceding odd-numbered column.

The restrictions now have the form

$$H'p = \xi^{(1)}. \tag{1.3}$$

All classical situations can be described in terms of H-matrices, but moreover an essential generalization is obtained.

## 2. INTERACTION

Let H be a configuration matrix of the form (1.2), $V_1, V_2, \dots, V_l$ the subset of L characterized by the columns of H.

DEFINITION 1: $p^{(0)}$ and $p \in P$ are said to have the same interaction with respect to H iff k-dimensional vectors $\rho^{(1)}, \rho^{(2)}, \dots, \rho^{(s)}$ exist,

$$\left.\begin{array}{ll} \rho_i^{(j)} = \lambda_j > 0 & \text{for } i \in V_j \\ \rho_i^{(j)} = 1 & \text{for } i \notin V_j \end{array}\right\} \quad (i=1,2,\dots,k)$$

such that

$$p_i = p_i^{(0)} \prod_{j=1}^{s} \rho_i^{(j)}.$$

Let $\tau_j = \log \lambda_j$, $j = 1,2,\dots,s$. The vectors $p^{(0)}$ and $p \in P$ have the same interaction with respect to H iff

$$(2.1) \qquad p = p^{(0)} e^{H\tau} \qquad \text{for some s-vector } \tau.$$

$p^{(0)}e^{H\tau}$ here indicates the column-vector, given by $p_i^{(0)} \exp(\sum_{j=1}^{s} h_{ij}\tau_j)$, $i = 1,2,\dots,k$.

It is clear from (2.1) that the property of having the same interaction defines an equivalence relation on $P \times P$.

The meaning of definition 1 will now be illustrated for a special case. Let $H_\alpha$ be the matrix consisting of the columns $t_{\alpha-1}+1, t_{\alpha-1}+2, \dots, t_\alpha$ of H ($1 \leq \alpha \leq l$). Vectors $p^{(\alpha-1)}, p^{(\alpha)} \in P$, satisfying $p^{(\alpha)} = p^{(\alpha-1)} e^{H_\alpha \tau^{(\alpha)}}$ for some $s_\alpha$-vector $\tau^{(\alpha)}$ have the same interaction. It is seen that $p^{(\alpha)}$ is obtained from $p^{(\alpha-1)}$ by dividing up the components $p_i^{(\alpha-1)}$ such that $p_i^{(\alpha-1)}$ and $p_h^{(\alpha-1)}$ are in the same group iff i and $h \in V_j$ for some j, ($t_{\alpha-1} < j \leq t_\alpha$), and for each group multiplying the components in that group by the same constant. As a result, interaction is independent of selection procedures, affecting the $p_{V_j}$ ($t_{\alpha-1} < j \leq t_\alpha$) only, or of prior distributions on the $p_{V_j}$. An application of this property is given by Joel H. Levine and described in Mosteller (1968), p.8 and 9.

If p and $p^{(0)}$ have the same interaction with respect to $H \in \mathcal{H}$, $p \overset{H}{\sim} p^{(0)}$ for short, and if $H^* \in \mathcal{H}$, where H is a (k×s)-matrix and $H^*$ a (k×t)-matrix, a (t×s)-matrix A exists such that $H^*A = H$. According to (2.1), $p = p^{(0)}e^{H\tau} = p^{(0)}e^{H^*A\tau}$ for some s-vector $\tau$, and consequently $p = p^{(0)}e^{H^*\sigma}$ for a t-vector $\sigma$. Therefore, (2.1) holds irrespective of the choice of $H \in \mathcal{H}$.

The relation (2.1) may be rewritten as

$$\text{(2.2)} \qquad \log p = \log p^{(0)} + H\tau.$$

Let r be the rank of H, that is, the dimension of $\mathcal{S}$ is r, then the linear space $\mathcal{S}_\perp$, the orthogonal complement of $\mathcal{S}$ has dimension k-r. Let $b^{(1)}, b^{(2)}, \ldots, b^{(k-r)}$ be a base for $\mathcal{S}_\perp$. The matrix B having the $b^{(j)}$, $(1 \leq j \leq k-r)$ as columns is a k×(k-r)-matrix of rank k-r, satisfying

$$\text{(2.3)} \qquad B'H = 0.$$

B is called a $\mathcal{S}_\perp$-matrix. From (2.2) and (2.3) we have that $p, p^{(0)} \in \mathcal{P}$, $p \overset{H}{\sim} p^{(0)}$ satisfy

$$\text{(2.4)} \qquad B' \log p = B' \log p^{(0)}.$$

If, on the other hand, (2.4) holds, then $\log p - \log p^{(0)} \in \mathcal{S}$, that is $\log p - \log p^{(0)}$ is of the form $H\tau$ for some s-vector $\tau$, and (2.2) holds. Summarizing we have:

$$p \overset{H}{\sim} p^{(0)} \longleftrightarrow p = p^{(0)}e^{H\tau} \longleftrightarrow B' \log p = B' \log p^{(0)}.$$

This allows of the following definition:

DEFINITION 2: Given a configuration matrix H and a $\mathcal{S}_\perp$-matrix B, $p \in \mathcal{P}$ has interaction $\gamma_B$ if $B' \log p = \gamma_B$. The components of $\gamma_B$ will be called interaction terms.

## 3. THE "ITERATIVE PROPORTIONAL FITTING PROCEDURE"

To given matrices H of the form (1.2) and B we have a transformation

$$\psi(p) = \begin{cases} H'p \\ B' \log p \end{cases} \qquad (3.1)$$

Let $\xi = \binom{\xi^{(1)}}{\xi^{(2)}}$, $\xi^{(1)} = H'p$, $\xi^{(2)} = B' \log p$.

$\psi$ has a uniquely defined inverse on $Y = \{\psi(p): p \in P\}$. (See van Nooten (1971)).

We shall try to find p given $\psi(p)$. To this end a method known as the "iterative proportional fitting procedure" (IPFP) is generalized. The procedure has been introduced by Deming and Stephan (1940), Fienberg (1970) gives a review of the literature. Ireland and Kullback (1968) give a proof of the convergence of the IPFP for a number of special cases. The main line of their method of proof is followed here.

(3.2) The IPFP is defined as follows:

1. Take $p^{(0,0)} \in P$ for which $B' \log p^{(0,0)} = \xi^{(2)}$.
2. A sequence $p^{(n,\alpha)} \in P$ ($\alpha=1,2,\ldots,l$; $n=0,1,2,\ldots$) is defined by

$$p^{(n,l)} = p^{(n+1,0)};$$

$$p^{(n,\alpha)} = p^{(n,\alpha-1)} e^{H_\alpha \tau^{(n,\alpha)}}, \text{ where}$$

$$\tau^{(n,\alpha)} = \log \lambda^{(n,\alpha)} = \log(\xi_j^{(1)}/p_{V_j}^{(n,\alpha-1)}), \text{ for } t_{\alpha-1} < j \leq t_\alpha.$$

Remarks:

ad 1: $B' \log \mu = \xi^{(2)}$ does have solutions $\mu$ for any $\xi^{(2)} \in R^{k-r}$.

Let $\mu^*$ be a solution, then $\mu^*$ is a probability vector, apart from a constant factor, $\sum_{i=1}^{k} \mu_i^* = c$. Let $p^{(0,0)} = \mu^* c^{-1}$. Then $B' \log p^{(0,0)} =$ $= B' \log \mu^* - B' \log C$ (where $\log C = (\log c, \log c, \ldots, \log c)'$).

But since the vector $(1,1,\ldots,1)'$, as the sum of columns $t_{\alpha-1}+1$ to $t_\alpha$ of H is in $S$, $B' \log C = 0$. Therefore $B' \log p^{(0,0)} = \xi^{(2)}$.

ad 2: for fixed n and $\alpha$ the transition of $p^{(n,\alpha-1)}$ to $p^{(n,\alpha)}$ consists of fitting the marginals $\xi_j^{(1)}$, $t_{\alpha-1} < j \leq t_\alpha$ without altering the interaction. Generally, the marginals already fitted are disturbed at each step. Therefore it must be shown that the sequence $p^{(n,\alpha)}$ converges and that the limit satisfies (3.1). If so, this limit is the desired p owing to the uniqueness of the inverse of $\psi$.

## 4. CONVERGENCE OF THE IPFP

Let a and b be two vectors of dimension t, where $a_i, b_i > 0$ $(1 \leq i \leq t)$. We define

$$(4.1) \qquad d_t(a,b) = \sum_{i=1}^{t} a_i \log \frac{a_i}{b_i} .$$

The right hand side will be denoted by $a' \log \frac{a}{b}$ .

If a and b are probability vectors, i.e. if $\sum_{i=1}^{k} a_i = \sum_{i=1}^{k} b_i = 1$, then $d_t(a,b)$ satisfies

(1) $d_t(a,b) \geq 0$

(2) $d_t(a,b) = 0$ iff $a = b$,

(3) $d_t(a,b)$ is a convex function both in

$a \in P_t = \{x\colon x_i \geq 0 \ (1 \leq i \leq t), \ \sum_{i=1}^{t} x_i = 1\}$ for fixed b and in

$b \in P_t$ for fixed a.

(4) Writing $\zeta_i = \frac{a_i}{b_i}$ we have

$d_t(a,b) = \sum_{i=1}^{t} (\zeta_i \log \zeta_i)\, b_i$, and $\min_{\zeta_i \in (0,\infty)} \zeta_i \log \zeta_i = -\frac{1}{e}$ .

For (1) and (2) see Kullback (1968), (3) and (4) are easily verified.

Let p be the solution of $\psi(p) = \xi$, $\xi \in Y$. We write $\xi_\alpha^{(1)}$ to abbreviate $(\xi_{t_{\alpha-1}+1}^{(1)}, \ldots, \xi_{t_\alpha}^{(1)})'$.

We then may state:

$$(4.2) \qquad d_k(p, p^{(n+1,0)}) = d_k(p, p^{(n,0)}) - \sum_{\alpha=1}^{l} d_{s_\alpha}(\xi_\alpha^{(1)}, H_\alpha' p^{(n,\alpha-1)}).$$

Proof:

$$d_k(p, p^{(n,0)}) - d_k(p, p^{(n+1,0)}) = p'(\log \frac{p}{p^{(n,0)}} - \log \frac{p}{p^{(n+1,0)}}) =$$

$$= p' \log \frac{p^{(n+1,0)}}{p^{(n,0)}} .$$

$$p^{(n+1,0)} = p^{(n,1)} = p^{(n,0)} \exp\left(\sum_{\alpha=1}^{l} H_\alpha \tau^{(n,\alpha)}\right), \text{ according to (3.2).}$$

Therfore, $\log \dfrac{p^{(n+1,0)}}{p^{(n,0)}} = \sum_{\alpha=1}^{l} H_\alpha \tau^{(n,\alpha)}$, and

$$p' \log \frac{p^{(n+1,0)}}{p^{(n,0)}} = \sum_{\alpha=1}^{l} p' H_\alpha \tau^{(n,\alpha)} = \sum_{\alpha=1}^{l} \xi_\alpha^{(1)} \log \frac{\xi_\alpha^{(1)}}{H_\alpha' p^{(n,\alpha-1)}} =$$

$$= \sum_{\alpha=1}^{l} d_{s_\alpha}(\xi_\alpha^{(1)}, H_\alpha' p^{(n,\alpha-1)}), \text{ and (4.2) follows immediately.}$$

Now $\sum_{j=t_{\alpha-1}+1}^{t_\alpha} \xi_j^{(1)} = (1,1,\ldots,1)\, H_\alpha' p = \sum_{i=1}^{k} p_i = 1$ and

$$\sum_{j=t_{\alpha-1}+1}^{t} H_\alpha' p^{(n,\alpha-1)} = \sum_{i=1}^{k} p_i^{(n,\alpha-1)} = 1.$$

By (1), therefore, $d_{s_\alpha}(\xi^{(1)}, H_\alpha' p^{(n,\alpha-1)}) \geq 0$ for all $\alpha$.

This, together with (4.2) leads to:

$$(4.3) \qquad d_k(p, p^{(n+1,0)}) \leq d_k(p, p^{(n,0)}).$$

In (4.3) the equality sign holds for some n iff for $\alpha = 1,2,\ldots,k$: $\xi_\alpha^{(1)} = H_\alpha' p^{(n,\alpha-1)}$ (property (2)). If this is the case, the process ends in a finite number of steps. $p^{(n,0)}$ is the result and we have both $B' \log p^{(n,0)} = \xi^{(2)}$ and $H' p^{(n,0)} = \xi^{(1)}$. Since $p^{(n,0)} \in \mathcal{P}$, $p^{(n,0)}$ must be the (uniquely defined) p.

Now suppose that for all n

$$(4.4) \qquad d_k(p, p^{(n+1,0)}) < d_k(p, p^{(n,0)}).$$

Then $c = \lim_{n\to\infty} d_k(p, p^{(n,0)}) \geq 0$ does exist (according to property (1) of $d_t$). Consequently

(4.5) $$\lim_{n\to\infty} d_{s_\alpha}(\xi_\alpha^{(1)}, H'_\alpha p^{(n,\alpha-1)}) = 0 \qquad \text{for } \alpha = 1,2,\ldots,l.$$

Let $\bar{P}_t = \{a\colon a_i \geq 0 \ (1 \leq i \leq t), \ \sum_{i=1}^{t} a_i = 1\}$.

LEMMA: If $a$ and $b^{(n)}$ $(n=0,1,2,\ldots) \in P_t$ and if $d_t(a,b^{(n)})$ converges to $\beta < \infty$ for $n \to \infty$, then the sequence $b^{(n)}$ has limitpoints $b \in P_t$, all satisfying $d_t(a,b) = \beta$ and no limitpoints in $\bar{P}_t \backslash P_t$.

Proof: $d_t(a,x)$ is continuous for $x \in P_t$. If therefore $b \in P_t$ is a limitpoint of $b^{(n)}$, $d_t(a,b) = \beta$. Now suppose $b \in \bar{P}_t \backslash P_t$ is a limitpoint of $b^{(n)}$. Then for at least one $i$: $b_i = 0$. Let $b^{(n_h)}$, $h = 1,2,\ldots$ be a subsequence of $b^{(n)}$ converging to $b$, then

$$d_t(a,b^{(n_h)}) = \sum_{j\neq i} (\zeta_j \log \zeta_j)\, b_j^{(n_h)} + a_i \log \frac{a_i}{b_i^{(n_h)}} \geq$$

$$\geq -\frac{1}{e} \sum_{j\neq i} b_j^{(n_h)} + a_i \log \frac{a_i}{b_i^{(n_h)}} \geq -\frac{1}{e} + a_i \log \frac{a_i}{b_i^{(n_h)}} .$$

Here $a_i$ is fixed, positive and $\lim_{h\to\infty} b_i^{(n_h)} = 0$.

Therefore, $\lim_{h\to\infty} d_t(a,b^{(n_h)}) = \infty$ contradicting $\lim_{h\to\infty} d_t(a,b^{(n)}) = \beta < \infty$.

In particular we have from (4.5) and property (2) that $\lim_{n\to\infty} H'p^{(n,\alpha-1)}$ exists and equals $\xi_\alpha^{(1)}$, for $\alpha = 1,2,\ldots,l$.

From $\lim_{n\to\infty} d_k(p,p^{(n,0)}) = c < \infty$ it follows that the limitpoints $p$ of the sequence $p^{(n,0)}$ all belong to $P$ and satisfy $d_k(p,p^{\infty}) = c$.

Let $p^*$ be such a limitpoint and let $p^{(n_h,0)}$ $(h=1,2,\ldots)$ be a subsequence of $p^{(n,0)}$ converging to it. We then have:

(i) $$B' \log p^* = \lim_{h\to\infty} B' \log p^{(n_h,0)} = \lim_{h\to\infty} \xi^{(2)} = \xi^{(2)}$$

(ii) $$H'p^{*} = \lim_{h\to\infty} H'p^{(n_h,0)} = \xi^{(1)}.$$

By uniqueness of the solution of $\psi(p) = \xi$ for $\xi \in Y$, $p^{*} = p$ and consequently the sequence $p^{(n,0)}$ has p as a limit.

REFERENCES:

Ireland, C.T. and Kullback, S.
Contingency tables with given marginals (1968), Biometrika 55, 1 (179-188).

Kullback, S.
Information theory and statistics (1968), Dover Publications Inc., New York.

Mosteller, F.
Association and estimation in contingency tables (1968), JASA 63, (1-29).

van Nooten, W.N.
Restricted multinomial distributions and interaction terms (1971), unpublished manuscript.