KYBERNETIKA — VOLUME 39 (2003), NUMBER 6, PAGES 731 – 737

A NOTE ON THE IPF ALGORITHM WHEN THE MARGINAL PROBLEM IS UNSOLVABLE

CLAUDIO ASCI AND MAURO PICCIONI

In this paper we analyze the asymptotic behavior of the IPF algorithm for the problem of finding a $2 \times 2 \times 2$ contingency table whose pair marginals are all equal to a specified 2×2 table, depending on a parameter. When this parameter lies below a certain threshold the marginal problem has no solution. We show that in this case the IPF has a "period three limit cycle" attracting all positive initial tables, and a bifurcation occur when the parameter crosses the threshold.

Keywords: contingency tables, hierarchical models, partial maximization algorithms
AMS Subject Classification: 62H17, 65C60

1. INTRODUCTION

The Iterated Proportional Fitting (IPF) algorithm has the purpose of constructing solutions to the marginal problem. This consists in constructing joint contingency tables with certain specified marginal tables [1], [5]. This problem arises in various instances in statistics, for example for the computation of maximum likelihood estimates for log-affine hierarchical models [2], [3], [6].

The IPF algorithm produces a sequence of contingency tables through the cyclic application of "fitting operators" T_i which impose to the contingency tables a fixed marginal $\mathbf{n}_i$ over the subset C_i of the variables, for $i = 1, ..., r$. A minimal requirement is that the family $\{\mathbf{n}_i\}_{i=1,2,\dots,r}$ is compatible, i.e. by marginalizing $\mathbf{n}_i$ and $\mathbf{n}_j$ over $C_i \cap C_j$ the same table is obtained. When $\{C_1, C_2, ..., C_r\}$ are the cliques of a triangulated graph (whose corresponding hierarchical model is then called a decomposable model) ordered in a suitable way, one iteration cycle of the IPF from a suitably chosen initial table (e.g. the uniform one) allows to solve the marginal problem for any compatible family of positive marginal contingency tables $\{\mathbf{n}_i\}_{i=1,2,\dots,r}$.

The simplest example of non decomposable model is given by the subsets $C_1 = \{1,2\}$, $C_2 = \{2,3\}$, $C_3 = \{1,3\}$ of three variables $\{1,2,3\}$. The corresponding marginal problem consists in the construction of three-way contingency tables with all the pair marginals fixed. The easiest example which illustrates nicely the possible lack of solutions to the marginal problem is obtained by imposing all 2×2 marginals

$\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$ equal to the following 2×2 contingency table

$$a_\varepsilon(i,j) = \begin{cases} \varepsilon & i = j \\ \frac{1}{2} - \varepsilon & i \neq j. \end{cases}$$

Let us check for which values of ε there exists a $2 \times 2 \times 2$ joint contingency table $\mathbf{n}$ with all 2×2 marginals equal to $\mathbf{a}_\varepsilon$. It is easy to see that $\mathbf{n}$ must be symmetric and

$$\begin{cases} n(1,1,1) = -n(1,1,0) + \varepsilon \\ n(1,0,0) = -n(1,1,0) + \frac{1}{2} - \varepsilon \\ n(0,0,0) = n(1,1,0) + 2\varepsilon - \frac{1}{2} \end{cases}, \tag{1}$$

from which:

$$\max\left(0, \frac{1}{2} - 2\varepsilon\right) \leq n(1,1,0) \leq \min\left(\varepsilon, \frac{1}{2} - \varepsilon\right). \tag{2}$$

For any $n(1,1,0)$ satisfying (2) the other entries of $\mathbf{n}$ can be obtained from (1) and symmetry. If $n(1,1,0)$ is strictly inside the above interval the table $\mathbf{n}$ is positive. Therefore, for $\frac{1}{6} < \varepsilon < \frac{1}{2}$, there exist infinitely many positive contingency tables $2 \times 2 \times 2$ with the required marginals. For $\varepsilon = \frac{1}{6}$ the unique solution left is degenerate (i.e. has some zero entries) and it is given by the following table $\mathbf{m}$

$$\begin{cases} m(0,0,0) = m(1,1,1) = 0 \\ m(i,j,k) = \frac{1}{6} \quad \forall \quad (i,j,k) \neq (0,0,0), (1,1,1) \end{cases}, \tag{3}$$

whereas for $\varepsilon < \frac{1}{6}$ the marginal problem has no solution.

The purpose of this note is to analyze the behavior of the IPF for this family of examples, starting from positive initial tables. It is natural to look at these initial tables, since a zero in the positive initial table is preserved in all the iterations, so it may lead to discard a priori the solutions to the marginal problem. It turns out that the algorithm exhibits a bifurcation at the critical point $\varepsilon = \frac{1}{6}$. In fact for larger values of ε, the algorithm converges to a positive solution of the marginal problem, which depends on the original table. All these solutions collapse into the single degenerate table $\mathbf{m}$ when $\varepsilon = \frac{1}{6}$, to which the algorithm converges independently of the positive starting table. For smaller values of ε the limit behavior of the algorithm is a cycle through three different degenerate tables, which are obtained one from the other through a corresponding cyclic shift of the coordinates. Thus it is seen that the bifurcation breaks the symmetry of the marginal problem.

The paper is organized as follows. In the next section our result is precisely stated. The subsequent section is entirely devoted to the proof of Theorem 2, which is our original contribution.

2. THE RESULT

Let us follow the usual convention of denoting marginal contingency tables by writing a + over those indices which are saturated. In this way the fitting operators over C_1, C_2 and C_3 can be written as

$$\begin{cases} T_1\left(\mathbf{p}\right)(i,j,k) = a_\varepsilon(i,j)\left(\frac{p(i,j,k)}{p(i,j,+)}\right) \\ T_2\left(\mathbf{p}\right)(i,j,k) = a_\varepsilon(j,k)\left(\frac{p(i,j,k)}{p(+,j,k)}\right) \\ T_3\left(\mathbf{p}\right)(i,j,k) = a_\varepsilon(i,k)\left(\frac{p(i,j,k)}{p(i,+,k)}\right) \end{cases}.$$

If

$$\mathbf{p} = T_1\left(\mathbf{p}\right) = T_2\left(\mathbf{p}\right) = T_3\left(\mathbf{p}\right)$$

then $\mathbf{p}$ is a solution of the marginal problem and conversely. The fitting operators T_1, T_2 and T_3 leave invariant the set of contingency tables with positive entries. From any such table $\mathbf{p}^{(0)} > \mathbf{0}$ define for $n = 0, 1, \ldots$ the IPF iterations

$$\begin{cases} \mathbf{p}^{(3n+1)} = T_1\left(\mathbf{p}^{(3n)}\right) \\ \mathbf{p}^{(3n+2)} = T_2\left(\mathbf{p}^{(3n+1)}\right) \\ \mathbf{p}^{(3n+3)} = T_3\left(\mathbf{p}^{(3n+2)}\right) \end{cases}. \tag{4}$$

If we define for $\mathbf{p} > \mathbf{0}$

$$\lambda(\mathbf{p}) = \log \frac{p(1,1,1)\,p(1,0,0)\,p(0,1,0)\,p(0,0,1)}{p(0,0,0)\,p(1,1,0)\,p(0,1,1)\,p(1,0,1)}, \tag{5}$$

then

$$\lambda(\mathbf{p}) = \lambda(T_i(\mathbf{p})), \quad i = 1,2,3. \tag{6}$$

This means that the iterations of the IPF algorithm do not change the value of λ. The following result explains the asymptotic behavior of the IPF in the case $\frac{1}{6} < \varepsilon < \frac{1}{2}$, when the marginal problem has a positive solution. It can be immediately obtained from the general convergence result in [6] (Theorem 4.13), whose essential ingredient is to consider the IPF as a partial maximization algorithm [4].

Proposition 1. Let $\mathbf{p}^{(0)} > \mathbf{0}$, and $\frac{1}{6} < \varepsilon < \frac{1}{2}$. Then $\lim_{n\to+\infty} \mathbf{p}^{(n)} = \mathbf{p}^{(\infty)}_{\lambda(\mathbf{p}^{(0)})}$, where $\mathbf{p}^{(\infty)}_{\lambda}$ is the unique solution $\mathbf{p} > \mathbf{0}$ to the marginal problem such that $\lambda(\mathbf{p}) = \lambda$.

The remaining cases are covered by the following

Theorem 2. Let $0 < \varepsilon \le \frac{1}{6}$ and $\mathbf{p}^{(0)} > \mathbf{0}$. Then $\lim_{n\to+\infty} \mathbf{p}^{(3n)} = \mathbf{p}^{(\infty)}$, where

$$\begin{cases} p^{(\infty)}(0,0,0) = p^{(\infty)}(1,1,1) = 0 \\ p^{(\infty)}(0,1,0) = p^{(\infty)}(1,0,1) = \varepsilon \\ p^{(\infty)}(0,0,1) = p^{(\infty)}(1,1,0) = \frac{-\varepsilon+1-\sqrt{-3\varepsilon^2+2\varepsilon}}{2} \\ p^{(\infty)}(1,0,0) = p^{(\infty)}(0,1,1) = \frac{-\varepsilon+\sqrt{-3\varepsilon^2+2\varepsilon}}{2} \end{cases} \tag{7}$$

Moreover $\lim_{n\to+\infty} \mathbf{p}^{(3n+1)} = \mathbf{r}^{(\infty)}$ and $\lim_{n\to+\infty} \mathbf{p}^{(3n+2)} = \mathbf{s}^{(\infty)}$, where

$$r^{(\infty)}(i,j,k) \equiv T_1(\mathbf{p}^{(\infty)})(i,j,k) = p^{(\infty)}(j,k,i), \tag{8}$$

$$s^{(\infty)}(i,j,k) \equiv T_2(\mathbf{r}^{(\infty)})(i,j,k) = p^{(\infty)}(k,i,j), \tag{9}$$

and

$$p^{(\infty)}(i,j,k) = T_3(\mathbf{s}^{(\infty)})(i,j,k).$$

It is $\mathbf{p}^{(\infty)} = \mathbf{r}^{(\infty)} = \mathbf{s}^{(\infty)}$ only for $\varepsilon = \frac{1}{6}$, in which case they are all equal to $\mathbf{m}$. For $\varepsilon < \frac{1}{6}$, $\mathbf{p}^{(\infty)}$ is not to invariant w.r.t. cyclic shifts of the coordinates, and $\mathbf{p}^{(\infty)}$, $\mathbf{r}^{(\infty)}$ and $\mathbf{s}^{(\infty)}$ are all different.

Proposition 1 and Theorem 2 together show that when ε decreases over the "bifurcation" point $\frac{1}{6}$ (in which case a unique degenerate limit table exists) the IPF passes from infinitely many positive limit tables ($\varepsilon > \frac{1}{6}$) to a "period three limit cycle" of degenerate tables ($\varepsilon < \frac{1}{6}$).

3. PROOF OF THEOREM 2

Define the function

$$\Lambda(\mathbf{p}) = p(0,0,0)^{3\varepsilon-\frac{1}{2}} [p(1,0,0)\,p(0,1,0)\,p(0,0,1)]^{\frac{1}{2}-2\varepsilon} \cdot [p(1,1,0)\,p(0,1,1)\,p(1,0,1)]^{\varepsilon}, \quad \mathbf{p} > \mathbf{0} \tag{10}$$

With a direct computation

$$\begin{aligned}\frac{\Lambda(T_1(\mathbf{p}))}{\Lambda(\mathbf{p})} &= \frac{\varepsilon^{2\varepsilon}\left(\frac{1}{2}-\varepsilon\right)^{1-2\varepsilon}}{p(0,0,+)^{\varepsilon}\,p(1,0,+)^{\frac{1}{2}-\varepsilon}\,p(0,1,+)^{\frac{1}{2}-\varepsilon}\,p(1,1,+)^{\varepsilon}} \\ &= \exp\left(\sum_{i,j} a_\varepsilon(i,j)\log\frac{a_\varepsilon(i,j)}{p(i,j,+)}\right) = \exp\left\{D\left(\mathbf{a}_\varepsilon,\mathbf{p}_{12}\right)\right\} \geq 1,\end{aligned}$$

where $\mathbf{p}_{12}(\cdot,\cdot) = \mathbf{p}(\cdot,\cdot,+)$, since D (the Kullback–Leibler divergence) is always nonnegative. Arguing in the same way for T_2 and T_3, we get that the sequence $\{\Lambda(\mathbf{p}^{(n)})\}$ is non decreasing. Moreover, since

$$\Lambda(T_1(\mathbf{p})) = \Lambda(\mathbf{p}) \Longleftrightarrow D(\mathbf{a}_\varepsilon,\mathbf{p}_{12}) = 0 \Longleftrightarrow \mathbf{a}_\varepsilon = \mathbf{p}_{12},$$

$\Lambda(T_1(\mathbf{p})) = \Lambda(\mathbf{p})$ implies that $T_1(\mathbf{p}) = \mathbf{p}$. Analogously, if $\Lambda(T_3 \circ T_2 \circ T_1(\mathbf{p})) = \Lambda(\mathbf{p})$ it follows that $T_1(\mathbf{p}) = T_2(\mathbf{p}) = T_3(\mathbf{p}) = \mathbf{p}$, i.e. $\mathbf{p}$ is a solution to the marginal problem. Looking at the expression (10) it is easy to realize that for $\varepsilon \in \left(0,\frac{1}{6}\right]$ and $\mathbf{p}^{(0)} > \mathbf{0}$

$$\lim_{n\to+\infty} p^{(n)}(0,0,0) = 0. \tag{11}$$

Suppose by contradiction that there exists a subsequence $p^{(n_l)}(0,0,0) \geq \gamma > 0$: then it must be $p^{(n_l)}(i,j,k) \geq \delta$, $\forall\ (i,j,k) \neq (0,0,0),(1,1,1)$, for some $0 < \delta \leq \gamma$; otherwise we could draw from $\mathbf{p}^{(n_l)}$ a further subsequence (unchanged for ease of notation) such that, for some $(i,j,k) \neq (0,0,0),(1,1,1)$, $p^{(n_l)}(i,j,k)$ converges to 0, which implies that $\Lambda(\mathbf{p}^{(n_l)})$ converges to 0, too. But this is impossible, since $\Lambda(\mathbf{p}^{(0)})$ is positive and $\Lambda(\mathbf{p}^{(n_l)})$ is nondecreasing. Now, since $[\delta,1]^7 \times [0,1]$ is compact, we can draw from $\mathbf{p}^{(n_l)}$ a further subsequence convergent to $\mathbf{p}^* \in [\delta,1]^7 \times [0,1]$. Since over this set the functions Λ and T_i, $i = 1,2,3$ are continuous, by indicating this subsequence again by $\mathbf{p}^{(n_l)}$, we have

$$\Lambda\left(F(\mathbf{p}^*)\right) = \Lambda\left(F\left(\lim_{l\to\infty}\mathbf{p}^{(n_l)}\right)\right)$$

$$= \lim_{l\to\infty}\Lambda\left(F(\mathbf{p}^{(n_l)})\right) = \lim_{l\to\infty}\Lambda\left(\mathbf{p}^{(n_l+3)}\right) = \sup_l \Lambda\left(\mathbf{p}^{(n_l+3)}\right)$$

$$= \lim_{l\to\infty}\Lambda\left(\mathbf{p}^{(n_l)}\right) = \Lambda\left(\lim_{l\to\infty}\mathbf{p}^{(n_l)}\right) = \Lambda\left(\mathbf{p}^*\right).$$

By the discussion made before, this implies that $F(\mathbf{p}^*) = T_1(\mathbf{p}^*) = T_2(\mathbf{p}^*) = T_3(\mathbf{p}^*) = \mathbf{p}^*$, and since $\mathbf{p}^*(0,0,0) > 0$ and $\varepsilon \leq \frac{1}{6}$, the existence of such a $\mathbf{p}^*$ is impossible. By exchanging 0 with 1 we can likewise prove that

$$\lim_{n\to+\infty} p^{(n)}(1,1,1) = 0. \tag{12}$$

Next define the function

$$g(x,t) = \left(\frac{1}{2} - \varepsilon\right)\frac{t}{x+t}, \quad \text{for} \quad 0 \leq x \leq \frac{1}{2} - \varepsilon \quad \text{and} \quad t > 0.$$

By the definition of the IPF algorithm, for $n = 0,1,\ldots$

$$\begin{cases} p^{(6n+1)}(1,0,1) = g\left(p^{(6n)}(1,0,0), p^{(6n)}(1,0,1)\right) \\ p^{(6n+2)}(0,0,1) = g\left(p^{(6n+1)}(1,0,1), p^{(6n+1)}(0,0,1)\right) \\ p^{(6n+3)}(0,1,1) = g\left(p^{(6n+2)}(0,0,1), p^{(6n+2)}(0,1,1)\right) \\ p^{(6n+4)}(0,1,0) = g\left(p^{(6n+3)}(0,1,1), p^{(6n+3)}(0,1,0)\right) \\ p^{(6n+5)}(1,1,0) = g\left(p^{(6n+4)}(0,1,0), p^{(6n+4)}(1,1,0)\right) \\ p^{(6(n+1))}(1,0,0) = g\left(p^{(6n+5)}(1,1,0), p^{(6n+5)}(1,0,0)\right). \end{cases} \tag{13}$$

Furthermore by (11) and (12) the following holds:

$$\begin{aligned} &\lim_{n\to+\infty} p^{(6n)}(1,0,1) = \lim_{n\to+\infty} p^{(6n+1)}(0,0,1) = \lim_{n\to+\infty} p^{(6n+2)}(0,1,1) = \\ &\lim_{n\to+\infty} p^{(6n+3)}(0,1,0) = \lim_{n\to+\infty} p^{(6n+4)}(1,1,0) = \lim_{n\to+\infty} p^{(6n+5)}(1,0,0) = \varepsilon. \end{aligned} \tag{14}$$

Since $p^{(6n)}(1,0,0) \in \left[0, \frac{1}{2} - \varepsilon\right]$, there is a subsequence $p^{(6n_l)}(1,0,0)$ convergent to $a \in \left[0, \frac{1}{2} - \varepsilon\right]$. From (13) and (14) and the continuity of g, it follows

$$\lim_{l \to +\infty} p^{(6n_l+1)}(1,0,1) = g(a,\varepsilon).$$

Setting

$$f(x) = g(x,\varepsilon) = \frac{\varepsilon(1-2\varepsilon)}{2(x+\varepsilon)}$$

and $f^{(i)} = \underbrace{f \circ \ldots \circ f}_{i}$, for $i = 2, \ldots, 6$, we have

$$\begin{cases} \lim_{l \to +\infty} p^{(6n_l+1)}(1,0,1) = f(a) \\ \lim_{l \to +\infty} p^{(6n_l+2)}(0,0,1) = f^{(2)}(a) \\ \lim_{l \to +\infty} p^{(6n_l+3)}(0,1,1) = f^{(3)}(a) \\ \lim_{l \to +\infty} p^{(6n_l+4)}(0,1,0) = f^{(4)}(a) \\ \lim_{l \to +\infty} p^{(6n_l+5)}(1,1,0) = f^{(5)}(a) \\ \lim_{l \to +\infty} p^{(6n_l+6)}(1,0,0) = f^{(6)}(a). \end{cases}$$

By iterating this argument, we see that $f^{(6h)}(a)$ is a limit point of the sequence $\{p^{(2n)}(1,0,0)\}$ for any integer h. Next observe that f is decreasing, $f(0) = \frac{1}{2} - \varepsilon > 0$, $f\left(\frac{1}{2} - \varepsilon\right) = 2\varepsilon\left(\frac{1}{2} - \varepsilon\right) < \frac{1}{2} - \varepsilon$; then, on $\left[0, \frac{1}{2} - \varepsilon\right]$ f has the unique fixed point

$$x^* = \frac{-\varepsilon + \sqrt{2\varepsilon - 3\varepsilon^2}}{2}.$$

Moreover $f^{(2)}(x) = \left(\frac{1}{2} - \varepsilon\right) \frac{x+\varepsilon}{\frac{1}{2}+x}$ is a contraction on $\left[0, \frac{1}{2} - \varepsilon\right]$, since

$$f^{(2)\prime}(x) = \frac{\left(\frac{1}{2} - \varepsilon\right)^2}{\left(\frac{1}{2} + x\right)^2} \le (1-2\varepsilon)^2 < 1$$

from which $f^{(6k)}(a)$ converge to the unique fixed point of $f^{(2)}$, which has to be x^* as well. By consequence x^* is a limit point of $\{p^{(2n)}(1,0,0)\}$.

We then come to a study of the local behavior of f near x^*. By a direct computation

$$|f'(x^*)| = \frac{1-2\varepsilon}{1-\varepsilon+\sqrt{2\varepsilon - 3\varepsilon^2}} < 1,$$

which implies that there exists $\delta > 0$ such that

$$x^* - \delta < f(x^* + \delta) < f(x^* - \delta) < x^* + \delta$$

and by the pointwise convergence of $g(x,t)$ to $f(x)$ as $t \to \varepsilon$ there exists $\gamma > 0$ such that for $|t - \varepsilon| < \gamma$

$$x^* - \delta < g(x^* + \delta, t) < g(x^* - \delta, t) < x^* + \delta.$$

This allows to obtain, from the fact that x^* is a limit point of $\{p^{(6n)}(1,0,0)\}$, the convergence of the whole sequence, since $(x^* - \delta, x^* + \delta)$ is invariant under all $g(\cdot, t)$ for $|t - \varepsilon| < \gamma$. The same is obviously true for $\{p^{(6n+1)}(1,0,1)\}$, $\{p^{(6n+2)}(0,0,1)\}$, $\{p^{(6n+3)}(0,1,1)\}$, $\{p^{(6n+4)}(0,1,0)\}$ and $\{p^{(6n+5)}(1,1,0)\}$, whereas by exchanging 1 with 0 and repeating the argument the same result is obtained for the sequences $\{p^{(6n)}(0,1,1)\}$, $\{p^{(6n+1)}(0,1,0)\}$, $\{p^{(6n+2)}(1,1,0)\}$, $\{p^{(6n+3)}(1,0,0)\}$, $\{p^{(6n+4)}(1,0,1)\}$ and $\{p^{(6n+5)}(0,0,1)\}$. It is then immediately checked that this implies (7) and (8).

ACKNOWLEDGEMENT

We thank an anonymous referee for having suggested to us the current proof of Theorem 2, which is considerably simplified compared to our original one.

(Received July 22, 2002.)

REFERENCES

[1] I. Csiszár: I-divergence geometry of probability distributions and minimization problems. Ann. Probab. *3* (1975), 146–158.

[2] W. E. Deming and F. F. Stephan: On a least square adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Statist. *11* (1940), 427–444.

[3] S. J. Haberman: The analysis of frequency data. The University of Chicago Press, Chicago 1974.

[4] S. T. Jensen, S. Johansen, and S. L. Lauritzen: Globally convergent algorithms for maximizing a likelihood function. Biometrika *78* (1991), 867–877.

[5] R. Jiroušek: Solution of the marginal problem and decomposable distributions. Kybernetika *27* (1991), 403–412.

[6] S. L. Lauritzen: Graphical Models. Clarendon Press, Oxford 1996.

Dr. Claudio Asci, Università degli Studi di Roma La Sapienza, Dipartimento di Studi Geoeconomici, Linguistici, Statistici e Storici per l'analisi regionale, Via del Castro Laurenziano, 9-00161 Roma. Italy.
e-mail: claudio.asci@uniroma1.it

Prof. Dr. Mauro Piccioni, Università degli Studi di Roma La Sapienza, Dipartimento di Matematica, Piazzale Aldo Moro, 2-00185, Roma. Italy.
e-mail: piccioni@mat.uniroma1.it