Claudio Asci; Mauro Piccioni
Asymptotic behaviour of a BIPF algorithm with an improper target

Kybernetika, Vol. 45 (2009), No. 2, 169--188

Persistent URL: http://dml.cz/dmlcz/140071

KYBERNETIKA — VOLUME 45 (2009), NUMBER 2, PAGES 169 – 188

ASYMPTOTIC BEHAVIOUR OF A BIPF ALGORITHM WITH AN IMPROPER TARGET

Claudio Asci and Mauro Piccioni

The BIPF algorithm is a Markovian algorithm with the purpose of simulating certain probability distributions supported by contingency tables belonging to hierarchical log-linear models. The updating steps of the algorithm depend only on the required expected marginal tables over the maximal terms of the hierarchical model. Usually these tables are marginals of a positive joint table, in which case it is well known that the algorithm is a blocking Gibbs Sampler. But the algorithm makes sense even when these marginals do not come from a joint table. In this case the target distribution of the algorithm is necessarily improper. In this paper we investigate the simplest non trivial case, i.e. the $2 \times 2 \times 2$ hierarchical interaction. Our result is that the algorithm is asymptotically attracted by a limit cycle in law.

Keywords: log-linear models, marginal problem, null Markov chains

AMS Subject Classification: 60J05, 65C40, 62F15

1. INTRODUCTION

In the book by Gelman et al. [9] the Bayesian Iterated Proportional Fitting (BIPF) was introduced as a Markovian sampling algorithm for certain probability distributions over contingency tables belonging to log-linear models. In the simplest instance, presented in detail in Schafer [16], the BIPF algorithm is a process $\{\boldsymbol{M}^{(n)}, n \in \mathbb{N}\}$ with values in the set $T_3 = (\mathbb{R}^+)^8$ of positive $2\times2\times2$ contingency tables, defined recursively with random inputs taking values in the set $T_2 = (\mathbb{R}^+)^4$ of positive 2×2 contingency tables. In order to define it, first introduce the transformation

$$T : \boldsymbol{T}_3 \times \boldsymbol{T}_2 \to \boldsymbol{T}_3,\ T(\boldsymbol{M}, \boldsymbol{V})(i, j, k) = V(i, j)\frac{M(i, j, k)}{M(i, j, +)}, \quad (i, j, k) \in \{0, 1\}^3, \quad (1)$$

where the symbol $+$ means summation over the corresponding index. Thus $T(M, V)$ is a $2 \times 2 \times 2$ table with the marginal over the first two variables equal to V and the conditionals on the remaining variable given the first two equal to those in M. In order to apply this transformation to different pair marginals the cyclic shift σ is defined as

$$\sigma(i, j, k) = (k, i, j),\ (i, j, k) \in \{0, 1\}^3 .$$

A single step of the BIPF algorithm is split into three "fractional" updates defined as

$$\boldsymbol{M}^{\left(n+\frac{l+1}{3}\right)} = T\left(\boldsymbol{M}^{\left(n+\frac{l}{3}\right)} \circ \sigma^{l}, \boldsymbol{V}^{\left(n+\frac{l+1}{3}\right)}\right) \circ \sigma^{-l}, \quad l = 0, 1, 2, \ n \in \mathbb{N}, \tag{2}$$

where, for $l = 0, 1, 2$, $\left\{\boldsymbol{V}^{\left(n+\frac{l+1}{3}\right)}, n \in \mathbb{N}\right\}$ are independent i.i.d. sequences in T_2 with independent entries

$$V^{\left(n+\frac{l+1}{3}\right)}(i,j) \sim \text{Gamma}(n_l(i,j), 1), \quad i, j = 0, 1. \tag{3}$$

Each "fractional" update sets a pair marginal: first the entries of the pair marginal $(1, 2)$ are drawn from Gamma laws with means specified by the table n_0, then the entries of the pair $(2, 3)$ are drawn from Gamma laws with means specified by the table n_1 and finally the entries of the pair $(3, 1)$ are drawn from Gamma laws with means specified by the table n_2. The tables n_0, n_1, n_2 are assumed to be *pairwise consistent, that is*

$$n_l(+, h) = n_{l+1}(h, +), \ h \in \{0, 1\}, \quad l = 0, 1, 2 \pmod 3.$$

It is immediately checked that the cross product ratio : $T_3 \to \mathbb{R}^+$, defined by

$$R(\boldsymbol{M}) = \frac{M(1,1,1)\, M(1,0,0)\, M(0,1,0)\, M(0,0,1)}{M(0,0,0)\, M(1,1,0)\, M(0,1,1)\, M(1,0,1)} \tag{4}$$

is invariant under all permutations of the arguments of the table and invariant under the action of T, i. e.

$$R(T(\boldsymbol{M}, \boldsymbol{V})) = R(\boldsymbol{M}), \ \forall \boldsymbol{V} \in \boldsymbol{T}_2.$$

As a consequence provided $M^{(0)}$ belongs to the hierarchical log-affine model (see e. g. Lauritzen [11])

$$\boldsymbol{T}_3^r \equiv \{\boldsymbol{M} \in \boldsymbol{T}_3 : R(\boldsymbol{M}) = r\},$$

then $M^{(n+\frac{l}{3})} \in T_3^r$ for all integer n, and $l = 0, 1, 2$.

In the "standard" case considered by Schafer the tables n_0, n_1, $n_2 \in T_2$ are obtained as two-dimensional marginals of a $2 \times 2 \times 2$ table $n \in T_3$

$$n(i,j,+) = n_0(i,j), n(+,j,k) = n_1(j,k), n(i,+,k) = n_2(k,i), \quad (i,j,k) \in \{0,1\}^3. \tag{5}$$

Then, for any $r \in \mathbb{R}^+$, $M^{(n)}$ has a unique probability distribution π_r supported by T_3^r which is stationary for each fractional update, i. e.

$$\boldsymbol{M}^{\left(n+\frac{l}{3}\right)} \sim \pi_r \Longrightarrow \boldsymbol{M}^{\left(n+\frac{l+1}{3}\right)} \sim \pi_r, \quad l = 0, 1, 2, \quad n \in \mathbb{N},$$

towards which $M^{(n+\frac{l}{3})}$ converge in law as $n \to \infty$, for $l = 0, 1, 2$. The reason is that a suitable change of variable reveals (2) to be an irreducible blocking Gibbs Sampler for the target π_r (see Schafer [16] and in greater generality Piccioni [14]). In applications to Bayesian statistics this density is a prior or a posterior density supported by the tables belonging to a hierarchical log-affine model. This density is "locally" specified, i. e. uniquely constructed from requirements on the distribution of marginal tables over the maximal sets of the hierarchical model (in our case all the pairs). For example, one can imagine that these distributions are specified by differents "experts" assigned to different parts of the model (see Asci and Piccioni [4]).

In the "non-standard" case n_0, n_1, n_2 are pairwise consistent but (5) is not fulfilled by any table $n \in T_3$. In these situations we have proved in Asci and Piccioni [4] that (for general hierarchical log-linear models) the BIPF algorithm can be still interpreted as an irreducible blocking Gibbs Sampler for an implicitly defined improper target distribution. As a consequence $\{\boldsymbol{M}^{(n)}, n \in \mathbb{N}\}$ is not positive recurrent. In this paper we focus on the case when the three tables n_l coincide, in which case necessarily

$$\boldsymbol{n}_0 = \boldsymbol{n}_1 = \boldsymbol{n}_2 = \begin{pmatrix} a_1 & b \\ b & a_2 \end{pmatrix}, \tag{6}$$

for some $a_1, a_2, b > 0$. In this case it is easy to check that they are the marginal tables of a joint $n \in T_3$ if and only if $a_1 + a_2 - b > 0$. For $a_1 + a_2 - b < 0$ and $a_1 = a_2$, Asci and Piccioni [3] analyzed the corresponding deterministic algorithm in which the random variables in (3) are replaced by their expected values (the celebrated IPF algorithm, see Csiszár [5]), showing the existence of a limit cycle over the period of the three fractional updates. The main contribution of the present paper is an analogous result for the BIPF algorithm.

The impossibility of finding a non-negative marginal three-way table with pair marginals all equal to (6), with $a_1 = a_2 = \frac{1}{8}$, $b = \frac{3}{8}$, has received a considerable interest in quantum mechanics, since it denies an explanation in terms of classical probability of the correlations predicted by quantum theory in the famous example of two-particle system proposed by Einstein, Podolsky and Rosen [8]. In this paradox three spin observables are considered on the two particles, for a total of six variables; but since each of these observables has opposite values on the two particles, it is possible to reduce to three as in our case. We refer the interested reader to the monograph by Albert [1].

The paper is essentially devoted to the case $a_1 + a_2 - b \le 0$. In Section 2, through the analysis of the invariant density, we prove that some of the entries of $M^{(n+\frac{l}{3})}$ go to zero as $n \to \infty$. It suggests the analysis of the reduced dynamics obtained by setting these elements to zero from the beginning. Section 3 is devoted to the study of the reduced dynamics: we determine explicitly the limit laws μ_l of $M^{(n+\frac{l}{3})}$, as $n \to \infty$, for $l = 0, 1, 2$, which are related by a cyclic shift

$$\mu_l = \mu_0 \sigma^l, \qquad l = 1, 2,$$

where, with a slight abuse of notation, we define $\sigma^l(M) = M \circ \sigma^l$. The law μ_0 is invariant under σ if and only if $a_1 + a_2 - b = 0$, in which case $M^{(n+\frac{l}{3})}$ converges in law to μ_0 as $n \to \infty$, for any $l = 0, 1, 2$, whereas for $a_1 + a_2 - b < 0$ the process $\left\{\boldsymbol{M}^{\left(n+\frac{l}{3}\right)},\ l = 0, 1, 2,\ n \in \mathbb{N}\right\}$ exhibits a period three limit cycle in law. In Section 4 we show that this asymptotic behaviour is preserved by starting from any initial table $M^{(0)} \in T_3$, independently of the value of $R(M^{(0)})$. Some of the results needed for the proofs are collected in the Appendix. Similar results for the deterministic case are sketched in Section 5. A discussion of some practical implications of these results for actual simulations is outlined in Section 6.

2. REDUCTION OF DIMENSIONALITY

We begin with the analysis of the recursion (2)–(3) with the choice of parameters (6). This suggests to define a new sequence of matrices

$$\widetilde{\boldsymbol{M}}^{(3n+l)} = \boldsymbol{M}^{\left(n+\frac{l}{3}\right)} \circ \sigma^l \tag{7}$$

which is a time-homogeneous Markov chain evolving for $n \in \mathbb{N}$, $l = 0, 1, 2$ according to

$$\widetilde{\boldsymbol{M}}^{(3n+l+1)} = T\left(\widetilde{\boldsymbol{M}}^{(3n+l)}, \boldsymbol{V}^{\left(n+\frac{l+1}{3}\right)}\right) \circ \sigma. \tag{8}$$

By the definitions of T and σ the $(3, 1)$ pair marginal of $\widetilde{\boldsymbol{M}}^{(3n+l+1)}$ is always equal to the random input $V^{\left(n+\frac{l+1}{3}\right)}$, whereas the family of "conditionals" of the label 2 given the pair $(3, 1)$ is equal to the family of "conditionals" of the label 3 given the pair $(1, 2)$ of $\widetilde{\boldsymbol{M}}^{(3n+l)}$. Therefore these conditionals remain a "reduced" Markov chain, from which $\widetilde{\boldsymbol{M}}$ and M can be recovered. They can be parametrized for example by the vector $X^{(3n+l)} = \psi\left(\widetilde{\boldsymbol{M}}^{(3n+l)}\right) \in (0, 1)^4$, with components

$$X_{i,j}^{(3n+l)} = \psi_{i,j}\left(\widetilde{\boldsymbol{M}}^{(3n+l)}\right) = \frac{\widetilde{M}^{(3n+l)}(i, j, i)}{\widetilde{M}^{(3n+l)}(i, j, +)}, \qquad i, j \in \{0, 1\}. \tag{9}$$

Using for short the following notation

$$X_1^{(3n+l)} = X_{0,0}^{(3n+l)},\ X_2^{(3n+l)} = X_{1,0}^{(3n+l)},\ X_3^{(3n+l)} = X_{0,1}^{(3n+l)},\ X_4^{(3n+l)} = X_{1,1}^{(3n+l)},$$

for $n \in \mathbb{N}$ and $l = 0, 1, 2$, the following result is easily established.

Proposition 2.1. The process $\left\{\boldsymbol{X}^{(m)}\right\}$ is a Markov chain satisfying the following equations

$$\boldsymbol{X}^{(m+1)} = F\left(\boldsymbol{X}^{(m)}, H_{m+1}, K_{m+1}\right), \quad m \in \mathbb{N} \tag{10}$$

where $F : (0,1)^4 \times (0,+\infty)^2 \to (0,1)^4$ is defined by

$$\begin{cases} F_1(\boldsymbol{x},h,k) = \frac{hx_1}{hx_1+1-x_2} \\ F_2(\boldsymbol{x},h,k) = \frac{k(1-x_4)}{k(1-x_4)+x_3} \\ F_3(\boldsymbol{x},h,k) = \frac{h(1-x_1)}{h(1-x_1)+x_2} \\ F_4(\boldsymbol{x},h,k) = \frac{kx_4}{kx_4+1-x_3} \end{cases} \tag{11}$$

and

$$H_{3n+l} = \frac{V^{\left(n+\frac{l}{3}\right)}(0,0)}{V^{\left(n+\frac{l}{3}\right)}(1,0)}, \quad K_{3n+l} = \frac{V^{\left(n+\frac{l}{3}\right)}(1,1)}{V^{\left(n+\frac{l}{3}\right)}(0,1)}, \qquad n \in \mathbb{N},\ l = 0,1,2.$$

Moreover, for any $l = 0,1,2$ and $n \in \mathbb{N}$

$$\widetilde{M}^{(3n+l+1)}(i,j,k) = V^{\left(n+\frac{l+1}{3}\right)}(i,j)\left(\delta_{ik}X_{i,j}^{(3n+l)} + (1-\delta_{ik})\left(1 - X_{i,j}^{(3n+l)}\right)\right). \tag{12}$$

Since the ratio of two independent Gamma variables with the same scale parameter and shape parameters a and b, respectively, has the beta density of second kind

$$\beta_{b,a}^{(2)}(\mathrm{d}h) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{h^{b-1}}{(1+h)^{a+b}} \mathbf{1}_{(0,+\infty)}(h), \tag{13}$$

we have that $H_m \sim \beta_{a_1,b}^{(2)}$ and $K_m \sim \beta_{a_2,b}^{(2)}$ for any $m \in \mathbb{N}$.

As already noticed, if $M^{(0)} \in T_3^r$ for some $r > 0$, then $M^{\left(n+\frac{l}{3}\right)}$ (and likewise $\widetilde{\boldsymbol{M}}^{(3n+l)}$) remains in T_3^r for any $l = 0,1,2$ and $n \in \mathbb{N}$. Define the three-dimensional surface of the space $(0,1)^4$

$$\boldsymbol{\Xi}^r \equiv \boldsymbol{\psi}(\boldsymbol{T}_3^r) = \left\{ \boldsymbol{x} \in (0,1)^4 : \phi(\boldsymbol{x} \equiv \frac{(1-x_1)(1-x_2)x_3x_4}{x_1x_2(1-x_3)(1-x_4)} = r \right\}.$$

The process $\left\{\boldsymbol{X}^{(m)}\right\}$, with $X^{(0)} \in \Xi^r$, is a Markov chain evolving in Ξ^r. Obviously Ξ^r can be parametrized by $(x_1,x_2,x_3) \in (0,1)^3$ if we take

$$x_4 = b(x_1,x_2,x_3;r) = \frac{rx_1x_2(1-x_3)}{(1-x_1)(1-x_2)x_3 + rx_1x_2(1-x_3)}. \tag{14}$$

Proposition 2.2. For any $r \in \mathbb{R}^+$, the measure

$$h_r(x_1,x_2,x_3)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\delta_{b(x_1,x_2,x_3;r)}(\mathrm{d}x_4), \tag{15}$$

where

$$h_r(x_1,x_2,x_3) = \frac{x_1^{a_1+a_2-b-1}\left[(1-x_1)(1-x_2)x_3\right]^{b-1}\left[x_2(1-x_3)\right]^{a_2-1}}{\left[(1-x_1)\,(1-x_2)\,x_3 + r x_1 x_2\,(1-x_3)\right]^{a_2}}, \tag{16}$$

is invariant for $\left\{\boldsymbol{X}^{(m)}\right\}$. This measure is finite if and only if $a_1+a_2-b>0$.

P r o o f. First notice that for a_1, a_2 and $b>0$ there always exists a $2\times 2\times 2$ table $n\in\mathbb{R}^8$ such that $n\circ\sigma = n$, with all the pair marginals equal to (6). The process $\widetilde{\boldsymbol{M}}$ has the invariant measure

$$g(\boldsymbol{m})\mathrm{d}\boldsymbol{m} = \prod_{i=0}^{1}\prod_{j=0}^{1}\prod_{k=0}^{1} m(i,j,k)^{n(i,j,k)-1}\exp(-m(i,j,k))\,\mathrm{d}\boldsymbol{m},\, \boldsymbol{m}\in\boldsymbol{T}_3. \tag{17}$$

Since n is invariant under σ, the same is true for g. Thus, it suffices to show that if $\widetilde{\boldsymbol{M}}$ has the law (17) and V has independent entries distributed as in (3) then $T(\widetilde{\boldsymbol{M}},V)$ has the same law as $\widetilde{\boldsymbol{M}}$. When $n\in T_3$ this results from the well known beta-gamma algebra (see Williams [18], page 250):

$$\widetilde{M}(i,j,0)\sim \mathrm{Gamma}(n(i,j,0),1)\perp \widetilde{M}(i,j,1)\sim\mathrm{Gamma}(n(i,j,1),1)$$

implies

$$\frac{\widetilde{M}(i,j,1)}{\widetilde{M}(i,j,+)}\sim \mathrm{Beta}(n(i,j,1),n(i,j,0))\perp\widetilde{M}(i,j,+)\sim\mathrm{Gamma}(n(i,j,+),1).$$

Therefore, if $\widetilde{M}(i,j,+)$ is replaced by $V(i,j)$ with the same distribution, then the joint distribution of $\left(\widetilde{M}(i,j,0),\widetilde{M}(i,j,1)\right)$ is retained. Being based on a change of variable formula, this argument works for any choice of $n\in\mathbb{R}^8$ with pair marginals equal to (6), hence for any choice of a_1, a_2 and $b>0$. The same arguments prove that the measure induced by the function ψ has the following form

$$\begin{aligned} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= x_1^{n(0,0,0)-1}(1-x_1)^{n(0,0,1)-1}x_2^{n(1,0,1)-1}(1-x_2)^{n(1,0,0)-1}\\ &\quad\cdot x_3^{n(0,1,0)-1}(1-x_3)^{n(0,1,1)-1}x_4^{n(1,1,1)-1}(1-x_4)^{n(1,1,0)-1}\mathrm{d}\boldsymbol{x}, \end{aligned}$$

and is invariant for the process $\left\{\boldsymbol{X}^{(m)}\right\}$.

Under the mapping $x\longmapsto(x_1,x_2,x_3,\phi(x))$ this measure is transformed into the measure on $(0,1)^3\times\mathbb{R}^+$ given by

$$\frac{f}{\frac{\partial\phi}{\partial x_4}}(x_1,x_2,x_3,b\,(x_1,x_2,x_3;r)\,)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}r = r^{n(1,1,1)-1}h_r(x_1,x_2,x_3)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}r,$$

and since $\phi(X^{(m)})=\phi(X^{(0)})$, for any $m\in\mathbb{N}$, this means that the measure $r^{n(1,1,1)-1}\mathrm{d}r$ can be replaced by any Dirac mass centered in any $r>0$, which is the first assertion of the theorem.

For $a_1 + a_2 - b > 0$ we have

$$\begin{aligned}&[(1-x_1)(1-x_2)x_3 + rx_1x_2(1-x_3)]^{a_2}\\ &\geq [(1-x_1)(1-x_2)x_3]^{\frac{a_2 b}{a_1+a_2}} [rx_1x_2(1-x_3)]^{\frac{a_2}{a_1+a_2}(a_1+a_2-b)},\end{aligned}$$

then

$$\begin{aligned}h_r(x_1,x_2,x_3) \leq{}& r^{-\frac{a_2}{a_1+a_2}(a_1+a_2-b)} x_1^{\frac{a_1}{a_1+a_2}(a_1+a_2-b)-1}\\ &\cdot [(1-x_1)(1-x_2)x_3]^{\frac{a_1 b}{a_1+a_2}-1} [x_2(1-x_3)]^{\frac{a_2 b}{a_1+a_2}-1}, \qquad (18)\end{aligned}$$

and it is immediately checked that the r.h.s. of (18) is integrable over $(0,1)^3$. Conversely, if $a_1 + a_2 - b \leq 0$, we have

$$[(1-x_1)(1-x_2)x_3 + rx_1x_2(1-x_3)]^{a_2} \leq (1+r)^{a_2}$$

and therefore

$$h_r(x_1,x_2,x_3) \geq (1+r)^{-a_2} x_1{}^{a_1+a_2-b-1} [(1-x_1)(1-x_2)x_3]^{b-1} [x_2(1-x_3)]^{a_2-1}. \qquad (19)$$

Since $a_1 + a_2 - b \leq 0$, the density h_r is not integrable over the whole cube $(0,1)^3$.□

Remark 2.3. The first three components of $X^{(m)}$ evolve as a Markov chain in $(0,1)^3$ described by the recursion (10) with the fourth component obtained from (14) with $r = \phi(X^{(0)})$. It is easy to check that the two-step transition density of this chain w.r.t. the Lebesgue measure is positive in the whole of $(0,1)^3$. This implies that, for $a_1+a_2-b > 0$, (15) is the unique invariant measure for $X^{(m)}$ supported by Ξ^r (up to a proportionality factor), this process is positive recurrent in Ξ^r and, for $X^{(0)} \in \Xi^r$, it converges in law to the probability measure obtained by normalizing (15). Plugging a random vector $X^{(3n+l)}$ with this distribution in (12) independent of $V^{\left(n+\frac{l+1}{3}\right)}$ we get a random matrix distributed as the limit law of $\widetilde{\boldsymbol{M}}^{(m)}$ when $M^{(0)} \in T_3^r$, say τ_r. The proof of Proposition 2.2 shows that τ_r is just the law (17) conditional to $R(\boldsymbol{m}) = r$: since (17) and R are invariant under σ, by (7) we can conclude that, when $a_1 + a_2 - b > 0$, $M^{\left(n+\frac{l}{3}\right)}$ converges in law to τ_r as $n \to \infty$, for $l = 0, 1, 2$, in agreement with the result of Example 2 in Asci and Piccioni [4].

For $a_1 + a_2 - b \leq 0$ the knowledge of an invariant measure still yields important information about the asymptotic behaviour of $\left\{\boldsymbol{X}^{(m)}\right\}$.

Corollary 2.4. If $a_1 + a_2 - b \leq 0$ then $X_1^{(m)} \xrightarrow{P} 0$ and $X_4^{(m)} \xrightarrow{P} 0$ for any $X^{(0)} \in (0,1)^4$.

P r o o f. It is clear from (16) that the function h_r is integrable over

$$A_1(d) = \left\{\boldsymbol{x} \in (0,1)^4 : x_1 > d\right\}.$$

It suffices to apply Thereom 6 and Proposition 10.3 by Tweedie [17] which state the probability that $X^{(m)}$ belongs to a set where a non finite invariant measure is finite vanishes as $m \to \infty$. These theorems work a.e. in $X^{(0)}$ (w.r.t. the invariant measure), but the positivity of the two-step transition density guarantees that the same result holds everywhere in Ξ^r. Since $r > 0$ is arbitrary the results hold for any $X^{(0)} \in (0,1)^4$. By exchanging $X_1^{(m)}$ with $X_4^{(m)}$ and $X_2^{(m)}$ with $X_3^{(m)}$ the same result can be obtained for

$$A_4(d) = \left\{\boldsymbol{x} \in (0,1)^4 : x_4 > d\right\}. \qquad \square$$

Finally we need to establish the tightness of the process $\left\{\left(X_2^{(m)}, X_3^{(m)}\right)\right\}$ for any fixed initial state $X^{(0)} \in (0,1)^4$. This holds trivially for $a_1 + a_2 - b > 0$ (since $X^{(m)}$ converges in law), but it is not difficult to prove for the case $a_1 + a_2 - b \le 0$ as well.

Proposition 2.5. Suppose $X^{(0)} \in (0,1)^4$, $a_1 + a_2 - b \le 0$. Then, for any $\delta > 0$, there exist $d > 0$ and $m_0 \in \mathbb{N}$ such that

$$P\left(d \le X_i^{(m)} \le 1 - d,\ i = 2,3\right) > 1 - \delta \quad \text{for } m \ge m_0.$$

Proof. For any $0 < d < 1/2$

$$\begin{aligned} \mathrm{P}\left(X_2^{(m)} < d\right) &\le \mathrm{P}\left(K_m\left(1 - X_4^{(m-1)}\right) < \frac{d}{1-d}\right) \\ &\le \mathrm{P}\left(\left\{K_m < \sqrt{\frac{d}{1-d}}\right\} \cup \left\{1 - X_4^{(m-1)} < \sqrt{\frac{d}{1-d}}\right\}\right) \\ &\le \mathrm{P}\left(K_m < \sqrt{\frac{d}{1-d}}\right) + \mathrm{P}\left(X_4^{(m-1)} > 1 - \sqrt{\frac{d}{1-d}}\right). \end{aligned}$$

Since K_m has a continuous distribution function and $X_4^{(m)} \xrightarrow{P} 0$ by Corollary 2.4, we have that for d sufficiently small and m sufficiently large

$$\mathrm{P}\left(X_2^{(m)} \ge d\right) > 1 - \delta.$$

With obvious changes we obtain the same result for the sequences $X_3^{(m)}$, $1 - X_2^{(m)}$ and $1 - X_3^{(m)}$. $\square$

3. THE REDUCED DYNAMICS

From (11) it is clear that, for any pair $(h,k) \in (0,+\infty)^2$, the function $F(\cdot,h,k)$ can be extended to a continuous function defined on the whole set $\Gamma = [0,1) \times (0,1)^2 \times [0,1)$, which we continue to denote by F. The set Γ is invariant, meaning that $F(\Gamma,h,k) \subset \Gamma$ for any $(h,k) \in (0,+\infty)^2$. This extension allows us to start the chain

$\{\boldsymbol{X}^{(m)}\}$ from any initial state vector $X^{(0)} = x_0 \in \Gamma$. In particular Corollary 2.4 suggests to study the behaviour of (10) for $x_0 \in C$, where

$$C = \left\{\boldsymbol{x} \in [0,1]^4 : x_1 = x_4 = 0,\, 0 < x_2 < 1,\, 0 < x_3 < 1\right\}.$$

Even if the analysis is motivated by the case $a_1 + a_2 - b \leq 0$, the results of this section hold for any choice of $a_1, a_2, b > 0$.

Notice that C is invariant under $F(\cdot; h, k)$ for any $h, k > 0$, hence $X_1^{(m)} = X_4^{(m)} = 0$ for $m \in N$, whereas for the remaining components we have

$$\begin{cases} X_2^{(m+1)} = \dfrac{K_{m+1}}{K_{m+1} + X_3^{(m)}} \\[2ex] X_3^{(m+1)} = \dfrac{H_{m+1}}{H_{m+1} + X_2^{(m)}}, \end{cases} \tag{20}$$

where $H_{m+1} \sim \beta^{(2)}_{a_1,b}$ and $K_{m+1} \sim \beta^{(2)}_{a_2,b}$.

The asymptotic behaviour of this process is easier to analyze.

Proposition 3.1. If $a_1, a_2, b > 0$, there exists a law υ on $(0,1)^2$ and constants $R > 0$, $r \in (0,1)$ such that

$$\left|\mathrm{P}_{\boldsymbol{x}}\left(\left(X_2^{(m)}, X_3^{(m)}\right) \in A\right) - \upsilon(A)\right| \leq Rr^m, m \in \mathbb{N},$$

for any $x \in C$ and for any Borel subset A of $(0,1)^2$.

Proof. The statement follows from the uniform ergodicity of (20) on $(0,1)^2$. In order to prove it, consider the joint density of

$$\begin{cases} X_2^{(2)} = \dfrac{K_2(H_1 + X_2^{(0)})}{K_2(H_1 + X_2^{(0)}) + H_1} \\[2ex] X_3^{(2)} = \dfrac{H_2(K_1 + X_3^{(0)})}{H_2(K_1 + X_3^{(0)}) + K_1} \end{cases} \tag{21}$$

together with the auxiliary variables H_1, K_1, for fixed $X_2^{(0)} = x_2$, $X_3^{(0)} = x_3$, which is easily seen to be

$$f_{(X_2^{(2)}, X_3^{(2)}, H_1, K_1)}(z_2, z_3, s, t; x_2, x_3)$$

$$= \frac{z_2^{a_2-1}(1-z_2)^{b-1} z_3^{a_1-1}(1-z_3)^{b-1} s^{a_1+a_2-1}\left(s + x_2\right)^b}{\left(B(a_1,b)B(a_2,b)\right)^2 \left(s + x_2\left(1 - z_2\right)\right)^{a_2+b}\left(1+s\right)^{a_1+b}}$$

$$\cdot \frac{t^{a_1+a_2-1}\left(t + x_3\right)^b}{\left(t + x_3\left(1 - z_3\right)\right)^{a_1+b}\left(1+t\right)^{a_2+b}}$$

$$\geq \frac{z_2^{a_2-1}(1-z_2)^{b-1}z_3^{a_1-1}(1-z_3)^{b-1}s^{a_1+a_2-1}t^{a_1+a_2-1}}{\left(B(a_1,b)B(a_2,b)\right)^2(1+s)^{a_1+a_2+b}(1+t)^{a_1+a_2+b}}.$$

Integrating w.r.t. s and t and recalling (13) we get that, for any Borel subset A of $(0,1)^2$,

$$\mathrm{P}^2((x_2,x_3),A) \geq \frac{\left(B(a_1+a_2,b)\right)^2}{B(a_1,b)B(a_2,b)}\left(\mathrm{Beta}(a_1,b)\otimes\mathrm{Beta}(a_2,b)\right)(A).$$

From this minorization condition the uniform ergodicity of (20) on $(0,1)^2$ follows (see Meyn and Tweedie [13], Theorem 16.2.4, page 392). □

Since the two equations in (21) are decoupled, it is clear that υ has the product form $\upsilon = \upsilon_1 \otimes \upsilon_2$. Quite surprisingly it is possible to determine υ_1 and υ_2 explicitly. For any a, b, $c > 0$, recall that the hypergeometric function is defined for $x \in (0,1)$ as

$$_2F_1(a,b;c;x) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{n!(c)_n}x^n, \tag{22}$$

where $(a)_0 = 1$ and $(a)_{n+1} = (a+n)(a)_n$. For a_1, a_2 and $b > 0$ it is then possible to define the probability measure (Asci, Letac, and Piccioni [2], Proposition 2.1)

$$\mu_{a_1,a_2,b}(dx) = C(a_1,a_2,b)x^{a_1-1}(1-x)^{b-1}\ {}_2F_1(a_1,b;a_1+a_2;x)\mathbf{1}_{(0,1)}(x)\,dx, \tag{23}$$

where $C(a_1,a_2,b)$ is a suitable constant.

Proposition 3.2. For any $a_1,a_2,b>0$, the law υ has the form $\mu_{a_2,a_1,b}\otimes\mu_{a_1,a_2,b}$.

P r o o f. We need to prove that $\mu_{a_2,a_1,b}\otimes\mu_{a_1,a_2,b}$ is invariant for the Markov chain $\left\{\left(X_2^{(m)},X_3^{(m)}\right)\right\}$. In fact the probability measure $\mu_{a_1,a_2,b}$ has the following property (see Asci, Letac and Piccioni [2]), Theorem 2.2): if $X\sim\mu_{a_1,a_2,b}$ and $W\sim\beta^{(2)}_{b,a_2}$ are independent then

$$\frac{1}{1+XW}\sim\mu_{a_2,a_1,b}. \tag{24}$$

Since $H_{m+1}^{-1}\sim\beta^{(2)}_{b,a_1}$ and $W_{m+1}^{-1}\sim\beta^{(2)}_{b,a_2}$, the statement follows by using (20), (24) and the analogous result obtained by exchanging a_1 with a_2. □

Theorem 3.3. Suppose that $M^{(0)} = m^{(0)}$ is such that

$$m^{(0)}(0,0,0) = m^{(0)}(1,1,1) = 0, m^{(0)}(i,j,k) > 0, (i,j,k) \neq (0,0,0),(1,1,1). \tag{25}$$

Then, for any $a_1,a_2,b>0$, the sequences $\left\{\boldsymbol{M}^{(n)}\right\}$, $\left\{\boldsymbol{M}^{\left(n+\frac{1}{3}\right)}\right\}$, $\left\{\boldsymbol{M}^{\left(n+\frac{2}{3}\right)}\right\}$ converge in law to $M^{(\infty)}$, $R^{(\infty)}$, $S^{(\infty)}$, respectively. These laws can be constructed as functions of mutually independent random variables G_1, G_2, G_3, G_4, $X_2^{(\infty)}$, $X_3^{(\infty)}$, where

$G_1 \stackrel{\mathcal{L}}{=} \text{Gamma}(a_1, 1)$, $G_2, G_3 \stackrel{\mathcal{L}}{=} \text{Gamma}(b, 1)$, $G_4 \stackrel{\mathcal{L}}{=} \text{Gamma}(a_2, 1)$, $X_2^{(\infty)} \sim \mu_{a_2,a_1,b}$ and $X_3^{(\infty)} \sim \mu_{a_1,a_2,b}$, as

$$\begin{pmatrix} M^{(\infty)}(0,0,0) & M^{(\infty)}(0,1,0) \\ M^{(\infty)}(0,0,1) & M^{(\infty)}(0,1,1) \\ M^{(\infty)}(1,0,0) & M^{(\infty)}(1,1,0) \\ M^{(\infty)}(1,0,1) & M^{(\infty)}(1,1,1) \end{pmatrix} \stackrel{\mathcal{L}}{=} \begin{pmatrix} 0 & G_1 \\ G_2\left(1 - X_2^{(\infty)}\right) & G_2 X_2^{(\infty)} \\ G_3 X_3^{(\infty)} & G_3\left(1 - X_3^{(\infty)}\right) \\ G_4 & 0 \end{pmatrix} \tag{26}$$

and

$$\boldsymbol{R}^{(\infty)} \stackrel{\mathcal{L}}{=} \boldsymbol{M}^{(\infty)} \circ \sigma^2, \ \boldsymbol{S}^{(\infty)} \stackrel{\mathcal{L}}{=} \boldsymbol{M}^{(\infty)} \circ \sigma. \tag{27}$$

If $b = a_1 + a_2$ these limit laws are equal, hence all the sequences $\left\{\boldsymbol{M}^{\left(n+\frac{l}{3}\right)}, n \in \mathbb{N}\right\}$, for $l = 0, 1, 2, \ldots$ converge to the same limit law. Otherwise, the limit laws are all different.

P r o o f. The assumption (25) corresponds to $X^{(0)} \in C$. All the results, except the last one, follow from Proposition 3.1, Proposition 3.2 and (12).

If $b = a_1 + a_2$, recalling that

$$_2F_1(a_1, a_1 + a_2; a_1 + a_2; x) = (1 - x)^{-a_1}$$

we get

$$\mu_{a_1,a_2,a_1+a_2} = \text{Beta}(a_1, a_2). \tag{28}$$

Then, by using again the beta-gamma algebra, it is easy to check that all the random matrices $M^{(\infty)}$, $R^{(\infty)}$, $S^{(\infty)}$ have the same distribution. The entries are independent and distributed as

$$\begin{cases} M^{(\infty)}(0,0,0) \sim M^{(\infty)}(1,1,1) \sim \delta_{\{0\}} \\ M^{(\infty)}(1,0,0) \sim M^{(\infty)}(0,1,0) \sim M^{(\infty)}(0,0,1) \sim \text{Gamma}\,(a_1, 1) \\ M^{(\infty)}(1,1,0) \sim M^{(\infty)}(0,1,1) \sim M^{(\infty)}(1,0,1) \sim \text{Gamma}\,(a_2, 1)\,. \end{cases} \tag{29}$$

Conversely, when $b \neq a_1 + a_2$, the laws of $M^{(\infty)}$, $R^{(\infty)}$ and $S^{(\infty)}$ are all different. If $M^{(\infty)} \stackrel{\mathcal{L}}{=} R^{(\infty)}$, we would have $M^{(\infty)}(0,1,0) \stackrel{\mathcal{L}}{=} R^{(\infty)}(0,1,0) \stackrel{\mathcal{L}}{=} M^{(\infty)}(1,0,0)$, hence

$$G_1 \stackrel{\mathcal{L}}{=} G_3 X_3^{(\infty)} \Rightarrow \text{E}\left(X_3^{(\infty)}\right) = \frac{a_1}{b}\ .$$

Analogously, from $M^{(\infty)}(1,1,0) \stackrel{\mathcal{L}}{=} R^{(\infty)}(1,1,0) \stackrel{\mathcal{L}}{=} M^{(\infty)}(1,0,1)$ we would have the contradiction

$$\begin{aligned} G_4 \stackrel{\mathcal{L}}{=} G_3\left(1 - X_3^{(\infty)}\right) &\Rightarrow \text{E}\left(1 - X_3^{(\infty)}\right) = \tfrac{a_2}{b} \\ \Rightarrow 1 = \text{E}\left(X_3^{(\infty)} + 1 - X_3^{(\infty)}\right) &= \tfrac{a_1+a_2}{b}\ . \end{aligned}$$

By using (27), it is immediately obtained that also $M^{(\infty)} \overset{\mathcal{L}}{\neq} S^{(\infty)}$, $R^{(\infty)} \overset{\mathcal{L}}{\neq} S^{(\infty)}$. □

It is worth noting that, for the standard case $a_1 + a_2 - b > 0$, Theorem 3.3 is not in contradiction with the positive recurrence of $\{\boldsymbol{M}^{\left(n+\frac{l}{3}\right)}\}$, for $l = 0, 1, 2$, in the state space $T_3 = (\mathbb{R}^+)^8$. In fact if the algorithm starts with the empty cells $(0, 0, 0)$ and $(1, 1, 1)$ it keeps them empty forever.

4. CONVERGENCE OF THE BIPF FROM POSITIVE MATRICES

In this section we focus on the case $a_1 + a_2 - b \leq 0$. After having proved that the random variables $X_1^{(m)}$, $X_4^{(m)}$ converge to 0, our aim is to study, as suggested by Corollary 2.4, the asymptotic behaviour of the remaining variables $X_2^{(m)}$, $X_3^{(m)}$. It will turn out that the asymptotic behaviour discussed in the previous section for $x_1^{(0)} = 0$, $x_4^{(0)} = 0$ holds for any starting point $x^{(0)} \in (0, 1)^4$.

Let A be any Borel subset of $(0, 1)^2$ and define the function

$$G_{m,A}(\boldsymbol{x}^{(0)}) = \mathrm{P}\left(\left(X_2^{(m)}, X_3^{(m)}\right) \in A | \boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}\right)$$

for $x^{(0)} \in \Gamma$. The following result holds for any $a_1, a_2, b > 0$.

Proposition 4.1. For any $m \geq 1$ the function $G_{m,A}$ is continuous in Γ.

P r o o f. See Appendix A. □

We are now ready to prove the main result of the paper.

Theorem 4.2. For $a_1 + a_2 - b \leq 0$ and any $x^{(0)} \in (0, 1)^4$

$$\lim_{m\to\infty} G_{m,A}(\boldsymbol{x}^{(0)}) = \left(\mu_{a_2,a_1,b} \otimes \mu_{a_1,a_2,b}\right)(A),$$

for any Borel subset A of $(0, 1)^2$.

P r o o f. Define the sets

$$B_d = [0, 1) \times [d, 1 - d]^2 \times [0, 1)\,, D_\rho = [0, \rho] \times (0, 1)^2 \times [0, \rho]\,,$$

for any $0 < d < \frac{1}{2}$ and $0 < \rho < 1$. Moreover define the projection p of $(0, 1)^4$ onto C as

$$p(x_1, x_2, x_3, x_4) = (0, x_2, x_3, 0)$$

and consider the process $\widetilde{\boldsymbol{X}}^{(m)} = p\big(\boldsymbol{X}^{(m)}\big) \in C$.

By Proposition 3.1 and Proposition 3.2, for any $x^{(0)} \in (0, 1)^4$ and $\delta > 0$, there exists $m_0 \in \mathbb{N}^*$ such that for any integer m

$$\left| G_{m_0,A}\left(\widetilde{\boldsymbol{X}}^{(m)}\right) - \upsilon(A)\right| < \frac{\delta}{4}, \tag{30}$$

for any Borel subset A of $(0,1)^2$, where $\upsilon = \mu_{a_2,a_1,b} \otimes \mu_{a_1,a_2,b}$.

By Proposition 2.5, for any $x^{(0)} \in (0,1)^4$ and $\delta > 0$, there exist $\overline{d} > 0$, $m_1 \in \mathbb{N}$ such that

$$\mathrm{P}\left(\boldsymbol{X}^{(m)} \in B_{\overline{d}} | \boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}\right) > 1 - \frac{\delta}{8}, \quad \forall\ m \geq m_1. \tag{31}$$

By Proposition 4.1 the function $G_{m_0,A}$ is uniformly continuous over the compact subsets $B_{\overline{d}} \cap D_\rho$. Therefore there exists $\overline{\rho} \in (0,1)$ such that whenever $X^{(m)} \in B_{\overline{d}} \cap D_{\overline{\rho}}$, it holds

$$\left| G_{m_0,A}(\boldsymbol{X}^{(m)}) - G_{m_0,A}\left(\widetilde{\boldsymbol{X}}^{(m)}\right)\right| \leq \frac{\delta}{4}\ . \tag{32}$$

By Corollary 2.4, there exists $m_2 \in \mathbb{N}$ such that

$$\mathrm{P}\left(\boldsymbol{X}^{(m)} \in D_{\overline{\rho}} | \boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}\right) > 1 - \frac{\delta}{8}, \quad \forall\ m \geq m_2. \tag{33}$$

Therefore, because of (31) and (33), for $m \geq \max\{m_1, m_2\}$

$$\mathrm{P}\left(\boldsymbol{X}^{(m)} \in D^c_{\overline{\rho}} \cup B^c_{\overline{d}} | \boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}\right) \leq \frac{\delta}{4}. \tag{34}$$

By the Chapman–Kolmogorov equations

$$\begin{aligned} & \mathrm{P}\left(\left(X_2^{(m+m_0)}, X_3^{(m+m_0)}\right) \in A | \boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}\right) \\ = \quad & G_{m+m_0,A}\left(\boldsymbol{x}^{(0)}\right) = E_{\boldsymbol{x}^{(0)}}\left(G_{m_0,A}\left(\boldsymbol{X}^{(m)}\right)\right), \end{aligned}$$

from which

$$\left|G_{m+m_0,A}\left(\boldsymbol{x}^{(0)}\right) - \upsilon(A)\right| = \left|E_{\boldsymbol{x}^{(0)}}\left(G_{m_0,A}\left(\boldsymbol{X}^{(m)}\right) - \upsilon(A)\right)\right|.$$

Writing the integrand as

$$\begin{aligned} & G_{m_0,A}\left(\boldsymbol{X}^{(m)}\right) - \upsilon(A) \\ = \quad & G_{m_0,A}\left(\boldsymbol{X}^{(m)}\right) - G_{m_0,A}\left(\widetilde{\boldsymbol{X}}^{(m)}\right) + G_{m_0,A}\left(\widetilde{\boldsymbol{X}}^{(m)}\right) - \upsilon(A) \end{aligned}$$

and using (30), (32) and (34)

$$\begin{aligned} & \left|\mathrm{P}\left(\left(X_2^{(m+m_0)}, X_3^{(m+m_0)}\right) \in A | \boldsymbol{X}^{(0)} = \boldsymbol{x}^{(0)}\right) - \upsilon(A)\right| \\ \leq & \left|\mathrm{E}_{\boldsymbol{x}^{(0)}}\left(1_{\left\{\boldsymbol{X}^{(m)} \in B_{\overline{d}} \cap D_{\overline{\rho}}\right\}}\left(G_{m_0,A}\left(\boldsymbol{X}^{(m)}\right) - G_{m_0,A}\left(\widetilde{\boldsymbol{X}}^{(m)}\right)\right)\right)\right| \\ + & \left|\mathrm{E}_{\boldsymbol{x}^{(0)}}\left(1_{\left\{\boldsymbol{X}^{(m)} \in D^c_{\overline{\rho}} \cup B^c_{\overline{d}}\right\}}\left(G_{m_0,A}\left(\boldsymbol{X}^{(m)}\right) - G_{m_0,A}\left(\widetilde{\boldsymbol{X}}^{(m)}\right)\right)\right)\right| \\ + & \left|\mathrm{E}_{\boldsymbol{x}^{(0)}}\left(G_{m_0,A}\left(\widetilde{\boldsymbol{X}}^{(m)}\right) - \upsilon(A)\right)\right| < \tfrac{\delta}{4} + 2 \cdot \tfrac{\delta}{4} + \tfrac{\delta}{4} = \delta, \end{aligned}$$

for $m \geq \max\{m_1, m_2\}$, which concludes the proof. □

As a consequence of Theorem 4.2 and Corollary 2.4, when $a_1 + a_2 - b \leq 0$ the sequence $\{\boldsymbol{X}^{(m)}\}$ converges in law to $\pi \equiv \delta_{\{0\}} \otimes \mu_{a_2,a_1,b} \otimes \mu_{a_1,a_2,b} \otimes \delta_{\{0\}}$, for any $x^{(0)} \in (0,1)^4$. This allows us to extend Theorem 3.3 to any positive initial table.

Theorem 4.3. When $a_1 + a_2 - b \leq 0$, for $M^{(0)} = m^{(0)} \in (\mathbb{R}^+)^8$ the random sequences $\left\{\boldsymbol{M}^{(n)}\right\}$,$\left\{\boldsymbol{M}^{(n+\frac{1}{3})}\right\}$ and $\left\{\boldsymbol{M}^{(n+\frac{2}{3})}\right\}$ converge in law to the respective limits $M^{(\infty)}$, $R^{(\infty)}$, $S^{(\infty)}$ defined by (26) and (27).

5. COMPARISON WITH THE IPF

Putting together the observations of Remark 2.3 with Theorem 3.3 and Theorem 4.3 we have proved a kind of bifurcation result for the matrix-valued processes $\left\{\boldsymbol{M}^{\left(n+\frac{l}{3}\right)}\right\}$, for $l = 0, 1, 2$, defined by (2) (with n_l all given by (6)) starting from an arbitrary $M^{(0)} \in T_3$.

(a) For $a_1 + a_2 - b > 0$ all the sequences $M^{\left(n+\frac{l}{3}\right)}$, for $l = 0, 1, 2$ converge to a limit law τ_r supported by $T_3^r \subset T_3$, depending on $r = R(M^{(0)}) > 0$.

(b) For $a_1 + a_2 - b = 0$ the convergence result still holds, but the limit law does not depend on r, and it is supported by the following subset of the boundary of T_3:

$$\boldsymbol{T}_3^0 = \left\{\boldsymbol{M} \in \mathbb{R}^8\colon M(0,0,0) = M(1,1,1) = 0,\ M(i,j,k) > 0 \text{ otherwise}\right\}.$$

(c) For $a_1 + a_2 - b < 0$ the three sequences $\left\{\boldsymbol{M}^{\left(n+\frac{l}{3}\right)}\right\}$, for $l = 0, 1, 2, \ldots$ have different limit laws supported by T_3^0, which do not depend on r. They are obtained one from the other by a cyclic shift of the arguments of the matrix.

This behaviour is analogous to that of the deterministic algorithm $\left\{\boldsymbol{m}^{\left(n+\frac{l}{3}\right)}, n \in \mathbb{N},\ l = 0, 1, 2\right\}$ defined by (1) and (2) with

$$V^{\left(n+\frac{l}{3}\right)}(i,j) = n_l(i,j), \qquad i, j = 0, 1,\ l = 0, 1, 2,$$

with all n_l given by (6), which is an example of the celebrated IPF algorithm. In the rest of the paper, we discuss this analogy.

When $a_1+a_2-b > 0$, from Lauritzen [11], Theorem 4.13, $m^{\left(n+\frac{l}{3}\right)}$ converges to the unique joint table $n_r^{(\infty)} \in T_3^r$ with all pair marginal equal to n_l, with $r = R(m^{(0)})$, for any $l = 0, 1, 2$. In fact the purpose of the IPF algorithm is precisely to obtain this table. The result of Lauritzen covers the case $a_1 + a_2 - b = 0$ as well: in this case there is only one table with non negative entries with pair marginals all equal to n_l, given by

$$\begin{cases} m^{(\infty)}(0,0,0) = m^{(\infty)}(1,1,1) = 0, \\ m^{(\infty)}(1,0,0) = m^{(\infty)}(0,1,0) = m^{(\infty)}(0,0,1) = a_1, \\ m^{(\infty)}(1,1,0) = m^{(\infty)}(0,1,1) = m^{(\infty)}(1,0,1) = a_2, \end{cases} \tag{35}$$

(the table of mean values of a matrix distributed as in (29)) to which all the three sequences $m^{\left(n+\frac{l}{3}\right)}$ converge as $n \to \infty$, for $l = 0, 1, 2$.

We have already considered the case $a_1 + a_2 - b < 0$ with $a_1 = a_2$ in Asci and Piccioni [3]. It is not difficult to extend the results in this paper in the following way.

Theorem 5.1. When $a_1 + a_2 - b \leq 0$, for $m^{(0)} \in (\mathbb{R}^+)^8$ the sequences $\{\boldsymbol{m}^{(n)}\}$, $\{\boldsymbol{m}^{(n+\frac{1}{3})}\}$ and $\{\boldsymbol{m}^{(n+\frac{2}{3})}\}$ converge to the limit tables $m^{(\infty)}$, $r^{(\infty)}$, $s^{(\infty)}$ defined by

$$\begin{pmatrix} m^{(\infty)}(0,0,0) & m^{(\infty)}(0,1,0) \\ m^{(\infty)}(0,0,1) & m^{(\infty)}(0,1,1) \\ m^{(\infty)}(1,0,0) & m^{(\infty)}(1,1,0) \\ m^{(\infty)}(1,0,1) & m^{(\infty)}(1,1,1) \end{pmatrix} = \begin{pmatrix} 0 & a_1 \\ b\left(1 - x_2^{(\infty)}\right) & bx_2^{(\infty)} \\ bx_3^{(\infty)} & b\left(1 - x_3^{(\infty)}\right) \\ a_2 & 0 \end{pmatrix}$$

and

$$\boldsymbol{r}^{(\infty)} = \boldsymbol{m}^{(\infty)} \circ \sigma^2, \ \boldsymbol{s}^{(\infty)} = \boldsymbol{m}^{(\infty)} \circ \sigma,$$

where

$$x_2^{(\infty)} = \frac{v - w - 1 + \sqrt{(v-w)^2 + 2(v+w) + 1}}{2v}, \tag{36}$$

$$x_3^{(\infty)} = \frac{w - v - 1 + \sqrt{(v-w)^2 + 2(v+w) + 1}}{2w}, \tag{37}$$

with $v = \frac{b}{a_1}$ and $w = \frac{b}{a_2}$. The tables $m^{(\infty)}$, $r^{(\infty)}$, $s^{(\infty)}$ are equal if and only if $a_1 + a_2 - b = 0$ (in which case they are equal to (35)), otherwise they are all different.

Sketch of the proof.

Step 1. We define the sequence $\{\boldsymbol{x}^{(m)}\}$ as in (7) and (9). It is possible to prove that the sequence $\{\Lambda\left(\boldsymbol{x}^{(m)}\right)\}$ is non decreasing in m, where

$$\Lambda(\boldsymbol{x}) = x_1^{a_1+a_2-b}[(1-x_1)(1-x_2)x_3]^{b-a_2}$$

$$\cdot [x_2(1-x_3)(1-x_4)]^{a_2}, \ \boldsymbol{x} \in (0,1)^4,$$

and an analogous result holds by exchanging x_1 with x_4 and x_2 with x_3. This implies

$$\lim_{m \to +\infty} x_1^{(m)} = \lim_{m \to +\infty} x_4^{(m)} = 0. \tag{38}$$

Step 2. With a proof similar to that of Proposition 2.5, it is possible to show that there exist $d > 0$ and $m_0 \in \mathbb{N}$ such that

$$d \leq x_j^{(m)} \leq 1 - d, \ j = 2, 3, \ \ m \geq m_0. \tag{39}$$

Step 3. By using a contraction argument to the deterministic counterpart of the equations (20) it follows that $x_1^{(0)} = x_4^{(0)} = 0$ implies

$$\left\| \boldsymbol{x}^{(m)} - \boldsymbol{x}^{(\infty)} \right\|_\infty \leq M r^m, \tag{40}$$

for some $M > 0$, $r \in (0, 1)$, where the non-zero components of $x^{(\infty)} \in C$ are given by (36) and (37).

Step 4. By using (38), (39), (40) and the uniform continuity of (11) over any compact set of the form $[0, \rho] \times [d, 1-d]^2 \times [0, \rho]$, where $\rho, d > 0$, we get that $\left\| \boldsymbol{x}^{(n)} - \boldsymbol{x}^{(\infty)} \right\|_\infty \overset{n}{\to} 0$ for any $x^{(0)} \in (0, 1)^4$. Plugging the vector $x^{(\infty)}$ in (12) for $l = 0, 1, 2$ we get the matrices $m^{(\infty)}$, $r^{(\infty)}$, $s^{(\infty)}$, respectively. The last statement of the theorem follows immediately. □

6. PRACTICAL IMPLICATIONS

There is some recent literature on the behaviour of a Gibbs Sampler with improper target. Lauritzen and Richardson [12] consider a general form of single variable conditional specifications, proving that there exists a joint from which they are obtained *(consistency)* if and only if the corresponding Gibbs Sampler (corresponding to any enumeration of variables) has a (proper) stationary distribution, which is preserved under all the intermediate stages of a single update. However, for the general BIPF algorithm with *pairwise consistent* tables of mean values, we always have the existence of a stationary measure, possibly improper [4]. In this paper positive contingency tables belonging to a hierarchical model are parametrized by interactions (see [11], page 246), which take arbitrary real values. The stationary measure has a density in the whole interaction space, and it is proper if and only if there is a positive joint table which produces the specified tables of mean values by marginalization. Since the stationary density is continuous, it is always integrable in any bounded subset of interactions. When it is improper, by applying Theorem 6 and Proposition 10.3 in Tweedie [17], the probability that the trajectory of the BIPF algorithm lies in any bounded subset of interactions (in particular the set of tables with all the entries and their reciprocals bounded by a constant) goes to zero as the number of updates increases. This does not imply that the algorithm is eventually "approaching the boundary" with probability 1; in fact in our examples we do not know if this is true; in fact we have established only that $X_i^{(m)} \overset{P}{\to} 0$ and not the stronger property $X_i^{(m)} \overset{\text{a.s.}}{\to} 0$, $i = 1, 4$. This makes more difficult the implementation of suitable diagnostics for detecting the improperness of the target, as shown in a number of examples presented in the paper by Hobert and Casella [10]. In the class of examples analyzed here, provided Φ is a cyclically invariant functional defined on T_3 (i. e. invariant by σ), its expected value is the same under all three limiting probability measures μ_l, $l = 0, 1, 2$, so by observing $\Phi\left(\boldsymbol{M}^{(n+\frac{l}{3})}\right)$ for $l = 0, 1, 2$ and all integers n, we will never see a failure of convergence to equilibrium even if $a_1 + a_2 - b \le 0$. On the other hand, this means that even if the target measure is improper, there is some information encoded in it about expected values of these cyclically invariant functionals. In a different context, van Dyk and Meng [7] have given some arguments supporting the use of improper targets for Gibbs Sampler computations. In their examples additional latent data are introduced within a statistical model together with auxiliary parameters having improper prior measures. Under suitable conditions, they prove (page 10 in [7]) that it is still possible to get samples from the proper posterior distribution on the original parameters of interest.

A. PROOF OF PROPOSITION 4.1

Define for $m \ge 2$

$$(\boldsymbol{H}_{m-1}, \boldsymbol{K}_{m-1}) = (H_1, \dots, H_{m-1}, K_1, \dots, K_{m-1}).$$

From (10)–(11), since $H_m \sim \beta^{(2)}_{a_1,b}$ and $K_m \sim \beta^{(2)}_{a_2,b}$ we see that for any $x^{(0)} \in (0,1)^4$ the law of the pair $\left(X_2^{(m)}, X_3^{(m)}\right)$ conditional to $(\boldsymbol{H}_{m-1}, \boldsymbol{K}_{m-1}) = (\boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}) \in (\mathbb{R}^+)^{2m-2}$ is absolutely continuous w.r.t. the law Beta$(a_2, b)\otimes$Beta(a_1, b) with density $\varphi\left(y_2, y_3, \boldsymbol{x}^{(m-1)}\right)$ given by

$$\frac{\left(1 - x_1^{(m-1)}\right)^b \left(x_2^{(m-1)}\right)^{a_1} \left(x_3^{(m-1)}\right)^{a_2} \left(1 - x_4^{(m-1)}\right)^b}{\left[\left(1 - x_1^{(m-1)}\right)(1 - y_3) + x_2^{(m-1)} y_3\right]^{a_1+b} \left[\left(1 - x_4^{(m-1)}\right)(1 - y_2) + x_3^{(m-1)} y_2\right]^{a_2+b}}, \tag{41}$$

where the vector $x^{(m-1)}$ is a function of $\boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}$ and $x^{(0)}$. In order to bound this function independently of $x^{(0)}$, we introduce the following lemma.

Lemma A.1. For any $0 < \varepsilon < \frac{1}{2}$ and $n = 1, 2, \dots$

$$\frac{1}{\left(1 - x_1^{(n)}\right) x_2^{(n)} x_3^{(n)} \left(1 - x_4^{(n)}\right)} \le \frac{1}{\varepsilon^6} \prod_{i=1}^{n} \frac{(h_i + 1)^3 (k_i + 1)^3}{h_i k_i},$$

for any $x^{(0)} \in [0, 1-\varepsilon) \times (\varepsilon, 1-\varepsilon) \times (\varepsilon, 1-\varepsilon) \times [0, 1-\varepsilon) \equiv U_\varepsilon$.

Proof. From the equations (10) and (11), it follows

$$\frac{1}{\left(1-x_1^{(n)}\right)\left(1-x_2^{(n)}\right)x_3^{(n)}}$$

$$= \frac{\left[h_n x_1^{(n-1)}+1-x_2^{(n-1)}\right]\left[k_n\left(1-x_4^{(n-1)}\right)+x_3^{(n-1)}\right]\left[h_n\left(1-x_1^{(n-1)}\right)+x_2^{(n-1)}\right]}{\left(1-x_1^{(n-1)}\right)\left(1-x_2^{(n-1)}\right)x_3^{(n-1)}h_n}$$

$$\leq \frac{1}{\left(1-x_1^{(n-1)}\right)\left(1-x_2^{(n-1)}\right)x_3^{(n-1)}}\cdot\frac{\left(h_n+1\right)^2\left(k_n+1\right)}{h_n}.$$

Iterating the argument, we have

$$\frac{1}{\left(1-x_1^{(n)}\right)\left(1-x_2^{(n)}\right)x_3^{(n)}} \leq \frac{1}{\varepsilon^3}\prod_{i=1}^{n}\frac{\left(h_i+1\right)^2\left(k_i+1\right)}{h_i}. \tag{42}$$

Repeating the same argument replacing $x_1^{(n)}$ with $x_4^{(n)}$, and exchanging $x_2^{(n)}$ with $x_3^{(n)}$ and H_i with K_i we get

$$\frac{1}{\left(1-x_4^{(n)}\right)\left(1-x_3^{(n)}\right)x_2^{(n)}} \leq \frac{1}{\varepsilon^3}\prod_{i=1}^{n}\frac{\left(h_i+1\right)\left(k_i+1\right)^2}{k_i}, \tag{43}$$

from which, multiplying both sides of (42) and (43), the statement of the lemma follows. □

Let us denote the two factors of the denominator in (41) by D_1 and D_2, respectively. Put δ be such that $0 < \delta < \min\{a_1, a_2, b/2\}$. The first factor is bounded as

$$D_1 \geq \left[\left(1-x_1^{(m-1)}\right)\left(1-y_3\right)\right]^{b-\delta}\left(x_2^{(m-1)}y_3\right)^{a_1-\delta}\left[\left(1-x_1^{(m-1)}\right)x_2^{(m-1)}\right]^{2\delta}, \tag{44}$$

whereas for the second factor we have

$$D_2 \geq \left[\left(1-x_4^{(m-1)}\right)\left(1-y_2\right)\right]^{b-\delta}\left(x_3^{(m-1)}y_2\right)^{a_2-\delta}\left[\left(1-x_4^{(m-1)}\right)x_3^{(m-1)}\right]^{2\delta}. \tag{45}$$

Now inserting (44) and (45) in (41), for $x^{(0)} \in U_\varepsilon$

$$\varphi\left(y_2, y_3, \boldsymbol{x}^{(m-1)}\right) \leq \frac{y_2^{\delta-a_2}y_3^{\delta-a_1}\left[\left(1-y_2\right)\left(1-y_3\right)\right]^{\delta-b}}{\left[\left(1-x_1^{(m-1)}\right)x_2^{(m-1)}x_3^{(m-1)}\left(1-x_4^{(m-1)}\right)\right]^{\delta}}$$

$$\leq \varepsilon^{-6\delta} y_2^{\delta-a_2} y_3^{\delta-a_1} \left[(1-y_2)(1-y_3)\right]^{\delta-b} \prod_{1=1}^{m-1} \left[(h_i+1)(k_i+1)\right]^{3\delta} \frac{1}{(h_i k_i)^{\delta}}, \tag{46}$$

by Lemma A.1 with $n = m-1$. Since $\boldsymbol{H}_{m-1}$ and $\boldsymbol{K}_{m-1}$ are independent vectors with i.i.d. components distributed as $\beta^{(2)}_{a_1,b}$ and $\beta^{(2)}_{a_2,b}$, respectively, the joint density of $\left(X_2^{(m)}, X_3^{(m)}, \boldsymbol{H}_{m-1}, \boldsymbol{K}_{m-1}\right)$ conditioned by $x^{(0)}$ has the following form:

$$f_{\left(X_2^{(m)}, X_3^{(m)}, \boldsymbol{H}_{m-1}, \boldsymbol{K}_{m-1}\right)} \left(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}; \boldsymbol{x}^{(0)}\right)$$

$$= \varphi\left(y_2, y_3, \boldsymbol{x}^{(m-1)}\right) g\left(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}\right),$$

where

$$g\left(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}\right)$$

$$= \frac{y_2^{a_2-1} y_3^{a_1-1} \left[(1-y_2)(1-y_3)\right]^{b-1} \prod_{i=1}^{m-1} h_i^{a_1-1} k_i^{a_2-1}}{\gamma \prod_{i=1}^{m-1} (h_i+1)^{a_1+b} (k_i+1)^{a_2+b}}, \qquad m \geq 1, \tag{47}$$

where $\gamma = (B(a_1,b)B(a_2,b))^m$ and the product terms are equal to 1 for $m = 1$. By (46)

$$f_{\left(X_2^{(m)}, X_3^{(m)}, \boldsymbol{H}_{m-1}, \boldsymbol{K}_{m-1}\right)} \left(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}; \boldsymbol{x}^{(0)}\right)$$

$$\leq \varepsilon^{-6\delta} \left[y_2(1-y_2)y_3(1-y_3)\right]^{\delta-1} \cdot \frac{\prod_{i=1}^{m-1} h_i^{a_1-\delta-1} k_i^{a_2-\delta-1}}{\gamma \prod_{i=1}^{m-1} (h_i+1)^{a_1+b-3\delta} (k_i+1)^{a_2+b-3\delta}} \tag{48}$$

$$\equiv h\left(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}\right),$$

where h is integrable in $(y_2,\ y_3,\ \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1})$, since it is proportional to a product of densities beta and beta of second kind.

Finally, because of the continuity of $F(x^{(0)}, h, k)$ in $x^{(0)} \in \Gamma$ for any $h, k > 0$, the density $f_{\left(X_2^{(m)}, X_3^{(m)}, \boldsymbol{H}_{m-1}, \boldsymbol{K}_{m-1}\right)} \left(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}; \boldsymbol{x}^{(0)}\right)$ is continuous in $x^{(0)} \in \Gamma$ for any $(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}) \in (0,1)^2 \times \left(\mathbb{R}^+\right)^{2m-2}$. By (48) and the bounded convergence theorem, it follows that

$$G_{m,A}\left(\boldsymbol{x}^{(0)}\right) = \int_{A \times (\mathbb{R}^+)^{2m-2}} f_{\left(X_2^{(m)}, X_3^{(m)}, \boldsymbol{H}_{m-1}, \boldsymbol{K}_{m-1}\right)}$$

$$\cdot \left(y_2, y_3, \boldsymbol{h}_{m-1}, \boldsymbol{k}_{m-1}; \boldsymbol{x}^{(0)}\right) \mathrm{d}y_2 \mathrm{d}y_3 \mathrm{d}\boldsymbol{h}_{m-1} \mathrm{d}\boldsymbol{k}_{m-1}$$

is continuous in U_ε for any $0 < \varepsilon < \frac{1}{2}$, therefore it is continuous in the whole $\Gamma = \bigcup_\varepsilon U_\varepsilon$.

(Received March 3, 2008.)

REFERENCES

[1] D. Z. Albert: Quantum Mechanics and Experience. Harvard College, Cambridge 1992.

[2] C. Asci, G. Letac, and M. Piccioni: Beta-hypergeometric distributions and random continued fractions. Statist. Probab. Lett. *78* (2008), 1711–1721.

[3] C. Asci and M. Piccioni: The IPF algorithm when the marginal problem is unsolvable: the simplest case. Kybernetika *39* (2003), 731–737.

[4] C. Asci and M. Piccioni: Functionally compatible local characteristics for the local specification of priors in graphical models. Scand. J. Statist. *34* (2007), 829–840.

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

[6] A. P. Dawid and S. L. Lauritzen: Hyper-Markov laws in the statistical analysis of decomposable graphical models. Ann. Statist. *21* (1993), 1272–1317.

[7] D. A. van Dyk and X. Meng: The art of data augmentation. With discussions, and a rejoinder by the authors. J. Comput. Graph. Statist. *10* (2001), 1–111.

[8] A. Einstein, B. Podolsky, and N. Rosen: Can the quantum-mechanical description of physical reality be considered complete? Phys. Rev. *47* (1935), 777–780.

[9] A. Gelman, J. B. Carlin, D. B. Rubin, and H. S. Stern: Bayesian Data Analysis. Chapman and Hall, London 1995.

[10] J. P. Hobert and G. Casella: Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors. J. Comput. Graph. Statist. *7* (1998), 42–60.

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

[12] S. L. Lauritzen and T. S. Richardson: Chain graph models and their causal interpretations (with discussion). J. Roy. Statist. Soc. Ser. B *64* (2002), 321–348.

[13] S. P. Meyn and R. L. Tweedie: Markov Chains and Stochastic Stability. Springer-Verlag, London 1993.

[14] M. Piccioni: Independence structure of natural conjugate densities to exponential families and the Gibbs' sampler. Scand. J. Statist. *27* (2000), 111–127.

[15] E. D. Rainville: Special Functions. MacMillan, New York 1960.

[16] J. L. Schafer: Analysis of Incomplete Multivariate Data. Chapman and Hall, London 1997.

[17] R. L. Tweedie: R-theory for Markov chains on a general state space I, II. Ann. Probab. *2* (1974), 840–878.

[18] D. Williams: Weighing the Odds. Cambridge University Press, Cambridge 2001.

Claudio Asci, Dipartimento di Matematica e Informatica, Università degli Studi di Trieste, Via Valerio 12/1, 34127 Trieste. Italy.
e-mail: asci@dmi.units.it

Mauro Piccioni, Dipartimento di Matematica, Sapienza Università di Roma, Piazzale A. Moro 2, 00185 Roma. Italy.
e-mail: piccioni@mat.uniroma1.it