Department of Applied Mathematics, University of Venice

WORKING PAPER SERIES

Riccardo Gusso and Uwe Schmock

Urn-based models for dependent credit risks and their calibration through EM algorithm

Working Paper n. 163/2008
Aprile 2008

ISSN: 1828-6887

Urn-based models for dependent credit risks and their calibration trough EM algorithm

Riccardo Gusso
Department of Applied Mathematics
University Ca' Foscari Venice
<rgusso@unive.it>

Uwe Schmock
Institute for Mathematical Methods in Economics
Vienna University of Technology
<schmock@fam.tuwien.ac.at>

(April 2008)

Abstract. In this contribution we analyze two models for the joint probability of defaults of dependent credit risks that are based on a generalisation of Pólya urn scheme. In particular we focus our attention on the problems related to the maximum likelihood estimation of the parameters involved, and to this purpose we introduce an approach based on the use of the Expectation-Maximization algorithm. We show how to implement it in this context, and then we analyze the results obtained, comparing them with results obtained by other approaches.

Keywords: Credit risk, credit rating, urn models, exchangeable random variables, maximum likelihood, EM algorithm.

JEL Classification Numbers: C13, C16.

MathSci Classification Numbers: 60E05, 60F99, 62H12.

Correspondence to:

Riccardo Gusso	Department of Applied Mathematics University Ca' Foscari Venice Dorsoduro 3825/E 30123 Venice, Italy
Phone:	[++39] 041 2346933
Fax:	[++39] 041 5221756
E-mail:	rgusso@unive.it

1 Introduction

One of the main topics in credit risk modeling and management is the problem of assigning the default probability for every obligor or group of obligors in a portfolio of risks. Moreover, when considering the point of view of a financial firm facing the risk of a portfolio of obligors, the principal issue is the occurrence of many joint defaults over a fixed time horizon. Joint defaults events also are important with respect to the performance of derivative securities whose payoff is linked to the profit and loss of a portfolio of underlying bonds. From this kind of considerations it emerges then that to measure the expected loss in a portfolio of credit risks, dependence between them cannot be ignored and that its specification is at least as important as the specification of individual default probabilities; moreover the concept of dependence cannot be interpreted only as linear correlation between the random variables involved.

To takle this issues, in the recent years various mathematical models have been developed both in the academic literature and in the financial industry; in this contribution we consider the case of a portofolio of firms belonging to several different rating classes; and we analyze two models which are based on a generalization of the Pólya urn scheme. In this way their dependence structure allows for dependence both in the same rating group and in different rating groups, introducing then some form of contagion between defaults. For both of them we derive the expression of the joint default probability for the number of defaults in the different rating groups; the complexity involved in the statistical estimation of its parameters lead us to introduce the Expectation-Maximization algorithm for iterative maximum likelihood estimations.

The Pólya urn scheme can be described as follows. Consider an urn that contains b black balls and r red ones. Then we make subsequent draws from the urn following this scheme:

i) draw a ball from the urn

ii) return the ball in the urn together with $c > 0$ balls of the same colour.

In this way we clearly introduce some dependence structure between the draws. This scheme can be used to model the number of defaults in a group of n firms each beloging to the same rating class. For each firm $i \in \{1, \dots, n\}$ we draw a ball from the urn; if the ball is red, then the firm defaults, if black it does not. Then we return the ball in the urn toghether with c balls of the same colour. If we denote with N_n the random variable describing the number of defaults within this group of companies, then using the theory of exchangeable sequences it can be easily shown that the distribution of N_n is given by a beta-binomial mixture:

$$\mathbb{P}[N_n = k] = \binom{n}{k} \frac{B(\alpha + k, n - k + \beta)}{B(\alpha, \beta)} \tag{1}$$

where α, β are two parameters which have to be estimated from historical data, and $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, with $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$, is the Euler beta function.

If you want to deal with the case of firms belonging to different rating classes, this model needs to be generalized, and we will discuss two possible extensions, and the problems related to the estimation of the parameters involved, in the next sections of this paper.

2 Multidimensional urn scheme for defaults

Suppose we are given k homogeneous groups of n_i, $1 \leq i \leq k$ companies, with credit ratings satisfying $r_1 \succ r_2 \succ \cdots \succ r_k$, where $r_i \succ r_j$ means that the rating r_i is higher than the rating r_j. To determine the joint probability of the defaults a multicolour urn scheme is introduced.

Consider an urn which contains balls of $k+1$ different colours, with $k \geq 2$ and with $b_j > 0$ balls for every $1 \leq j \leq k+1$. Originally then there are $b = \sum_{j=1}^{k+1}$ balls in the urn. The scheme works through these steps:

1. draw a ball at random from the urn
2. note the colour of the ball
3. return the ball in the urn together with other $c \geq 0$ balls of the same colour.

The mechanism is then the following: to determine if a company of any rating class defaults or not we draw randomly a ball from the urn. If its rating is r_j, and the ball has a colour from $1, \dots, j$ then the firm defaults; otherwise if the colour is from $j+1, \dots, k+1$ it does not default. Then we return the ball in the urn together with other c balls of the same colour, introducing in this way some dependence between the defaults both in the same and in different rating groups.

The results about exchangeable sequences of random variables are then used to determine the default probabilities.

Definition 2.1. We define the random vector $\mathbf{X}_n = (X_{n,1}, \dots X_{n,k+1}) \in \{0,1\}^{k+1}$ indicating the colour of the ball of the n-th draw in the following way:

$$X_{n,j} := \begin{cases} 1 & \text{if the } n\text{th ball drawn has colour } j \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that this sequence of random variable is an exchangeable one; we recall here the definition:

Definition 2.2. The finite set $(X_1, X_2, \dots, X_n)$ of random variables is said to be *exchangeable* if the joint distribution is invariant under all n-permutations:

$$(X_1, X_2, \dots, X_n) \stackrel{d}{=} (X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)}), \tag{2}$$

for every permutation π of $\{1, 2, \dots, n\}$.

An infinite sequence of random variables $(X_n)_{n\geq 1}$ is said to be *exchangeable* if $(X_1, \dots, X_n)$ is exchangeable for each $n \geq 2$.

If we denote $\mathbf{e}_j = (\delta_{1,j}, \dots, \delta_{d,j})$ with $\delta_{j,j} = 1$ and $\delta_{i,j} = 0$ if $i \neq j$, we have then the following result (see [8]):

Proposition 2.3. *The sequence of random vectors $(\mathbf{X}_n)_{n\geq 1}$ is exchangeable: in fact for every $n \geq 1$ we have that*

$$\mathbb{P}[\mathbf{X}_1 = \mathbf{e}_{j_1}, \dots, \mathbf{X}_n = \mathbf{e}_{j_n}] = \frac{\prod_{j=1}^{k+1} \prod_{i=0}^{l_{n,j}-1} (b_j + ic)}{b(b+c)\cdots(b+(n-1)c)} \tag{3}$$

where $\mathbf{l}_n = (l_{n,1}, \dots, l_{n,k+1}) = \sum_{i=1}^{n} \mathbf{e}_{j_i}$.

The main result about exchangeable sequences of randoms variables is the following theorem, here stated for the case of an exchangeable sequence of random vectors, whose proof is shown in the appendix:

Theorem 2.4 (De Finetti's theorem)**.** *Let $(\mathbf{X}_n)_{n\geq 1}$ be an exchangeable sequence of random vectors from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the measurable space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, where $\mathscr{B}(\mathbb{R}^d))$ is the Borel σ-algebra. Then there exists a sub σ-field $\mathcal{F}_\infty \subseteq \mathcal{F}$ conditioned on which the $\mathbf{X}_n$'s are independent and identically distributed.*

An immediate corollary of De Finetti's theorem is the following one:

Theorem 2.5. *Let $(\mathbf{X}_n)_{n\geq 1}$ be an exchangeable sequence of random vectors taking values in $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$. Then there exists a random vector $(P_1, \dots, P_d)$ taking values in $\Delta^d = \{(p_1, \dots, p_d) \in [0,1]^d | \sum_{j=1}^{d} p_j = 1\}$ such that:*

1. *for all $\mathbf{l} \in \mathbb{N}_0^d$ with $\sum_{j=1}^{d} l_j = n$ it holds that*

$$\mathbb{P}\left[\sum_{i=1}^{n} \mathbf{X}_i = \mathbf{l} \middle| P_1, \dots, P_d\right] \overset{a.s.}{=} \frac{n!}{(l_1!)(l_2!)\cdots(l_d!)} P_1^{l_1} P_2^{l_2} \cdots P_d^{l_d}$$

2. *for $1 \leq j \leq d$,*

$$P_j \overset{a.s.}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_{i,j}.$$

Using the previous results it is now possible to determine the joint distribution of the random vector $\mathbf{N} = (N_1, N_2, \dots, N_k)$ of the number of defaults within each group. Since the sequence $(\mathbf{X}_n)_{n\geq 1}$ is exchangeable, it does not matter in which order we draw the balls for the companies, hence we can choose a special order to facilitate calculations; that is we first consider the firms with the best rating, than the next lower, and so on. The number of defaults in the j-th rating is given by how many times a ball of color from 1 to j has been drawn within n_j subsequent draws. We can then determine the joint default probabilities using Theorem 2.5:

$$
\begin{aligned}
&\mathbb{P}[N_1 = l_1, \dots, N_k = l_k] \\
&= \mathbb{P}\left[\sum_{i=1}^{n_1} X_{i,1} = l_1, \sum_{i=n_1+1}^{n_1+n_2} (X_{i,1} + X_{i,2}) = l_2, \dots, \sum_{i=n_1+\cdots+n_{k-1}+1}^{n_1+\cdots+n_k} (X_{i,1} + \cdots + X_{i,k}) = l_k\right] \\
&= \mathbb{E}\Bigg[\mathbb{P}\Bigg[\sum_{i=1}^{n_1} X_{i,1} = l_1, \sum_{i=n_1+1}^{n_1+n_2} (X_{i,1} + X_{i,2}) = l_2, \\
&\qquad \dots, \sum_{i=n_1+\cdots+n_{k-1}+1}^{n_1+\cdots+n_k} (X_{i,1} + \cdots + X_{i,k}) = l_k \Bigg| P_1, \dots, P_{k+1}\Bigg]\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{P}\Bigg[\sum_{i=1}^{n_1} X_{i,1} = l_1 \Bigg| P_1, \dots, P_{k+1}\Bigg] \mathbb{P}\Bigg[\sum_{i=n_1+1}^{n_1+n_2} (X_{i,1} + X_{i,2}) = l_2 \Bigg| P_1, \dots, P_{k+1}\Bigg] \\
&\qquad \cdots \mathbb{P}\Bigg[\sum_{i=n_1+\cdots+n_{k-1}+1}^{n_1+\cdots+n_k} (X_{i,1} + \cdots + X_{i,k}) = l_k \Bigg| P_1, \dots, P_{k+1}\Bigg]\Bigg] \\
&= \mathbb{E}\Bigg[\binom{n_1}{l_1} P_1^{l_1} (1 - P_1)^{n_1 - l_1} \binom{n_2}{l_2} (P_1 + P_2)^{l_2} (1 - P_1 - P_2)^{n_2 - l_2} \\
&\qquad \cdots \binom{n_k}{l_k} (P_1 + \cdots + P_k)^{l_k} (1 - P_1 - \cdots - P_k)^{n_k - l_k}\Bigg]
\end{aligned} \tag{4}
$$

since, being $\sigma(P_1, \dots, P_{k+1}) \subseteq \mathcal{F}_\infty$, the $\mathbf{X}_n$'s are, conditioned on $(P_1, \dots, P_{k+1})$, independent and identically distributed.

To obtain an explicit form for the joint default probabilities it is necessary to compute the distribution of $(P_1, \dots, P_{k+1})$; using the previous results it is possible to show (see [8]) that it has to be Dirichlet distributed $D_{k+1}(\alpha_1, \dots, \alpha_{k+1})$ with density

$$
f_{k+1}(p_1, \dots, p_{k+1}) = \frac{\Gamma\left(\sum_{j=1}^{k+1} \alpha_j\right)}{\prod_{j=1}^{k+1} \Gamma(\alpha_j)} \prod_{j=1}^{k+1} p_j^{\alpha_j - 1}, \tag{5}
$$

with the constraint $\sum_{j=1}^{k+1} p_j = 1$ and parameters $\alpha_1, \dots, \alpha_{k+1} > 0$.

It is then possible to compute exactly the joint default probabilities in equation (4):

$$\begin{aligned}
&\mathbb{P}[N_1 = l_1, \ldots, N_k = l_k] \\
&= \mathbb{E}\Bigg[\binom{n_1}{l_1} P_1^{l_1}(1-P_1)^{n_1-l_1}\binom{n_2}{l_2}(P_1+P_2)^{l_2}(1-P_1-P_2)^{n_2-l_2}\cdots \\
&\quad \cdots\binom{n_k}{l_k}(P_1+\cdots+P_k)^{l_k}(1-P_1-\cdots-P_k)^{n_k-l_k}\Bigg] \\
&= \binom{n_1}{l_1}\cdots\binom{n_k}{l_k}\int_0^1\int_0^{1-p_1}\cdots\int_0^{1-p_1-\cdots-p_{k-1}} p_1^{l_1}(1-p_1)^{n_1-l_1}(p_1+p_2)^{l_2}(1-p_1-p_2)^{n_2-l_2} \\
&\quad \times\cdots\times(p_1+p_2+\cdots+p_k)^{l_k}(1-p_1-p_2-\cdots-p_k)^{n_k-l_k} \\
&\quad \times\frac{\Gamma(\sum_{j=1}^{k+1}\alpha_j)}{\prod_{j=1}^{k+1}\Gamma(\alpha_j)}p_1^{\alpha_1-1}p_2^{\alpha_2-1}\cdots p_k^{\alpha_k-1}(1-p_1-p_2-\cdots-p_k)^{\alpha_{k+1}-1}dp_k dp_{k-1}\cdots dp_1 \\
&= \binom{n_1}{l_1}\cdots\binom{n_k}{l_k}\frac{\Gamma(\sum_{j=1}^{k+1}\alpha_j)}{\prod_{j=1}^{k+1}\Gamma(\alpha_j)}\sum_{j_1=0}^{l_k}\binom{l_k}{j_1}B(\alpha_k+j_1,\alpha_{k+1}+n_k-l_k) \\
&\quad \times\sum_{j_2=0}^{l_{k-1}+l_k-j_1}\binom{l_{k-1}+l_k-j_1}{j_2}B(\alpha_{k-1}+j_2,\alpha_k+\alpha_{k+1}+n_{k-1}+n_k-l_{k-1}-l_k+j_1) \\
&\quad \cdots\times\sum_{j_{k-1}=0}^{\substack{l_2+\cdots+l_k\\-j_1-\cdots-j_{k-2}}}\binom{l_2+\cdots+l_k-j_1-\cdots-j_{k-2}}{j_{k-1}} \\
&\quad \times B(\alpha_2+j_{k-1},\alpha_3+\cdots+\alpha_{k+1}+n_2+\cdots+n_k-l_2-\cdots-l_k+j_1+\cdots+j_{k-2}) \\
&\quad \times B(\alpha_1+l_1+\cdots+l_k-j_1-\cdots-j_{k-1},\alpha_2+\cdots+\alpha_{k+1} \\
&\qquad +n_1+\cdots+n_k-l_1-\cdots-l_k+j_1\cdots+j_{k-1}),
\end{aligned} \tag{6}$$

where $j_{-1} = j_0 = 0$.

We remark here that even if an analytic expression of the joint default probabilities is obtained, the number of terms involved in its calculation is very large and this will cause numerical problems in estimation of the parameters involved. This is the reason that will address us to the use of the EM algorithm.

3 Iterative urn scheme for defaults

This second model is inspirated by an iterative urn scheme: the number of defaults in the best rating group is determined with a Pólya urn scheme, hence its random default frequency is $P_1 \sim beta(\alpha_1, \beta_1)$. The number of defaults in the worse ratings are then determined by the number of firms that would have defaulted in the next better rating plus a certain part of the group that would have survived in the next better rating, and this additional part is determined again via Pólya's urn scheme. This allows again for dependence both between defaults within the rating groups, given the use of the Pólya's urn scheme, and for monotone dependence between the defaults of the different rating groups, given by the way in which the default frequencies are built. Formally the model is built as follows: we start from a random vector $\tilde{\mathbf{P}} = (\tilde{P}_1, \dots, \tilde{P}_k)$ where $\tilde{P}_1 \sim beta(\alpha_1, \beta_1), \dots, \tilde{P}_k \sim beta(\alpha_k, \beta_k)$ are independent, and we define the random default probabilities in the following way:

$$\begin{aligned} P_1 &= \tilde{P}_1 \\ P_2 &= P_1 + (1 - P_1)\tilde{P}_2 \\ &\dots \\ P_k &= P_{k-1} + (1 - P_{k-1})\tilde{P}_k. \end{aligned} \tag{7}$$

To calculate the joint distribution of the number of defaults it is useful to see how things work for the marginal distributions. Starting from the group of n_1 firms with the best rating r_1, from the Pólya urn scheme we have:

$$\mathbb{P}[N_1 = l_1] = \binom{n_1}{l_1} \frac{B(\alpha_1 + l_1, \beta_1 + n_1 - l_1)}{B(\alpha_1, \beta_1)}, \qquad 0 \le l_1 \le n_1$$

The joint probability of l_1 defaults in the group of firms with credit rating r_1 and l_2 defaults in the group of firms with credit rating r_2 can be calculated using the binomial expansion for the power of a sum and the fact that $1 - (x + (1 - x)y) = (1 - x)(1 - y)$. Hence we have:

$$\begin{aligned} &\mathbb{P}[N_1 = l_1, N_2 = l_2] = \mathbb{E}[\mathbb{P}[N_1 = l_1, N_2 = l_2 | P_1, P_2]] \\ &= \mathbb{E}\left[\binom{n_1}{l_1} P_1^{l_1}(1 - P_1)^{n_1 - l_1} \binom{n_2}{l_2} P_2^{l_2}(1 - P_2)^{n_2 - l_2}\right] \\ &= \binom{n_1}{l_1}\binom{n_2}{l_2} \mathbb{E}\left[\tilde{P}_1^{l_1}(1 - \tilde{P}_1)^{n_1 - l_1}(\tilde{P}_1 + (1 - \tilde{P}_1)\tilde{P}_2)^{l_2}(1 - (\tilde{P}_1 + (1 - \tilde{P}_1)\tilde{P}_2))^{n_2 - l_2}\right]. \end{aligned} \tag{8}$$

Since

$$\begin{aligned} &\mathbb{E}\left[\tilde{P}_1^{l_1}(1 - \tilde{P}_1)^{n_1 - l_1}(\tilde{P}_1 + (1 - \tilde{P}_1)\tilde{P}_2)^{l_2}(1 - (\tilde{P}_1 + (1 - \tilde{P}_1)\tilde{P}_2))^{n_2 - l_2}\right] \\ &= \mathbb{E}\left[\tilde{P}_1^{l_1}(1 - \tilde{P}_1)^{n_1 - l_1}\left(\sum_{j=0}^{l_2}\binom{l_2}{j}\tilde{P}_1^{j}(1 - \tilde{P}_1)^{l_2 - j}\tilde{P}_2^{l_2 - j}\right)(1 - \tilde{P}_1)^{n_2 - l_2}(1 - \tilde{P}_2)^{n_2 - l_2}\right] \\ &= \sum_{j=0}^{l_2}\binom{l_2}{j}\int_0^1 \tilde{p}_1^{l_1 + j}(1 - \tilde{p}_1)^{n_1 + n_2 - l_1 - j} f_{\tilde{P}_1}(\tilde{p}_1)d\tilde{p}_1 \int_0^1 \tilde{p}_2^{l_2 - j}(1 - \tilde{p}_2)^{n_2 - l_2} f_{\tilde{P}_2}(\tilde{p}_2)d\tilde{p}_2 \end{aligned} \tag{9}$$

we get at the end

$$\begin{aligned} &\mathbb{P}[N_1 = l_1, N_2 = l_2] = \\ &= \binom{n_1}{l_1}\binom{n_2}{l_2}\sum_{j=0}^{l_2}\binom{l_2}{j}\frac{B(\alpha_1 + l_1 + j, \beta_1 + n_1 + n_2 - l_1 - j)}{B(\alpha_1, \beta_1)}\frac{B(\alpha_2 + l_2 - j, \beta_2 + n_2 - l_2)}{B(\alpha_2, \beta_2)}. \end{aligned} \tag{10}$$

It is clear now how, through successive iterations of the procedure seen above, it is possible to compute the general formula for the joint default probabilities:

$$
\begin{aligned}
&\mathbb{P}[N_1 = l_1, \dots, N_k = l_k] = \mathbb{E}[\mathbb{P}[N_1 = l_1, \dots, N_k = l_k | P_1, \dots, P_k]] \\
&= \mathbb{E}\left[\binom{n_1}{l_1} P_1^{l_1}(1-P_1)^{n_1-l_1}\binom{n_2}{l_2} P_2^{l_2}(1-P_2)^{n_2-l_2}\cdots\binom{n_k}{l_k} P_k^{l_k}(1-P_k)^{n_k-l_k}\right] \\
&= \mathbb{E}\left[\binom{n_1}{l_1}\tilde{P}_1^{l_1}(1-\tilde{P}_1)^{n_1-l_1}\binom{n_2}{l_2}(\tilde{P}_1+(1-\tilde{P}_1)\tilde{P}_2)^{l_2}(1-\tilde{P}_1)^{n_2-l_2}(1-\tilde{P}_2)^{n_2-l_2}\cdots\right] \\
&= \binom{n_1}{l_1}\cdots\binom{n_k}{l_k}\frac{1}{B(\alpha_1,\beta_1)\cdots B(\alpha_k,\beta_k)} \times \sum_{j_1=0}^{l_k}\binom{l_k}{j_1} B(\alpha_k + l_k - j_1, \beta_k + n_k - l_k) \\
&\cdots \times \sum_{j_i=0}^{l_{k+1-i}+j_{i-1}} \binom{l_{k+1-i}+j_{i-1}}{j_i} \\
&\qquad \times B(\alpha_{k+1-i} + l_{k+1-i} + j_{i-1} - j_i, \beta_{k+1-i} + n_k + \dots + n_{k+1-i} - l_{k+1-i} - j_{i-1}) \\
&\cdots \times \sum_{j_{k-1}=0}^{l_2+j_{k-2}} \binom{l_2+j_{k-2}}{j_{k-1}} \\
&\quad \times B(\alpha_2 + l_2 + j_{k-2} - j_{k-1}, \beta_2 + n_k + \dots + n_2 - l_2 - j_{k-2}) \\
&\quad \times B(\alpha_1 + l_1 + j_{k-1}, \beta_1 + n_k + \dots + n_1 - l_1 - j_{k-1})
\end{aligned} \tag{11}
$$

We can do here the same remarks as before about computational complexity in the maximum likelihood estimation of the parameters involved because of the large number of terms present in this analytic expression.

3.1 The Iterative Urn Scheme and the Generalized Dirichlet Distribution

As we have just seen to calculate the joint default probabilities of the number of defaults it is not necessary to know the joint distribution of $P_1, \dots, P_k$ since we can reduce to work with the $\tilde{P}_j$ that are independent. However to exploit the similarities between this model and the previous one, it has been shown in [8] that introducing the new random variables $Q_1 = P_1$ and $Q_i = P_i - P_{i-1} = (1 - P_{i-1})\tilde{P}_i$ for $i = 2, \dots, k$, it is possible to write the default probabilities as $P_i = Q_1 + \dots + Q_i$ so that the default probabilities for the iterative urn scheme can be written again as:

$$
\begin{aligned}
&\mathbb{P}[N_1 = l_1, \dots, N_k = l_k] \\
&\qquad = \mathbb{E}\Bigg[\binom{n_1}{l_1} Q_1^{l_1}(1-Q_1)^{n_1-l_1}\binom{n_2}{l_2}(Q_1+Q_2)^{l_2}(1-Q_1-Q_2)^{n_2-l_2} \\
&\qquad\quad \cdots \binom{n_k}{l_k}(Q_1+\dots+Q_k)^{l_k}(1-Q_1-\dots-Q_k)^{n_k-l_k}\Bigg]
\end{aligned} \tag{12}
$$

Since

$$
\begin{aligned}
Q_1 &= \tilde{P}_1 \\
Q_i &= \tilde{P}_i \left(\prod_{j=1}^{i-1} (1 - \tilde{P}_j) \right) \qquad \text{for } i = 2, \dots, k
\end{aligned}
$$

it is then possible to prove (see [1]) that $(Q_1, \dots, Q_k)$ has a generalized Dirichlet distribution $GD_k(\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k)$ with density

$$
f(q_1, \dots, q_k) = \frac{1}{\prod_{i=1}^{k} B(\alpha_i, \beta_i)} \left(1 - \sum_{i=1}^{k} q_i\right)^{\beta_k - 1} \prod_{i=1}^{k} \left[q_i^{\alpha_i - 1} \left(1 - \sum_{j=0}^{i-1} q_j\right)^{\beta_{i-1} - (\alpha_i + \beta_i)} \right]
$$

for $(q_1, \dots, q_k) \in [0,1]^k$ with the constraint $\sum_{i=1}^{k} q_k \leq 1$, where $q_0 = 0$ and β_0 is arbitrary.

Since for the special choice of the parameters $\beta_{j-1} = \alpha_j + \beta_j$ for $j = k, k-1, \dots, 2$ with β_k arbitrarily chosen it holds that $(Q_1, \dots, Q_k) \sim D_{k+1}(\alpha_1, \dots, \alpha_k, \beta_k)$, this allows to conclude that the multidimensional urn scheme, described in the previous section is embedded in the iterative urn scheme model.

4 EM algorithm

The Expectation-Maximization algorithm is a tool for the iterative computation of maximum-likelihood estimates that is very useful to apply in situations where the estimation of ML can be simplified by (artificially) considering the observed data as incomplete data, so that the complete-data likelihood has a nice form and the complexity of the estimation can be reduced with respect to the one required by the incomplete-data likelihood. In the following we will briefly describe the theory behind the algorithm and some of its properties; more details can be found in [3], [10] and [6].

Let $\mathbf{Y}$ be the random vector corresponding to the observed (incomplete) data $\mathbf{y} \in \mathcal{Y}$ and suppose that it has density $g(\mathbf{y}\,; \boldsymbol{\phi})$, where $\boldsymbol{\phi}$ is a vector of unknown parameters to determine in the parameter space $\boldsymbol{\Phi}$. Let $\mathbf{x} \in \mathcal{X}$ be the vector of the (augmented) complete data, and let $f(\mathbf{x}\,; \boldsymbol{\phi})$ be the density of the random vector $\mathbf{X}$ corresponding to $\mathbf{x}$. We assume that there is a mapping $\pi : \mathcal{X} \mapsto Y$ that express the fact that we don't observe directly $\mathbf{x}$, but, for an observed $\mathbf{y}$, we have a subset $\mathcal{X}(\mathbf{y}) = \{\mathbf{x} \in \mathcal{X} \mid \pi(\mathbf{x}) = \mathbf{y}\}$ of possible outcomes of the complete data. Then the relation between the incomplete-data and the complete data densities is given by:

$$
g(\mathbf{y}\,; \boldsymbol{\phi}) = \int_{\mathcal{X}(\mathbf{y})} f(\mathbf{x}\,; \boldsymbol{\phi})\, d\mathbf{x}. \tag{13}
$$

The goal of the EM algorithm is to find a value of $\boldsymbol{\phi}$ which maximizes $L(\boldsymbol{\phi}) := g(\mathbf{y}\,; \boldsymbol{\phi})$ using iteratively the complete data log-likelihood function $\log L_c(\boldsymbol{\phi}) := \log f(\mathbf{x}\,; \boldsymbol{\phi})$. Since $\log L_c(\boldsymbol{\phi})$ is not determined because we don't have $\mathbf{x}$, we replace it with its conditional expectation given the observed data $\mathbf{y}$, and using an initial value for the parameter $\boldsymbol{\phi}$.

More precisely, suppose that $\boldsymbol{\phi}^{(p)}$ denotes the current values of the parameters after p iterations of the algorithm. Then the $p+1$ values are computed using the following two steps:

Expectation-step: compute

$$Q(\boldsymbol{\phi}, \boldsymbol{\phi}^{(p)}) := \mathbb{E}[\log L_c(\boldsymbol{\phi}) \,|\, \mathbf{y}\,; \boldsymbol{\phi}^{(p)}] = \mathbb{E}[\log f(\mathbf{X}\,; \boldsymbol{\phi}) \,|\, \mathbf{y}\,; \boldsymbol{\phi}^{(p)}] \tag{14}$$

Maximization-step: choose $\boldsymbol{\phi}^{(p+1)}$ such that:

$$Q(\boldsymbol{\phi}^{(p+1)}, \boldsymbol{\phi}^{(p)}) \geq Q(\boldsymbol{\phi}, \boldsymbol{\phi}^{(p)}) \quad \forall \boldsymbol{\phi} \in \Phi. \tag{15}$$

Proposition 4.1. *The sequence* $\{L(\boldsymbol{\phi}^{(k)})\}_{k\geq 0}$ *is monotone increasing.*

Proof. The conditional density of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$ is:

$$k(\mathbf{x} \,|\, \mathbf{y}\,; \boldsymbol{\phi}) := \frac{f(\mathbf{x}\,; \boldsymbol{\phi})}{g(\mathbf{y}\,; \boldsymbol{\phi})}$$

So if we introduce $H(\boldsymbol{\phi}', \boldsymbol{\phi}) := \mathbb{E}[\log k(\mathbf{X} \,|\, \mathbf{y}\,; \boldsymbol{\phi}') \,|\, \mathbf{y}\,; \boldsymbol{\phi}]$ we have, for all $\boldsymbol{\phi}, \boldsymbol{\phi}' \in \boldsymbol{\Phi}$:

$$H(\boldsymbol{\phi}', \boldsymbol{\phi}) = \mathbb{E}[\log f(\mathbf{X}\,; \boldsymbol{\phi}') \,|\, \mathbf{y}\,; \boldsymbol{\phi}] - \mathbb{E}[\log g(\mathbf{y}\,; \boldsymbol{\phi}') \,|\, \mathbf{y}\,; \boldsymbol{\phi}] = Q(\boldsymbol{\phi}', \boldsymbol{\phi}) - \log L(\boldsymbol{\phi}')$$

and so

$$\begin{aligned}\log L(\boldsymbol{\phi}^{(k+1)}) - \log L(\boldsymbol{\phi}^{(k)}) = {} & \underbrace{[Q(\boldsymbol{\phi}^{(k+1)}, \boldsymbol{\phi}^{(k)}) - Q(\boldsymbol{\phi}^{(k)}, \boldsymbol{\phi}^{(k)})]}_{\geq 0 \text{ by M-step}} \\ & - [H(\boldsymbol{\phi}^{(k+1)}, \boldsymbol{\phi}^{(k)}) - H(\boldsymbol{\phi}^{(k)}, \boldsymbol{\phi}^{(k)})].\end{aligned}$$

Now observe that for any $\boldsymbol{\phi}, \boldsymbol{\phi}' \in \boldsymbol{\Phi}$:

$$\begin{aligned}& H(\boldsymbol{\phi}', \boldsymbol{\phi}) - H(\boldsymbol{\phi}, \boldsymbol{\phi}) \\ & = \mathbb{E}\left[\log \frac{k(\mathbf{X} \,|\, \mathbf{y}\,; \boldsymbol{\phi}')}{k(\mathbf{X} \,|\, \mathbf{y}\,; \boldsymbol{\phi})} \;\middle|\; \mathbf{y}\,; \boldsymbol{\phi}\right] \\ & \leq \log \mathbb{E}\left[\frac{k(\mathbf{X} \,|\, \mathbf{y}\,; \boldsymbol{\phi}')}{k(\mathbf{X} \,|\, \mathbf{y}\,; \boldsymbol{\phi})} \;\middle|\; \mathbf{y}\,; \boldsymbol{\phi}\right] \\ & = \log \int_{\mathcal{X}(\mathbf{y})} k(\mathbf{x} \,|\, \mathbf{y}\,; \boldsymbol{\phi}')\, d\mathbf{x} \\ & = 0\end{aligned}$$

and we have completed the proof (using Jensen inequality). □

So if $\{L(\boldsymbol{\phi}^{(k)})\}_{k\geq 0}$ is bounded, we have that $L(\boldsymbol{\phi}^{(k)}) \uparrow L^*$. Which conditions ensure that $\{L(\boldsymbol{\phi}^{(k)})\}_{k\geq 0}$ is bounded, and is L^* a global or local maximum of $L(\boldsymbol{\phi})$, or only a stationary point? The main result to this questions is the following theorem (see [10, Theorem 2]):

Theorem 4.2. *Suppose the following conditions hold:*

i) $\mathbf{\Phi} \subseteq \mathbb{R}^d$*;*

ii) $\mathbf{\Phi}_{\boldsymbol{\phi}_0} := \{\boldsymbol{\phi} \in \mathbf{\Phi} : L(\boldsymbol{\phi}) \geq L(\boldsymbol{\phi}_0)\} \subseteq int(\mathbf{\Phi})$ *and is compact for any* $L(\boldsymbol{\phi}_0) > -\infty$*;*

iii) L *is continuous in* $\mathbf{\Phi}$ *and differentiable in* $int(\mathbf{\Phi})$*.*

Then $\{L(\boldsymbol{\phi}^{(k)})\}_{k\geq 0}$ *is bounded above for any* $\boldsymbol{\phi}_0 \in \mathbf{\Phi}$*.*

If in addition $Q(\boldsymbol{\phi}', \boldsymbol{\phi})$ *is continuous in both* $\boldsymbol{\phi}'$ *and* $\boldsymbol{\phi}$*, then* $L(\boldsymbol{\phi}) \uparrow L^* = L(\boldsymbol{\phi}^*)$ *for some stationary point* $\boldsymbol{\phi}^*$*.*

Convergence to local maxima of $L(\boldsymbol{\phi})$ can be obtained with more restrictive conditions; see again [10] for details. A special case is the following:

Theorem 4.3. *Suppose that* $L(\boldsymbol{\phi})$ *is unimodal in* $\mathbf{\Phi}$ *with* $\boldsymbol{\phi}^*$ *being the only stationary point, and that* $\frac{\partial Q(\boldsymbol{\phi}', \boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}$ *is continuous in* $\boldsymbol{\phi}'$ *and* $\boldsymbol{\phi}$*. Then* $\{\boldsymbol{\phi}^{(k)}\}_{k\geq 0}$ *converges to the unique maximizer* $\boldsymbol{\phi}^*$ *of* $L(\boldsymbol{\phi})$*.*

4.1 The regular exponential family case

In the case that $f(\mathbf{x}\,; \boldsymbol{\phi})$ has the regular exponential family form

$$f(\mathbf{x}\,; \boldsymbol{\phi}) = b(\mathbf{x}) \exp(\boldsymbol{\phi}^\intercal \mathbf{t}(\mathbf{x}))/a(\boldsymbol{\phi})$$

where $\mathbf{t}(\mathbf{x})$ is a vector of complete-data sufficient statistics, $b(\mathbf{x})$ and $a(\boldsymbol{\phi})$ are scalar functions, and $\mathbf{\Phi} = \{\boldsymbol{\phi} \in \mathbb{R}^d : \int_{\mathcal{X}} b(\mathbf{x}) \exp(\boldsymbol{\phi}^\intercal \mathbf{t}(\mathbf{x}))\, d\mathbf{x} < +\infty\}$, it is possible then to simplify the Expectation and Maximization steps. In fact observe that:

$$\begin{aligned}
&\mathbb{E}[\mathbf{t}(\mathbf{X}); \boldsymbol{\phi}] \\
&= \frac{1}{a(\boldsymbol{\phi})} \int_{\mathcal{X}} b(\mathbf{x}) \exp(\boldsymbol{\phi}^\intercal \mathbf{t}(\mathbf{x}))\mathbf{t}(\mathbf{x})\, d\mathbf{x} \\
&= \frac{1}{a(\boldsymbol{\phi})} \int_{\mathcal{X}} b(\mathbf{x}) \frac{\partial}{\partial \boldsymbol{\phi}} \exp(\boldsymbol{\phi}^\intercal \mathbf{t}(\mathbf{x}))\, d\mathbf{x} \\
&= \frac{1}{a(\boldsymbol{\phi})} \frac{\partial}{\partial \boldsymbol{\phi}} a(\boldsymbol{\phi}) \\
&= \frac{\partial}{\partial \boldsymbol{\phi}} \log a(\boldsymbol{\phi}).
\end{aligned} \tag{16}$$

Now assume that $\boldsymbol{\phi}^{(p)}$ denotes the current values of the parameters after p iterations of the algorithm. To compute the $p+1$ values we have to maximize w.r.t $\boldsymbol{\phi}$ the function $Q(\boldsymbol{\phi}, \boldsymbol{\phi}^{(p)})$ that in this case is given by:

$$\begin{aligned}
Q(\boldsymbol{\phi}, \boldsymbol{\phi}^{(p)}) &= \mathbb{E}\left[\log\left(\frac{b(\mathbf{X}) \exp(\boldsymbol{\phi}^\intercal \mathbf{t}(\mathbf{X}))}{a(\boldsymbol{\phi})} \right) \middle| \, \mathbf{y}; \boldsymbol{\phi}^{(p)} \right] \\
&= \underbrace{\mathbb{E}[\log b(\mathbf{X}) |\, \mathbf{y}; \boldsymbol{\phi}^{(p)}]}_{\text{independent from } \boldsymbol{\phi}} + \mathbb{E}[\boldsymbol{\phi}^\intercal \mathbf{t}(\mathbf{X}) |\, \mathbf{y}; \boldsymbol{\phi}^{(p)}] - \mathbb{E}[\log a(\boldsymbol{\phi}) |\, \mathbf{y}; \boldsymbol{\phi}^{(p)}] \\
&= \mathbb{E}[\log b(\mathbf{X}) |\, \mathbf{y}; \boldsymbol{\phi}^{(p)}] + \boldsymbol{\phi}^\intercal \mathbb{E}[\mathbf{t}(\mathbf{X}) |\, \mathbf{y}; \boldsymbol{\phi}^{(p)}] - \log a(\boldsymbol{\phi}).
\end{aligned} \tag{17}$$

Then we see, by differentiating (17) taking into account (16), that maximization is equivalent to solve for $\boldsymbol{\phi}$ the system of equations

$$\mathbb{E}[\mathbf{t}(\mathbf{X})|\,\mathbf{y};\boldsymbol{\phi}^{(p)}] = \mathbb{E}[\mathbf{t}(\mathbf{X});\boldsymbol{\phi}] \tag{18}$$

The two steps of the algorithm become then:
Expectation-step: estimate $\mathbf{t}^{(p)}$ by:

$$\mathbf{t}^{(p)} = \mathbb{E}[\mathbf{t}(\mathbf{X})|\,\mathbf{y};\boldsymbol{\phi}^{(p)}]. \tag{19}$$

Maximization-step: determine $\boldsymbol{\phi}^{(p+1)}$ solving the equation:

$$\mathbb{E}[\mathbf{t}(\mathbf{X})|\,\boldsymbol{\phi}] = \mathbf{t}^{(p)}. \tag{20}$$

4.2 Application to multidimensional urn scheme

Following the notation of the the previous section, we consider our observed data given by $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_m)$, where $\mathbf{y}_i = (l_{i1}, \dots, l_{ik})$ for $i = 1, \dots, m$ and m is the total number of observations. As complete data we take $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$, where $\mathbf{x}_i = (p_{i1}, \dots, p_{ik}, l_{i1}, \dots, l_{ik})$ with p_{ij} the i-th (unknown) realization of the random variable P_j. Since our goal is the MLE of the parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{k+1})$ of $g(\mathbf{y}\,;\boldsymbol{\alpha}) = \prod_{i=1}^m g_i(\mathbf{y}_i\,;\boldsymbol{\alpha})$ where g_i is given by equation (6)[1], to fit our situation in the EM algorithm, we consider as sampling density $f(\mathbf{x}\,;\boldsymbol{\alpha}) = \prod_{i=1}^m f_i(\mathbf{x}_i\,;\boldsymbol{\alpha})$ where

$$\begin{aligned} f_i(\mathbf{x}_i\,;\boldsymbol{\alpha}) &= \frac{\Gamma(\sum_{j=1}^{k+1}\alpha_j)}{\prod_{j=1}^{k+1}\Gamma(\alpha_j)} p_{i1}^{\alpha_1-1} p_{i2}^{\alpha_2-1}\cdots p_{ik}^{\alpha_k-1}(1-p_{i1}-p_{i2}-\cdots-p_{ik})^{\alpha_{k+1}-1} \\ &\times \binom{n_{i1}}{l_{i1}}\cdots\binom{n_{ik}}{l_{ik}} p_{i1}^{l_{i1}}(1-p_{i1})^{n_{i1}-l_{i1}}\cdots(p_{i1}+\dots p_{ik})^{l_{ik}}(1-p_{i1}-\cdots-p_{ik})^{n_{ik}-l_{ik}} \end{aligned} \tag{21}$$

We can so immediately observe that f belongs to the regular exponential family: in fact we can write:

$$\begin{aligned} &\log(f_i(\mathbf{x}_i\,;\boldsymbol{\alpha}) = \\ &\log\left[\frac{\Gamma(\sum_{j=1}^{k+1}\alpha_j)}{\prod_{j=1}^{k+1}\Gamma(\alpha_j)}\right] \\ &\quad \underbrace{+\log\left[\binom{n_{i1}}{l_{i1}}\cdots\binom{n_{ik}}{l_{ik}}\frac{p_{i1}^{l_{i1}}(1-p_{i1})^{n_{i1}-l_{i1}}\cdots(p_{i1}+\dots p_{ik})^{l_{ik}}(1-p_{i1}-\cdots-p_{ik})^{n_{ik}-l_{ik}}}{p_{i1}\cdots p_{ik}(1-p_{i1}-\cdots-p_{ik})}\right]}_{b_i(p_{i1},\cdots,p_{ik},l_{i1},\cdots,l_{ik})} \\ &\quad +(\alpha_1,\dots,\alpha_{k+1})(\log p_{i1},\dots,\log p_{ik},\log(1-p_{i1}-\cdots-p_{ik}))^\intercal \end{aligned}$$

[1] We are assuming here that our data are realizations of independent identically distributed random vectors $(N_1, \dots, N_k)$.

So we have that:

$$
\begin{aligned}
&f(\mathbf{x}\,;\boldsymbol{\alpha}) = \exp\left(\sum_{i=1}^{m} \log f_i(\mathbf{x}_i\,;\boldsymbol{\alpha})\right)\\
&=\left(\frac{\Gamma(\sum_{j=1}^{k+1}\alpha_j)}{\prod_{j=1}^{k+1}\Gamma(\alpha_j)}\right)^m \exp\left[\sum_{i=1}^{m} b_i(p_{i1},\dots,p_{im},l_{i1},\dots,l_{im})\right]\\
&\times\exp\left[(\alpha_1,\dots,\alpha_{k+1})\left(\sum_{i=1}^{m}\log p_{i1},\dots,\sum_{i=1}^{m}\log p_{ik},\sum_{i=1}^{m}\log(1-p_{i1}-\cdots-p_{ik})\right)^{\intercal}\right]
\end{aligned}
$$

so that the statistics of our interest are given by $t_j(\mathbf{x}) = \sum_{i=1}^{m}\log p_{ij}$ for $j = 1,\dots,k$ and $t_{k+1}(\mathbf{x}) = \sum_{i=1}^{m}\log(1-p_{i1}-\cdots-p_{ik})$.

For the E-step of the algorithm we have then to compute the conditional expectations $\mathbb{E}[t_j(\mathbf{x})|\mathbf{y},(\alpha_1,\dots,\alpha_{k+1})] = \sum_{i=1}^{m}\mathbb{E}[\log p_{ij}|\mathbf{y},(\alpha_1,\dots,\alpha_{k+1})]$ for $j = 1,\dots,k$ and $\mathbb{E}[t_{k+1}(\mathbf{x})|\mathbf{y},(\alpha_1,\dots,\alpha_{k+1})] = \sum_{i=1}^{m}\mathbb{E}[\log(1-p_{i1}-\cdots-p_{ik})|\mathbf{y},(\alpha_1,\dots,\alpha_{k+1})]$

The conditional density in our case is given by

$$
\begin{aligned}
&k(\mathbf{x}|\mathbf{y},(\alpha_1,\dots,\alpha_{k+1})) = \frac{f(\mathbf{x}|(\alpha_1,\dots,\alpha_{k+1}))}{g(\mathbf{y}|(\alpha_1,\dots,\alpha_{k+1}))} = \prod_{i=1}^{m}\frac{f_i(\mathbf{x}|(\alpha_1,\dots,\alpha_{k+1})}{g_i(\mathbf{y}|(\alpha_1,\dots,\alpha_{k+1})}\\
&=\prod_{i=1}^{m}\frac{k_i(\mathbf{x}|(\alpha_1,\dots,\alpha_{k+1}))}{\int_0^1\int_0^{1-p_{i1}}\cdots\int_0^{1-p_{i1}-\cdots-p_{i,k-1}} k_i(\mathbf{x}|(\alpha_1,\dots,\alpha_{k+1}))dp_{ik}\cdots dp_{i1}}
\end{aligned}
$$

where

$$
k_i(\mathbf{x}|(\alpha_1,\dots,\alpha_{k+1})) = p_{i1}^{\alpha_1-1}\cdots(1-p_{i1}-\cdots-p_{ik})^{\alpha_{k+1}-1}p_{i1}^{l_{i1}}\cdots(1-p_{i1}-\dots p_{ik})^{n_{ik}-l_{ik}}.
$$

Hence we obtain:

$$
\begin{aligned}
&\mathbb{E}[\log p_{ij}|\mathbf{y},(\alpha_1,\dots,\alpha_{k+1})] =\\
&=\frac{\int_0^1\int_0^{1-p_{i1}}\cdots\int_0^{1-p_{i1}-\cdots-p_{i,k-1}}\log p_{ij}k_i(\mathbf{x}|(\alpha_1,\dots,\alpha_{k+1}))dp_{ik}\cdots dp_{i1}}{\int_0^1\int_0^{1-p_{i1}}\cdots\int_0^{1-p_{i1}-\cdots-p_{i,k-1}} k_i(\mathbf{x}|(\alpha_1,\dots,\alpha_{k+1}))dp_{ik}\cdots dp_{i1}}
\end{aligned}
\tag{22}
$$

To see how things work in practice, we consider the case $k = 3$. We use the facts, analogous to what we used in equation (6), that:

$$
\begin{aligned}
&\int_0^{1-a}\log p\; p^{k-1}(a+p)^m(1-a-p)^{n-1}dp\\
&=\sum_{j=1}^{m}\binom{m}{j}(1-a)^{k+j+n-1}a^{m-j}B(k+j,n)\,[\log(1-a)+D(k+j,n)]
\end{aligned}
\tag{23}
$$

and

$$
\begin{aligned}
&\int_0^{1-a}\log(1-a-p)\; p^{k-1}(a+p)^m(1-a-p)^{n-1}dp\\
&=\sum_{j=1}^{m}\binom{m}{j}(1-a)^{k+j+n-1}a^{m-j}B(k+j,n)\,[\log(1-a)+D(n,k+j)]
\end{aligned}
\tag{24}
$$

where $D(\alpha,\beta) \coloneqq \dfrac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \dfrac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)}$. Since we have already computed the denominator of (22) when we have computed the joint default probability in (6) we can proceed in the same way to compute the numerator; we do it explicitly for $j = 3$ (and suppress index i for ease of notation):

$$
\begin{aligned}
&\int_0^1 \int_0^{1-p_1} \int_0^{1-p_1-p_2} \log p_3 \; p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_3^{\alpha_3-1} (1-p_1-p_2-p_3)^{\alpha_4-1} \\
&\times p_1^{l_1}(1-p_1)^{n_1-l_1}(p_1+p_2)^{l_2}(1-p_1-p_2)^{n_2-l_2}(p_1+p_2+p_3)^{l_3}(1-p_1-p_2-p_3)^{n_3-l_3} dp_3 dp_2 dp_1 \\
&= \int_0^1 \int_0^{1-p_1} p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_1^{l_1}(1-p_1)^{n_1-l_1}(p_1+p_2)^{l_2}(1-p_1-p_2)^{n_2-l_2} \\
&\times \sum_{j_1=0}^{l_3} \binom{l_3}{j_1} (p_1+p_2)^{l_3-j_1}(1-p_1-p_2)^{\alpha_3+\alpha_4+n_3-l_3+j_1-1} \\
&\times B(\alpha_3+j_1, \alpha_4+n_3-l_3) D(\alpha_3+j_1, \alpha_4+n_3-l_3) dp_1 dp_2 \\
&+ \int_0^1 \int_0^{1-p_1} p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_1^{l_1}(1-p_1)^{n_1-l_1}(p_1+p_2)^{l_2}(1-p_1-p_2)^{n_2-l_2} \\
&\times \sum_{j_1=0}^{l_3} \binom{l_3}{j_1} (p_1+p_2)^{l_3-j_1}(1-p_1-p_2)^{\alpha_3+\alpha_4+n_3-l_3+j_1-1} \\
&\times B(\alpha_3+j_1, \alpha_4+n_3-l_3) \log(1-p_1-p_2) dp_1 dp_2
\end{aligned} \tag{25}
$$

We see that the first summand leads to the same result as in (6) with the additional term $D(\alpha_3 + j_1, \alpha_4 + n_3 - l_3)$, while for the second in evaluating the integral with respect to p_2 we get that it splits again in a sum of two terms, and at the end we get:

$$
\begin{aligned}
&\int_0^1 \int_0^{1-p_1} \int_0^{1-p_1-p_2} \log p_3 \; p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_3^{\alpha_3-1} (1-p_1-p_2-p_3)^{\alpha_4-1} \\
&\times p_1^{l_1}(1-p_1)^{n_1-l_1}(p_1+p_2)^{l_2}(1-p_1-p_2)^{n_2-l_2}(p_1+p_2+p_3)^{l_3}(1-p_1-p_2-p_3)^{n_3-l_3} dp_3 dp_2 dp_1 \\
&= A_1 + A_2 + A_3
\end{aligned} \tag{26}
$$

where

$$
\begin{aligned}
A_1 = &\sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3+j_1, \alpha_4+n_3-l_3) D(\alpha_3+j_1, \alpha_4+n_3-l_3) \\
&\times \sum_{j_2=0}^{l_2+l_3-j_1} \binom{l_2+l_3-j_1}{j_2} B(\alpha_2+j_2, \alpha_3+\alpha_4+n_2+n_3-l_2-l_3+j_1) \\
&\times B(\alpha_1+l_1+l_2+l_3-j_1-j_2, \alpha_2+\alpha_3+\alpha_4+n_1-l_1+n_2-l_2+n_3-l_3+j_1+j_2)
\end{aligned}
$$

$$
\begin{aligned}
A_2 = &\sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3 + j_1, \alpha_4 + n_3 - l_3) \\
&\times \sum_{j_2=0}^{l_2+l_3-j_1} \binom{l_2 + l_3 - j_1}{j_2} B(\alpha_2 + j_2, \alpha_3 + \alpha_4 + n_2 + n_3 - l_2 - l_3 + j_1) \\
&\times D(\alpha_3 + \alpha_4 + n_2 + n_3 - l_2 - l_3 + j_1, \alpha_2 + j_2) \\
&\times B(\alpha_1 + l_1 + l_2 + l_3 - j_1 - j_2, \alpha_2 + \alpha_3 + \alpha_4 + n_1 - l_1 + n_2 - l_2 + n_3 - l_3 + j_1 + j_2)
\end{aligned}
$$

$$
\begin{aligned}
A_3 = &\sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3 + j_1, \alpha_4 + n_3 - l_3) \\
&\times \sum_{j_2=0}^{l_2+l_3-j_1} \binom{l_2 + l_3 - j_1}{j_2} B(\alpha_2 + j_2, \alpha_3 + \alpha_4 + n_2 + n_3 - l_2 - l_3 + j_1) \\
&\times B(\alpha_1 + l_1 + l_2 + l_3 - j_1 - j_2, \alpha_2 + \alpha_3 + \alpha_4 + n_1 - l_1 + n_2 - l_2 + n_3 - l_3 + j_1 + j_2) \\
&\times D(\alpha_2 + \alpha_3 + \alpha_4 + n_1 - l_1 + n_2 - l_2 + n_3 - l_3 + j_1 + j_2, \alpha_1 + l_1 + l_2 + l_3 - j_1 - j_2)
\end{aligned}
$$

Analogously for the other statistics of our interests we get:

$$
\begin{aligned}
&\int_0^1 \int_0^{1-p_1} \int_0^{1-p_1-p_2} \log p_2 \; p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_3^{\alpha_3-1} (1 - p_1 - p_2 - p_3)^{\alpha_4-1} \\
&\times p_1^{l_1} (1 - p_1)^{n_1-l_1} (p_1 + p_2)^{l_2} (1 - p_1 - p_2)^{n_2-l_2} (p_1 + p_2 + p_3)^{l_3} (1 - p_1 - p_2 - p_3)^{n_3-l_3} dp_3 dp_2 dp_1 \\
&= B_1 + B_2
\end{aligned}
\tag{27}
$$

where

$$
\begin{aligned}
B_1 = &\sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3 + j_1, \alpha_4 + n_3 - l_3) \\
&\times \sum_{j_2=0}^{l_2+l_3-j_1} \binom{l_2 + l_3 - j_1}{j_2} B(\alpha_2 + j_2, \alpha_3 + \alpha_4 + n_2 + n_3 - l_2 - l_3 + j_1) \\
&\times D(\alpha_2 + j_2, \alpha_3 + \alpha_4 + n_2 + n_3 - l_2 - l_3 + j_1) \\
&\times B(\alpha_1 + l_1 + l_2 + l_3 - j_1 - j_2, \alpha_2 + \alpha_3 + \alpha_4 + n_1 - l_1 + n_2 - l_2 + n_3 - l_3 + j_1 + j_2)
\end{aligned}
$$

and $B_2 = C_3$;

$$
\begin{aligned}
&\int_0^1 \int_0^{1-p_1} \int_0^{1-p_1-p_2} \log p_1 \; p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_3^{\alpha_3-1} (1 - p_1 - p_2 - p_3)^{\alpha_4-1} \\
&\times p_1^{l_1} (1 - p_1)^{n_1-l_1} (p_1 + p_2)^{l_2} (1 - p_1 - p_2)^{n_2-l_2} (p_1 + p_2 + p_3)^{l_3} (1 - p_1 - p_2 - p_3)^{n_3-l_3} dp_3 dp_2 dp_1 \\
&= C
\end{aligned}
\tag{28}
$$

where

$$
\begin{aligned}
C = & \sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3 + j_1, \alpha_4 + n_3 - l_3) \\
& \times \sum_{j_2=0}^{l_2+l_3-j_1} \binom{l_2 + l_3 - j_1}{j_2} B(\alpha_2 + j_2, \alpha_3 + \alpha_4 + n_2 + n_3 - l_2 - l_3 + j_1) \\
& \times B(\alpha_1 + l_1 + l_2 + l_3 - j_1 - j_2, \alpha_2 + \alpha_3 + \alpha_4 + n_1 - l_1 + n_2 - l_2 + n_3 - l_3 + j_1 + j_2) \\
& \times D(\alpha_1 + l_1 + l_2 + l_3 - j_1 - j_2, \alpha_2 + \alpha_3 + \alpha_4 + n_1 - l_1 + n_2 - l_2 + n_3 - l_3 + j_1 + j_2);
\end{aligned}
$$

and the last:

$$
\begin{aligned}
& \int_0^1 \int_0^{1-p_1} \int_0^{1-p_1-p_2} \log(1 - p_1 - p_2 - p_3)\, p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_3^{\alpha_3-1} (1 - p_1 - p_2 - p_3)^{\alpha_4-1} \\
& \times p_1^{l_1} (1 - p_1)^{n_1-l_1} (p_1 + p_2)^{l_2} (1 - p_1 - p_2)^{n_2-l_2} (p_1 + p_2 + p_3)^{l_3} (1 - p_1 - p_2 - p_3)^{n_3-l_3} dp_3 dp_2 dp_1 \\
& = E_1 + E_2 + E_3
\end{aligned}
\tag{29}
$$

where

$$
\begin{aligned}
E_1 = & \sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3 + j_1, \alpha_4 + n_3 - l_3) \\
& \times D(\alpha_4 + n_3 - l_3, \alpha_3 + j_1) \\
& \times \sum_{j_2=0}^{l_2+l_3-j_1} \binom{l_2 + l_3 - j_1}{j_2} B(\alpha_2 + j_2, \alpha_3 + \alpha_4 + n_2 + n_3 - l_2 - l_3 + j_1) \\
& \times B(\alpha_1 + l_1 + l_2 + l_3 - j_1 - j_2, \alpha_2 + \alpha_3 + \alpha_4 + n_1 - l_1 + n_2 - l_2 + n_3 - l_3 + j_1 + j_2)
\end{aligned}
$$

and $E_2 = A_2$, $E_3 = A_3$.

We see that when we compute the conditional expectation of the j-th statistic, with $j = 1, \dots, k$, we obtain a sum of j terms, the first containing the D function in the $k+1-j$ entry of the nested sums, with the same arguments of the B function, and the other terms containing the D function respectively in the $k+2-j, \dots, k$ entries of the nested sums, with arguments exchanged w.r.t. the B function. For the conditional expectation of the $k+1$ statistic we obtain a sum of k terms, each containing the D function in the $1, \dots, k$ entries of the nested sums, with arguments exchanged w.r.t. the B function.

Now for the M-step of the algorithm we have to compute the unconditional expectations $\mathbb{E}[t_j(\mathbf{x})|(\alpha_1, \dots, \alpha_{k+1})]$ for $j = 1, \dots, k$ and $\mathbb{E}[t_{k+1}(\mathbf{x})|(\alpha_1, \dots, \alpha_{k+1})] = \sum_{i=1}^m \mathbb{E}[\log(1 - p_{i1} - \cdots - p_{ik})|\mathbf{y}, (\alpha_1, \dots, \alpha_{k+1})]$. Since the statistics don't depend on the variables $(l_{11}, \dots, l_{mk})$

we have that:

$$
\begin{aligned}
&\mathbb{E}[t_j(\mathbf{x})|(\alpha_1,\ldots,\alpha_{k+1})] = \sum_{i=1}^{m} \mathbb{E}[\log p_{ij}|(\alpha_1,\ldots,\alpha_{k+1})] \\
&= \sum_{i=1}^{m} \int_0^{p_{i1}} \cdots \int_0^{1-p_{i1}-\cdots-p_{i,k-1}} \log p_{ij} \\
&\times \frac{\Gamma\left(\sum_{j=1}^{k+1}\alpha_j\right)}{\prod_{i=1}^{k+1}\Gamma(\alpha_j)} p_{i1}^{\alpha_1-1}\cdots p_{ik}^{\alpha_k-1}(1-p_{i1}-\cdots-p_{ik})^{\alpha_{k+1}} dp_{ik}\cdots dp_{i1} \\
&= mD(\alpha_j,\alpha_1+\cdots+\alpha_{k+1})
\end{aligned}
\tag{30}
$$

Analogously we have that $\mathbb{E}[t_{k+1}(\mathbf{x})|(\alpha_1,\ldots,\alpha_{k+1})] = mD(\alpha_{k+1},\alpha_1+\cdots+\alpha_{k+1})$.

So, the algorithm in our case works in this way: suppose that $(\alpha_1^{(0)},\ldots,\alpha_{k+1}^{(0)})$ are our starting values of the parameters, for example given by moment estimation; then the new values of the parameters $(\alpha_1^{(1)},\ldots,\alpha_{k+1}^{(1)})$ are given by the solutions of the following system:

$$
\begin{cases}
mD(\alpha_1^{(1)},\alpha_2^{(1)}+\alpha_3^{(1)}+\cdots+\alpha_{k+1}^{(1)}) = f_1(\alpha_1^{(0)},\ldots,\alpha_{k+1}^{(0)}) \\
mD(\alpha_2^{(1)},\alpha_1^{(1)}+\alpha_3^{(1)}+\cdots+\alpha_{k+1}^{(1)}) = f_2(\alpha_1^{(0)},\ldots,\alpha_{k+1}^{(0)}) \\
\qquad\cdots \qquad\qquad\qquad\qquad\qquad \cdots \\
mD(\alpha_{k+1}^{(1)},\alpha_1^{(1)}+\alpha_2^{(1)}+\cdots+\alpha_k^{(1)}) = f_{k+1}(\alpha_1^{(0)},\ldots,\alpha_{k+1}^{(0)})
\end{cases}
\tag{31}
$$

where $f_j(\alpha_1^{(0)},\ldots,\alpha_{k+1}^{(0)})$ are the values of the conditional expectations f the statistics with respect to the starting values of the parameters, that can be computed using formulas analogous to the ones seen for the 3-dimensional case in equations (26), (27), (28) and (29).

4.3 Application to iterative urn scheme

Here we don't give all the details of the application of the EM algorithm to the iterative urn scheme, since the steps are quite similar to the ones in the previous section. We limitate to the case $k=3$ to show how the procedure works, the generalization for $k>3$ is obvious.

As complete data in this case we consider $\mathbf{x}_i = (\tilde{p}_{i1},\ldots,\tilde{p}_{i3},l_{i1},\ldots,l_{i3})$ with $\tilde{p}_{ij}$ the i-th (unknown) realization of the random variable $\tilde{P}_j$. The sampling density is then given as before by the product of m terms $f_i(\mathbf{x}_i|(\alpha_1,\beta_1,\ldots,\alpha_3,\beta_3))$ defined by:

$$
\begin{aligned}
&f_i(\mathbf{x}_i|(\alpha_1,\beta_1,\ldots,\alpha_3,\beta_3)) = \\
&\frac{1}{B(\alpha_1,\beta_1)}\tilde{p}_{i1}^{\alpha_1-1}(1-\tilde{p}_{i1})^{\beta_1-1}\frac{1}{B(\alpha_2,\beta_2)}\tilde{p}_{i2}^{\alpha_2-1}(1-\tilde{p}_{i2})^{\beta_2-1}\frac{1}{B(\alpha_3,\beta_3)}\tilde{p}_{i3}^{\alpha_3-1}(1-\tilde{p}_{i3})^{\beta_3-1} \\
&\times\binom{n_{i1}}{l_{i1}}\binom{n_{i2}}{l_{i2}}\binom{n_{i3}}{l_{i3}}\sum_{j_1=0}^{l_{i3}}\binom{l_{i3}}{j_1}\sum_{j_2=0}^{l_{i2}+j_1}\binom{l_{i2}+j_1}{j_2}\tilde{p}_{i1}^{l_{i1}+j_2}(1-\tilde{p}_{i1})^{n_{i1}+n_{i2}+n_{i3}-l_{i1}-j_2} \\
&\times\tilde{p}_{i2}^{l_{i2}+j_1-j_2}(1-\tilde{p}_{i2})^{n_{i3}+n_{i2}-l_{i2}-j_1}\tilde{p}_{i3}^{l_{i3}-j_1}(1-\tilde{p}_{i3})^{n_{i3}-l_{i3}}
\end{aligned}
\tag{32}
$$

It is easy to show that $f_i(\mathbf{x}_i|(\alpha_1, \beta_1, \dots, \alpha_3, \beta_3))$ still belongs to the regular exponential family, and that the sufficient statistics are given by $t_{2j+1}(\mathbf{x}) = \sum_{i=1}^m \log \tilde{p}_{ij}$ for $j = 0, 1, 2$, corresponding to the parameters $\alpha_1, \alpha_2, \alpha_3$, and $t_{2j}(\mathbf{x}) = \sum_{i=1}^m \log(1 - \tilde{p}_{ij})$ for $j = 1, 2, 3$ corresponding to the parameters $\beta_1, \beta_2, \beta_3$.

For the computation of the conditional expectations of the statistics it is easy to see that, analogously to what we have done in the multidimensional case, for example for $t_5(\mathbf{x})$ we have to compute, for the term corresponding to the numerator in (22) (we suppress index i for ease of notation):

$$\begin{aligned}
&\sum_{j_1=0}^{l_3} \binom{l_3}{j_1} \sum_{j_2=0}^{l_2+j_1} \binom{l_2+j_1}{j_2} \int_0^1 \int_0^1 \int_0^1 \tilde{p}_1^{l_1+j_2} (1-\tilde{p}_1)^{n_1+n_2+n_3-l_1-j_2} \\
&\times \tilde{p}_2^{l_2+j_1-j_2} (1-\tilde{p}_2)^{n_3+n_2-l_2-j_1} \tilde{p}_3^{l_3-j_1} (1-\tilde{p}_3)^{n_3-l_3} \log \tilde{p}_3 d\tilde{p}_1 d\tilde{p}_2 d\tilde{p}_3 \\
&= \sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3 + l_3 - j_1, \beta_3 + n_3 - l_3) D(\alpha_3 + l_3 - j_1, \beta_3 + n_3 - l_3) \qquad (33) \\
&\times \sum_{j_2=0}^{l_2+j_1} \binom{j_2}{l_2+j_1} B(\alpha_2 + l_2 + j_1 - j_2, \beta_2 + n_3 + n_2 - l_2 - j_1) \\
&\times B(\alpha_1 + l_1 + j_2, \beta_1 + n_1 + n_2 + n_3 - l_1 - j_2)
\end{aligned}$$

This is much easier than in the multidimensional urn scheme, and analogously for $t_6(\mathbf{x})$ we get:

$$\begin{aligned}
&\sum_{j_1=0}^{l_3} \binom{l_3}{j_1} \sum_{j_2=0}^{l_2+j_1} \binom{l_2+j_1}{j_2} \int_0^1 \int_0^1 \int_0^1 \tilde{p}_1^{l_1+j_2} (1-\tilde{p}_1)^{n_1+n_2+n_3-l_1-j_2} \\
&\times \tilde{p}_2^{l_2+j_1-j_2} (1-\tilde{p}_2)^{n_3+n_2-l_2-j_1} \tilde{p}_3^{l_3-j_1} (1-\tilde{p}_3)^{n_3-l_3} \log(1-\tilde{p}_3) d\tilde{p}_1 d\tilde{p}_2 d\tilde{p}_3 \\
&= \sum_{j_1=0}^{l_3} \binom{l_3}{j_1} B(\alpha_3 + l_3 - j_1, \beta_3 + n_3 - l_3) D(\beta_3 + n_3 - l_3, \alpha_3 + l_3 - j_1) \qquad (34) \\
&\times \sum_{j_2=0}^{l_2+j_1} \binom{j_2}{l_2+j_1} B(\alpha_2 + l_2 + j_1 - j_2, \beta_2 + n_3 + n_2 - l_2 - j_1) \\
&\times B(\alpha_1 + l_1 + j_2, \beta_1 + n_1 + n_2 + n_3 - l_1 - j_2)
\end{aligned}$$

We see then that computing conditional expectations of the statistics requires only to multiply one term of the nested sums, the one with parameters "pointed" by the statistics, by the D function with the same arguments of the B function if the statistic corresponds to an α_j parameter, or by the D function with arguments exchanged w.r.t. the B function if the statistic corresponds to a β_j parameter.

For the unconditional expectations of the statistics again, as for the multidimensional urn scheme, since they don't depend on the variables $(l_{11}, l_{12}, \dots, l_{m2}, l_{m3})$, we get that

$$\mathbb{E}[t_{2j+1}(\mathbf{x})|(\alpha_1, \dots, \beta_3)] = mD(\alpha_{2j+1}, \beta_{2j+1})$$

for $j = 0, \ldots, k-1$ and

$$\mathbb{E}[t_{2j}(\mathbf{x})|(\alpha_1, \ldots, \beta_3)] = mD(\beta_{2j}, \alpha_{2j})$$

for $j = 1, \ldots, k$.

So, the algorithm in our case works in this way: suppose that $(\alpha_1^{(0)}, \ldots, \beta_3^{(0)})$ are our starting values of the parameters, for example given by moment estimation; then the new values of the parameters $(\alpha_1^{(1)}, \ldots, \beta_3^{(1)})$ are given by the solutions of the following system:

$$\begin{cases} mD(\alpha_1^{(1)}, \beta_1^{(1)}) = f_1(\alpha_1^{(0)}, \ldots, \beta_3^{(0)}) \\ mD(\beta_1^{(1)}, \alpha_1^{(1)}) = f_2(\alpha_1^{(0)}, \ldots, \beta_3^{(0)}) \\ \quad \cdots \qquad\qquad \cdots \\ mD(\alpha_3^{(1)}, \beta_3^{(1)}) = f_5(\alpha_1^{(0)}, \ldots, \beta_3^{(0)}) \\ mD(\beta_3^{(1)}, \alpha_3^{(1)}) = f_6(\alpha_1^{(0)}, \ldots, \beta_3^{(0)}) \end{cases} \tag{35}$$

where $f_j(\alpha_1^{(0)}, \ldots, \beta_3^{(0)})$ are the values of the conditional expectations of the statistics with respect to the starting values of the parameters, that can be computed using the formulas just seen. We see then that this system is much simpler than the one seen for the multi-dimensional urn scheme, since instead of solving a system of $2k$ equations in $2k$ unknowns we have only to solve k independent systems of two equations in two unknowns.

5 Calibration of the models and conclusions

In this section we discuss the calibration of the two models presented in the previous sections, using the data from Standard & Poor's report [9]. For every rating we are given the number of firms and defaults for the years 1981-2002, see table 1. Since the AAA-rating showed no default in the data, we considered only the six rating class from AA to CCC in our estimations. Our main purpose has been to try to do a maximum likelihood estimation of the parameters of the models through the use of the EM algorithm shown in the previous chapter.

5.1 Multidimensional Urn Scheme

Since the speed of the convergence of the EM algorithm is very sensible with respect to the starting values, it is very important to choose them carefully. For the multidimensional urn scheme we observe that since

$$P_1 + \cdots + P_j = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (X_{i,1} \cdots + X_{i,j}) \,.$$

we can take, for n_j large, l_j/n_j as realization of $P_1 + \cdots + P_j$, $1 \le j \le k$. Hence

$$\left(\frac{l_1}{n_1}, \frac{l_2}{n_2} - \frac{l_1}{n_1}, \ldots, \frac{l_k}{n_k} - \frac{l_{k-1}}{n_{k-1}} \right)$$

		1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
AAA	firms	93	94	116	138	103	125	145	159	171	171	183
	defaults	0	0	0	0	0	0	0	0	0	0	0
AA	firms	204	225	247	299	341	371	377	384	388	412	423
	defaults	0	0	0	0	0	0	0	0	0	0	0
A	firms	485	477	458	462	505	560	518	515	570	592	609
	defaults	0	1	0	0	0	1	0	0	0	0	0
BBB	firms	273	290	303	300	278	302	321	331	343	359	389
	defaults	0	1	1	2	0	1	0	0	2	2	3
BB	firms	222	167	171	172	198	224	264	285	276	283	237
	defaults	0	7	2	2	3	3	1	3	2	10	6
B	firms	90	161	156	180	207	293	358	415	416	367	289
	defaults	2	5	7	6	12	24	12	16	14	31	39
CCC	firms	11	13	15	19	18	17	63	58	55	46	60
	defaults	0	3	0	3	2	3	6	12	16	14	19
		1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
AAA	firms	202	208	206	210	205	199	203	193	188	181	184
	defaults	0	0	0	0	0	0	0	0	0	0	0
AA	firms	480	514	530	550	561	586	612	632	643	604	600
	defaults	0	0	0	0	0	0	0	1	0	0	0
A	firms	688	768	847	1025	1089	1161	1198	1227	1223	1234	1260
	defaults	0	0	1	0	0	0	0	1	1	3	1
BBB	firms	409	475	544	650	732	846	1010	1089	1160	1282	1383
	defaults	0	0	0	2	0	3	4	2	4	5	16
BB	firms	244	291	379	433	477	557	663	793	884	920	902
	defaults	0	1	1	4	3	1	5	9	10	26	26
B	firms	226	237	344	406	442	479	701	903	960	933	838
	defaults	16	6	9	17	12	16	32	62	74	100	71
CCC	firms	50	49	26	29	29	28	33	74	87	116	184
	defaults	12	6	4	7	1	3	11	23	26	50	80

Table 1: Data from Standard & Poor's Special Report [9]

are realizations of $(P_1, \ldots, P_k) \sim D_{k+1}(\alpha_1, \ldots, \alpha_{k+1})$.
We can so try to get our starting values by statistical calibration of the underlying Dirichlet distribution.

We did it in two ways:

1) First we estimated $\alpha_1, \ldots, \alpha_{k+1}$ by the method of moments. For $j = 1, \ldots, k$ we denote the first order sample moments from the observations as

$$M_j^1 = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{l_{i,j}}{n_{i,j}} - \frac{l_{i,j-1}}{n_{i,j-1}} \right)$$

and the second order sample moments as

$$M_j^2 = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{l_{i,j}}{n_{i,j}} - \frac{l_{i,j-1}}{n_{i,j-1}} \right)^2,$$

where $l_{i,0}/n_{i,0} = 0$ by definition.

We can then equate the theoretical moments given by $\mathbb{E}[P_j] = \alpha_j/\alpha$ and $\mathbb{E}[P_j^2] = \alpha_j(\alpha_j + 1)/(\alpha(\alpha + 1))$ to the sample moments to get the moment estimates.

2) Then we estimated $\alpha_1, \ldots, \alpha_{k+1}$ by the maximum likelihood method. Since the density of the Dirichlet distribution is given by equation 5, then the log-likelihood function is:

$$l(\alpha_1, \ldots, \alpha_{k+1}) = \sum_{i=1}^{m} \log \left[\frac{\Gamma(\alpha)}{\prod_{j=1}^{k+1} \Gamma(\alpha_j)} \prod_{j=1}^{k+1} \left(\frac{l_{i,j}}{n_{i,j}} - \frac{l_{i,j-1}}{n_{i,j-1}} \right)^{\alpha_j - 1} \right],$$

where $l_{i,0}/n_{i,0} = 0$ and $l_{i,k+1}/n_{i,k+1} = 1$.

We checked that the better estimates as starting values were the ones obtained by the method of moments; the values obtained are shown in table 2. Note that since we have to estimate $k+1$ parameters and we have to our disposal $2k$ equations, we have chosen to do k different estimation of the parameters using in each of them the k equations for the first order moments and one equation for the r-th second order moment.

2^{nd} Mo	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$	$\hat{\alpha}_7$
6th	0.00078	0.00414	0.02840	0.10888	0.45526	1.68344	8.6315
5th	0.00443	0.02333	0.16018	0.61408	2.56768	9.49468	48.6821
4th	0.00706	0.03720	0.25547	0.97936	4.09506	15.1426	77.6405
3th	0.02057	0.10842	0.74452	2.85421	11.9344	44.1307	226.271
2nd	0.04595	0.24212	1.66265	6.37398	26.6518	98.552	505.306
1st	0.04754	0.25053	1.72046	6.59561	27.5785	101.979	522.876

Table 2: Comparison of the moment estimates for the multidimensional urn scheme with different second order sample moments.

We then applied the EM algorithm to maximize the log-likelihood function deduced by equation (6) using the procedure described in section 4.2. First we tried with groups of three ratings, using the amalgamation property of the Dirichlet distribution, so that for example $(P_{AA}, P_A, P_{BBB}) \sim D_4(\alpha_1, \alpha_2, \alpha_3, \alpha_4 + \cdots + \alpha_7)$.

We noticed that there is a kind of monotonicity relation between the group of rating classes and the second order moment used for the parameters estimation; that is the estimates obtained using the first component second order sample moment were better for the maximum likelihood estimation of *AA-A-BBB* group, and the ones obtained using the sixth component second order sample moment were better for *BB-B-CCC* group. For the numerical computation, we first tried to use Mathematica 5.0®, but we soon saw that this was not feasible due to the time required for the computation of right end side of system (31). So we implemented the algorithm in a C++ program. For the computation of values of Gamma and related functions and for the solution of the system (31) we used some routines provided by the GNU Scientific Library(see [5]); in particular for the solution of the system we used an implementation of M. J. D. Powell hybrid method for nonlinear equations (see [7]). As stopping criteria for the algorithm, we checked at every step both the difference between the values of the log-likelihood and the Euclidean distance between the vectors of expectation and covariances of underlying Dirichlet distribution. For the simple case of groups of three rating classes we could so get very fast results, and they are shown in table 3.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	Loglike
AA-A-BBB	0.145251	0.527336	3.26885	1197.91	-59.1917
A-BBB-BB	0.292419	1.00181	3.12502	352.522	-112.106
BBB-BB-B	0.664552	1.41154	6.67936	146.846	-171.191
BB-B-CCC	1.07227	3.08588	12.9567	53.7531	-200.273

Table 3: Maximum likelihood estimates for the multidimensional urn scheme via the EM algorithm for groups of three different ratings.

We tried then to do the same for the groups *BBB-BB-B-CCC* and *A-BBB-BB-B-CCC* rating class, that have the biggest number of defaults, which is the parameter that mostly influence the time of the computations, and we have been successful in a "reasonable" time in these cases too; the results are shown in table 4. The times required for the computation has been three hours for the *BBB-BB-B-CCC* group and one day and a half for the *A-BBB-BB-B-CCC* group.

5.2 Iterative Urn Scheme

As for the multidimensional urn scheme, we first estimated the parameters $\alpha_1, \beta_1, \ldots, \alpha_k, \beta_k$ trough the underlying generalized Dirichlet distribution. In analogy to what we have done in section 5.1 we can take

$$\left(\frac{l_1}{n_1}, \frac{l_2}{n_2} - \frac{l_1}{n_1}, \ldots, \frac{l_k}{n_k} - \frac{l_{k-1}}{n_{k-1}}\right)$$

	A-BBB-BB-B-CCC	BBB-BB-B-CCC
$\hat{\alpha}_1$	0.175989	0.49368
$\hat{\alpha}_2$	0.444428	0.940949
$\hat{\alpha}_3$	1.13056	4.06962
$\hat{\alpha}_4$	5.01342	17.6662
$\hat{\alpha}_5$	21.6164	71.4711
$\hat{\alpha}_6$	88.4868	

Table 4: MLE estimates for the multidimensional urn scheme via the EM algorithm for the groups of *A-BBB-BB-B-CCC* and *BBB-BB-B-CCC* ratings.

as realizations of $(Q_1, \ldots, Q_k)$.

Inspirated by the results seen in previous section we did only the method of moments estimation. For $j = 1, \ldots, k$, the sample mean and variance are given by

$$E_j = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{l_{i,j}}{n_{i,j}} - \frac{l_{j-1}}{n_{j-1}} \right)$$
$$V_j = \frac{1}{m-1} \sum_{i=1}^{m} \left[\left(\frac{l_{i,j}}{n_{i,j}} - \frac{l_{i,j-1}}{n_{i,j-1}} \right) - E_j \right]^2,$$

where $l_{i,0}/n_{i,0} = 0$

By equating the sample moments with the theoretical ones of the Generalized Dirichlet distribution and solving for α_j and β_j we get:

$$\alpha_j = E_j \frac{E_j^2 - E_j M_{j-1} + V_j}{E_j^2(M_{j-1} - K_{j-1}) - K_{j-1} V_j} \tag{36}$$
$$\beta_j = (K_{j-1} - E_j) \frac{E_j^2 - E_j M_{j-1} + V_j}{E_j^2(M_{j-1} - K_{j-1}) - K_{j-1} V_j},$$

where $K_j = \prod_{m=1}^{j} \beta_m/(\alpha_m + \beta_m)$, $K_0 = 1$ and $M_j = \prod_{m=1}^{j} [(\beta_m + 1)/(\alpha_m + \beta_m + 1)]$, $M_0 = 1$. Since every equation for α_j and β_j depends only on $\alpha_1, \ldots, \alpha_{j-i}$ and $\beta_1, \ldots, \beta_{j-1}$, we can iteratively solve them beginning with $j = 1$.

We did the estimation for four groups of consecutive rating class (*AA-A-BBB-BB-B-CCC*, *A-BBB-BB-B-CCC*, *BBB-BB-B-CCC* and *BB-B-CCC*) since, by the way in which the parameters are calculated through equations (36), this allows to obtain also moment estimates for the other groups of three rating classes that we later maximized trough the EM algorithm. The results are shown in table 5.

Then we applied the EM algorithm to maximize the log-likelihood function deduced from equation (11) using the procedure described in section 4.3. Again we wrote a C++

	AA-A-BBB-BB-B-CCC	A-BBB-BB-B-CCC	BBB-BB-B-CCC	BB-B-CCC
$\hat{\alpha}_1$	0.0453794	0.344364	0.974863	1.21783
$\hat{\beta}_1$	630.909	763.35	318.281	92.2378
$\hat{\alpha}_2$	0.289133	0.764108	0.96496	2.68283
$\hat{\beta}_2$	698.256	286.875	94.94	60.6776
$\hat{\alpha}_3$	0.764108	0.964959	2.68282	1.82676
$\hat{\beta}_3$	286.763	94.9366	60.6742	9.633317
$\hat{\alpha}_4$	0.964933	2.68282	1.82673	
$\hat{\beta}_4$	94.7508	60.718	9.63242	
$\hat{\alpha}_5$	2.68237	1.8267		
$\hat{\beta}_5$	60.1925	9.63165		
$\hat{\alpha}_6$	1.80991			
$\hat{\beta}_6$	9.12122			

Table 5: Moment estimates for the iterative urn scheme via the generalized Dirichlet distribution for the groups of AA-A-BBB-BB-B-CCC, A-BBB-BB-B-CCC and BBB-BB-B-CCC ratings.

program to implement the procedure, and we first tried with groups of three rating classes. The results are shown in table 6.

	AA-A-BBB	A-BBB-BB	BBB-BB-B	BB-B-CCC
$\hat{\alpha}_1$	1.06397	2.37793	1.67204	1.77055
$\hat{\beta}_1$	10424.6	4635.29	493.504	134.237
$\hat{\alpha}_2$	1.92413	1.27116	1.18086	4.11124
$\hat{\beta}_2$	4735.66	443.612	123.711	90.1223
$\hat{\alpha}_3$	1.71035	1.31964	4.42703	2.82681
$\hat{\beta}_3$	613.042	139.403	96.2997	12.7833

Table 6: Maximum likelihood estimates for the iterative urn scheme via the EM algorithm for the groups of AA-A-BBB, A-BBB-BB, BB-B-CCC and BB-B-CCC ratings.

Then we tried to do the same for the groups BBB-BB-B-C and A-BBB-BB-B-C, using as starting values the maximum likelihood estimates of the corresponding parameters in view of the similarity seen in table 5. The results are shown in table 7. The use of the EM algorithm has produced a great improvement in terms of the time needed for the computation: in fact, in comparison with [8], where it was needed one day for the group BBB-BB-B-C and one week and a half for the group A-BBB-BB-B-C, in our case it took respectively only two hours and two days. On one hand we have to say that probably the computer used in this case was faster than the one used in [8]; on the other hand we remark that the data that we used for the calibration of the model showed a greater number of defaults in almost every year and every rating class, due to the yearly update of the internal Standard & Poors database, and and the models are very sensible to this parameter in terms of the complexity of the log-likelihood function and consequently of the functions involved in the EM algorithm procedure.

	A-BBB-BB-B-CCC	*BBB-BB-B-CCC*
$\hat{\alpha}_1$	1.43955	1.66868
$\hat{\beta}_1$	2821.65	492.468
$\hat{\alpha}_2$	1.25883	1.18321
$\hat{\beta}_2$	441.523	123.992
$\hat{\alpha}_3$	1.18784	4.10329
$\hat{\beta}_3$	124.159	89.5257
$\hat{\alpha}_4$	4.10343	2.82995
$\hat{\beta}_4$	89.5244	12.7955
$\hat{\alpha}_5$	2.82984	
$\hat{\beta}_5$	12.7951	

Table 7: MLE estimates for the iterative urn scheme via the EM algorithm for the groups of *A-BBB-BB-B-CCC* and *BBB-BB-B-CCC* ratings.

Appendix. Proof of De Finetti theorem

Theorem 5.1 (De Finetti's theorem)**.** *Let $(\mathbf{X}_n)_{n\geq 1}$ be an exchangeable sequence of random vectors from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the measurable space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, where $\mathscr{B}(\mathbb{R}^d))$ is the Borel σ-algebra. Then there exists a sub σ-field $\mathcal{F}_\infty \subseteq \mathcal{F}$ conditioned on which the $\mathbf{X}_n$'s are independent and identically distributed.*

Proof. Since $(\mathbf{X}_n)_{n\geq 1}$ is exchangeable it follows directly that $(\mathbf{X}_n)_{n\geq 1}$ is identically distributed. Conditional independence is shown in two steps:

Step 1. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Borel-measurable function with $\mathbb{E}\left[\,|f(\mathbf{X}_1)|\,\right] < \infty$. Since $(\mathbf{X}_n)_{n\geq 1}$ is exchangeable and hence identically distributed, it holds for every $n \geq 1$ that

$$\mathbb{E}\left[f(\mathbf{X}_j)I_A\right] = \mathbb{E}\left[f(\mathbf{X}_1)I_A\right] \qquad 1 \leqslant j \leqslant n, \quad \forall A \in \mathcal{F}_n,$$

and hence

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} f(\mathbf{X}_j)I_A\right] = \mathbb{E}\left[f(\mathbf{X}_1)I_A\right] \qquad \forall A \in \mathcal{F}_n.$$

Since $\frac{1}{n}\sum_{j=1}^{n} f(\mathbf{X}_j)$ is n-symmetric and hence $\mathcal{F}_n$-measurable, we get, by the definition of conditional expected value,

$$\frac{1}{n}\sum_{j=1}^{n} f(\mathbf{X}_j) = \mathbb{E}\left[f(\mathbf{X}_1) \mid \mathcal{F}_n\right] \qquad \text{a.s.}$$

Note that $(\mathbb{E}\left[f(\mathbf{X}_1) \mid \mathcal{F}_n\right], \mathcal{F}_n)_{n\geq 1}$ is a reverse martingale[2]. Now, applying the martingale convergence theorem, we get:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} f(\mathbf{X}_j) = \mathbb{E}\left[f(\mathbf{X}_1) \mid \mathcal{F}_\infty\right] \qquad \text{a.s.} \tag{37}$$

where $\mathcal{F}_\infty = \bigcap_{n\geq 1} \mathcal{F}_n$.

If $f(t_1, \ldots, t_d)$ is chosen as the indicator function $I_{(-\infty,\mathbf{x}]} := I_{(-\infty,x_1]\times\cdots\times(-\infty,x_d]}(t_1, \ldots, t_d)$, where $\mathbf{x} = (x_1, \ldots, x_d)$ then previous equation gives

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} I_{\{X_{j,1}\leqslant x_1\}\cap\cdots\cap\{X_{j,d}\leqslant x_d\}} \overset{\text{a.s.}}{=} \mathbb{P}\left[\mathbf{X}_1 \leqslant \mathbf{x} \mid \mathcal{F}_\infty\right]. \tag{38}$$

Step 2. The same argument as in step 1 can be applied for $f : (\mathbb{R}^d)^k \to \mathbb{R}$ defined as $f = \prod_{i=1}^{k} I_{(-\infty,\mathbf{x}_i]}$; then $f(\mathbf{X}_1, \ldots, \mathbf{X}_k) = \prod_{i=1}^{k} I_{\{\mathbf{X}_i\leqslant\mathbf{x}_i\}}$ due to exchangeability is n-symmetric for $k \leq n$, and so for every $(j_1, \ldots, j_k) \in S_n^* = \{(j_1, \ldots, j_k) \in \mathbb{N}^k \ : \ 1 \leq j_i \leq n, \ j_i \neq j_{i'}\}$ and for every $A \in \mathcal{F}_n$ we have:

$$\mathbb{E}[I_{\{\mathbf{X}_{j_1}\leqslant\mathbf{x}_1\}} \cdots I_{\{\mathbf{X}_{j_k}\leqslant\mathbf{x}_k\}} I_A] = \mathbb{E}[I_{\{\mathbf{X}_1\leqslant\mathbf{x}_1\}} \cdots I_{\{\mathbf{X}_k\leqslant\mathbf{x}_k\}} I_A].$$

Since $|S_n^*| = n(n-1)\cdots(n-k+1)$, we have then:

$$\frac{1}{n(n-1)\ldots(n-k+1)} \sum_{(j_1,\ldots,j_k)\in S_n^*} \prod_{i=1}^{k} I_{\{\mathbf{X}_{j_i}\leqslant\mathbf{x}_i\}} = \mathbb{E}\left[\prod_{i=1}^{k} I_{\{\mathbf{X}_i\leqslant\mathbf{x}_i\}} \,\middle|\, \mathcal{F}_n\right]. \tag{39}$$

Again from the martingale convergence theorem it follows that

$$\begin{aligned}\mathbb{P}\left[\bigcap_{i=1}^{k}\{\mathbf{X}_i \leqslant \mathbf{x}_i\} \,\middle|\, \mathcal{F}_\infty\right] &= \mathbb{E}\left[\prod_{i=1}^{k} I_{\{\mathbf{X}_i\leqslant\mathbf{x}_i\}} \,\middle|\, \mathcal{F}_\infty\right] \\ &\overset{(39)}{=} \lim_{n\to\infty} \frac{1}{n(n-1)\ldots(n-k+1)} \sum_{(j_1,\ldots,j_k)\in S_n^*} \prod_{i=1}^{k} I_{\{\mathbf{X}_{j_i}\leqslant\mathbf{x}_i\}}.\end{aligned}$$

Now observe that

$$0 \leq \prod_{i=1}^{k}\sum_{l=1}^{n} I_{\{\mathbf{X}_l\leqslant\mathbf{x}_i\}} - \sum_{(j_1,\ldots,j_k)\in S_n^*} \prod_{i=1}^{k} I_{\{\mathbf{X}_{j_i}\leqslant\mathbf{x}_i\}} \leq n^k - n(n-1)\cdots(n-k+1) \sim n^{k-1}$$

[2] $(X_n, \mathcal{F}_n)_{n\geq 1}$ with $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ is called a reverse martingale if for all $1 \leq m \leq n$ it holds that X_n is $\mathcal{F}_n$-measurable, $\mathbb{E}[|X_n|] < \infty$ and $\mathbb{E}[X_m|\mathcal{F}_n] = X_n$.

so that

$$\lim_{n\to\infty} \frac{1}{n(n-1)\dots(n-k+1)} \sum_{(j_1,\dots,j_k)\in S_n^*} \prod_{i=1}^{k} I_{\{\mathbf{X}_{j_i}\leqslant \mathbf{x}_i\}}$$
$$= \lim_{n\to\infty} \frac{1}{n(n-1)\dots(n-k+1)} \prod_{i=1}^{k} \sum_{l=1}^{n} I_{\{\mathbf{X}_l\leqslant \mathbf{x}_i\}}$$
$$= \lim_{n\to\infty} \prod_{i=1}^{k} \frac{1}{n} \sum_{l=1}^{n} I_{\{\mathbf{X}_l\leqslant \mathbf{x}_i\}} \overset{(38)}{=} \prod_{i=1}^{k} \mathbb{P}\left[\mathbf{X}_1 \leqslant \mathbf{x}_i \mid \mathcal{F}_\infty\right]$$

that gives the independence of $(\mathbf{X}_n)_{n\geq 1}$ conditional on $\mathcal{F}_\infty$.

□

References

[1] R. J. Connor and J.E. Mosimann, *Concepts of independence for proportions with a generalization of the dirichlet distribution*, Journal of the American Statistical Association **64** (1969), 194–206.

[2] I. D. Currie, *Maximum likelihood estimation and mathematica*, Applied Statistics **44** (1995), 379–394.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society. Series B (Methodological) **39** (1977), 1–38.

[4] S. N. Ethier and T. G. Kurtz, *Markov processes, characterization and convergence*, John Wiley & Sons, New York, 1986.

[5] *The GNU scientific library*, `http://www.gnu.org/software/gsl/`.

[6] G. J. Mclachlan and T. Krishnan, *The EM algorithm and extensions*, Wiley, New York, 1997.

[7] M.J.D. Powell, *A hybrid method for nonlinear equations*, Numerical Methods for Nonlinear Algebraic Equations (P. Rabinowitz, ed.), Gordon and Breach, 1970.

[8] D. Seiler, *Modelling dependent credit risks*, Master's thesis, ETH Zürich, 2001.

[9] Standard & Poor's, New York, *Rating performance 2002*, 2003.

[10] C. F. J. Wu, *On the convergence properties of the EM algorithm*, The Annals of Statistics **11** (1983), 95–103.