Population Review

Volume 53, Number 2, 2014

Type: Article pp. 56-71

Modeling of Marriage Preferences: Baysian Approach

Authors: Babulal Seal and Sk. Jakir Hossain

Affiliations: Department of Statistics, Burdwan University, India (Seal); Department of Statistics, Burdwan University, India (Hossain)

Corresponding author/address: Babulal Seal, Department of Statistics, Burdwan University, Burdwan, W.B.; e-mail: babulal_seal@yahoo.com

Abstract

Empirical Bayes estimation for marriage preferences are worked under two kinds of beliefs regarding the parameters, e.g., where the marriage preferences seem to be equal or not. In both cases priors are in accordance with beliefs. Computational methods for different hyperparameters of the empirical Bayes are developed. It is seen from the data that a male of age i has preferences of marrying a woman in age category j. A general model of preferences is given using empirical Bayes procedure. Real life data are considered, preference patterns are analysed and a smooth estimate is given.

Keywords

Empirical Bayes, EM algorithm, Dirichlet prior

Note: The authors would like to express their gratitude to Dr. Joseph Roberts, Executive Director of *Population Review*, for his many helpful comments.

1. Introduction

Marriage and divorce rates directly measure changes in population composition characteristics rather than changes in population size as do the other rates of population dynamics, such as fertility, mortality and migration. But if the change in population size due to births is considered as a function of the broader process of one generation reproducing another - a process that in most societies takes place largely through formation of families that produce offspring, then the rates of family formation and dissolution are appropriate for consideration under population dynamics. Theoretical model and inference procedures of internal migration which is important for understanding some characteristics of a population were considered only in Seal and Hossain (2013a, 2013b, 2013c). In this paper, the current characteristic deserves attention to understand population.

Data on marriage are obtained directly from the national vital statistics registration system or, less often, from the continuous population register. In addition, data on marital status, available from censuses and sample surveys in most countries, may be used to estimate the number of marriages and marriage rates. This indirect method represents the only source in countries where census data are satisfactory but registrations of data are lacking or inadequate. Only a few of the economically developed countries in Africa, Asia and the United States of America publish registration statistics on marriages.

Perhaps the best single source of international marriage statistics is the *Demographic Yearbook* of the United Nations. Two tables are shown annually; one on marriages and crude marriage rates for the last several years, and the other on age of bride and age of groom from the latest year record.

Now suppose $p_{i1}, p_{i2},..., p_{ik}$ are the probabilities of marrying a female of age category $j = 1,2,..., k$ at age group $i = 1,2,..., l$ of male. Suppose further that n_i data are available. Then estimates of p_{ij} using a general model are to be developed.

For modeling this this we are to suggest a prior distribution for a problem. Actually the system will speak the general form of prior.The form of prior can be understood from the form of the problem. We seek to make the prior finer and finer from only the data set. Some researchers give emphasis on data and not on prior, which is not the spirit of Bayes procedure. So, once we have feeling about prior, we should take its general form and then from data we make it concrete as possible.Clearly, in this case, i.e., marriage preference at an age, the general form of the prior must be multivariate Beta or Dirichlet prior. With the data, this can be made as concrete as possible using Bayes procedure. Now, from this general form in terms of variations of hyperparameters and data, empirical Bayes procedure is obtained.

In the following sections empirical Bayes estimations for marriage preferences are considered under two different beliefs: 1) when the hyperparameters are identical for different ages, i.e. the preferences are equal at all ages and 2) when one does not have such beliefs. Following this, unknown hyperparameters are estimated,computational methods are worked out using EM algorithm, real life data are In section considered and estimations using this method are given.The relevant R- codes are presented in the appendix.

2. Empirical Bayes Estimation of marriage preferences of different age groups

In this section, the estimators of the unknown parameters involved in the prior distribution are obtained.The Bayes estimator $\delta(n_{ij})$ of p_{ij} depends on prior distribution $\pi(p \mid \alpha)$. When α_{ij}'s are unknown, the empirical Bayes approach is employed to combine information from observations n_i, *at age level* $i = 1,2,\cdots,k$.

For general form of prior distribution of p it may be possible to compute the posterior distribution of p given n without the complete knowledge of the distribution function of p. The estimate in such situations may involve unknown parameters which are estimated from the marginal distribution of n.

Case 1: Priors for different age groups are same

Let

$$n_i = (n_{i1}, n_{i2}, \cdots, n_{ik}) \sim Multinomial(n_{i\cdot}; p_{i1}, p_{i2}, \cdots, p_{ik})$$

and

$$p_i = (p_{i1}, p_{i2}, \cdots, p_{ik}) \sim Dirichlet(\alpha_1, \alpha_2, \cdots, \alpha_k), \; for\ all\ i = 1,2,\cdots,k$$

i.e., the k groups have common prior distribution or they have same preferences.

Therefore the Bayes estimator of p_{ij} under sum of squared error loss is

$$\delta(n_{ij}) = \frac{n_{ij} + \alpha_i}{k\alpha_i + n_{i\cdot}},$$

where

$$n_{i\cdot} = \sum_{j=1}^{k} n_{ij}, \; for\ all\ i = 1,2,\cdots,k$$

If the Dirichlet parameters are unknown, we compute the ML estimate of $(\alpha_1, \alpha_2, \cdots, \alpha_k)$ based on available observations $n_i, \; for\ all\ i = 1,2,\cdots,k$. We have $n_{i\cdot} = \sum_{j=1}^{k} n_{ij}, \; for\ all\ i = 1,2,\cdots,k.$

The maximum likelihood estimators of $(p_{i1}, p_{i2}, \cdots, p_{ik})$ become the relative frequencies

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^{k} n_{ij}} = \frac{n_{ij}}{n_{i\cdot}}, \; for\ all\ i, j = 1,2,\cdots,k$$

Let us write

$$y_{ij} = \frac{n_{ij}}{n_{i\cdot}}, \; for\ all\ i, j = 1,2,\cdots,k \; and \; y_i = (y_{i1}, y_{i2}, \cdots, y_{ik})^T, \; for\ all\ i = 1,2,\cdots,k$$

Then the vector y_i approximately follow Dirichlet distribution (Johnson and Kotz (1969).As $Y_1, Y_2, \cdots, Y_k$ are k random vectors from Dirichlet $(\alpha_1, \alpha_2, \cdots, \alpha_k)$, the likelihood function becomes

$$L(\alpha) = \left(\frac{\Gamma(\alpha_{\cdot})}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \right)^k \prod_{i=1}^{k} \left(y_{i1}^{\alpha_1 - 1} y_{i2}^{\alpha_2 - 1} \cdots y_{ik}^{\alpha_k - 1} \right),$$

where $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)^T$ and $\alpha_{\cdot} = \sum_{j=1}^{k} \alpha_j$. (2.1)

Now we have, $$\frac{\delta \log L(\alpha)}{\delta \alpha_j} = k\psi(\alpha_{\cdot}) - k\psi(\alpha_j) + \sum_{i=1}^{k} \log y_{ij},$$

where $\psi(t)=\frac{\delta}{\delta t}\log\Gamma(t)$ is a digamma function.

Then the maximum-likelihood equations for estimators of $(\alpha_1,\alpha_2,\cdots,\alpha_k)$ are given by

$$\psi(\hat{\alpha}_j)-\psi(\hat{\alpha}_.)=\frac{1}{k}\sum_{i=1}^{k}\log(y_{ij}),\ \mathit{for\ all}\ \ j=1,2,\cdots,k(2.2)$$

The equations (2.2) must be solved by numerical methods.

Case 2: Priors for different age groups have different choices

Now

$$n_i \sim \mathit{Multinomial}(n_{i.};p_{i1},p_{i2},\cdots,p_{ik}),\ \mathit{for\ all}\ i=1,2,...,k$$

such that

$$n_{i.}=\sum_{j=1}^{k}n_{ij}\ ,\ \sum_{j=1}^{k}p_{ij}=1,\ \mathit{for\ all}\ i=1,2,...,k$$

and p_i have Dirichlet prior distribution i.e.,

$$p_i \sim \mathit{Dirichlet}(\alpha_{i1},\alpha_{i2},\cdots,\alpha_{ik}),\ \mathit{for\ all}\ i=1,2,\cdots,k$$

Then the Bayes estimator of p_{ij} (Seal and Hossain, Oct 2013) under squared error loss is

$$\delta(n_{ij})=\frac{n_{ij}+\alpha_{ij}}{n_{i.}+\alpha_{i.}},\ \mathit{for\ all}\ i,j=1,2,\cdots,k$$

3. Computational method when hyperparameters are the same for k age groups (i.e priors are identical)

Now using the approximation

$$\psi(t)\equiv\log(t-\frac{1}{2})$$

the equations in(2.2) become,

$$\log(\alpha_j-\frac{1}{2})-\log(\alpha_.-\frac{1}{2})=\frac{1}{k}\sum_{i=1}^{k}\log(y_{ij}),\ \mathit{for\ all}\ \ j=1,2,\cdots,k$$

$$\Rightarrow\log\left(\frac{\alpha_j-\frac{1}{2}}{\alpha_.-\frac{1}{2}}\right)=\log\left(\prod_{i=1}^{k}y_{ij}^{\frac{1}{k}}\right),\ \mathit{for\ all}\ \ j=1,2,\cdots,k$$

Thus the approximate values of $(\hat{\alpha}_j-\frac{1}{2})/(\hat{\alpha}_.-\frac{1}{2}),\ \mathit{for\ all}\ \ j=1,2,\cdots,k$ are given by

$$\frac{\alpha_j - \frac{1}{2}}{\alpha_. - \frac{1}{2}} = \prod_{i=1}^{k} y_{ij}^{\frac{1}{k}}, \ for\ all\ \ j = 1,2,\cdots,k \qquad (3.1)$$

From these, the following identity is obtained after summing over all j,

$$\frac{\alpha_. - \frac{k}{2}}{\alpha_. - \frac{1}{2}} = \sum_{j=1}^{k}\prod_{i=1}^{k} y_{ij}^{\frac{1}{k}}$$

$$\Rightarrow 1 - \frac{\frac{k-1}{2}}{\alpha_. - \frac{1}{2}} == \sum_{j=1}^{k}\prod_{i=1}^{k} y_{ij}^{\frac{1}{k}}$$

$$\Rightarrow \alpha_. - \frac{1}{2} = \frac{\frac{k-1}{2}}{1 - \sum_{j=1}^{k}\prod_{i=1}^{k} y_{ij}^{\frac{1}{k}}}$$

Thus, from (3.1) ,the estimators of α_j, i.e. $\hat{\alpha}_j$ are given by

$$\hat{\alpha}_j = \frac{1}{2} + \frac{\frac{k-1}{2}\prod_{i=1}^{k} y_{ij}^{\frac{1}{k}}}{1 - \sum_{j=1}^{k}\prod_{i=1}^{k} y_{ij}^{\frac{1}{k}}}, \ for\ all\ \ j = 1,2,\cdots,k (3.2)$$

Starting from these values of $\hat{\alpha}_j$ from (3.2), solutions of (2.2) can be obtained by an iterative process.

Let us denote the value of α at the r th subsequent iteration of Newton-Raphson method by $\alpha^{(r)}$. Then the value of α in the next iteration is given by

$$\alpha^{(r+1)} = \alpha^{(r)} + I(\alpha^{(r)})^{-1} S(\alpha^{(r)})$$

where $I(\alpha) = -\frac{\delta^2}{\delta\alpha\delta\alpha^{'}} \log L(\alpha)$ is the observed information matrix and $S(\alpha) = \frac{\delta}{\delta\alpha} \log L(\alpha)$ is the score vector.

From the likelihood function (2.1), the score has entries

$$\frac{\delta}{\delta\alpha_j} \log L(\alpha) = k\psi(\alpha_.) - k\psi(\alpha_j) + \sum_{i=1}^{k} \log y_{ij}$$

The observed information has entries

$$-\frac{\delta^2}{\delta\alpha_j \delta\alpha_{j^{'}}} \log L(\alpha) = k\{1_{\{j=j^{'}\}}\psi^{'}(\alpha_j) - \psi^{'}(\alpha_.)\}$$

where $1_{\{j=j'\}}$ is the indicator function of the event $\{j = j'\}$, and $\psi'(t)$ is the trigamma function $\frac{\delta^2}{\delta t^2}\log\Gamma(t)$.

The observed information can be summarised in matrix form by

$$-\frac{\delta^2}{\delta\alpha\delta\alpha'}\log L(\alpha) = I(\alpha) = k(D - c11^T)$$

where D is a diagonal matrix with j th diagonal entry

$$d_j = \psi'(\alpha_j),\ j = 1,2,\cdots,k$$

c is the constant $\psi'(\alpha_.)$, and 1 is a column vector of all 1's.

Thus the limiting value of $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)$ as $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_k^*)$ gives the empirical Bayes estimate in this case.

4. Computational method when hyperparameters are different for k age groups (i.e the priors are not identical)

Consider the case where α_{ij} 's are unknown parameters, for $i, j = 1,2,\cdots,k$. Here,

$$\alpha = (\alpha_1^T, \alpha_2^T, \cdots, \alpha_k^T)^T = (\alpha_{11}, \cdots, \alpha_{1k}; \alpha_{21}, \cdots, \alpha_{2k}; \cdots; \alpha_{k1}, \cdots, \alpha_{kk})^T$$

Now,let us find out the estimate of the Dirichlet parameter α from observed data ***n***. But it is not possible to start with an estimate α from observed data for each component using equation (3.2) with $k = 1$. In this case of individual state, it is seen that $\hat{\alpha}_{ij} = 1/2$ which is absurd. It is possible to compute iteratively the MLE of α from observed data by using Quasi-Newton accelerated EM algorithm.

Now, we estimate α_i individually, since estimates of $\alpha_i = (\alpha_{i1}, \cdots, \alpha_{ik})$ depend only on $n_i = (n_{i1}, \cdots, n_{ik})$ and independent of $n_j = (n_{j1}, \cdots, n_{jk})$, for $(j \neq i)$.

Thus after iterating each α_i individually and taking the limiting value of α_i as α_i^*, for all $i = 1,2,\cdots,k$, we have finally

$$\alpha^* = (\alpha_1^{*T}, \cdots, \alpha_k^{*T})^T.$$

Let us denote the vector containing ***p*** and ***n***,by ***x*** i.e.

$$x = (n^T, p^T)^T.$$

The variables $(p_{i1}, p_{i2}, \cdots, p_{ik})$ are related by

$$f(n_{i1}, n_{i2}, \cdots, n_{ik} \mid p_{i1}, p_{i2}, \cdots, p_{ik}) = \frac{n_{i\cdot!}}{n_{i1!} n_{i2!} \cdots n_{ik!}} p_{i1}^{n_{i1}} p_{i2}^{n_{i2}}, \cdots, p_{ik}^{n_{ik}} \quad (4.1)$$

for all $i = 1,2,\cdots,k$ and

$$\pi(p_{i1}, p_{i2}, \cdots, p_{ik}) = \frac{\Gamma\left(\sum_{j=1}^{k} \alpha_{ij}\right)}{\prod_{j=1}^{k} \Gamma(\alpha_{ij})} p_{i1}^{\alpha_{i1}-1} p_{i2}^{\alpha_{i2}-1}, \cdots, p_{ik}^{\alpha_{ik}-1}, \ for\ all\ i = 1,2,\cdots,k \quad (4.2)$$

After integrating w.r.t. $(p_{i1}, p_{i2}, \cdots, p_{ik})$ from joint density function of $(n_{i1}, n_{i2}, \cdots, n_{ik})$ and $(p_{i1}, p_{i2}, \cdots, p_{ik})$, from (4.1) and (4.2), the density function of $(n_{i1}, n_{i2}, \cdots, n_{ik}),\ for\ all\ i = 1,2,\cdots,k$ is given by

$$f(n_{i1}, n_{i2}, \cdots, n_{ik} \mid \alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{ik}) = \frac{n_{i\cdot}!}{\prod_{j=1}^{k} n_{ij}!} \frac{\Gamma\left(\sum_{j=1}^{k} \alpha_{ij}\right)}{\prod_{j=1}^{k} \Gamma(\alpha_{ij})} \int_{S_i} \prod_{j=1}^{k} p_{ij}^{n_{ij}+\alpha_{ij}-1} \prod_{j=1}^{k} dp_{ij}$$

where the integration is carried out over the region

$$S_i = \left\{ p_{ij} : p_{ij} \geq 0\ and\ \sum_{j=1}^{k} p_{ij} = 1 \right\},\ for\ all\ i = 1,2,\cdots,k$$

Thus marginal density of $(n_{i1}, n_{i2}, \cdots, n_{ik}),\ for\ all\ i = 1,2,\cdots,k$ becomes

$$f(n_{i1}, n_{i2}, \cdots, n_{ik} \mid \alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{ik}) = \frac{n_{i\cdot!}}{\prod_{j=1}^{k} n_{ij!}} \frac{\Gamma\left(\sum_{j=1}^{k} \alpha_{ij}\right)}{\prod_{j=1}^{k} \Gamma(\alpha_{ij})} \frac{\prod_{j=1}^{k} \Gamma(n_{ij} + \alpha_{ij})}{\Gamma\left(\sum_{j=1}^{k} (n_{ij} + \alpha_{ij})\right)} \quad (4.3)$$

which is Multinomial-Dirichlet distribution.

Now, $(n_{i1}, n_{i2}, \cdots, n_{ik})$ denotes an observed random sample from Multinomial-Dirichlet distribution with parameters $(\alpha_{i1}, \cdots, \alpha_{ik})$ for $i = 1,2,\cdots,k$. Then like incomplete-data (i.e. instead of ***n*** and ***p*** it involves only ***n***) the log likelihood function becomes

$$\log L(\alpha) = \sum_{i=1}^{k} \log L(\alpha_i) = \sum_{i=1}^{k} \log(n_{i\cdot!}) - \sum_{i=1}^{k}\sum_{j=1}^{k} \log(n_{ij}!) + \sum_{i=1}^{k} \log\Gamma(\alpha_{i\cdot})$$

$$-\sum_{i=1}^{k}\sum_{j=1}^{k} \log\Gamma(\alpha_{ij}) + \sum_{i=1}^{k}\sum_{j=1}^{k} \log\Gamma(n_{ij} + \alpha_{ij}) - \sum_{i=1}^{k} \log\Gamma(n_{i\cdot} + \alpha_{i\cdot}) \quad (4.4)$$

where $n_{i\cdot} = \sum_{j=1}^{k} n_{ij}$, and $\alpha_{i\cdot} = \sum_{j=1}^{k} \alpha_{ij}$, for all $i = 1,2,\cdots,k$.

The complete-data log likelihood is

$$\log L_c(\alpha) = \sum_{i=1}^{k} \log L_c(\alpha_i) = \sum_{i=1}^{k} \log(n_{i\cdot!}) - \sum_{i=1}^{k}\sum_{j=1}^{k} \log(n_{ij}!) + \sum_{i=1}^{k} \log\Gamma(\alpha_{i\cdot})$$

$$-\sum_{i=1}^{k}\sum_{j=1}^{k} \log\Gamma(\alpha_{ij}) + \sum_{i=1}^{k}\sum_{j=1}^{k} (n_{ij} + \alpha_{ij} - 1) \log p_{ij}$$

The EM algorithm approaches the problem of solving the incomplete-data likelihood equation (4.4) indirectly by proceeding iteratively in terms of the complete-data log likelihood function $\log L_c(\alpha)$. The obstacle, due to unobservability is overcome by averaging the complete-data likelihood over its conditional distribution given the observed data ***n***. But in order to calculate this conditional expectation, a value for α is to be given.

Let $\alpha_i^{(0)}$ denote the starting value of α_i and $\alpha_i^{(r)}$, the value of α_i on the r th subsequent iteration of the EM algorithm.

E-step: Then on the first iteration of the EM algorithm, the E-step requires the computation of the conditional expectation of $\log L_c(\alpha)$ given ***n***, using $\alpha_i^{(0)}$ for α_i, which can be written as

$$Q(\alpha_i, \alpha_i^{(0)}) = E_{\alpha_i^{(0)}}\{\log L_c(\alpha_i) \mid n_i\}$$

where Q-function is used to denote the conditional expectation of the complete-data log likelihood function, $\log L_c(\alpha)$, given the observed data **n**, and using the current fit for α .Thus, the $(r+1)$ th iteration of the E-step becomes,

$$Q(\alpha_i, \alpha_i^{(r)}) = E_{\alpha_i^{(r)}}\{\log L_c(\alpha_i) \mid n_i\}$$

Therefore, $Q(\alpha_i, \alpha_i^{(r)})$ can be written as

$$Q(\alpha_i, \alpha_i^{(r)}) = \log(n_{i\cdot!}) - \sum_{j=1}^{k} \log(n_{ij}!) + \log\Gamma(\alpha_{i\cdot}) - \sum_{j=1}^{k} \log\Gamma(\alpha_{ij})$$

$$+ \sum_{j=1}^{k} (n_{ij} + \alpha_{ij} - 1) E_{\alpha_i^{(r)}}\{\log p_{ij} \mid n_{i1}, \cdots, n_{ik}\} (4.5)$$

It is seen from (4.5) that, in order to carry out M-step,it is necessary to calculate the term

$$E_{\alpha_i^{(r)}}\{\log p_{ij} \mid n_{i1}, \cdots, n_{ik}\}$$

The calculation of the above term can be avoided if we make use of the identity (Lange (1995b)),

$$S_i(n_i; \alpha_i^{(r)}) = [\delta Q(\alpha_i, \alpha_i^{(r)}) / \delta\alpha_i]_{\alpha_i = \alpha_i^{(r)}}, \ for\ all\ i = 1,2,\cdots,k$$

where $S_i(n_i; \alpha_i^{(r)})$ are the score statistics given by

$$S_i(n_i; \alpha_i^{(r)}) = \delta \log L(\alpha_i^{(r)}) / \delta\alpha_i^{(r)}, \ for\ all\ i = 1,2,\cdots,k$$

On evaluating $S_{ij}(n_i; \alpha_i^{(r)})$, the derivative of (4.4) with respect to α_{ij} at the point $\alpha_i = \alpha_i^{(r)}$,it becomes,

$$S_{ij}(n_i; \alpha_i^{(r)}) = \delta \log L(\alpha_i^{(r)}) / \delta\alpha_{ij}^{(r)} = \delta \log\Gamma(\alpha_{i\cdot}^{(r)}) / \delta\alpha_{ij}^{(r)} - \delta \log\Gamma(\alpha_{ij}^{(r)}) / \delta\alpha_{ij}^{(r)}$$

$$+ \delta \log\Gamma(n_{ij} + \alpha_{ij}^{(r)}) / \delta\alpha_{ij}^{(r)} - \delta \log\Gamma(n_{i\cdot} + \alpha_{i\cdot}^{(r)}) / \delta\alpha_{ij}^{(r)}, \ for\ all\ i, j = 1,2,\cdots,k (4.6)$$

On equating $S_{ij}(n_i; \alpha_i^{(r)})$ equal to the derivative of (4.5) with respect to α_{ij} at the point $\alpha_i = \alpha_i^{(r)}$, the following identity is obtained.

$$E_{\alpha_i^{(r)}} \log(p_{ij} \mid n_{i1}, \cdots, n_{ik}) = \delta \log\Gamma(n_{ij} + \alpha_{ij}^{(r)}) / \delta\alpha_{ij}^{(r)} - \delta \log\Gamma(n_{i\cdot} + \alpha_{i\cdot}^{(r)}) / \delta\alpha_{ij}^{(r)}$$

$$= \psi(n_{ij} + \alpha_{ij}^{(r)}) - \psi(n_{i\cdot} + \alpha_{i\cdot}^{(r)})$$

where

$$\psi(s) = \delta \log \Gamma(s) / \delta s$$

is the digamma function of s.

Therefore from (4.5) one gets,

$$Q(\alpha_i, \alpha_i^{(r)}) = \log(n_{i\cdot}!) - \sum_{j=1}^{k} \log(n_{ij}!) + \log \Gamma(\alpha_{i\cdot}) - \log \Gamma(\alpha_{ij})$$

$$+ \sum_{j=1}^{k} (n_{ij} + \alpha_{ij} - 1)\left[\delta \log \Gamma(n_{ij} + \alpha_{ij}^{(r)}) / \delta \alpha_{ij}^{(r)} - \delta \log \Gamma(n_{i\cdot} + \alpha_{i\cdot}^{(r)}) / \delta \alpha_{ij}^{(r)}\right] \quad (4.7)$$

M-step: On the M-step at the $(r+1)$ th iteration of EM algorithm is to maximise $Q(\alpha_i, \alpha_i^{(r)})$, i.e., $\alpha_{ij}^{(r+1)}$ are obtained after solving

$$\delta Q(\alpha_i, \alpha_i^{(r)}) / \delta \alpha_{ij} = 0, \ \textit{for all} \ i = 1, \cdots, k.$$

It follows that $\alpha_{ij}^{(r+1)}$ is a solution of the equation

$$\psi(\alpha_{i\cdot}) - \psi(\alpha_{ij}) + \psi(n_{ij} + \alpha_{ij}^{(r)}) - \psi(n_{i\cdot} + \alpha_{i\cdot}^{(r)}) = 0$$

The E-step and M-step are alternated repeatedly until the difference $L(\alpha_i^{(r+1)}) - L(\alpha_i^{(r)})$ changes by an arbitrary small amount in the case of convergence of the sequence of likelihood values $\{L(\alpha_i^{(r)})\}$.

Finally,the limiting solutions are obtained.

$$\alpha^* = (\alpha_1^{*T}, \alpha_2^{*T}, \cdots, \alpha_k^{*T})^T,$$

where α_i^* the limiting value of α_i , for all $i = 1, 2, \cdots, k$. Thus , we have the empirical Bayes estimator of p_{ij} as,

$$\delta^{\hat{\pi}}(n_{ij}) = \frac{n_{ij} + \alpha_{ij}^*}{n_{i\cdot} + \alpha_{i\cdot}^*}, \ \textit{for all} \ i, j = 1, 2, \cdots, k$$

where α_{ij}^* is the limiting value of α_{ij} for all $i, j = 1, \cdots, k$, and

$$\alpha_{i\cdot}^* = \sum_{j=1}^{k} \alpha_{ij}^*, \ \textit{for all} \ i = 1, \cdots, k$$

5. Working with real data set

Table 1 corresponds to the data of the age of bride by age of groom for marriages in Israel (Jews Only):1966 (Shryock and Siegel 1976; using Table C-10 of the Israeli Central Bureau of Statistics, *Statistical Abstract of Israel*, 1968, NO. 19). Tables 2 and 3 give the estimates of preferences respectively under two cases as specified earlier. The respective program codes are also given in the appendix.

Table 1. Age of bride by age of groom for marriages in Israel (Jews only):1966

Age of	Age of Bride						
Groom	Under 20	20-24	25-29	30-34	35-39	40-44	Total
Under 20	582	127	10	3	0	0	722
20-24	3,998	4,124	210	17	7	3	8,359
25-29	1,644	2,821	616	75	16	2	5,174
30-34	269	831	486	167	51	11	1,815
35-39	40	170	229	180	80	39	738
40-44	6	37	78	116	128	71	436
Total	6,539	8,110	1,629	558	282	126	17,244

Table 2. Empirical Bayes estimate of marriage preferences when prior parameters are same for all age groups

0.80037666	0.17578907	0.01518083	0.005571793	0.002826353	0.0028263532
0.47821695	0.49328245	0.02529548	0.002218960	0.001023285	0.0005450155
0.31761927	0.54490602	0.11910545	0.014634656	0.003241353	0.0005378573
0.14810246	0.45698180	0.26736726	0.092042510	0.028288055	0.0063037599
0.05433681	0.22942553	0.30888888	0.242893898	0.108210261	0.0529899705
0.01422229	0.08459994	0.17768007	0.263949462	0.291192426	0.1617883444

Table 3. Empirical Bayes estimate of marriage preferences when prior parameters are different for all age groups

0.74930030	0.21119739	0.02756923	0.009261183	0.0025891781	8.271551e-05
0.47821867	0.49324024	0.02520825	0.002074008	0.0008713648	3.874675e-04
0.31788703	0.54465563	0.11928947	0.014578370	0.0031589814	4.305259e-04
0.14834426	0.45717471	0.26767415	0.092213221	0.0283338553	6.259794e-03
0.05473975	0.23010312	0.30941796	0.243557530	0.1087946602	5.338697e-02
0.01464156	0.08563294	0.17901881	0.265222785	0.2923801960	1.631037e-01

6. Concluding remarks

In this paper empirical Bayes estimate of marriage preferences for different age groups has been demonstrated. Even fitting the priors having equal choices, the estimate of the probabilities are not same for different age groups, which is reasonable to guess. This is evident from table 2. But table 3 is realistic. The choices for different age groups are different, as row vectors are different in this table. Moreover, table 3 is interesting e.g. bride within age group 20-24 is most preferable. More importantly, we have worked out in this paper for a class of Dirichlet priors, which are reasonable under two cases where hyperparameters are the same for all age groups and different for all age groups. The preference pattern may be useful in vital statistics and to understand the nature of a family.Using this method together with R-codes, this kind of data may be handled.

References

Israeli Central Bureau of Statistics, *Statistical Abstract of Israel*, 1968, No. 19, Jerusalem, September 1968, table C-10.

Johnson, N.L., and S. Kotz, (1969): *Distributions in Statistics: Discrete Disributions*. John Willy & Sons, New york.

Johnson, N. L., S. Kotz, and N. Balakrishnan, (1994): *Continuous Univariate Distributions*. Vol. 2, John Willy & Sons, New York.

Lange, K. (1995b): "A quasi-newton accleration of the EM algorithm." *Statistica Sinica*, 5, 1-18.

Seal, B. and Hossain, S. J.(2013a): "Testing of hypotheses: LRT test of some important hypothesis for internal migration model". *Slovak Statistics and Demography- Scientific Journal*, Vol. 23, Issue 2, 35-47

Seal, B. and S. J. Hossain, (2013b): "Measure of state's importance and its estimation from migration data". *ProbStat Forum*, Vol. 6, April 2013, pages 10-17

Seal, B. and S. J. Hossain, (2013c): "Bayes and Minimax Estimation of Parameters of Markov Transition Matrix". *ProbStat Forum*, Vol. 6,October 2013.

Shryock, H. S., et al. (1976): The Methods and Materials of Demography. Academic Press, New York.

Appendix

Program for EM Bayes estimator of marriage preferences when all age groups the same prior (Dirichlet) parameter.

The *est.embayes()* function needs an argument. Value of x is the sample transition matrix.

```
est.embayes < - function(x)
{
if(ncol(x)!= nrow(x))
stop("The given matrix is not a square matrix")
k < - ncol(x)
obs.mat < - x
prob.mat < - matrix(0, ncol=k, nrow= k)
Q < - matrix(0, ncol=k, nrow= k)
embayes.est < - matrix(0, ncol=k, nrow= k)
I < - diag(k)
alpha < - rep(0,k)
g < - rep(1,k)
for(i in 1:k){
for(j in 1:k){
if(x[i, j] != 0.0) obs.mat[i, j] < - obs.mat[i, j]
else obs.mat[i, j] < - obs.mat[i, j]+1
prob.mat[i, j] < - obs.mat[i, j]/sum(obs.mat[i, ])
Q[i, j] < - prob.mat[i, j] ^( 1/k)
}
}
for(j in 1:k){
for(i in 1:k){
g[ j] < - g[j]*Q[i, j]
}
}
for(j in 1:k){
alpha[j] < - (k-1)/2*g[j]/(1-sum(g))+0.5
}
cat("Initial value of dirichlet parameters")
print(alpha) repeat{
fisher.inf < - matrix(0, ncol=k, nrow= k)
inv.finf < - matrix(0, ncol=k, nrow= k)
S < - rep(0,k)
if(any(alpha < = 0.0001)) alpha < - alpha - min(alpha)+1
for(i in 1:k){
for(j in 1:k){
if(j!= i) fisher.inf[i, j] < - -k*trigamma(sum(alpha))
else fisher.inf[i, j] < - k*trigamma(alpha[i]) -k*trigamma(sum(alpha))
}
S[i] < - k*digamma(sum(alpha)) -k*digamma(alpha[i]) + sum(log(prob.mat[ ,i]))
}
if(det(fisher.inf) < = 0.001) fisher.inf < - fisher.inf+I
else fisher.inf < - fisher.inf
inv.finf < -solve(fisher.inf)
```

```
c < -inv.finf%*%S
alpha < -alpha+c
if(any(abs( c) < 0.00001, na.rm=F)) break
}
cat("Estimated value of Dirichlet parameters")
print(alpha)
for(i in 1:k){
for(j in 1:k){
embayes.est[i, j] < -(obs.mat[i, j]+ alpha[i])/(sum(obs.mat[i,])+sum(alpha))
}
}
cat("Empirical Bayes estimate of marriage preferences when prior parameters are same for all age groups")
print(embayes.est)
}
```

Program for empirical Bayes estimator of marriage preferences when all age groups have different prior (Dirichlet) parameters using EM algorithm.

The *embayes.est()* function needs an argument. Value of "N" is the sample transition matrix and "alpha" is the initial value of the dirichlet parameters.

```
embayes.est < - function(N, alpha)
{
if(ncol(N)!=nrow(N))
stop("Given matrix is not a square one")
if(ncol(alpha)!=nrow(alpha))
stop("Given parameter matrix is not asquare one")
if(ncol(N)!=ncol(alpha))
stop("Given matrices are not of same dimention")
k < -ncol(N)
bayes.est < - matrix(0,ncol=k,nrow=k)
embayes.est < - matrix(0,ncol=k,nrow=k)
fisher.inf < - matrix(0,ncol=k,nrow=k)
inv.finf < - matrix(0,ncol=k,nrow=k)
B0 < - diag(k)
B < - matrix(0,ncol=k,nrow=k)
D < - matrix(0,ncol=k,nrow=k)
P < - matrix(0,ncol=k,nrow=k)
S < - rep(0,k)
d < - rep(0,k)
v < - rep(0,k)
h < - rep(0,k)
hd < - matrix(0,100,k)
w < - rep(1,k)
z < - 0
g < - 0
m < - 0
u < - rep(0,k)
eps < - 0.00001
#Loop for Bayes estimate#
for(i in 1:k){
```

```
for(j in 1:k){
bayes.est[i,j]=(N[i,j]+alpha[i,j])/(sum(N[i,])+sum(alpha[i,]))
}
}
cat("Bayes estimate of transition pobability matrix")
print(bayes.est)
#To calculate Fisher information matrix#
for(i in 1:k){
for(j in 1:k){
D[j,j] < - trigamma(alpha[i,j])-trigamma(N[i,j]+alpha[i,j])
u[j] < - digamma(alpha[i,j])-digamma(N[i,j]+alpha[i,j])
hd[1,j] < - digamma(N[i,j]+alpha[i,j])-digamma(sum(N[i,])+sum(alpha[i,]))
}
z < - trigamma(sum(alpha[i,]))-trigamma(sum(N[i,])+sum(alpha[i,]))
g < - digamma(sum(alpha[i,]))-digamma(sum(N[i,])+sum(alpha[i,]))
S < - g-u
fisher.inf < - D-z*(w%*%t(w))
inv.finf < - solve(fisher.inf+B0)
d < - inv.finf%*%S
alpha[i,]<-alpha[i,]+d
if(any(abs(d) < eps,na.rm=F)) break
for(r in 2:100){
if(any(alpha[i, ] <= 0.0)) alpha[i, ] < - alpha[i, ]-min(alpha[i,])+1
for(j in 1:k){
D[j,j] < - trigamma(alpha[i,j])-trigamma(N[i,j]+alpha[i,j])
u[j] < - digamma(alpha[i,j])-digamma(N[i,j]+alpha[i,j])
hd[r,j] < - digamma(N[i,j]+alpha[i,j])-digamma(sum(N[i,])+sum(alpha[i,]))
}
z < - trigamma(sum(alpha[i,]))-trigamma(sum(N[i,])+sum(alpha[i,]))
g < - digamma(sum(alpha[i,]))-digamma(sum(N[i,])+sum(alpha[i,]))
S < - g-u
fisher.inf < - D-z*(w%*%t(w))
h < - hd[r, ]-hd[r-1, ]
v < - h+B0%*%d
m < - t(v)%*%d
q < - matrix(m, ncol=k, nrow=k)
P < - v%*%t(v)
B < - B0+P/q inv.finf < - solve(fisher.inf+B)
d < - inv.finf%*%S
alpha[i,] < - alpha[i,]+d
if(any(abs(d) < eps, na.rm=F)) break
}
}
cat( "Estimate of dirichlet parameter.")
print(alpha)
for(i in 1:k){
for(j in 1:k){
embayes.est[i, j] < - (N[i, j]+alpha[i, j])/(sum(N[i, ])+sum(alpha[i, ]))
}
}
cat("Empirical Bayes estimate of marriage preferences when prior parameters are different for all
```

```
age groups")
print(embayes.est)
}
```