



## Empirical Bayes Estimation in Nonstationary Markov chains

M. R. Meshkani<sup>†,\*</sup> and L. Billard<sup>‡</sup>

<sup>†</sup>Shahid Beheshti University

<sup>‡</sup>University of Georgia

**Abstract.** Estimation procedures for nonstationary Markov chains appear to be relatively sparse. This work introduces empirical Bayes estimators for the transition probability matrix of a finite nonstationary Markov chain. The data are assumed to be of a panel study type in which each data set consists of a sequence of observations on  $N \geq 2$  independent and identically distributed chains recorded collectively.

**Keywords.** Bayes estimates; empirical Bayes estimates; natural conjugate priors; nonstationary Markov chains.

### 1 Introduction

Markov chains have been with us for a long time and provided the transition probability matrix (t.p.m.) and the initial probability vector are known, a wealth of results can be found. However, there are relatively few results available when it comes to estimating that t.p.m.

A basic study in which maximum likelihood estimates are obtained is provided by Anderson and Goodman (1957). However, there are many situations when the available data are ideally suited to the use of empirical Bayes procedures. Typically, the "experimenter" has at his disposal  $(n + 1)$ , say, sets of

---

\* Corresponding author.

data each being a sequence of observations from a Markov chain process. These  $(n + 1)$  data sets may occur concurrently or sequentially. By using the sets of "past" data, estimates of unknown prior parameters can be found. These in conjunction with the "current" data set allow us to obtain the so-called empirical Bayes estimates for the t.p.m.

Preston (1971) has briefly considered this problem for a stationary two-state Markov chain. We propose here to derive some empirical Bayes estimators for the more general case of  $s \geq 2$  states and for a non-stationary Markov chain. The related questions pertaining to the properties of such estimators are being deferred at this time.

There are numerous situations, especially in the medical and social sciences, where the t.p.m. is not constant over time. For example, consider a process regarding changes in people's behavior when a definite decision about some issue (such as voting on a proposal or electing a candidate) must be made by a due date. It seems more realistic that the t.p.m. of the penultimate week will be different from those of earlier weeks.

There are examples in other areas, too. Suppose one is looking for an object in a place (such as a document in a file), which is certain to be present there. Assume initially that the searcher has a certain probability of finding the object on each search, and that he continues to search until he finds it. Now, consider a tiring search where the searcher tires as the search progresses. The same process might serve as a model for a student trying to learn but becoming discouraged, Howard (1971). In these cases, it seems reasonable to assume that the t.p.m. will change over time.

A concrete example of this model is the people's intention to vote for various presidential candidates. If one monitors the intention of people in various geographical regions and/or ethnic groups as the candidates' campaign go on then clearly one has a set of independent processes evolving which culminate to the final election day.

Estimation for Markov Processes, specially for stationary ones, has been treated by many authors; Billingsley (1961), Basawa and Prakaso Rao (1980), Bhat and Miller (2002), Billard and Meshkani (1992), Meshkani and Billard (1995) are among the others. Works on non-stationary processes are rare, specially from Bayesian perspective.

We treat this problem in the context of panel studies. By "panel study", we mean that each data set consists of a sequence of observations on  $N \geq 2$  independent and identically distributed chains recorded collectively. No specific functional relation between the t.p.m.,  $\Lambda(t)$ , will be assumed. The only stipulation will be that if observations are made at each time  $t \in \tau = \{1, \dots, T\}$ ,

$$\Lambda(t) \neq \Lambda \quad \text{for all } t \in \tau,$$

and that  $\Lambda(t)$ ,  $t \in \tau$ , are independently distributed.

The basic probability results are obtained in Section 2. From these, we are able to derive the Bayes estimate for  $\Lambda(t)$ , given in Section 3. For this, the natural conjugate priors are used. Then, the empirical Bayes estimates are derived in Section 4. Section 5 presents the optimality properties of the estimators, providing the asymptotic distributions.

## 2 Theoretical Moments

### 2.1 Some Definitions and Assumptions

Let  $\tau = \{1, \dots, T\}$  and  $\tau_0 = \{0\} \cup \tau$ . Let  $\{X(t), t \in \tau_0\}$  be a simple Markov chain of states with transition probability matrix (t.p.m.)  $\Lambda(t)$  with elements  $\Lambda_{jk}(t)$ ,  $j, k \in S$  for  $S = \{1, \dots, s\}$ . We shall refer to the data sets  $i \in N$ , for  $N = \{1, \dots, n\}$  as the "past data" and the set  $i = n + 1$  as the "current data". Let  $N_1 = \{n + 1\} \cup N$ .

**Definition 1** *The frequency count vector (f.c.v.),  $\mathbf{G}(t)$ , is the vector whose elements,  $G_j(t)$ ,  $j \in S$ , are the number of individuals in state  $j$  at time  $t$ ,  $t \in \tau_0$ .*

For reasons that will become clear later, at  $t = 1$ , we need to have  $\mathbf{G} \neq \mathbf{0}$ . This is not a confining restriction as this is usually the case in panel studies where  $N$  is generally large and the initial distribution  $\boldsymbol{\theta} = (\theta_j)$ ,  $j \in S$ , is such that  $\theta_j \neq 0$ .

**Definition 2** *The frequency count matrix (f.c.m.),  $\mathbf{F}(t)$ , is the matrix whose elements,  $F_{jk}(t)$ ,  $j, k \in S$ , are the number of individuals moving from state  $j$  at time  $t - 1$  to state  $k$  at time  $t$ .*

Note there will be a f.c.v.  $\mathbf{G}_i(t)$  and f.c.m.  $\mathbf{F}_i(t)$ , for each  $i \in N_1$ . Ultimately we are interested in some moments of the  $\mathbf{F}_i(t)$ . Hence, it is sufficient to consider the properties of  $\mathbf{F}_{n+1}(t) \equiv \mathbf{F}(t)$ .

From Anderson and Goodman (1957), it follows that the conditional distribution of  $\mathbf{F}(t)$  for a given t.p.m.  $\Lambda(t)$  and initial f.c.v.  $\mathbf{G}(0)$ , is

$$P\{\mathbf{F}(t)|\Lambda(t), \mathbf{G}(0), t \in \tau\} = A_0\{\mathbf{F}(t)\} \prod_{t \in \tau} \prod_{j, k \in S} \Lambda_{jk}^{F_{jk}(t)}, \quad (1)$$

where

$$A_0\{\mathbf{F}(t)\} = \prod_{t \in \tau} \left\{ \prod_{j \in S} \frac{F_{j+}(t)!}{\prod_{k \in S} F_{jk}(t)!} \right\},$$

with

$$F_{j+}(t) = \pi_{k \in S} F_{jk}(t).$$

The distribution of  $\mathbf{\Lambda}(t)$  will be assumed to follow the natural conjugate prior, viz., the matrix beta distribution  $MB[p(t)]$ ,  $t \in \tau$ . That is,

$$q\{\mathbf{\Lambda}(t)\} = \pi_{j \in S} \left[ \Gamma(p_{j+}(t)) \pi_{k \in S} \left\{ \frac{\Lambda_{jk}^{p_{jk}(t)-1}(t)}{\Gamma(p_{jk}(t))} \right\} \right], \tag{2}$$

with

$$p_{j+}(t) = \sum_{k \in S} p_{jk}(t).$$

Throughout, we shall use independent priors for  $\mathbf{\Lambda}(t)$ ,  $t \in \tau$ . The reason for this assumption is twofold. First, we do not have a clear idea about the variation of  $\mathbf{\Lambda}(t)$  in  $t$  to incorporate it appropriately in the model. Secondly, it provides mathematical tractability in this initial attempt of the subject.

### 2.2 Moments of $F(t)$ and Some Functions of it

Now, from the definitions of  $\mathbf{G}(t)$  and  $\mathbf{F}(t)$ , we can readily see that

$$\mathbf{G}(t-1) = \mathbf{F}(t) \cdot \mathbf{1}, \tag{3}$$

$$\mathbf{G}(t) = \mathbf{F}'(t) \cdot \mathbf{1}, \tag{4}$$

$$\mathbf{1}'\mathbf{G}(t) = \mathbf{1}'\mathbf{F}(t)\mathbf{1} = N, \quad t \in \tau. \tag{5}$$

These relationships suggest it is more practical to establish a recursive formula in  $t$  for  $E\{\mathbf{F}(t)\}$ . Let

$$\mathbf{L}(t) = E\{\mathbf{\Lambda}(t)\}, \quad t \in \tau, \tag{6}$$

with elements

$$l_{jk}(t) = \frac{p_{jk}(t)}{p_{j+}(t)}, \quad j, k \in S.$$

let  $\mathbf{L}(u, \nu)$ , for  $u, \nu \in \tau$ , be a matrix with elements

$$l_{jk}(u, \nu) = \begin{cases} (j, k)\text{th element of } \frac{\nu}{t=u} \mathbf{L}(t), & u < \nu \\ l_{jk}(\nu), & u = \nu \\ \delta_{jk}, & u > \nu, \end{cases} \tag{7}$$

where  $\delta_{jk}$  is the Kronecker's delta.

Let

$$\phi(t) = E\{\mathbf{F}(t)|\mathbf{G}(0)\}, \quad t \in \tau, \tag{8}$$

with elements

$$\phi_{jk}(t) = E\{F_{jk}(t)|\mathbf{G}(0)\}, \quad k \in S,$$

and let

$$\Psi(t) = E\{\mathbf{G}(0)\}, \quad t \in \tau, \quad (9)$$

where the elements  $\psi_j(t)$  of the vector  $\Psi(t)$  are given by

$$\psi_j(t) = \begin{cases} G_j(0), & t = 1 \\ G'_j(0) \sum_{k \in S} \mathbf{L}_{.k}(1) l_{kj}(2, t-1), & t \geq 2, \end{cases}$$

where  $\mathbf{L}_{.k}(t)$  is the  $k$ th column of  $\mathbf{L}(t)$ .

Later in the process of estimation of the hyperparameters of the prior distribution, we shall need some explicit equations for the moments of  $\mathbf{F}(t)$ .

Taking expectation of (8) leads to

$$\phi(t) = \text{diag}\{\psi_j(t-1)\} \mathbf{L}(t), \quad t \in \tau$$

which is too involved to be useful in our estimation procedure. We shall use the following simpler expressions given in Theorem (1) below.

Let us first define

$$\mathbf{M}(t) = \text{diag}\{F_{j+}^{-1}(t)\} \cdot \mathbf{F}(t), \quad t \in \tau, \quad (10)$$

with elements

$$M_{jk}(t) = \frac{F_{jk}(t)}{F_{j+}(t)}, \quad j, k \in S.$$

we note that since each f.c.v.  $\mathbf{G}(t-1)$  is assumed to be non-zero,  $F_{j+}(t) \neq 0$  for all  $j \in S$ . Hence, (10) is well-defined. We also define the matrix  $\mathbf{Z}_j(t)$  with elements

$$Z_{j;kh}(t) = \frac{F_{jk}(t)\{F_{jh}(t) - \delta_{kh}\}}{\{F_{j+}(t) - 1\}} \quad j, k, h \in S, \quad t \in \tau. \quad (11)$$

**Theorem 1** Let  $\mathbf{M}(t) = \{M_{jk}(t)\}$ ,  $t \in \tau$ , be defined as in (10). Then, for  $j, k, g, h \in S$ ,

$$E\{M_{jk}(t)\} = l_{jk}(t), \quad t \in \tau, \quad (12)$$

and

$$\text{cov}\{M_{jk}(t), M_{gh}(t)\} = \delta_{jg} \cdot \frac{p_{jk}(t)\{\delta_{kh}p_{j+}(t) - p_{jh}(t)\}}{p_{j+}^2(t)\{p_{j+}(t) + 1\}} [1 + E\{G_j^{-1}(t-1)\}]. \quad (13)$$

**Proof.** We have for  $j, k \in S, t \in \tau,$

$$E\{M_{jk}(t)\} = E_2 \left[ E_1 \left\{ \frac{F_{jk}(t)}{F_{j+}(t)|G_j(t-1)} \right\} \right] \\ = E_2\{\Lambda_{jk}(t)\} = l_{jk}(t),$$

by definition (6) giving the required result. Here, the subscript 1 indicates the expectations is for a given  $\Lambda(t)$  and 2 means the expectation is taken with respect to the distribution of  $\Lambda(t)$ .

For the conditional covariance we have,

$$\text{cov}_1\{M_{jk}(t), M_{gh}(t)\} = E_1 \left\{ \frac{F_{jk}(t)}{G_j(t-1) - \Lambda_{jk}(t)} \cdot \frac{F_{gh}(t)}{G_g(t-1) - \Lambda_{gh}(t)} \right\}. \quad (14)$$

when  $j = g,$  (14) becomes

$$\text{cov}_1\{M_{jk}(t), M_{jh}(t)\} = \Lambda_{jk}(t)\{\delta_{kh} - \Lambda_{jh}(t)\}E\{G_j^{-1}(t-1)|\mathbf{\Lambda}(u), u \leq t-1\}.$$

Thus, the result (13) for  $j = g$  follows. When  $j \neq g,$  (14) becomes

$$E_1[G_j^{-1}(t-1)G_g^{-1}(t-1)E_1\{(F_{jk}(t) - G_j(t-1)\Lambda_{jk}(t)) \\ \cdot (F_{gh}(t) - G_g(t-1)\Lambda_{gh}(t))|G_j(t-1)G_g(t-1)\}] \\ = E_1[G^{-1}(t-1)G_g^{-1}(t-1) \cdot E_1\{F_{jk}(t) - G_j(t-1)\Lambda_{jk}(t)|G_j(t-1)\} \\ \cdot E_1\{F_{gh}(t) - G_g(t-1)\Lambda_{gh}(t)|G_g(t-1)\}].$$

Clearly, this is zero. Also,

$$\text{cov}_2[E_1\{M_{jk}(t)\}, E_1\{M_{gh}(t)\}] = \text{cov}_2\{\Lambda_{jk}(t), \Lambda_{gh}(t)\} = 0, \quad j \neq g$$

Hence, the required result follows.

### 3 Bayes Estimate of $\Lambda$

#### 3.1 The Loss Function

As is usual, we shall assume an squared error loss function for the estimation of  $\Lambda(t)$  by  $\tilde{\Lambda}(t) = \{\tilde{\Lambda}_{jk}(t)\}, j, k \in S.$  Following DeGroot (1970), the corresponding loss function for the matrix  $\Lambda(t)$  is given by

$$L\{\tilde{\Lambda}(t)\} = \sum_{j,k \in S} \{\tilde{\Lambda}_{jk}(t) - \Lambda_{jk}(t)\}^2.$$

It can be shown that  $L\{\Lambda(t), \tilde{\Lambda}(t)\}$  is minimized when each  $\Lambda_{jk}(t), j, k \in S,$  achieves least possible risk. Thus, the Bayes estimate for each  $\Lambda_{jk}(t), j, k \in S,$  will give us the Bayes estimate for  $\Lambda(t).$  However, the Bayes estimate of  $\Lambda(t)$  relative to the squared error loss function is just the posterior mean of  $\Lambda(t).$

### 3.2 The Bayes Estimate

**Theorem 2** Let  $\mathbf{F}(t)$  be the f.c.m. of a collection of i.i.d. non-stationary Markov chains at time  $t$ . Let  $\mathbf{\Lambda}(t)$  have a  $MB\{\mathbf{p}(t)\}$  prior distribution independent of  $\mathbf{\Lambda}(u)$ ,  $u \neq t$ . Then the posterior distribution of  $\mathbf{\Lambda}(t)$  given  $\mathbf{F}(t)$  is a  $MB\{\mathbf{F}(t) + \mathbf{p}(t)\}$ .

**Proof.** From Anderson and Goodman (1957), we have, for each  $t \in \tau$ ,

$$q^*\{\mathbf{\Lambda}(t)|\mathbf{F}(t), \mathbf{G}(t-1)\} = \left[ \frac{B\{\mathbf{p}(t)\}}{B\{\mathbf{p}(t) + \mathbf{F}(t)\}} \right] \prod_{j,k \in S} \Lambda_{jk}^{F_{jk}(t) + p_{jk}(t) - 1}(t)$$

$$\mathbf{\Lambda}(t) \in \Omega_S, \quad \mathbf{F}(t) \in F(t),$$

where

$$B\{\mathbf{p}(t)\} = \prod_{j \in S} \left[ \frac{\Gamma\{p_j + (t)\}}{\prod_{k \in S} \Gamma\{p_{jk}(t)\}} \right]$$

and  $F(t)$  is the set of all matrixes  $\mathbf{F}(t)$  satisfying (3), (4) and (5).

Since the righthand side does not depend on  $\mathbf{G}(t-1)$ , the desired result follows readily.

**Theorem 3** Let  $F(t)$  be the f.c.m. of a collection of  $N$  i.i.d. non-stationary Markov chains at time  $t \in \tau$ . Let  $\mathbf{\Lambda}(t)$  have a  $MB\{\mathbf{p}(t)\}$  prior distribution. For an initial f.c.v.  $\mathbf{G}(0)$ , the Bayes estimate of  $\mathbf{\Lambda}(t)$  relative to the squared error loss function is

$$\mathbf{\Lambda}_B(t) = \mathbf{\Lambda}_B\{\mathbf{F}(t), \mathbf{p}(t)\} = \{\Lambda_{B;jk}(t)\} \quad (15)$$

where

$$\Lambda_{B;jk}(t) = \frac{F_{jk}(t) + p_{jk}(t)}{F_{j+}(t) + p_{j+}(t)}, \quad j, k \in S.$$

**Proof.** Since we only need to find the posterior mean of each  $\Lambda_{jk}(t)$ ,  $j, k \in S$ , the result follows immediately from the properties of  $MB\{\mathbf{F}(t) + \mathbf{p}(t)\}$ .

Anderson and Goodman (1957) gave the maximum likelihood estimate of  $\Lambda_{jk}(t)$  as  $\frac{F_{jk}(t)}{F_{j+}(t)}$ ,  $j, k \in S$ . Thus the Bayes estimate of  $\mathbf{\Lambda}(t)$  is a convex combination of the maximum likelihood estimate and  $E\{\mathbf{\Lambda}(t)\}$ .

The posterior covariance of  $\Lambda_{jk}(t)$  and  $\Lambda_{j'k'}(t)$  is obtained as

$$\text{cov}\{\Lambda_{jk}(t), \Lambda_{j'k'}(t)|\mathbf{F}(t)\} =$$

$$\begin{cases} \frac{[F_{jk}(t)+p_{jk}(t)]\{\delta_{kk'}[F_{j+}(t)+p_{j+}(t)]-[F_{j'k'}(t)+p_{j'k'}(t)]\}}{[F_{j+}(t)+p_{j+}(t)]^2[F_{j+}(t)+p_{j+}(t)+1]} & j = j', \quad k, k' \in S \\ 0 & j \neq j', \quad k, k' \in S \end{cases} \quad (16)$$

where  $\delta_{kk'}$  is the Kronecker's delta.

## 4 Empirical Bayes Estimate of $\Lambda(t)$

### 4.1 Preliminaries

If we knew the value of  $\mathbf{p}(t)$ , then we would use (15) as the best estimate of  $\Lambda(t)$  and incur the least possible risk relative to the squared error loss function. Not knowing  $\mathbf{p}(t)$ , we shall estimate it from the "past data" and substitute it in (15). We shall use the method of moments to estimate  $\mathbf{p}(t)$ .

We recall that the availability of  $\{\mathbf{F}_i(t) : i \in N\}$ , a set of realizations of  $n$  i.i.d. random matrices is assumed. For each  $t \in \tau$ , the p.m.f. of  $\mathbf{F}(t)$  contains  $s^2$  parameters, namely, elements of  $\mathbf{p}(t)$ . In the following subsection, we give a procedure to estimate them from  $\{\mathbf{F}_i(t) : i \in N\}$ .

### 4.2 Method of moments estimate of $\mathbf{p}(t)$

We may rewrite (12) and (13) as

$$\frac{p_{jk}(t)}{p_{j+}(t)} = E\{M_{jk}(t)\}, \quad j, k \in S \quad (17)$$

and

$$\begin{aligned} p_{j+}(t)[E\{Z_{j;kh}(t)\} - E\{M_{jk}(t)\} \cdot E\{M_{jk}(t)\}] \\ = \delta_{kh} E\{M_{jk}(t)\} - E\{Z_{j;kh}(t)\}, \quad j \in S. \end{aligned} \quad (18)$$

Since (17) is actually  $s(s - 1)$  independent equations, we need (18) to furnish the remaining  $s$  equations.

We introduce the matrices  $\mathbf{\Gamma}_j(t) = \{\gamma_{j;kh}(t)\}$  and  $\mathbf{\eta}_j(t) = \{\eta_{j;kh}(t)\}$  where the elements  $\gamma_{j;kh}(t)$  and  $\eta_{j;kh}(t)$  are defined for  $j, k, h \in S, t \in \tau$ , by

$$\gamma_{j;kh}(t) = E\{Z_{j;kh}(t)\} - E\{M_{jk}(t)\} \cdot E\{M_{jk}(t)\}, \quad (19)$$

$$\eta_{j;kh}(t) = \delta_{kh} E\{M_{jk}(t)\} - E\{Z_{j;kh}(t)\}, \quad (20)$$

respectively.

Since

$$\sum_{h \in S} \gamma_{j;kh}(t) = \sum_{h \in S} \eta_{j;kh}(t) = 0, \quad (21)$$



$\Gamma(t)$  and  $\boldsymbol{\eta}(t)$  are singular. Let  $\Gamma_j^*(t)$  and  $\boldsymbol{\eta}_j^*(t)$  denote the matrices of the first  $(s-1)$  rows and columns of  $\Gamma_j(t)$  and  $\boldsymbol{\eta}_j(t)$ , respectively.

Then,

$$p_{j+}(t) = \left\{ \frac{|\boldsymbol{\eta}_j^*(t)|}{|\Gamma_j^*(t)|} \right\}^{\frac{1}{(s-1)}}, \quad j \in S \quad (22)$$

Now, we define the statistics, for  $j, k, h \in S$ ,  $t \in \tau$ ,

$$C_{j;kh}(t) = \bar{Z}_{j;kh}(t) - \bar{M}_{jk}(t) \cdot \bar{M}_{jh}(t),$$

and

$$d_{j;kh}(t) = \delta_{kh} \bar{M}_{jk}(t) - \bar{Z}_{j;kh}(t),$$

as  $(k, h)$ th elements of  $\mathbf{C}_j(t)$  and  $\mathbf{D}_j(t)$ , respectively where \* has the same meaning as in (22). The system of equations which provides the estimates of  $p_{jk}(t)$ , denoted by  $r_{jk}(t)$ , for each  $t \in \tau$ , is

$$r_{j+}(t) = \left\{ \frac{|\mathbf{D}_j^*(t)|}{|\mathbf{C}_j^*(t)|} \right\}^{\frac{1}{(s-1)}}, \quad j \in S, \quad (23)$$

and

$$r_{jk}(t) = r_{j+}(t) \bar{M}_{jk}(t), \quad j, k \in S. \quad (24)$$

The EB estimate of  $\Lambda(t)$  obtained by the method of moments is the matrix  $\Lambda_{EB}(t)$  whose elements  $\Lambda_{EB;jk}(t)$ ,  $j, k \in S$ , are given by

$$\Lambda_{EB;jk}(t) = \frac{F_{jk}(t) + r_{jk}(t)}{F_{j+}(t) + \Gamma_{j+}(t)}.$$

We note  $\Lambda_{EB}(t)$  may also be written in matrix notation as

$$\Lambda_{EB}(t) = \text{diag}\{\{F_{j+}(t) + r_{j+}(t)\}^{-1}\} \{\mathbf{F}(t) + \mathbf{R}(t)\}, \quad (25)$$

where  $\mathbf{R}(t)$  has elements  $r_{jk}(t)$ ,  $j, k \in S$ .

### 4.3 Maximum Likelihood Estimate of $\mathbf{p}(t)$

The joint p.m.f. of  $\mathbf{F}(t)$ ,  $t \in \tau$ , is derived from (1) by integrating it with respect to the joint distribution of  $\Lambda(t)$ ,  $t \in \tau$ .

That is,

$$p\{\mathbf{F}(t) | \mathbf{G}(0), t \in \tau\} = A_0\{\mathbf{F}(t)\} B_0\{\mathbf{p}(t), \mathbf{F}(t)\}, \quad (26)$$

where

$$B_0\{\mathbf{p}(t), \mathbf{F}(t)\} = \prod_{t \in \tau} \prod_{j \in S} \left[ \frac{\Gamma\{p_{j+}(t)\} \prod_{k \in S} \Gamma\{p_{jk}(t) + F_{jk}(t)\}}{\Gamma\{p_{j+}(t) + F_{j+}(t)\} \prod_{k \in S} \Gamma\{p_{jk}(t)\}} \right].$$

Since  $A_0\{F(t)\}$  is free of  $p_{jk}(t)$ , then  $L\{\mathbf{p}(t)|\mathbf{F}(t)\}$ ,  $t \in \tau$ , the likelihood of  $\mathbf{F}_i(t)$ ,  $i \in N$  and  $t \in \tau$ , is proportional to  $B_0\{\mathbf{p}(t), \mathbf{F}(t)\}$ .

However, we can show that

$$L\{\mathbf{p}(t)|\mathbf{F}(t)\}, t \in \tau = \prod_{j \in S} L\{p_{j+}(t)|F_{j+}(t)\}, t \in \tau,$$

and that it suffices to maximize  $L\{p_{j+}(t)|F_{j+}(t)\}$ ,  $t \in \tau$  for each  $j \in S$ . Details of this maximization can be found in Meshkani (1978).

Let the final solution of the maximization problem be  $\hat{\mathbf{p}}(t)$ . Then, we have:

The *EB* estimate of  $\Lambda(t)$  obtained by the method of maximum likelihood is the matrix  $\hat{\Lambda}_{EB}(t)$  whose elements  $\hat{\Lambda}_{EB;jk}(t)$ ,  $j, k \in S$ , are given by

$$\hat{\Lambda}_{EB;jk}(t) = \frac{F_{jk}(t) + \hat{p}_{jk}(t)}{F_{j+}(t) + p_{j+}(t)}, \quad j, k \in S. \tag{27}$$

The posterior variance corresponding to (16) is

$$\text{cov}\{\Lambda_{jk}(t), \Lambda_{j'k'}(t)|\mathbf{F}(t)\} = \begin{cases} \frac{[F_{jk}(t) + \hat{p}_{jk}(t)][\delta_{kk'}(F_{j+}(t) + \hat{p}_{j+}(t)) - [F_{j'k'}(t) + \hat{p}_{j'k'}(t)]]}{[F_{j+}(t) + \hat{p}_{j+}(t)]^2 [F_{j+}(t) + \hat{p}_{j+}(t) + 1]} & j = j', \quad k, k' \in S \\ 0 & j \neq j', \quad k, k' \in S \end{cases} \tag{28}$$

### 5 Optimality Properties of the EB Estimators

The *EB* estimators (27) and their covariance estimate (28) are continuous functions of the moment estimators of hyperparameters  $\hat{p}_{jk}(t)$ ,  $j, k \in S$ .

Since moment estimators are consistent, i.e.  $\hat{p}_{jk}(t) \xrightarrow{p} p_{jk}(t)$  by straightforward arguments we can claim that  $\hat{\Lambda}_{EB;jk}(t) \xrightarrow{p} \Lambda_B(t)$ . That is, the *EB* estimators (27) are asymptotically optimal.

The same argument applies to (28). With regard to the distribution of *EB* estimators (27), we note that, if one ab initio uses the Dirichlet prior with hyper parameters  $\hat{p}_{jk}(t)$ , then the posterior distribution of  $\hat{\Lambda}_{EB}(t)$  turns out to be a  $MB\{\mathbf{F}(t) + \hat{\mathbf{p}}(t)\}$  in the spirit of theorem 2.

However, one can appeal to the asymptotic properties of  $\hat{p}(t)$  and state the following result.

Downloaded from jsri.srtc.ac.ir at 17:05 +0330 on Monday March 1st 2021 [ DOI: 10.18869/acadpub.jsri.2.1.77 ]

Here, again we are dealing with continuous functions of various sample moments of  $F_{jk}(t)$ . Appealing to a generalized version of the central limit theorem followed by the  $\delta$ -method, Meshkani (1978, chapter 4) has shown that

**Theorem 4** *Let  $\mathbf{\Lambda}_{EB}(t)$  be an EB estimator of  $\mathbf{\Lambda}(t)$ , obtained by the method of moments, given in (27), then, each row of  $\mathbf{\Lambda}_{EB}(t)$  is asymptotically multivariate normal, i.e.,*

$$\text{vec}\mathbf{\Lambda}'_{EB}(t) \sim N\{\text{ve}\mathbf{\Lambda}'_B(t), n^{-1} \sum_R \hat{\Sigma}_R(t)c\} \quad (29)$$

where

$$\hat{\Sigma}_R(t) = \text{diag} \left\{ \frac{r_{j+}^2(t)}{(r_{j+}(t) + F_{j+}(t))^2 \cdot \hat{\Sigma}_{jj}(t)} \right\}$$

and  $\hat{\Sigma}_{jj}(t)$  is the sample covariance of  $j$ th row of  $\mathbf{M}(t) = \text{diag}\{F_{j+}^{-1}(t)\}_{t \in \tau} \mathbf{F}(t)$ .

## Conclusions

We have obtained Bayes and empirical Bayes estimators for the t.p.m. of a nonstationary finite Markov chain. This is just a first step in this direction. Properties of these estimators, as well as asymptotic properties have been considered. The asymptotic normal distribution has been derived.

It would also be of interest to consider the situation of a single chain as distinct from the panel study considered here. A difference in the distributions of the frequency counts  $\mathbf{F}(t)$  necessitates that this case be considered separately.

## References

- Anderson, T.W.; Goodman, L.A. (1957). Statistical inference about Markov chains. *Ann. Math. Statist.* **28**, 89-110.
- Basawa, I.V.; Prakasa Rao, B.L.S. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, New York.
- Bhat, U.N.; Miller, G.K. (2002). *Elements of Applied Stochastic Processes 3rd ed.* Wiley, New York.
- Billard, L.; Meshkani, M.R. (1992). Empirical Bayes estimation for a finite Markov Chain. *Biometrika*, **79**, 185-193.

Billingsley, P. (1961). *Statistical Inference for Markov Processes*, The University of Chicago Press, Chicago.

DeGroot, M.H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

Howard, R.A. (1971). *Dynamic Probabilistic Systems*, vol. 1: Markov Models. Wiley, New York.

Meshkani, M.R. (1978). *Empirical Bayes estimation of transition probabilities for Markov chains*. Ph.D. dissertation, Florida State University.

Meshkani, M.R.; Billard, L. (1995). Estimation of a finite Markov chain. *J. Amer. Statist. Assoc.*, **90**, 224-243.

Preston, P.F. (1971). An empirical Bayes problem with a Markovian parameter. *Biometrika* **58**, 535-543.

| Received: Jun 19, 2005

**M. R. Meshkani**

Department of Statistics,  
Faculty of Mathematical Sciences,  
Shahid Beheshti University,  
Tehran, Iran.  
e-mail: [rmeshkani@hotmail.com](mailto:rmeshkani@hotmail.com)

**L. Billard**

Department of Statistics,  
University of Georgia,  
Athens GA,  
USA,  
e-mail: [lynn@stat.uga.edu](mailto:lynn@stat.uga.edu)