Statistical Inference for a
Finite Markov Chain

by

L. Billard¹
Florida State University

and

M. R. Meshkani²
College of Statistics and Informatics, Tehran, Iran

ABSTRACT

While the theory for a known transition probability matrix for a Markov chain is fairly extensive, knowledge of how to estimate these probabilities is sparse. A probability model, based on that of Anderson and Goodman (1957), is presented for a finite stationary Markov chain. Then, empirical Bayes estimates for the transition probabilities are developed. The two cases in which the initial count vector is either fixed or random are considered.

Keywords: Transition probabilities, estimation, finite stationary chain, Bayes estimates, empirical Bayes estimates.

AMS Classification: 62F10, 62F15, 62M05.

¹Research supported by National Science Foundation Grant No. MCS76-10453.

²Work performed at Florida State University.
1. INTRODUCTION

This paper is concerned with estimation of the transition probabilities of a first order stationary Markov chain. Basically, a Markov chain is a
simple model for a sequential experiment involving a process in which an observation is classified according to the state in
which it falls, with the property that the value of an observation at a
certain time is dependent on the previous value. For a complete treatise on
the theory of Markov chains, the reader is referred to any of the numerous
books available on the subject [for example, Feller (1968)].

While the theory for a known transition probability matrix (t.p.m.) is
fairly extensive, knowledge of how to estimate these probabilities is sparse.

The traditional methods of estimation, that is, maximum likelihood and
Bayesian, have been used by various authors. Among others, we refer to
Bartlett (1951), Anderson and Goodman (1957), Billingsley (1961), Martin
(1967), and Lee, Judge and Zellner (1968). Unfortunately, there is virtually
no record of empirical Bayes procedures for Markov chains despite the fact
that the literature and the real world are replete with applications which
lend themselves ideally to empirical Bayes procedures. For example, suppose
there are 5 brands of a certain product on the market. Customers switch from
one brand to another brand as time progresses with some probability model.

There are many quantities of interest to the producers and for retailers
such as the percentage of customers using a particular brand in the long run,
average staying time and average return time to a certain brand, and so on.

A problem of this type has been given by Draper and Nolin (1964). A similar
problem dealing with voter preferences was considered by Anderson (1954).

Since knowledge of the initial distribution and the t.p.m. is sufficient to
completely describe the behavior of the underlying Markov chain, these quantities can be found once estimates of the t.p.m. are available.

It is supposed that the data falls into a panel study type. Thus, in the brand switching example, the producer chooses \((n + 1)\) cities within each of which a sample of \(N\) individuals is taken. These individuals are classified according to which brand they are using at each time \(t \in T_0 = \{0, \ldots, T\}\). Hence, for each city there is a frequency count matrix showing the number of individuals switching from one brand to another for consecutive time periods.

It is clear that the concepts of empirical Bayes estimation can be used advantageously to these situations. Hence, we aim to meet such a need as it pertains to Markov chains. First, in Section 2, the probability model is described. This probability model is formulated in a manner similar to that used by Anderson and Goodman (1957). The Bayes estimator of the t.p.m. is given in Section 3, with the corresponding empirical Bayes estimator given in Section 4. In these sections, it is assumed that the initial count vector is fixed. Thus, in Section 5, the case in which this initial count vector is random is considered.
2. THE PROBABILITY MODEL

2.1. Notation and assumptions.

Let \( N = \{1, \ldots, n\} \), \( N_1 = \{1, \ldots, n + 1\} \), \( S = \{1, \ldots, s\} \), \( T_0 = \{0, \ldots, T\} \) and \( T = \{1, \ldots, T\} \). Let \( \{X_t, t \in T_0\} \) be a simple Markov chain of \( s \) states with transition probability matrix (t.p.m.) \( \Lambda(t) \) whose \((j, k)\)th element, \( \Lambda_{jk}(t), j, k \in S \), is the probability an individual moves from state \( j \) at time \( t - 1 \) to state \( k \) at time \( t \), \( t \in T \). We will assume the chain is stationary so that \( \Lambda(t) = \Lambda \) for all \( t \in T \). It is assumed that for each \( i \in N_1 \), there are \( N \) independent and identically distributed (i.i.d.) chains observed collectively. An observation on a given individual is the sequence of states in which the individual falls for each \( t \in T_0 \). Let this sequence be represented by \( j(0), \ldots, j(T) \). Given the initial state \( j(0) \), the number of possible sequences is \( s^T \). These sequences represent mutually exclusive events with probabilities

\[
(2.1) \quad \prod_{t \in T} \Lambda_{j(t-1), j(t)}
\]

Before proceeding further, the following definitions are required.

Definition 2.1.

The frequency count vector (f.c.v.), \( G(t) \), is the vector whose \( j \)th element, \( G_j(t), j \in S \), is the number of individuals in the \( j \)th state at time \( t \), \( t \in T_0 \).

It is noted that the initial f.c.v., \( G(0) \), may be either fixed or random. The case in which \( G(0) \) is fixed is considered first with treatment of the random case being deferred to Section 5.
Definition 2.2.

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

We observe that there will be a f.c.v. $G_i(t)$ and a f.c.m. $F_i(t)$, for each $i \in \mathbb{N}_1$. In the sequel, we shall suppress the $i$ subscript whenever $i = n + 1$. The overall objective of this section is to find the probability mass function (p.m.f.) of $F_i(t)$. However, since the $F_i(t)$ are i.i.d., it is only necessary to find the distribution of $F(t)$.

Let $\underline{1}$ be the s-vector of ones, then the following relationships hold:

\[
\begin{align*}
G(t - 1) &= F(t) \cdot \underline{1}, \\
G(t) &= F'(t) \cdot \underline{1}, \\
1'G(t) &= 1'F(t) \underline{1} = \mathbb{N}, \quad t \in T.
\end{align*}
\]

It also follows that, for $j \in S$,

\[
G_j(t - 1) = \sum_{k \in S} F_{jk}(t) = F_{j+}(t), \quad \text{say},
\]

and

\[
G_k(t) = \sum_{j \in S} F_{jk}(t) = F_{+k}(t), \quad \text{say}, \quad t \in T.
\]

Suppose further that $F_{j(0)}, \ldots, j(T)$ is the number of individuals whose sequence of states is $j(0), \ldots, j(T)$. Then,

\[
F_{jk}(t) = \sum_{j(0), \ldots, j(t)} F_{j(0), \ldots, j(t)}
\]

where
When the chain is stationary, this reduces to

\[ F_{jk} = \sum_{t \in T} F_{jk}^{(t)}, \quad j, k \in S. \]  

(2.6)

Let \( F(t) \) denote the collection of all matrices satisfying (2.2). Define

\[ F = \{ F: F = \sum_{t \in T} F(t), F(t) \in F(t) \}. \]

For a given \( G_j(t-1) \), the \( j \)th row of \( F(t) \), viz., \( F_j(t) \), follows a multinomial distribution, that is,

\[ P[F_j(t)|A_j(t), G(t)] = G_j(t-1)! \prod_{k \in S} \frac{[A_{jk}(t)/F_{jk}(t)]^k}{k!}, \quad j \in S. \]

Using the relationship (2.3), the joint distribution over all \( j \in S \) and \( t \in T \) becomes

\[ P[F(1), \ldots, F(T)|A(1), \ldots, A(T), G(0)] \]

(2.7)

\[ = P[F(t)|A(t), G(0), t \in T] \]

\[ = A_0[F(t)] \prod_{t \in T} \prod_{j \in S} \frac{F_{jk}(t)}{I(t) F_{jk}(t)!} \prod_{j \in S} \frac{F_{jk}(t)!}{\prod_{j \in S} F_{jk}(t)!}, \]

(2.8)

where

\[ A_0[F(t)] = \prod_{t \in T} \prod_{j \in S} \prod_{k \in S} \frac{F_{jk}(t)!}{F_{jk}(t)!}. \]

Finally, let us define

**Definition 2.3.**

For a stationary Markov chain, the total frequency count matrix up to time \( T \), \( F \), has \((j, k)\)th element given by (2.6). That is, \( F_{jk} \) is the total
number of times an individual has moved from state $j$ to state $k$. When there
is no risk of confusion, we shall simply call $F$ the f.c.m.

Then, the p.m.f. of $F$ when $G(0)$ is fixed, for a given $\Lambda$, is

$$
P_0 (F|\Lambda) = P(F|\Lambda, G(0))$$

$$= \Lambda_0 [F(t)] \prod_{j, k \in S}^{F} \Lambda_{jk}^F, F \in F.
$$

Clearly, from (2.7) and (2.8), the set of statistics $\{F(t), t \in T\}$ and

$F$ is sufficient for $\Lambda$. We confine our attention here to the stationary

chain.

2.2. Conditional moments of $F(t)$ and $F$.

Let the elements of the $t$-step t.p.m. $\Lambda_{jk}[t]$ be denoted by $\Lambda_{jk}[t]$, $j, k \in S$.

Let $F_{h; jk}(t)$ be the number of sequences which started in state $h$ at time 0
and moved from state $j$ at time $t - 1$ to state $k$ at time $t$, $h \in S$. The random
variables $F_{h; jk}(t)$, for each $h \in S$, follow a multinomial distribution with
sample size $G_h(0)$ and cell probabilities $\Lambda_{hj}[t-1] \Lambda_{jk}$. Therefore, for each

$h \in S$, we can show that

$$
E_i[F_{h; jk}(t)] = G_h(0) \Lambda_{hj}[t-1] \Lambda_{jk}, j, k \in S,
$$

and

$$
Cov_i[F_{h; jk}(t), F_{h; g\ell}(t)] = \begin{cases}
G_h(0) \Lambda_{hj}[t-1] \Lambda_{jk}(1 - \Lambda_{hj}[t-1] \Lambda_{jk}), (j, k) = (g, \ell), \\
-C_h(0) \Lambda_{hj}[t-1] \Lambda_{jk} \Lambda_{hg}, (j, k) = (g, \ell),
\end{cases}
$$

$$
= C_h(0)(\delta_{gj} \delta_{k\ell} \Lambda_{hj}[t-1] \Lambda_{jk} - \Lambda_{hj}[t-1] \Lambda_{jk} \Lambda_{hg} \Lambda_{g\ell}).
$$
where the subscript \( l \) is used to indicate that the expectations have been taken for a given \( \Lambda \).

Let us now define, for \( h, j, k \in S \),

\[
U_{h;jk}(t) = F_{h;jk}(t) - F_{h;j+}(t)\Lambda_{jk}
\]

where

\[
F_{h;j+}(t) = \sum_{k \in S} F_{h;jk}(t).
\]

Since

\[
E[I_{F_{h;jk}(t)}] = \sum_{j' \in A(j)} \sum_{k' \in A(k)} I_{F_{h;jk}(t)} \Delta_{jk}
\]

and

\[
E[I_{F_{h;jk}(t)}] = E[I_{F_{h;jk}(t)}|F_{h;j+}(t)] = F_{h;j+}(t)\Lambda_{jk}
\]

it follows that

\[
E[I_{U_{h;jk}(t)}] = E[I_{E[U_{h;jk}(t)]|F_{h;j+}(t)}] = 0,
\]

\[
\text{Var}_l[U_{h;jk}(t)] = E[I_{E[U_{h;jk}(t)^2]|F_{h;j+}(t)}] = G_h(0)\Lambda_{[t-1]}\Lambda_{jk}(1 - \Lambda_{jk}).
\]

Similarly, it can be shown that

\[
\text{Cov}_l[U_{h;jk}(t), U_{h;jg}(t+r)] = \begin{cases} 
-G_h(0)\Lambda_{[t-1]}\Lambda_{jk}\Lambda_{jk}, & g = j, k \neq l, r = 0, \\
0, & g \neq j, \text{and/or } r \neq 0.
\end{cases}
\]

Thus, the random variables \( F_{h;jk}(t) \), given \( F_{h;j+}(t) \), \( h \in S \), satisfy a multinomial distribution with sample size \( G_h(0)\Lambda_{[t-1]} \) and cell probabilities \( \Lambda_{jk} \).
Finally, we may now find the moments of \( F_{jk}(t) \) and \( F_{jk} \), respectively, for \( j, k \in S \). Let consider \( F_{jk}(t) \). It is observed that \( F_{jk}(t) \) may be written as:

\[
F_{jk}(t) = \sum_{h \in S} h_{jk}(t), \quad j, k \in S.
\]

Then, using the above results, we have

\[
E_1 [F_{jk}(t) - F_{jk}(t) \Lambda_{jk}] = 0,
\]

\[
\text{Var} [F_{jk}(t) - F_{jk}(t) \Lambda_{jk}] = \sum_{h \in S} h_{jk}(0)^{\Lambda_{hj}} \Lambda_{jk}(1 - \Lambda_{jk}),
\]

and

\[
\text{Cov}_1 \{[F_{jk}(t) - F_{jk}(t) \Lambda_{jk}], [F_{g\ell}(t) - F_{g\ell}(t) \Lambda_{g\ell}]\}
\]

\[
= \begin{cases} 
- \sum_{h \in S} h_{jk}(0)^{\Lambda_{hj}} \Lambda_{jk}(1 - \Lambda_{jk}), & t = r, \quad g = j, \; k = \ell, \\
0, & t \neq r \; \text{or} \; g \neq j.
\end{cases}
\]

In order to find the moments for \( F_{jk}, \; j, k \in S \), we note that,

\[
F_{jk} = \sum_{k \in S} F_{jk} = \sum_{k \in S} \sum_{t \in T} F_{jk}(t) = \sum_{t \in T} G_{j}(t - 1) = G_{j}.
\]

Therefore, we have

\[
E_1 (F_{jk} - F_{jk} \Lambda_{jk}) = 0,
\]

\[
\text{Var} (F_{jk} - F_{jk} \Lambda_{jk}) = \sum_{t \in T} \sum_{h \in S} h_{jk}(0)^{\Lambda_{hj}} \Lambda_{jk}(1 - \Lambda_{jk}),
\]

and

\[
\text{Cov}_1 \{(F_{jk} - F_{jk} \Lambda_{jk}), (F_{g\ell} - F_{g\ell} \Lambda_{g\ell})\}
\]

\[
= \begin{cases} 
- \sum_{t \in T} \sum_{h \in S} h_{jk}(0)^{\Lambda_{hj}} \Lambda_{jk}(1 - \Lambda_{jk}), & g = j, \; k = \ell, \\
0, & g \neq j.
\end{cases}
\]
2.3. Unconditional Moments.

Later, we shall require unconditional moments of functions of \( F \). More specifically, we are interested in the random variable

\[ M_{jk} = \frac{F_{jk}}{F_{jj}}, \quad j, k \in S, \]

and

\[ G_{j+} = \sum_{k \in S} F_{jk}, \quad j \in S. \]

We assume that for all \( j \in S, G_j(0) \neq 0 \). Then, \( F_{j+} \neq 0 \) for all \( j \in S \).

In order to find the unconditional moments of \( M_{jk} \) and \( G_{j+} \), we need first to derive the unconditional p.m.f. of \( F \) for fixed \( G(0) \). This is achieved by assuming a conjugate prior distribution for \( \Lambda \), namely, the matrix beta distribution \( MB(p) \). That is, the joint probability density function of \( \Lambda \) is

\[ q(\Lambda) = C(\rho) \prod_{j, k \in S} \Lambda_{jk}^{p_{jk} - 1}, \quad \Lambda \in \Omega_S, \]

where

\[ \Omega_S = \{ \Lambda: \Lambda \text{ is an } s \times s \text{ matrix } \exists \Lambda_{jk} \geq 0, j, k \in S, \] \[ \sum_{k \in S} \Lambda_{jk} = 1, j \in S \} \]

and

\[ C(\rho) = \prod_{j \in S} \Gamma(p_{j+}) / \prod_{k \in S} \Gamma(p_{jk}) \]

with

\[ \rho_{j+} = \sum_{k \in S} \rho_{jk}, j \in S. \]
Then, the unconditional p.m.f. of $F$ for fixed $G(0)$ is

\[(2.20) \quad P_0(F) = \int_{\Omega} P_0(F|\Delta) \cdot q(\Delta) d(\Delta) = \Lambda_0[F(t)] \cdot B(\rho, F), \quad F \in F,\]

where

\[B(\rho, F) = \prod_{j \in S} \{\Gamma(\rho_{j+})/\Gamma(\rho_{j+} + F_{j+})\} \prod_{k \in S} [\Gamma(\rho_{jk} + F_{jk})/\Gamma(\rho_{jk})\}.\]

Hence, we may readily show that

\[(2.21) \quad E(M_{jk}) = E_2[E_1(M_{jk})] = \rho_{jk}/\rho_{j+}, \quad j, k \in S,\]

and

\[(2.22) \quad E(G_j) = E_2\{E_1[\sum_{k \in S} \sum_{t \in T} \sum_{h \in S} h_{j,k}(t)]\}\]

\[= \sum_{h \in \mathcal{H}(0)} \sum_{k \in S} E_2[\sum_{t \in T} \Lambda_{hk}^{[t-1]} \Lambda_{kj}], \quad j \in S,\]

where the subscript 2 indicates the expectation has been taken w.r.t. the distribution of $\Lambda$. 
3. BAYES ESTIMATE OF $\Lambda$

3.1. Posterior distribution of $\Lambda$.

We shall assume squared error loss function so that, following DéGroot (1970), the loss function associated with estimating $\Lambda$ by $d = (d_{jk})$, $j, k \in S$, is given by

$$L(d, \Lambda) = \sum_{j, k \in S} (d_{jk} - \Lambda_{jk})^2.$$  \hfill (3.1)

It is readily shown that for this loss function, the minimum Bayes risk for $d$ is achieved by having minimum Bayes risk for all $d_{jk}$, $j, k \in S$ [see, Billard and Meshkani (1978)]. Thus, the Bayes estimate of $\Lambda$ is found by finding the Bayes estimate of each $\Lambda_{jk}$, $j, k \in S$. However, the Bayes estimate of $\Lambda$ is just the posterior mean of $\Lambda$ given $F$. We first derive the posterior distribution of $\Lambda$ given $F$.

Lemma 3.1.

Let $F$ be the f.c.m. of a collection of N i.i.d. stationary Markov chains up to time $T$. Let $\Lambda$ be the t.p.m. of the chain. Assume $\Lambda$ has a MB($\rho$) prior distribution. Then, the posterior distribution of $\Lambda$ given $F$ is a MB($\rho + F$).

Proof.

The posterior distribution of $\Lambda$, $q^*(\Lambda)$, say, is obtained from (2.8) and (2.20) as

$$q^*(\Lambda) = P_0(F|\Lambda)q(\Lambda)/P_0(F), \quad F \in F,$$

$$= [B^{-1}(\rho, F) \prod_{j, k \in S} \Lambda_{jk}^{-1}] \cdot [C(\rho) \prod_{j, k \in S} \rho_{jk}]^{-1} \cdot [C(\rho + F) \prod_{j, k \in S} \rho_{jk}]^{-1} \Lambda \in \Omega_T, F \in F.$$
Now we can easily find the Bayes estimate of $\Lambda$ which we shall denote by $\Lambda_B$.

**Theorem 3.2.**

Let $\mathcal{F}$ be the f.c.m. of a collection of $N$ i.i.d. stationary Markov chains up to time $T$. Let $\Lambda$ be the t.p.m. of the chain. Assume $\Lambda$ has a MB($\rho$) prior distribution and $G(0)$, the f.c.v. is fixed. Then, the Bayes estimate of $\Lambda$ relative to the squared error loss function is

$$
\Lambda_B = \Lambda_B(\mathcal{F}, \rho) = (\Lambda_{B;jk})
$$

where

$$
\Lambda_{B;jk} = \frac{\left(F_{jk} + \rho_{jk}\right)}{\left(F_{jk} + \rho_{jk}\right)}, \ j, k \in S.
$$

**Proof.**

By the nature of the square error loss function in (3.1), we need only find the Bayes estimate for each $\Lambda_{jk}, \ j, k \in S$, viz., the marginal posterior mean. That is, from Lemma 3.1,

$$
\Lambda_{B;jk} = \Lambda_{B;jk}(\mathcal{F}, \rho) = \mathbb{E}(\Lambda_{jk} | \mathcal{F}) = \int_{\Omega_s} \Lambda_{jk}^* (\Lambda) \ d(\Lambda)
$$

$$
= \frac{\left(F_{jk} + \rho_{jk}\right)}{\left(F_{jk} + \rho_{jk}\right)}. \square
$$

From Anderson and Goodman (1957), it is known that the maximum likelihood estimate of $\Lambda$ based on $\mathcal{F}$, denoted by $\Lambda_{\text{ML}}$, is

$$
\Lambda_{\text{ML};jk} = \Lambda_{\text{ML};jk}(\mathcal{F}) = \frac{F_{jk}}{F_{jk} + \rho_{jk}}, \ j, k \in S.
$$

Hence, $\Lambda_{B;jk}$ is a convex combination of $\Lambda_{\text{ML};jk}$ and $\mathbb{E}(\Lambda_{jk}) = \rho_{jk}/\rho_{jk}$, i.e.,
\[ \Lambda_{E;jk} = \frac{F_{j+}(F_{j+} + \rho_{j+})}{\Lambda_{ML;jk} + \frac{\rho_{j+}}{(F_{j+} + \rho_{j+})}} \cdot \frac{\rho_{j+}}{\rho_{j+}}. \]

Thus, [see Bishop, Fienberg and Holland (1975)] there are points along the line connecting \( \Lambda_{ML;jk} \) and \( E(\Lambda_{jk}) \) that are closer to \( \Lambda_{jk} \) than is the maximum likelihood estimate. Hence, we are lead to consider estimates of \( \Lambda_{jk} \) that are formed by shrinking \( \Lambda_{ML;jk} \) towards the origin \( E(\Lambda_{jk}) \).
4. **Empirical Bayes Estimate of $\Lambda$**

4.1. Preliminaries.

For a squared error loss function, the best estimate in terms of least possible Bayes risk one can obtain is (3.2). However, it has a drawback in that we do not usually have a clear conception of $\rho_{jk}$, $j$, $k \in S$, to incorporate the right prior. Moreover, the right prior may change with experiments. Therefore, we shall estimate these values utilizing the empirical Bayes (EB) procedure.

In the sequel, we shall use the phrase "past data" to refer to the set of observations $\{F_i: i \in N\}$ as opposed to "current data", viz., $F_{n+1}$, although they may be concurrent.

A basic assumption of the EB procedure is that $\{F_i: i \in N\}$ are i.i.d. random matrices whose p.m.f. is given by (2.20). This p.m.f. contains $s^2$ parameters $\rho_{jk}$, $j$, $k \in S$. Estimation of these parameters from observations $\{F_i, i \in N\}$ is in general a routine problem. However, the involved nature of (2.20), due to the presence of the Gamma functions, makes it difficult in this case.

4.2. Method of moments estimate of $\rho$.

We want to estimate the set of parameters $\rho_{jk}$, $j$, $k \in S$, which specify the prior distribution of $\Lambda$.

From (2.21) and (2.22), we have

\[
\rho_{jk} = \rho_{j+} E(M_{jk}), \quad j, k \in S,
\]

\[
E(G_j) = \sum_{h \in S} G_h(0) E_2[A_{hj}] + \sum_{k \in S} A_{hk} A_{kj} + \ldots
\]

\[
+ \sum_{\ell_1 \in S} \sum_{\ell_2 \in S} \sum_{\ell_3 \in S} A_{\ell_1} A_{\ell_2} A_{h_1} A_{h_2} A_{h_3} + \sum_{k \in S} \sum_{\ell_1 \in S} \sum_{\ell_2 \in S} \sum_{\ell_3 \in S} A_{\ell_1} A_{\ell_2} A_{\ell_3} A_{k_1} A_{k_2} A_{k_3}, \quad j \in S.
\]
Since $\rho_{jk} = \sum_{k \in S} \rho_{jk}$ and $\sum_{k \in S} \rho_{jk} = 1$, there is one redundant equation in (4.1) for each $j \in S$. Thus, one additional equation is necessary, and this is supplied by (4.2). The equation (4.2) can be written as:

\begin{equation}
(4.3) \sum_{h \in S} \frac{G_{bh}(0)}{\rho_{bh}} \left( \frac{\rho_{bh}}{\rho_{bh} + \epsilon} \right) + \sum_{k \in S} \frac{G_{kh}(0)}{\rho_{kh}} \left( \frac{\rho_{kh}}{\rho_{kh} + \epsilon} \right) = E(G_j) + \phi(\rho) = 0,
\end{equation}

where $\phi(\rho)$ equals the remaining terms of (4.2) for $t > 3$. Theoretically, it should be possible to evaluate $\phi(\rho)$ for every $T$. However, for large $T$ it does not render a neat closed form answer.

Let us take $T = 2$ to illustrate the procedure used to solve (4.1) and (4.2) for the $s^2$ unknowns, namely, $\rho_{jk}, j, k \in S$. Substituting from (4.2) above and setting

\begin{equation}
(4.4) \xi_h = \frac{\rho_{bh}}{\rho_{bh} + \epsilon}, h \in S,
\end{equation}

(4.3) for each $j \in S$, becomes

\begin{equation}
(4.5) \sum_{h \in S} \frac{G_{bh}(0)}{\rho_{bh}} E(M_{bh}) - \delta_{bh} \left( \xi_h \right) = E(G_j) - \sum_{h \in S} \frac{G_{bh}(0)}{\rho_{bh}} \left[ E(M_{bh}) + E(M_{hh}) \delta_{bh} + \sum_{k \in S} E(M_{bh}) E(M_{kh}) \right].
\end{equation}

Let $\Gamma = (\gamma_{jk})$ be a $s \times s$ matrix, $\xi = (\xi_h)$ a $s$-vector, and $\beta = (\beta_j)$ a $s$-vector where $\xi_h$ is defined as in (4.4) above and

\begin{equation}
\gamma_{jh} = \frac{G_{bh}(0)}{\rho_{bh}} \left[ E(M_{bh}) - \delta_{bh} \right], j, h \in S,
\end{equation}

and

\begin{equation}
\beta_j = E(G_j) - \sum_{h \in S} \frac{G_{bh}(0)}{\rho_{bh}} \left[ E(M_{bh}) + E(M_{hh}) \delta_{bh} + \sum_{k \in S} E(M_{bh}) E(M_{kh}) \right], j \in S.
\end{equation}
Writing (4.5) in matrix form, we have

\[(4.6) \quad \Gamma \xi = \beta.\]

Since \(1'\Gamma = 0\) and \(1'\beta = 0\), the system is consistent but \(\Gamma\) is singular.

To solve (4.6) for \(\xi\), we may use a generalized inverse of \(\Gamma\), say \(\Gamma^{-}\). Thus,

\[(4.7) \quad \xi^* = \Gamma^{-}\beta + (\Gamma^{-}\Gamma - I)\eta\]

where \(\eta\) is any arbitrary \(s\)-vector, [see Searle (1971)]. We shall choose \(\eta\) such that the \(\xi^*_h, \ h \in S\), of \(\xi^*\) are in \((0, 1)\). This ensures us of an admissible solution for \(\rho_{h^+}\), since

\[(4.8) \quad \rho_{h^+} = \xi^*_h / (1 - \xi^*_h), \ h \in S.\]

The above procedure produces a method of moments estimate of \(\xi = (\xi_h)\) and hence, from (4.8) and (4.1), the method of moments estimate of \(\rho = (\rho_{jk})\).

Let these estimates of \(\xi\) and \(\rho\) be denoted by \(\bar{x} = (x_h)\) and \(\bar{R} = (r_{jk})\), respectively. From the choice of \(\eta\), we have \(\bar{x} \neq \bar{1}\).

Let \(C, b, \overline{M}_{jk}\) and \(\overline{G}_j\) be the sample counterparts of \(\Gamma, \beta, E(M_{jk})\), and \(E(G_j)\), respectively. Then,

\[(4.9) \quad C \cdot \bar{x} = \bar{b}.\]

Also, it follows

\[(4.10) \quad r_{j^+} = x_j / (1 - x_j), \ j \in S.\]

Now, we write the equations (4.1) for each \(j \in S\) as
(4.11) \[ r_{jk} \frac{M_{jk} - 1}{\bar{M}_{jk}} + \sum_{h \neq k} r_{jk} = 0, \]

where from (4.10),

(4.12) \[ \sum_{h \in S} r_{jk} = x_j/(1 - x_j), \]

which will give

(4.13) \[ r_{jk} = \frac{M_{jk} x_j}{(1 - x_j)}, \ j, k \in S. \]

This then is our required estimate for \( \rho_{jk} \), \( j, k \in S \).

Having estimated the matrix \( \rho \), we are now able to construct an EB estimate of \( \Lambda \), denoted by \( \Lambda_{\text{EB}} \). This is done by using (3.2) and replacing \( \rho_{jk} \) by its estimate \( r_{jk} \), \( j, k \in S \).

**Definition 4.1.**

The EB estimate of \( \Lambda \) obtained by the method of moments is the matrix \( \Lambda_{\text{EB}} \) whose \((j, k)\)th element is given by

(4.14) \[ \Lambda_{\text{EB};jk} = \frac{F_{jk} + r_{jk}}{(F_{j+} + r_{j+})}, \ j, k \in S. \]

This EB estimate can be expressed as a convex combination of \( \Lambda_{\text{ML};jk} \) and \( (r_{jk}/r_{j+}) \), the latter being an estimate of \( E_2(\Lambda_{jk}) \). That is,

\[ \Lambda_{\text{EB};jk} = \left[ \frac{F_{j+}/(F_{j+} + r_{j+})}{(F_{j+} + r_{j+})} \right] \Lambda_{\text{ML};jk} + \left[ \frac{r_{j+}/(F_{j+} + r_{j+})}{(F_{j+} + r_{j+})} \right] (r_{jk}/r_{j+}). \]

We note that usually in panel studies, \( T \) is not large. In the example reported by Anderson and Goodman (1957) \( T = 6 \). If we do not want to go through the cumbersome evaluation of (4.3), we may use the whole data in the following way.
Let \( r_{j+}(u) \) be an answer to (4.8) based on \( F(2u - 1) \) and \( F(2u) \). Let \( I(x) \) be the integer part of \( x \). Then,

\[
    r_{j+} = \sum_{u=1}^{I(T/2)} r_{j+}(u)/I(T/2),
\]

may be used in (4.12) and (4.13). This is justified by the stationarity of the chain.

4.3. Maximum likelihood estimate of \( \rho \).

We shall estimate \( \rho \) from the past data by the method of maximum likelihood. From (2.20) the likelihood of \( F \) is proportional to \( B[\rho, F] \). Maximizing \( L(\rho|F) \) in terms of \( \rho_{jk}, j, k \in S \), will result in \( \hat{\rho} \), the maximum likelihood estimate (MLE) of \( \rho \). Since

\[
    L(\rho|F) = \prod_{j \in S} L(\rho_{j+}|F_{j+})
\]

where

\[
    L(\rho_{j+}|F_{j+}) \propto \frac{\Gamma(\rho_{j+})}{\Gamma(\rho_{j+} + F_{j+})} \cdot \prod_{k \in S} \frac{\Gamma(\rho_{jk} + F_{jk})}{\Gamma(\rho_{jk})},
\]

and since there is no functional relation between \( \rho_{jk} \) and \( \rho_{hl} \) for \( j \neq h \), it would suffice to maximize \( L(\rho_{j+}|F_{j+}) \) for each \( j \in S \). This problem is similar to one discussed by Billard and Nesknani (1978) for the multinomial distribution.

Let the maximizing values of \( \rho_{jk} \) be \( \beta_{jk} \). Then, we have the EB estimate of \( \Lambda \), obtained by this procedure as in

**Definition 4.2.**

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

\[
    \hat{\Lambda}_{EB;jk} = \frac{(F_{jk} + \beta_{jk})/(F_{j+} + \beta_{j+})}.
\]
5. THE CASE OF RANDOM $G(0)$

In the previous sections, it was assumed that the initial count vector $G(0)$ was fixed. However, in some situations $G(0)$ may be random. Suppose such a random vector $G(0) = [G_j(0)]$ takes values in the space

$$G = \{ (G(0) = G_j(0) \geq 0, j \in S, \sum_{j \in S} G_j(0) = N) \}.$$  \hspace{1cm} (2.8)

Suppose the random stochastic vector $\theta = (\theta_j)$ takes values in the space

$$\Theta = \{ \theta : \theta_j > 0, j \in S, \sum_{j \in S} \theta_j = 1 \}.$$  \hspace{1cm} (2.9)

For a given $\theta \in \Theta$, the random vector $G(0)$ has a multinomial distribution, denoted by $M_\Delta(N, \theta)$.

The joint conditional distribution of $F$, given $\theta$ and $A$ is obtained from (2.8),

$$P(F, G(0) | \Theta, A) = M_\Delta(N, \theta) P_0(F | A), F \in F \text{ and } G(0) \in G.$$  \hspace{1cm} (2.10)

Suppose that $\theta$ has the natural conjugate prior, viz., the Dirichlet distribution with parameter $\alpha = (\alpha_j), j \in S$. That is,

$$q_1(\theta) = g(\alpha) \prod_{j \in S} \theta_j^{\alpha_j-1}, \theta \in \Theta,$$

where

$$g(\alpha) = \Gamma(\alpha_+) / \prod_{j \in S} \Gamma(\alpha_j)$$

with $\alpha_+ = \sum_{j \in S} \alpha_j$.

Then, the unconditional distribution of $F$ is
\[ P(F, G(0)) = \int \int P(F|\theta, \Lambda) \cdot q(\Lambda) \cdot q_1(\theta) \ d(\Lambda) d(\theta) \]

\[ = K[\Gamma(N + \alpha_+)/\Pi \Gamma(G_j(0) + \alpha_j)] B[\rho, F], \]

\[ F \in F, \ G(0) \in G, \]

where \( K[\cdot] \) is a constant free of \( \alpha \) and \( \rho, \ \alpha_+ = \Gamma'\alpha, \) and \( B[\rho, F] \) is given in (2.20). Using (5.1), we obtain the posterior distributions of \( \theta \) and \( \Lambda \) denoted by \( q_1^*(\theta) \) and \( q^*(\Lambda) \), respectively. They are

\[ q_1^*(\theta) = K_1 \Pi_{j \in S} \theta_j^{G_j(0)+\alpha_j-1}, \ \theta \in \Theta, \ G(0) \in G, \]

where \( K_1 \) is free of \( \theta \) and

\[ q^*(\Lambda) = C[\rho + F] \Pi_{j \in S} \Lambda_j^{F_j+k_0-1}, \ \Lambda \in \Omega_\Lambda, \ F \in F, \]

where \( q^*(\Lambda) \) was derived in Lemma (3.1).

We note that the posterior distribution of \( \Lambda, q^*(\Lambda) \), is the same regardless of whether \( G(0) \) is fixed or random. Hence, the Bayes estimate of \( \Lambda \) in the present situation will also be the same as we had for \( G(0) \) fixed.

The estimation of the \( \theta \) has been considered in Billard and Meshkani (1978). Thus, by their Theorem 3.7, the Bayes estimate of \( \theta \) is given by

\[ \theta_B; j = [G_j(0) + \alpha_j]/(N + \alpha_+), \ j \in S. \]

Finally, the estimation of \( \Lambda \) associated with a random \( G(0) \) follows from (2.21) and (2.24) where now (2.22) may be written as
(5.5) \[ E(G_j) = E_3[E_1(G_j | G(0)))] \]

\[ = \sum_{h \in S} N(\alpha_h | \alpha_+ ) \sum_{k \in S} E_2[ \sum_{t \in T} \Lambda_{hk}^{[t-1]} \Lambda_{kj} ], \ j \in S, \]

where \( E_3 \) indicates the expectation is taken w.r.t. the distribution of \( G(0) \) for a given \( \theta \).

Therefore, the impact of \( G(0) \) being random is to replace \( G_h(0) \) by its expectation \( N\alpha_h / \alpha_+ \), \( h \in S \). Thus, the results of Sections 3 and 4 apply but with \( G_h(0) \) replaced by its expectation. We omit the details.
REFERENCES


