برآورد ماتریس احتمال انتقال در یک زنجیر مارکف ایستا

دکتر محمدرضا مشکانی
گروه آمار دانشکدهٔ جامع انفرماتیك و مدیریت؛ دانشگاه شهید بهشتی، اوین ۱۹۸۳۲، تهران

چکیده

در این مقاله به برآورد بیز تجربی ماتریس احتمال انتقال دریک زنجیر مارکف ایستامی‌پردازیم. توزیع احتمال ماتریس شمارش انتقال را با تکیه برنتایج کار ویتل و توزیعهای ویتل به دست می‌آوریم. سپس بااستفاده از اصول روشهای بیزی وبه کارگیری توزیعهای پیشین مزدوج، توزیع پسین ماتریس شمارش انتقال را تعیین می کنیم. برآورد بیشینهٔ درستنمائی، برآورد بیزی و برآورد بیز تجربی راکه مبتنی برداده‌ها و این توزیع پسین هستند به دست می‌آوریم. نشان می‌دهیم که برآورد بیز تجربی ترکیبی خطی از برآورد بیزی و برآورد بیشینهٔ درستنمائی است.

Estimation of a single stationary Markov chain

Dr. M. Meshkani

Department of statistics, Shahid Beheshti University, Eveen 19832, *Tehran. Iran*

Abstract

Applications of markov chain models in science and technology are varied and numerous. Successful applications depend on the accurate knowledge of chain para - meters, i. e. initial distribution of states, transition probability matrix, and stationary distribution.

In this work, we intend to provide good estimates of such parameters. In order to avoid the vagaries of maximum likelihood estimators which heavily depend on the frequency counts of consecutive states visited by the chain, as well as the subjective nature of Bayesian estimators, we propose the empirical Bayes procedure which utilizes the information contained in the past data to identify the prior distribution of parameters.

1. **Introduction**

Theoretical results for Markov chains with known transition probability matrix (t. p. m.) are extensive. However, knowledge of how to estimate these transition probabilities and their attendant - properties is relatively sparse. Thus, we address ourselves to the question of the estimation of the t. p. m. of a finite stationary Markov chain, c. f. Romanof (1982) and Frydman (1984).

We are concerned with the case in which one observed sequence occurs. The problem of inference for a single Markov chain which is easily repeated occurs frequently. For examrle, Lowry (1970) describes some such situations arising in polymer chemistry and physics. The assumptions of the empirical Bayes (EB) procedure as introduced by Robbins (1955) are well suited to these cases. Therefore, we shall concern ourselves specifically with the EB estimation of the t. p. m. of our Markov chain.

The basic underlying probability model is - presented in Section 2. These results rely heavily on the Whittle distribution results derived in Whittle (1955). Then, in Section 3, the Bayes estimate of the t. p. m. is obtained. If, as is often the case, some or all of the parameters of the prior distribution are unknown, it is necessary to find their estimates thus giving rise to the EB estimates. These procedures are presented in Section 4.

2. **The Probability model**

2. 1. **Preliminaries.**

For the sake of brevity, we shall not repeat the well - known results pertaining to single Markov chains. These are found in, e. g., Feller (1968).

Suppose $\{\mathfrak{X}_t, t \varepsilon \mathcal{T}_o\}$ is a Markov chain with values in the finite state space $S = (1, \ldots, s)$ where $\mathcal{T} = (1, \ldots, T)$ and $\mathcal{T}_o = \{o\} U \mathcal{T}$. We assume the chain is simple, i. e., its order of dependency is 1.

Furthermore, it is ststionary and has an irreducible t. p. m. $\wedge$ with elements $\wedge_{jk}$, $j, k \varepsilon S$. Let the initial distribution be θ with elements θ_j, $j \varepsilon S$. .

The data are outcomes of $(n+1)$ repeated expe - riments. In each experiment, we observe and record the states visited by the chain during a fixed period of time, $T > 1$. The outcomes of the first n experiments will be referred to as the «past data» . Let a realization of an experiment be $X_T = (x_o, x_1, \ldots, x_T)$, where the subscripts refer to the order in which the observations were taken and not to their values.

Definition 2. 1.

Let F be an s x s matrix whose (j, k) *th* element F_{jk} is the number of times that the state k has followed the state j in a sequence of states visited by a Markov chain $\{X_t, t \varepsilon \mathcal{T}_o\}$. That is, F_{jk} is the number of times the event $\{X_{t-1} = j, X_t = k; t \varepsilon \mathcal{T}\}$ has occurred. For each fixed $T > 1$, F is called the *frequency count matrix* (f. c. m.) of the chain up to time T.

Then, the probability of observing a particular ordered sequence of states is

$$P(X_0 = u, X_1 = x_1, \ldots, X_T = x_T)$$

$$= P(X_0 = u) \prod_{t \varepsilon T} P(X_t = x \mid X_{t-1} = x_{t-1}) =$$

$$= \theta_u \prod_{j, k \varepsilon S} \wedge^{F_{jk}}, \qquad (2.1)$$

where $\theta_u \varepsilon \theta$ with

$$\theta = \{\theta : \theta_j > 0, j \varepsilon S, \sum_{j \varepsilon S} \theta_j = 1\},$$

and $\wedge_{jk} \varepsilon \Omega_s$ with

$$\Omega_s = \{\wedge : \wedge_{jk} \geq 0, j, k = 1, \ldots, s, \sum_{k \varepsilon S} \wedge_{jk} = 1, j \varepsilon S\},$$

and where $x_0 = u$ is the initial state of the chain.

It is clear that F is a sufficient statistic for $\wedge$ and θ. In the sequel we deal mainly with F.

Before observing the outeome x_T the integer X_0 and the matrix F are random quantities. The conditional distribution of F given the initial state is u and the t. p. m. is $\wedge$, was first derived by Whittle (1955). This and some other related distributons have been discussed in detail by Martin (1967).

2.2. **Conditional distributions.**

we are interested in the unconditional distribution of F given $X_0 = u$, and in the posterior distribution of $\wedge$ given F. we shall derive these distributions. utilizing Martin's results on conditional distribution.

Let $x_0 = u$, $x_T = v$. Then by the definition of F.

$$F_{j+} - F_{+j} = \delta_{ju} - \delta_{jv},\ j \varepsilon S,$$

where

$$F_{j+} = \sum_{k \varepsilon S} F_{jk},\ F_{+k} = \sum_{j \varepsilon S} F_{jk}.$$

For a given F and a fixed u, the equations (2.2) uniquely determine v and vice versa. The restrictiction on F is essentially the defining characteristic of the space of values of F.

Let N be the ste set of positive integers and $N_0 = N \cup \{0\}$. For fixed u, $u \varepsilon S$, $\wedge \varepsilon \Omega_s$ and $T \varepsilon N$, we define the following sets :

$$\Phi_s(u, v, T, \wedge) = \{F: F_{jk} \varepsilon T_0,\ 1'F1 = T,\ F_{j+} - F_{+j} = \delta_{ju} - \delta_{jv},\ F_{jk} = 0 \text{ if } \wedge_{jk} = 0,\ j, k \varepsilon S\}, \quad (2.3)$$

where 1 denotes, as usual, the columm matirces of one.

$$\Phi_s(u, A, \wedge) = \bigcup_{v \varepsilon S} \Phi_s(u, v, T, \wedge) \quad (2.4)$$

$$\Phi_s^*(T, \wedge) = \bigcup_{u \varepsilon S} \Phi_s(u, T, \wedge), \quad (2.5)$$

$$\Phi_{s1}^*(T, \wedge) = \{F: F \varepsilon \Phi_s^*(T, \wedge), F_{j+} = F_{+j}, j, \varepsilon S\}, \quad (2.6)$$

and

$$\Phi_{s2}^*(T, \wedge) = \Phi_s^*(T, \wedge) - \Phi_{s1}^*(T, \wedge) \quad (2.7)$$

For each f. c. m. F. we define $F^* = (F_{jk}^*)$ - where, for j, k ε S

$$F_{jk}^* = \begin{cases} \delta_{jk} - F_{jk}/F_{j+}, & F_{j+} > 0, \\ \delta_{jk} & F_{j+} = 0. \end{cases} \quad (2.8)$$

The (v. u) *th* cofactor of F^* will be denoted by $F^*_{(vu)}$.

The conditional p. m. f. of F given u and $\wedge$, - known as the Whittle distribution, is

$$P^{(s)}(F \mid u, T, \wedge) \quad (2.9)$$

$$= F^*_{(vu)} A(F) \prod_{j, k \varepsilon S} \wedge_{jk}^{F_{jk}},\ F \varepsilon \Phi_s(u, T, \wedge),$$

where v is unique solution of (2.2) and

$$A(F) = \prod_{j \varepsilon S} (F_{j+}! / \prod_{k \varepsilon S} F_{jk}!). \quad (2.9)'$$

Here and elsewhere, the convention $0^0 = 1$ will be observed.

The joint distribution of F and X_0 which is called the Whittle - 1 distribution, is

$$P_1^{(s)}(F, u \mid T, \wedge, \theta)$$

$$= \theta_u P(F \mid u, \wedge),\ u \varepsilon S,\ \Phi_s(u, T, \wedge). \quad (2.10)$$

The marginal distribution of U for a given probability vector $\theta = (\theta_1, \ldots, \theta_s)$ is a multinomial distribution, $M_s(1, \theta)$. The marginal distribution of F for a given $\wedge$ is given as follows.

There are exactly s pairs of integers $(x, y) = (u, u)$, $u \varepsilon S$, which satisfy the equations

$$F_{j+} - F_{+j} = \delta_{jx} \quad \delta_{jy},\ j \varepsilon S, \quad (2.11)$$

if $F \varepsilon \Phi_{s1}^*(T, \wedge)$. There is a unique solution $(x, y) = (u, v)$, $u \neq v$ to these equations if $F \varepsilon \Phi^*_{s2}(T, \wedge)$, see Martin (1967, Lemma 6.1.5). Then, the marginal ditstribution of F for a given t. p. m. $\wedge$ known as the Whittle - 2 distribution, is

$$P_2^{(s)}(P \mid T, \wedge, \theta) = \begin{cases} A(F)\left(\sum_{j \varepsilon S} \theta_j F^*_{(jj)}\right) \prod_{j, k \varepsilon S} \wedge_{jk}^{F_{jk}}, & F \varepsilon \Phi^*_{s1}(T, \wedge), \\ A(F)\, \theta_u F^*_{(vu)} \prod_{j, k \varepsilon S} \wedge_{jk}^{F_{jk}}, & F \varepsilon \Phi^*_{s2}(T, \wedge), \end{cases} \quad (2.12)$$

where (u, v) is the unique solution to (2.11) when F - $\varepsilon \Phi_{s2}^*(T, \wedge)$.

2. 3. Unconditional distributions.

We shall assume the «natural conjugate priors» for θ and $\wedge$ to be independent of each other. The «natural conjugate prior» for θ is a Dirichlet distrib - ution and for $\wedge$ is a matrix beta distribution. We denote these distributions by D(α) and MB (ρ), respectively. The resultant unonditional distribution will be named the Beta - Whittle distribution.

To specify the space of values of F, we define

$$\Phi_s(u, v, T) = \bigcup_{\wedge \varepsilon \Omega_s} \Phi_s(u, v, T, \wedge), \quad (2.13)$$

$$\Phi_s(u, T) = \bigcup_{\wedge \varepsilon \Omega_s} \Phi_s(u, T, \wedge), \quad (2.14)$$

$$\Phi^*_s(T) = \bigcup_{\wedge \varepsilon \Omega_s} \Phi^*_s(T, \wedge), \quad (2.15)$$

$$\Phi_{s1}^*(T) = \bigcup_{\wedge \varepsilon \Omega_s} \Phi_{s1}^*(T, \wedge), \quad (2.16)$$

and

$$\Phi_{s2}^*(T) = \Phi_s^*(T) - \Phi_{s1}^*(T). \quad (2.17)$$

Now, we derive the unconditional distributions by integrating the conditional ones with respect to $q_1(\theta)$ and $q(\wedge)$, the prior distributions of θ and $\wedge$, respectively. That is,

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

where he parameter $\alpha = (\alpha_j)$, $\alpha_j > 0$, $j \varepsilon S$, and

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

nad $\alpha_+ = \sum_{j \varepsilon S} \alpha_j$; and

$$q(\wedge) = C(\rho) \prod_{j, k \varepsilon S} \wedge_{jk}^{\rho_{jk} - 1}, \quad \wedge \varepsilon \Omega_s,$$

where the parameter $\rho = (\rho_{jk})$, $\rho_{jk} > 0$, j, k ε 2, and

$$C(\rho) = \prod_{j \varepsilon S} \{\Gamma(\rho_+) / \prod_{k \varepsilon S} \Gamma(\rho_{jk})\},$$

and $\rho_{j+} = \sum_{k \varepsilon S} \rho_{jk}$,j ε S.Thus, form (2. 9), he Beta - Whittle distribution for a MB (ρ) prior and known u (while $F \varepsilon \Phi_s(u, T)$ is

$$P(F|u) = F_{(vu)}^* A(F) \int_{\Omega_s} \Big(\prod_{j,k \varepsilon S} \wedge_{jk}^{F_{jk}}\Big) q(\wedge) d(\wedge)$$

$$= F^*_{(vu)} \cdot A(F)\, B(\rho, F), \quad (2.18)$$

where v is the unique solution to (2. 2) and

$$B(\rho, F) = \prod_{j \varepsilon S} \{[\Gamma(\rho_{j+} + F_{j+})] \prod_{k \varepsilon S} [\Gamma(\rho_{jk} + F_{jk}) / \Gamma(\rho_{jk})]\}. \quad (2.18)'$$

Similarly, from (2. 10) when assuming a D (α) prior for θ, we obtain the Beta - Whittle - 1 distribution, for

$$P_1(F, u) = \int_\theta \int_{\Omega_s} \theta_u q_1(\theta) P(F | u, \wedge) q(\wedge) d(\wedge) d(\theta)$$

$$= A(F) \cdot B(\rho, F) \cdot C(F^*_{vu}, \alpha), \ u \varepsilon S, F \varepsilon \Phi_s(u, T) \quad (2.19)$$

where

$$C(F^*_{(vu)}, \alpha) = [\Gamma(\alpha_+) / \prod_{j \varepsilon S} \Gamma(\alpha_j)] \cdot [F^*_{(vu)} \cdot \Gamma(\alpha_u + 1) \prod_{\substack{k \varepsilon S \\ k \neq u}} \Gamma(\alpha_k(/\Gamma(\alpha_+ + 1)]. \quad (2.19)'$$

Finally, the Beta - Whittle - 2 distribution is derived from (2. 12). Thus,

$$P_2(F) = \int_\theta \int_{\Omega_s} P_2(F | \wedge, \theta) q_1(\theta) q(\wedge)\, d(\wedge) d(\theta),$$

$$\begin{cases} A(F \int_\theta \Big(\sum_{j \varepsilon S} \theta_j F^*_{(jj)}\Big) q_1(\theta)\, d(\theta) \\ \int_\Omega \Big(\prod_{j \varepsilon S} \prod_{k \varepsilon S} \wedge_{jk}^{F_{jk}}\Big) q(\wedge)\, d(\wedge), F \varepsilon \Phi_{s1}^*(T), \\ F^*_{(vu)} \cdot A(F) \int_\theta \theta_u q(\theta)\, d(\theta) \cdot \\ \int_{\Omega_s} \Big(\prod_{j \varepsilon S} \prod_{k \varepsilon S} \wedge_{jk}^{F_{jk}}\Big) q(\wedge)\, d(\wedge), F \varepsilon \Phi_{s2}^*(T), \end{cases}$$

where (u, v) is the unique solution to (2. 11) when

$F\varepsilon\Phi^*_{s2}(T)$. Therefore, the distribution is

$$_2(F)=\begin{cases} A(F)\cdot B(\rho,F)\cdot C(F^*,\alpha),\ F\varepsilon\Phi^*_{s2}(T),\\ A(F)\cdot B(\rho,F)\cdot C(F^*_{(vu)},\alpha),F\varepsilon\Phi^*_{s2}(T)\end{cases} \quad (2.20)$$

where $A(F)$, $B(\rho, F)$ and $C(F^*_{(vu)}, \alpha)$ have been defined in (2. 9)′, (2. 18)′ and (2. 19)′, respectively, and where

$$C(F^*,\alpha)=[\Gamma(\alpha_+)/\prod_{j\varepsilon S}\Gamma(\alpha_j)]\cdot$$

$$\times[\sum_{j\varepsilon S}F_{(jj)}{}^*\Gamma(\alpha_j+1)\prod_{\substack{k\varepsilon S\\ k\pm j}}\Gamma(\alpha_k)/\Gamma(\alpha_++1)] \quad (2.20)'$$

When u is known, $P(X_o=u\,|\,\theta)=\theta_u=1$ Then, (2. 19) reduces to (2.18). In the sequel, we shall consider both cases and treat them simltaneously.

3. Bayes estimate of $\wedge$

3. 1. Posterior distribution of $\wedge$.

We assume squared error loss. Hence, the loss function associated with the estimation of $\wedge$ by $d=(d_{jk})$, j, k ε S, is given by, from De Groot (1970)

$$L(\wedge,d)=\sum_{j,\,k\varepsilon S}(d_{jk}-\wedge_{jk})^2.$$

It can be easily shown that the minimum risk is achived when each $\wedge_{jk}$, j, k ε S has least possible risk. Thus, the Bayes estimate of $\wedge$ is found by finding the Bayes estimate for each $\wedge_{jk}$, j k ε S. This in turn is given by the posterior mean of $\wedge$ for given F.

Theorem 3. 1.

Let F be the f. c. m. of a single stationary Marov chain up to time T. Let $\wedge$ be the t. p. m. of the chain. Assume $\wedge$ has a MB(ρ) prior distribution. Then, the posterior distribution of $\wedge$ given F is a MB $(\rho+F)$. Furthermore, the conclusion is true whether the initial state $X_0=u$ is known or unknown.

Proof.

First, we suppose u is not known. Then, from (2. 10) and (2. 20), we have

$$q^*(\wedge,\theta)=K(F,\alpha,\rho)\,\theta_u^{\alpha_u}\prod_{\substack{k\varepsilon S\\ k\neq u}}\theta_k^{\alpha_{k-1}}\times$$

$$\prod_{j,k\varepsilon S}\wedge_{jk}^{F_{jk}+\rho_{jk}-1},\ \theta\,\varepsilon\,\theta,\ \wedge\,\varepsilon\,\Omega_s,$$

where $K(\cdot)$ is free of θ_j and $\wedge_{jk}$, j, k ε S. Then,

$$q^*(\wedge)=\int_\theta q^*(\wedge,\theta)\,d\theta\propto\prod_{j,\,k\varepsilon S}\wedge_{jk}^{F_{jk}+\rho_{jk}-1},\ \wedge\,\varepsilon\,\Omega_s \quad (3.1)$$

It is obvious that (3. 1) is a $MB(\rho+F)$.

When u is known, we have $\theta_u=1$, u ε S and the above derivation more easily gives (3. 1). ■

Theorem 3. 2.

Let F be the f. c. m. of a single stationary Markov chain up to time T. Let $\wedge$ have a prior distribution MB (ρ). Then, *the Bayes estimate of* $\wedge$ relative to the squated error loss function, whether the initial state $X_0=u$ is known or unknown, is

$$\wedge_B=\wedge_B(F,\rho)=(\wedge_{B;jk}) \quad (3.2)$$

where

$$\wedge_{B;jk}=(F_{jk}+\rho_{jk})/(F_{j+}+\rho_{j+}),\ j,k\,\varepsilon\,S.$$

Proof.

It is enough to find the Bayes estimate of $\wedge_{jk}$. For the squated error loss function, the posterior mean is the Bayes estimate. Thus,

$$\wedge_{B;jk}=\int_{\Omega_s}\wedge_{jk}\cdot q^*(\wedge)\,d(\wedge)=$$

$$=(F_{jk}+\rho_{jk})/(F_{j+}+\rho_{j+}).\ ■$$

The maximum likelihood estimate (MLE) of $\wedge$ based on F, which will be denoted by $\wedge_{ML}=(\wedge_{ML;jk})$, is

$\wedge_{ML;jk} = F_{jk}/F_{j+}$, j, k ε S

[See Bartlett (1951) or Billingsley (1961)]. Note that $\wedge_{B;jk}$ is a convex combination of $\wedge_{ML;jk}$ and $E(\wedge_{jk}) = \rho_{jk}/\rho_{j+}$.

4. Emwirical Bayes estimate of $\wedge$.

4. 1. Preliminaries.

In this section, we shall estimate ρ_{jk} from the «past data». Then, we shall substitute these values in (3. 2). The resultant value will be called an EB estimate of $\wedge$.

Let $N = \{1, \dots, n\}$. Here, the «past data» refers to the set $\{F_i, i \in N\}$ which are independent of $F \equiv F_{n+1}$ which represents the «current data», but they are - identically distributed as F.

We have seen in (2. 19) that the pair (F, u) is distributed according to a Beta - Whittle -1 distribution. The marginal distribution of U is identical to a Diri - chlet - Multinomial distribution. The EB procedure for esimation of parameters of this distribution has been considered in Billard and Meshkani (1978).

Now, we address ourselves to the estimation of ρ_{jk} from $\{F_i, i \in N\}$. The marginal distribution of F was given in (2. 20) which contains s (s+1) parameters α and ρ. We can readily estimate s parameters α by methods proposed in Billard and Meshkani (1978). Therefore, in the rest of this section, we concentrate only on the estimation of ρ.

4. 2. Method of moments estimate of ρ.

Exact formulae for moments of F are too com - plicated to be useful in estimating . Using some results of Martin (1967), we have

$$E(F_{jk}) = E_2[E_1(F_{jk})] = \sum_{t=o}^{T=1} E_2(\wedge_{uj}^{[t]} \wedge_{jk}), \; j, k \in S$$

where the subscript 1 or (2) indicates the expectations have been taken for a given $\wedge$ (with respect to the dist - ribution of $\wedge$). We also have

$$E(F_{jk}F_{gh}) = \begin{cases} (\delta_{jg}\delta_{kh} E(F_{jk}), & T=1 \\ \delta_{jg}\delta_{kh} E(F_{jk}) + \sum_{t=1}^{T-1} E_2(\wedge_{uj}^{[T-1-t]}\wedge_{jk} \times \\ \sum_{m=0}^{t-1} \wedge_{kg}^{[m]}\wedge_{gh} + \wedge_{ug}^{[T-1-t]}\wedge_{gh} \sum_{m=0}^{t-1} \wedge_{hj}\wedge_{jk}), \\ T \geq 2, j, k, g, h \in S. \end{cases}$$

Evaluation of the expectations in the above equations will lead to polynomials of degree (T - 1) in ρ_{jk}, j, kεS. When $T \geq 3$, the resultant equations will be almost intractable. Since for singl chains, T is usually far greater than 3, setting $T \geq 3$ above to obtain solvable equations, would be a waste of available information. Moreover, the estimates would not be very efficient.

We shall seek some functions of F which render simpler expressions for their moments. One of these functions is

$$M_{jk} = F_{jk}/F_{j+} \; j, k \in S. \qquad (4.\ 1)$$

Since $\wedge$ is assumed to be irreducible, $\wedge_{j+} \neq 0$, all j ε S. Thus, from the condition (2. 2), for T large enough, $F_{j+} > 0$, j ε S. We assume $F_{j+} > 0$, j ε S so that we can use M_{jk} to estimate ρ_{jk}.

Whittle (1955), under the assumption that $F_{j+} > 0$, j ε S gave

$$E_1(M_{jk} \mid u) = \wedge_{jk}(T + a_{jk})/T + 0(T^{-3/2}), \qquad (4.\ 2)$$

and

$$Cov_1(M_{jk}, M_{gh} \mid u) = \delta_{jg}(\delta_{kh} - \wedge_{jk}\wedge_{gh}) \times E_1(F_{j+}^{-1} \mid u) + 0(T^{-3/2}), \qquad (4.\ 3)$$

where a_{jk} is the (j, k) *th* element of the matrix of right eigenvectors. By appropriate normalization of a, we can make $0 \leq a_{jk} \leq 1$

Now, using (4. 2) and (4. 3) and $a_{jk}=1$, we shall find the unconditional expectations and covariances relative to the MB(ρ) prior for $\wedge$. In the sequel, we shall assume T is large enough so that we can ignore $0(T^{-3/2})$, Thus,

$$E(M_{jk}) = [(T+1)/T]\,\rho_{jk}/\rho_{j+}, j, k \varepsilon S, \quad (4.4)$$

and

$$Cov(M_{jk}, M_{gh}) = \omega_j \delta_{jg} \rho_{jk}(\delta_{kh}\rho_{j+} - \rho_j) \,/\, \rho_{j+}^2,$$

$$j,k,g,h \varepsilon S \quad (4.5)$$

$$\omega_j = \{\rho_{j+}E[F_{j-}^{+1}] + [(T+1)/T]^2\}/(\rho_{j+}+1), j \varepsilon S. \quad (4.6)$$

The tesult (4. 5) indicates that different rows of the matrix $M = (M_{jk})$ are uncorrelated. Since $M1 = 1$ we shall delete its last column to avcid singularity in the civariance matrix of M. The covariance natrix of the first (s—1) columns of M will be denoted by $\sum^*$. Then, $\sum^*$ is a block diagonal matrix of order $s(s-1) \times s(s-1)$. That is,

$$\sum^* = \text{Diag}\{\sum_{jj}\}$$

where the elements $\sigma^*_{jk,jh} = Cov(M_{jk}, M_{jh})$ of $\sum_{jj}^*$ are degined in (4.5).

We observe that for each $j \varepsilon S$, the relations - (4. 4) give (s — 1) linearly independent equations in s unknowns , ρ_{jk}, $k \varepsilon S$. We need one more equation. This is established as follows.

From (4. 5), we may write

$$Cov(M_{jk}, M_{jh}) = \omega_j E(M_{jk})[\delta_{kh} - E(M_{jh})], k, h \varepsilon S.$$

In matrix form, we have

$$\sum_{jj}^* = \omega_j \sum_{jj}, j \varepsilon S \quad (4.7)$$

where we define the elements of $\sum_{jj}$ to be

$$\sigma_{jk;jh} = E(M_{jk})[\delta_{kh} - E(M_{jh})], k, h \varepsilon S.$$

We can solve (4. 7) for ω_j to obtain

$$\omega_j = \{|\sum_{jj}^*| / |\sum_{jj}|\}^{1/(s-1)}, j \varepsilon S$$

Therefore, substituting for ω_j in (4. 6) and solving for ρ_{j+}, we have

$$\rho_{j+} = \{[(T+1)/T]^2 - \omega_j\}/[\omega_j - E(F_{j+}^{-1})], j \varepsilon S. \quad (4.8)$$

This, together with (4. 4) which is rearranged into

$$\rho_{jk} = T\rho_{j+}E(M_{jk})/(T+1), j, k \varepsilon S$$

allows us to solve for ρ_{jk}, $j, k \varepsilon S$.

The equations (4. 4) and (4. 8) give the para - meters in terms of the moments of M_{jk} and F_{j+}^{-1}. Now, we substitute the sample moments obtained from the «past data» in (4. 4) and (4. 8) to obtain the method of moments estimates of ρ_{jk}, $j, k \varepsilon S$. These estimates will be denoted by r_{jk}, $j, k \varepsilon S$.

For each $j \varepsilon S$ and $k, h \varepsilon S$, let us define the sample means $M = (M_{jk})$ and $G = (G_j)$, and sample covariances $\hat{\sum}_{jj}^* = (\hat{\sigma}_{jk,jh})$ and $\hat{\sum}_{jj} = (\hat{\sigma}_{jk;jk})$ where the - elements are respectively defined by

$$\overline{M}_{jk} = n^{-1}\sum_{i \varepsilon N}(F_{i;jk}/F_{i;j+}), \quad (4.9)$$

$$G_j = n^{-1}\sum_{i \varepsilon N} F_{i;j+}^{-1}, \quad (4.10)$$

$$\hat{\sigma}^*_{jk,jh} = (n-1)^{-1}\sum_{i \varepsilon N}(M_{i;jk} - \overline{M}_{jk})(M_{i;jk} - \overline{M}_{jh}) \quad (4.11)$$

and

$$\hat{\sigma}_{jk,jh} = M_{jk}(\delta_{kh} - M_{jh}). \quad (4.12)$$

Then, the estimates of ω_j, ρ_{j+}, and ρ_{jk} respectively are

$$c_j = \{|\hat{\sum}_{jj}^*| / |\hat{\sum}_{jj}|\}^{1/(s-1)}, j \varepsilon S \quad (4.13)$$

and

$$r_{j+} = \{[(T+1)/T]^2 - c_j\}/[c_j - G_j]\}, j \varepsilon S, \quad (4.14)$$

and

$$r_{jk} = Tr_{j+}M_{jk}/(T+1), j, k \varepsilon S. \quad (4.15)$$

Consequently,

$$r_{js}=r_{j+}-\sum_{k\varepsilon S} r_{jk}=r_{j+}(T\bar{M}_{js}+1)/(T+1), j\varepsilon S. \qquad (4.16)$$

Therefore, from (3. 2), the EB estimate of $\wedge$, dentoted by $\wedge_{EB}$, is obtained by replacing ρ_{jk} by r_{jk}. Thus, we have the following result.

Result 4. 1

The *EB estimate of* $\wedge$ *obtained by the method of moments* is the matrix $\wedge_{EB}$ whose elements $\wedge_{EB;jk}$ are given by

$$\wedge_{EB;jk}=(F_{jk}+r_{jk})/(F_{j+}+r_{j+}), j,k\varepsilon S. \qquad (4.17)$$

4. 2. Maximum likelihood estimate of . P .

The unconditional distribution of F, viz., the Beta - Whittle - 2 distribution, was given in (2. 20). Hence, the likelihood of *F,* when we observed the chain until time T, is

$$L(\rho \mid F)=K(F, \alpha) . B(\rho, F)$$

where $K(F, \alpha)$ is free of ρ, We note that $L(\rho \mid F)$ may be written as

$$L(\rho \mid F)=\prod_{j\varepsilon S} L(\rho_{j+} \mid F_{j+})$$

$$\propto \prod_{j\varepsilon S} B(\rho_{j+}, F_{j+})$$

Hence, it is sufficient to maximise $B(\rho_{j+}, F_{j+})$ for each $j \varepsilon S$. However, this is simply the case of the multinomial distribution of Billard and Meshkani (1978). Using their approach, we obtain a maxiumm likelihood estimate ρ_{jk}, for each ρ_{jk}, j, k ε S. That is, we have the - foll ovaing result,

Result 4. 2.

The *EB estimate of* $\wedge$ *obtained by the maximum likelihood method* is the matrix $\wedge_{EB}$ whose elements $\wedge_{EB;jk}$ are given by

$$\hat{\wedge}_{EB;jk}=(F_{jk}+\hat{\rho}_{j+})/(F_{j+}+\hat{\rho}_{j+}), j, k \varepsilon S. \qquad (4.18)$$

This estimator, although requires more complicated computations, is more efficient than (4. 17).

It can be observed that our both estimators are linear Combinations of maximum likelihood and - Bayes estimators, i. e.,

$$\wedge_{EB;jk}=a\wedge_{ML,jk}+b\wedge_{B;jk}.$$

5. Concluding remarks

There are many interesting questions remaining especially those pertaining to the properties of these estimators as the number of «past pata» n increasees. A detailed study of some rainfall data in which the tech - niques of the work have been applied is currently being prepared. We note that we have been considering the case of a single sequence of observations. This is in contrast to the related but different case in which sveral independent sequences are obeserved collectively. That case was not attempted here.

References

Bartlett, M. S. (1951). The frequency goodness of fit test for probability chains. *Proc. Cambridge - Philoso. Soc.*, **47**, 36 - 95.

Billard, L. and Meshkani, M. R. (1978). Empirical Bayes estimation for the multinomial distribution. *Florida State University, Department of Statistics Technical Report* **No M475.**

Billingsley, P. (1961). *Satistical inference for Markov processes.* Univ. of Chicago press, Chicago.

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

Feller, W. (1968), *An Introduction to Probabillity Thecory and its Applications* **Vol. 1,** 3rd edition, John Wiley, New York.

Frydman, H. (1984), Maximum likelihood estimation in the mover - stayer model. *J. Amer. Statis. Assoc.*, **79**, 632 - 638.

Lowry, G. G. (1970), *Markov Chains and Monte Carlo Calculations in Polymer Science.* Marecl Dekker, New York.

Martin, J. J. (1967). *Bayesian Decision Problems and Markov Chains.* John Wiley, New York.

Robbins, H. (1955), An empirical Bayes approach to statistics. *Proc. Third Berkey Symp. Math. Statis. Prob.* **1**, 157 - 164.

Romanof, N. (1982). A Markov Chain model for the mean daily so concentration. *Atmospheric Environment*, **16**, 1895 - 97.

Whittle, P. (1955). Some distribution and moment formulae for the Markov chain. *J. Royal Statis. Soc.* **B 17**, 235 - 242.