Modelling Weather Data as a Markov Chain

L. Billard
Florida State University, USA

and

M. R. Meshkani
College of Statistics & Informatics, Iran.

ABSTRACT

Weather data is taken and by classifying the various degrees of wetness and dryness, it can be modelled as a Markov chain. Data are recorded over several years. Estimates for the transition probabilities are obtained which utilize the experience (data) of previous years giving so-called empirical Bayes estimates. These contrast with the maximum likelihood estimates which use the data for the year under discussion alone.

Keywords: Wet and dry days, stationary Markov chain, estimates of transition probabilities.
1. INTRODUCTION

Some time ago Gabriel and Neuman (1962) showed how weather data could be modelled as a Markov chain. They modelled days as either dry or wet (corresponding to chain with states 0 or 1). Obviously it is possible to subclassify the wet days into \((s - 1)\) classes designating the degree of wetness as measured by the amounts of rainfall actually received. This then gives a Markov chain with \(s\) states. As is well known once the appropriate transition probabilities of the chain are known together with the initial distribution everything is completely known about the Markov process. In this case then we are able to ascertain the transition probabilities of moving from one weather classification type to another in any given number of days. We can determine the long run probabilities of certain weather types and so forth.

In the Gabriel and Neuman work, estimates for the transition probabilities were determined via maximum likelihood techniques. However, it would be advantageous to be able to obtain estimates which incorporate the available data from previous years as well. Thus, in the following sections we first outline a theoretical approach which allows a researcher to utilise past experience represented by previous data when estimating these transition probabilities yielding so-called empirical Bayes estimates. This approach is quite general and therefore it is presented within a general framework.

Then, in Sections 5-7 we apply these techniques to some weather data from Tallahassee, Florida. For ease of computation we take the classes of the Markov chain to be simply "dry" and "wet." Exactly the same techniques apply to the cases of a larger number of classes (that is, \(s > 2\)) as well
as to data from broader regions such as other cities, states, localities, or whatever. It should be remarked that some underlying assumptions are made about the process such as stationarity of the chain, independence of the transition probabilities from year to year, etc. Quite clearly, the techniques discussed here can only be used for data sets for which such assumptions hold true. With weather data there will be some regions for which these assumptions are invalid. Thus, in Section 6, verification of these assumptions for the Tallahassee data is carried out.
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 \( \{X_t, \ t \in T_0\} \) is a Markov chain with values in the finite state space \( S = \{1, \ldots, s\} \) where \( T = \{1, \ldots, T\} \) and \( T_0 = \{0\} \cup T \). We assume the chain is simple, i.e., its order of dependency is 1. Furthermore, it is stationary and has an irreducible t.p.m. \( \Lambda \) with elements \( \Lambda_{jk}, j, k \in S \). Let the initial distribution be \( \theta \) with elements \( \theta_j, j \in S \).

The data are outcomes of \( (n + 1) \) repeated experiments. 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 \( x_T = (x_0, x_1, \ldots, x_T) \), 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 \( s \times s \) 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 \( \{X_t, \ t \in T_0\} \). That is, \( F_{jk} \) is the number of times the event \( \{X_{t-1} = j, X_t = k; t \in T\} \) 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

\[
(2.1) \quad P(X_0 = u, X_1 = x_1, \ldots, X_T = x_T)
\]

\[
= P(X_0 = u) \prod_{t \in T} P(X = x_t | X_{t-1} = x_{t-1}) = \theta_u \prod_{j,k \in S} \Lambda_{jk}^{F_{jk}},
\]
where $\theta_u \in \Theta$ with

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

and $\Lambda_{jk} \in \Omega_S$ with

$$\Omega_S = \{\Lambda: \Lambda_{jk} \geq 0, j, k = 1, \ldots, s, \sum_{k \in S} \Lambda_{jk} = 1, j \in S\},$$

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

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

Before observing the outcome $x_T$, 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 $A$, was first derived by Whittle (1955). This and some other related distributions 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 $\Lambda$ given $F$. We shall derive these distributions utilizing Martin's results on conditional distributions.

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

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

(2.2)

where

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

For a given $F$ and a fixed $u$, the equations (2.2) uniquely determine $v$, and vice versa. The restriction on $F$ is essentially the defining characteristic of the space of values of $F$. 
Let $M$ be the set of positive integers and $M_0 = M \cup \{0\}$. For fixed $u$, $u \in S$, $\Lambda \in \Omega_s$ and $T \in M$, we define the following sets:

(2.3) $\Phi_s(u, v, T, \Lambda)$

$$= \{F : F_{jk} \in T_0, 1'F_1 = T, F_{j+} - F_{+j} = \delta_{ju} - \delta_{jv}, F_{jk} = 0 \text{ if } \Lambda_{jk} = 0, j, k \in S\},$$

(2.4) $\Phi_s(u, T, \Lambda) = \bigcup_{v \in S} \Phi_s(u, v, T, \Lambda)$, $u \in S$, $T \in M$, $\Lambda \in \Omega_s$,

(2.5) $\Phi^*_s(T, \Lambda) = \bigcup_{u \in S} \Phi^*_s(u, T, \Lambda)$, $T \in M$, $\Lambda \in \Omega_s$,

(2.6) $\Phi^*_{s1}(T, \Lambda) = \{F : F \in \Phi^*_s(T, \Lambda), F_{j+} = F_{+j}, j \in S\},$

and

(2.7) $\Phi^*_{s2}(T, \Lambda) = \Phi^*_s(T, \Lambda) - \Phi^*_{s1}(T, \Lambda)$.

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

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

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

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

(2.9) $P(F|u, \Lambda) \equiv p^{(s)}(F|u, T, \Lambda)$

$$= F^*_{(vu)} A(F)^T \prod_{j, k \in S} \Lambda_{jk}, F \in \Phi_s(u, T, \Lambda).$$
where \( v \) is the unique solution of (2.2) and

\[
A(F) = \prod_{j \in S} (F_j^+ \prod_{k \in S} F_{jk}^+).
\]

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

\[
P_1(F, u|\Lambda, \theta) = P_{1}^{s}(F, u|T, \Lambda, \theta)
\]

\[
= \theta_u P(F|u, \Lambda), u \in S, F \in \phi_s(u, T, \Lambda).
\]

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

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

\[
F_j^+ - F_j = \delta_{jx} - \delta_{jy}, j \in S,
\]

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

\[
P_2(F|\Lambda, \theta) = P_{2}^{s}(F|T, \Lambda, \theta)
\]

\[
= \begin{cases} 
A(F)\left( \sum_{j \in S} \theta_j F_j^* \prod_{j, k \in S} F_{jk}^* \right) \prod_{j \in S} F_{jk}^* , F \in \phi_{s1}^*(T, \Lambda), \\
A(F)\theta_u F_{uu}^* \prod_{j, k \in S} F_{jk}^* , F \in \phi_{s2}^*(T, \Lambda),
\end{cases}
\]

where \( (u, v) \) is the unique solution to (2.11) when \( F \in \phi_{s2}^*(T, \Lambda) \).
2.3. Unconditional distributions.

We shall assume the "natural conjugate priors" for \( \theta \) and \( \Lambda \) to be independent of each other. The "natural conjugate prior" for \( \theta \) is a Dirichlet distribution and for \( \Lambda \) is a matrix beta distribution. We denote these distributions by \( D(\alpha) \) and \( MB(\phi) \), respectively. The resultant unconditional distribution will be named the Beta-Whittle distribution.

To specify the space of values of \( \mathcal{F} \), we define the following sets:

\[
\phi_s(u, v, T) = \bigcup_{\Lambda \in \Omega_S} \phi_s(u, v, T, \Lambda),
\]

\[
\phi_s(u, T) = \bigcup_{\Lambda \in \Omega_S} \phi_s(u, T, \Lambda),
\]

\[
\phi_s^*(T) = \bigcup_{\Lambda \in \Omega_S} \phi_s^*(T, \Lambda),
\]

\[
\phi_{s1}(T) = \bigcup_{\Lambda \in \Omega_S} \phi_{s1}(T, \Lambda),
\]

and

\[
\phi_{s2}(T) = \phi_s^*(T) - \phi_{s1}(T).
\]

Now, we derive the unconditional distributions by integrating the conditional ones w.r.t. \( q_1(\theta) \) and \( q(\Lambda) \), the prior distributions of \( \theta \) and \( \Lambda \), respectively. That is,

\[
q_1(\theta) = g(\alpha) \prod_{j \in S} \int_{0}^{\infty} \theta_j^{\alpha_j-1} \, d\theta_j, \quad \theta \in \Theta,
\]

where the parameter \( \alpha = (\alpha_j), \alpha_j > 0, j \in S \), and
\[ g(\alpha) = \frac{\Gamma(\alpha_\ast)}{\prod_{j \in S} \Gamma(\alpha_j)}, \]

and \( \alpha_\ast = \sum_{j \in S} \alpha_j \); and

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

where the parameter \( \rho = (\rho_{jk}) \), \( \rho_{jk} > 0 \), \( j, k \in S \), and

\[ C(\rho) = \prod_{j \in S} \{ \Gamma(\rho^+_j)/\Gamma(\rho_{jk}) \}, \]

and \( \rho^+_j = \sum_{k \in S} \rho_{jk}, j \in S \). Thus, from (2.9), the Beta-Whittle distribution for a \( \text{MB}(\rho) \) prior and known \( u \) is

\[ P(F|u) = F^*_0(vu) A(F)^{\sum_{j, k \in S} F_{jk}} q(\Lambda) d(\Lambda) \]

\[ = F^*_0(vu) \cdot A(F^*_0) \cdot B(\rho; F), F \in \Phi_s(u, T), \]

where \( v \) is the unique solution to (2.2) and where

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

Similarly, from (2.10) when assuming a \( D(\alpha) \) prior for \( \theta \), we obtain the Beta-Whittle-1 distribution,

\[ P_1(F, u) = \int_{\Theta} \int_{\Omega_s} \theta u q_1(\theta) P(F|u, \Lambda) c(\Lambda) d(\Lambda) d(\theta) \]

\[ = A(F^*_0) \cdot B(\rho; F) \cdot C(F^*_0, \alpha), u \in S, F \in \Phi_s(u, T), \]
where
\[
C(F_{(vu)}, \alpha) = [\Gamma(\alpha_+)/ \prod_{j \in S} \Gamma(\alpha_j)] \cdot \left[ \prod_{k \in S} \Gamma(\alpha_k)/\Gamma(\alpha_+ + 1) \right].
\]

Finally, the Beta-Whittle-2 distribution is derived from (2.12). Thus,
\[
P_2(F) = \int \int p_2(\Omega, \theta)q_1(\theta)q(\alpha)d(\alpha)d(\theta),
\]
\[
= \begin{cases} 
A(F) \int \textstyle \sum_{j \in S} F_{(jj)}^* q_1(\theta)d(\theta) \int (\prod_{j \in S} \Lambda_{jk}^{F_{jk}})q(\alpha)d(\alpha), & F \in \Phi_{s1}^*(T), \\
F_{(vu)}^* \cdot A(F) \int \theta_u q(\theta)d(\theta) \int (\prod_{j \in S} \Lambda_{jk}^{F_{jk}})q(\alpha)d(\alpha), & F \in \Phi_{s2}^*(T),
\end{cases}
\]

where \((u, v)\) is the unique solution to (2.11) when \(F \in \Phi_{s2}^*(T)\). Therefore, the distribution is
\[
P_2(F) = \begin{cases} 
A(F) \cdot B(\rho, F) \cdot C(F^*, \alpha), & F \in \Phi_{s1}^*(T), \\
A(F) \cdot B(\rho, F) \cdot C(F_{(vu)}^*, \alpha), & F \in \Phi_{s2}^*(T),
\end{cases}
\]

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
\[
C(F^*, \alpha) = [\Gamma(\alpha_+)/ \prod_{j \in S} \Gamma(\alpha_j)] \cdot \left[ \prod_{k \in S} \Gamma(\alpha_k)/\Gamma(\alpha_+ + 1) \right].
\]

When \(u\) is known, \(P(X_0 = u|\theta) = \theta_u = 1\). Then, (2.19) reduces to (2.18).

In the sequel, we shall consider both cases and treat them simultaneously.
3. BAYES ESTIMATE OF $\Lambda$

3.1. Posterior distribution of $\Lambda$.

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

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

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

**Theorem 3.1.**

Let $F$ be the f.c.m. of a single stationary Markov chain 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($\rho + F$). 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

$$q^*(\Lambda, \theta) = K(F, \alpha, \rho_\theta u u \prod_{\theta, j, k \in S} \Pi_{\Lambda, j, k}^{F_jk + \rho_{jk} - 1}, \theta \in \Theta, \Lambda \in \Omega_s),$$

where $K(\cdot)$ is free of $\theta, j$ and $\Lambda_{jk}, j, k \in S$. Then,

$$q^*(\Lambda) = \int q^*(\Lambda, \theta)d\theta \propto \prod_{\Lambda, j, k \in S}^{F_jk + \rho_{jk} - 1}, \Lambda \in \Omega_s.$$ 

(3.1)

It is obvious that (3.1) is a MB($\rho + F$).
When $u$ is known, we have $\theta_u = 1$, $u \in S$, 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 $\Lambda$ have a prior distribution $MB(\rho)$. Then, the Bayes estimate of $\Lambda$ relative to the squared error loss function, whether the initial state $X_0 = u$ is known or unknown, is

\begin{equation}
\Lambda_B = \Lambda_B(F, \rho) = (\Lambda_{B;jk})
\end{equation}

where

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

Proof.

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

$$
\Lambda_{B;jk} = \int_{\Omega_s} A \cdot \pi^*(A) d(A) = \frac{F_{jk} + \rho_{jk}}{F_{j+} + \rho_{j+}}. \quad \Box
$$

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

$$
\Lambda_{ML;jk} = \frac{F_{jk}}{F_{j+}}, j, k \in S.
$$

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

4.1. Preliminaries.

In this section, we shall estimate $\rho_{jk}$ from the "past data". Then, we shall substitute these values in (3.2). The resultant value will be called an EB estimate of $\Lambda$.

Let $N = \{1, \ldots, n\}$. Here, the "past data" refers to the set $\{F_i, i \in N\}$ which are independent of $F_i \equiv F_{n+1}$ 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 Dirichlet-Multinomial distribution. The EB procedure for estimation of parameters of this distribution has been considered in Billard and Meshkani (1978).

Now, we address ourselves to the estimation of $\rho_{jk}$ from $\{F_i, i \in N\}$. The marginal distribution of $F$ was given in (2.20) which contains $s(s+1)$ parameters $\alpha$ and $\rho$. We can readily estimate $s$ parameters $\alpha$ by methods proposed in Billard and Meshkani (1978). Therefore, in the rest of this section, we concentrate only on the estimation of $\rho$.

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

Exact formulae for moments of $F$ are too complicated to be useful in estimating $\rho$. Using some results of Martin (1967), we have

$$E(F_{jk}) = E_2[E_1(F_{jk})] = \sum_{t=0}^{T-1} E_2(\Lambda_{uj} \Lambda_{jk}), \ j, k \in S,$$

where the subscript 1 (2) indicates the expectations have been taken for a given $\Lambda$ (w.r.t. the distribution of $\Lambda$). We also have
\[
E(F_{jkF_{gh}}) = \begin{cases} 
\delta_{jg} \delta_{kh} E(F_{jk}), & T = 1, \\
\delta_{jg} \delta_{kh} E(F_{jk}) + \sum_{t=1}^{T-1} E_{jg}(A_{t}^{T-1-t}A_{jk})^{t-1} \sum_{m=0}^{T-1} A_{k}^{m} A_{gh} + 
A_{t}^{T-1-t}A_{gh} \sum_{m=0}^{T-1} A_{h}^{m} A_{jk}, & T \geq 2, j, k, g, h \in S.
\end{cases}
\]

Evaluation of the expectations in the above equations will lead to polynomials of degree \((T - 1)\) in \(\rho_{jk}\), \(j, k \in S\). When \(T \geq 3\), the resultant equations will be almost intractable. Since for single chains, \(T\) is usually far greater than 3, setting \(T \leq 3\) 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

\[(4.1) \quad M_{jk} = \frac{F_{jk}}{F_{j+}}, j, k \in S.\]

Since \(A\) is assumed to be irreducible, \(A_{j+} \neq 0\), all \(j \in S\). Thus, from the condition (2.2), for \(T\) large enough, \(F_{j+} > 0\), \(j \in S\). We assume \(F_{j+} > 0\), \(j \in S\), so that we can use \(M_{jk}\) to estimate \(\rho_{jk}\).

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

\[(4.2) \quad E_{1}(M_{jk}|u) = \Lambda_{jk}(T + a_{jk})/T + O(T^{-3/2}),\]

and

\[(4.3) \quad \text{Cov}_{1}(M_{jk}, M_{gh}|u) = \delta_{jg} \delta_{kh} - \Lambda_{jk} \Lambda_{gh} \cdot E_{1}(F_{j+}^{-1}|u) + O(T^{-3/2}),\]

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 \( \mathcal{HB}(\rho) \) prior for \( \Lambda \). In the sequel, we shall assume \( T \) is large enough so that we can ignore \( O(T^{-3/2}) \). Thus,

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

and

\[
\text{Cov}(M_{jk}, M_{gh}) = \omega_j \delta_{jg} \rho_{jk} (\delta_{kh} \rho_{j+} - \rho_{jh})/\rho_{j+}^2, \ j, k, g, h \in S,
\]

where

\[
\omega_j = (\rho_{j+} E[M_{j+}] + [(T + 1)/T]^2)/(\rho_{j+} + 1), \ j \in S.
\]

The result (4.5) indicates that different rows of the matrix \( \mathbf{M} = (M_{jk}) \) are uncorrelated. Since \( \mathbf{M} \mathbf{1} = \mathbf{1} \), we shall delete its last column to avoid singularity in the covariance matrix of \( \mathbf{M} \). The covariance matrix of the first \( (s - 1) \) columns of \( \mathbf{M} \) will be denoted by \( \mathbf{I}^* \). Then, \( \mathbf{I}^* \) is a block diagonal matrix of order \( s(s - 1) \times s(s - 1) \). That is,

\[
\mathbf{I}^* = \text{Diag}([\mathbf{I}_{jj}^*])
\]

where the elements \( \sigma_{jk, jh}^* = \text{Cov}(M_{jk}, M_{jh}) \) of \( \mathbf{I}_{jj}^* \) are defined in (4.5).

We observe that for each \( j \in S \), the relations (4.4) give \( (s - 1) \) linearly independent equations in \( s \) unknowns \( \rho_{jk}, k \in S \). We need one more equation. This is established as follows.

From (4.5), we may write

\[
\text{Cov}(M_{jk}, M_{jh}) = \omega_j E(M_{jk})[(\delta_{kh} - E(M_{jh})], k, h \in S.
\]
In matrix form, we have

\begin{equation}
\Sigma^*_{jj} = \omega_j \Sigma_{jj}, \quad j \in S,
\end{equation}

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

\[\sigma_{jk,jh} = \mathbb{E}(M_{jk})[\delta_{kh} - \mathbb{E}(M_{jh})], \quad k, h \in S.\]

We can solve (4.7) for $\omega_j$ to obtain

\[\omega_j = \left\{ \frac{1}{\Sigma_{jj}} \right\}^{1/(s-1)}, \quad j \in S.\]

Therefore, substituting for $\omega_j$ in (4.6) and solving for $\rho_{jk}^+$, we have

\begin{equation}
\rho_{jk}^+ = \left\{ \left[ \frac{(T + 1)}{T} \right]^2 - \omega_j \right\} / \left\{ \omega_j - \mathbb{E}(F_{jk}^{-1}) \right\}, \quad j \in S.
\end{equation}

This, together with (4.4) which is rearranged into

\[\rho_{jk} = T \rho_{jk}^+ \mathbb{E}(M_{jk}) / (T + 1), \quad j, k \in S,
\]

allows us to solve for $\rho_{jk}$, $j, k \in S$.

The equations (4.4) and (4.8) give the parameters in terms of the moments of $M_{jk}$ and $F_{jk}^{-1}$. 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 $\rho_{jk}$, $j, k \in S$.

These estimates will be denoted by $\hat{\rho}_{jk}$, $j, k \in S$.

For each $j \in S$ and $k, h \in S$, let us define the sample means $\overline{M} = (\overline{M}_{jk})$ and $\overline{G} = (\overline{G}_j)$, and sample covariances $\hat{\Sigma}^*_{jj} = (\hat{\sigma}_{jk,jh}^*)$ and $\hat{\Sigma}_{jj} = (\hat{\sigma}_{jk,jh})$ where the elements are respectively defined by

\begin{equation}
\overline{M}_{jk} = n^{-1} \sum_{i \in N} (F_{i;jk} / F_{i;j+}),
\end{equation}
\begin{align}
(4.10) \quad \bar{g}_j &= n^{-1} \sum_{i \in N} F_{i;j}^T \\
(4.11) \quad \hat{g}_{j, j}^* &= (n - 1)^{-1} \sum_{i \in N} (M_{i;j} - \bar{M}_{j}) (M_{i;j} - \bar{M}_{j}) \\
\text{and} \\
(4.12) \quad \hat{g}_{j, j} &= \bar{M}_{j}(\delta_{kh} - \bar{M}_{j}).
\end{align}

Then, the estimates of $\omega_j$, $\rho_j$, and $\rho_{jk}$ respectively are
\begin{align}
(4.13) \quad c_j &= \{\hat{g}_{j, j}^*/|\hat{g}_{j, j}^*|\}^{1/(s-1)}, \quad j \in S, \\
(4.14) \quad r_{j^+} &= \{[(T + 1)/T]^2 - c_j]/[c_j - \bar{g}_j]\}, \quad j \in S,
\end{align}
and
\begin{align}
(4.15) \quad r_{jk} &= Tr_{j^+} \bar{M}_{jk}/(T + 1), \quad j, k \in S.
\end{align}

Consequently,
\begin{align}
(4.16) \quad r_{js} &= r_{j^+} - \sum_{k \in S} r_{jk} = r_{j^+}(\bar{M}_{js} + 1)/(T + 1), \quad j \in S.
\end{align}

Therefore, from (3.2), the EB estimate of $\Lambda$, denoted by $\Lambda_{EB}$, is obtained by replacing $\rho_{jk}$ by $r_{jk}$.

**Definition 4.1.**

The EB estimate of $\Lambda$ obtained by the method of moments is the matrix $\Lambda_{EB}$ whose elements $\Lambda_{EB;jk}$ are given by
\begin{align}
\Lambda_{EB;jk} &= (F_{jk} + r_{jk})/(F_{j^+} + r_{j^+}), \quad j, k \in S.
\end{align}
5. OUR PROBLEM

Our interest is to apply the theoretical results of the previous section to some rainfall data compiled from the government publication Local Climatographical Data (1961-1977). Summer days (June through August) with a measurable amount of precipitation (that is, at least 0.01 inches) at Tallahassee were counted for the years 1961 through 1977 inclusive. A day with measurable precipitation is called a wet day. The sequence of wet and dry days is assumed to form a two-state stationary simple Markov chain. The frequency count matrix for each year is given in Table 1. We assume there is independence between the different years.

Since we have a two-state Markov chain, it is completely specified by just two parameters $\Lambda_1$ and $\Lambda_2$, say, where

$$\Lambda_1 \equiv \Lambda_{11} = P(X_t = 1 | X_{t-1} = 1)$$

and

$$\Lambda_2 \equiv \Lambda_{21} = P(X_t = 1 | X_{t-1} = 2).$$

Then,

$$\Lambda = \begin{pmatrix} \Lambda_1 & 1 - \Lambda_1 \\ \Lambda_2 & 1 - \Lambda_2 \end{pmatrix}.$$

Our objective is to estimate $\Lambda_1$ and $\Lambda_2$ for the year 1977, say, using the past data (years 1961-1976) and the current data (year 1977) according to the empirical Bayes procedure.
6. VERIFICATION OF THE ASSUMPTIONS

A crucial assumption is that the $\Lambda_i$, $i = 1, \ldots, n+1$, are independent and identically distributed, where in our case $n = 16$. In this example, it may appear that this assumption does not hold. However, in personal communications, meteorologists Gleeson and Stuart at the Florida State University believe that due to shower activity in summer in the Tallahassee area, the degree of dependence between the $\Lambda_i$, $i = 1, \ldots, n+1$, from year to year is very small, if any.

To substantiate this belief of independence we applied a Runs Test. If there is not a pattern of variation among the $\Lambda_i$, $i = 1, \ldots, n+1$, they should fluctuate around their mean or median in a random manner. We first consider the signs of $\Lambda_1(i) - \bar{\Lambda}_1$. They are $- + - + - - + - + + + + + + +$. Thus the sample size is $n = 17$, the number of runs is $u = 12$, the number of minuses is $n_1 = 8$, and the number of plusses is $n_2 = 9$. The null hypothesis $H_0$ is that the signs are randomly arranged. For these observations $P(U \leq 12 | H_0) = 0.939$ and $P(U \geq 9 | H_0) = 0.157$. Therefore, $H_0$ cannot be rejected. Similarly, for $\Lambda_2(i) - \bar{\Lambda}_2$, we have $+ - + + - - + + + + + +$. In this case, $n = 17$, $u = 9$, $n_1 = 9$, $n_2 = 8$ and hence $P(U \leq 9 | H_0) = 0.50$. Thus again, $H_0$ cannot be rejected. Therefore, $\Lambda_i$, $i = 1, \ldots, n+1$, can be regarded as independent random variables.

In addition to the Runs Test we determined the maximum likelihood estimates $\hat{\Lambda}_1(i)$ and $\hat{\Lambda}_2(i)$ for each year $i = 1, \ldots, n+1$, and plotted $\hat{\Lambda}_1(i - 1)$ against $\hat{\Lambda}_1(i)$, as well as $\hat{\Lambda}_2(i - 1)$ against $\hat{\Lambda}_2(i)$. These are shown in Figure 1. There does not appear to be any detectable relationship between consecutive $\Lambda_i$. Thus, these plots also suggest the assumption of independence of $\Lambda$ for different years is valid.
Another assumption was that the Markov chain was stationary. To avoid lengthy computations we chose one year at random, 1967, to test for stationarity. The chain contained 92 observations and was broken into 6 consecutive pieces each of 15 days with the final 2 observations being ignored. The resulting frequency count matrices were obtained. The null hypothesis $H_0$ is that the chain is stationary, that is, $A(t) = \Lambda$, $t = 1, \ldots, 6$. From Anderson and Goodman (1957), if $H_0$ is true the test statistic

$$Q = \sum_{t=1}^{6} \sum_{j,k=1}^{2} F_{jk}(t) \ln \frac{\hat{A}_{jk}}{\Lambda_{jk}(t)}$$

is asymptotically chi-square distributed with $d = (T - 1)s(s - 1) = 5 \times 2 \times (2 - 1) = 10$ degrees of freedom. Our observed value of $Q$ upon substituting becomes 9.8934 while the tabulated value at the 5% level of significance is 18.307. Thus, $H_0$ cannot be rejected and the chain is stationary.

Finally we test the assumption that the Markov chain is simple, that is, the order of dependency is one. We first test the null hypothesis $H_0$ that the chain is of order zero, that is, $\{X_t\}$, $t = 1, \ldots, T$, are independent and identically distributed binomial variables against the alternative that the order of the chain is not zero. A chi-square statistic

$$Q = \sum_{i=1}^{17} \chi_i^2$$

is used where $\chi_i^2$ is the chi-square value within the $i$th year and where $Q$ has a chi-square distribution with 17 degrees of freedom. The observed value for $Q$ is 78.77 while the tabulated value for the 5% level of significance is 27.59. Thus, $H_0$ is rejected and the chain has an order of dependence greater than or equal to one.
We now test the null hypothesis $H_0$ that the order of the chain is one against the alternative hypothesis that it is two. The transition probability matrix for the second order chain can be represented as

$$
\begin{bmatrix}
A_{111} & A_{112} & 0 & 0 \\
0 & 0 & A_{121} & A_{122} \\
A_{211} & A_{212} & 0 & 0 \\
0 & 0 & A_{221} & A_{222}
\end{bmatrix}
$$

The test statistic based on the likelihood ratio is

$$Q = \sum_{i=1}^{17} \sum_{j,k,l}^2 F_{i;jk} e^{\ln \frac{\Lambda_{i;kl}}{\hat{\Lambda}_{i;kl}}}$$

where $Q$ is chi-square distributed with $17(s-1)^2 = 34$ degrees of freedom. The observed value of $Q$ is 34.55 while the tabulated value at the 5% level of significance is 48.57. Thus, $H_0$ cannot be rejected, that is, we can assume safely we have a first order Markov chain.
7. THE ESTIMATES

Once the basic assumptions have been verified, it is a simple matter to substitute the data values into the formulae of Section 4. Thus, we obtain

\[ \hat{\sigma}_{11}^2 = 474794 \times 10^{-8}, \quad \hat{\sigma}_{21}^2 = 353397 \times 10^{-8} \]
\[ \sigma_{11}^2 = 122905 \times 10^{-8}, \quad \sigma_{21}^2 = 553235 \times 10^{-8} \]
\[ \overline{G}_1 = 0.02186, \quad \overline{G}_2 = 0.02307, \]
\[ c_1 = 3.8635, \quad c_2 = 0.6388, \]
and
\[ r_{11} = 8.9633, \quad r_{21} = 0. \]

Therefore,
\[ \Lambda_{EB;1} = 0.622, \quad \Lambda_{EB;2} = 0.419. \]

That is,
\[ \Lambda_{EB} = \begin{bmatrix} 0.622 & 0.375 \\ 0.419 & 0.581 \end{bmatrix} \]

We could compare this empirical Bayes estimate with the maximum likelihood estimate for 1977 which is
\[ \Lambda_{ML;1} = \frac{30}{48} = 0.625, \quad \Lambda_{ML;2} = 0.419. \]

That is,
\[ \Lambda_{ML} = \begin{bmatrix} 0.625 & 0.375 \\ 0.419 & 0.581 \end{bmatrix} \]

The advantage of the empirical Bayes estimate is that all the data from previous years is being used, that is, we are benefiting from past experience whereas the maximum likelihood estimate is found by using the data of 1977 alone.
8. CONCLUSION

Once $\hat{\Lambda}$, the estimate of $\Lambda$, has been obtained there are many applications and quantities of interest that can be further estimated. One such example relates to the work of Gabriel and Neuman (1962) in which they used $\hat{\Lambda}$ to determine the distribution of weather cycles. For this purpose a wet (dry) spell of $k$ days is defined as a sequence of $k$ wet (dry) days preceded and followed by a dry (wet) day. Let $W$ denote the length of a wet spell. Then

$$P(W = k) = \hat{\Lambda}_2(1 - \hat{\Lambda}_2)^{k-1}, \quad k = 1, 2, \ldots.$$  

A weather cycle is defined as combinations of a wet spell and an adjacent dry spell. Let $C$ denote the length of a weather cycle. Then,

$$P(C = m) = m \hat{\Lambda}_2(1 - \hat{\Lambda}_2)/(1 - \hat{\Lambda}_1 - \hat{\Lambda}_2) \quad (1 - \hat{\Lambda}_2)^{m-1} - \hat{\Lambda}_1^{m-1}, \quad m = 1, 2, \ldots.$$  

We refer the reader to the paper cited for more applications of this type.
Table 1
Frequency count matrix of summer days precipitation at Tallahassee 1961-1977.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1961</td>
<td>28</td>
<td>19</td>
</tr>
<tr>
<td>1962</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>1963</td>
<td>28</td>
<td>18</td>
</tr>
<tr>
<td>1964</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>1965</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>1966</td>
<td>41</td>
<td>50</td>
</tr>
<tr>
<td>1967</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>1968</td>
<td>51</td>
<td>40</td>
</tr>
<tr>
<td>1969</td>
<td>33</td>
<td>18</td>
</tr>
<tr>
<td>1970</td>
<td>42</td>
<td>49</td>
</tr>
<tr>
<td>1971</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>1972</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>1973</td>
<td>28</td>
<td>19</td>
</tr>
<tr>
<td>1974</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>1975</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>1976</td>
<td>26</td>
<td>19</td>
</tr>
<tr>
<td>1977</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>1978</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>1979</td>
<td>46</td>
<td>45</td>
</tr>
<tr>
<td>1980</td>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td>1981</td>
<td>33</td>
<td>58</td>
</tr>
<tr>
<td>1982</td>
<td>56</td>
<td>55</td>
</tr>
<tr>
<td>1983</td>
<td>37</td>
<td>54</td>
</tr>
<tr>
<td>$t/t^1$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>$F_{j^+}$</td>
<td>54</td>
<td>37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t/t^1$</th>
<th>1</th>
<th>2</th>
<th>$F_{j^+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>18</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>25</td>
<td>43</td>
</tr>
<tr>
<td>$F_{j^+}$</td>
<td>48</td>
<td>43</td>
<td>91</td>
</tr>
</tbody>
</table>
Figure 1a - Plot of $\hat{\Lambda}_1(i)$ against $\hat{\Lambda}_1(i - 1)$

Figure 1b - Plot of $\hat{\Lambda}_2(i)$ against $\hat{\Lambda}_2(i - 1)$
REFERENCES


