Computing Probabilities Associated with the Markov Chain Model for Precipitation

RICHARD W. KATZ

Dept. of Statistics, Pennsylvania State University, University Park 16802

9 November 1973 and 19 August 1974

1. Introduction

The two-state Markov chain has been shown in many instances to be an appropriate model for sequences of wet and dry days (see, for example, Gabriel and Neumann, 1962). In such a model the distribution of the number of wet days in a fixed period of time is often of interest. Gabriel (1959) has derived an exact formula for this distribution, and Gabriel and Neumann used this formula to compute the distribution of the number of wet days per week. However, Gabriel notes that it is a difficult task to calculate the desired probabilities using his formula.

The purpose of this note is to present a method by which this probability distribution can be computed with relative ease. This approach involves the use of a recurrence relation for the probabilities. As an example, the distribution of wet days is calculated by this method using parameters estimated from State College, Pennsylvania, precipitation data.

2. Markov chain model

Let

$$X_n=\begin{cases}1, & \text{if } n\text{th day is wet}\\ 0, & \text{if } n\text{th day is dry}\end{cases}$$

Assume that $(X_n, n=0, 1, \ldots)$ is a two-state Markov chain, with transition probabilities $P_{ij}=\Pr(X_n=j|X_{n-1}=i)$, for $i=0, 1$, $j=0, 1$, and $n=1, 2, \ldots$, and initial probability of a wet day $p=\Pr(X_0=1)$.

In N consecutive days, the number of wet days is

$$S_N=\sum_{n=1}^{N} X_n, \qquad \text{for} \qquad N=1, 2, \ldots.$$

We define probabilities $W(k;N)=\Pr(S_N=k)$, $W_0(k;N)=\Pr(S_N=k|X_0=0)$, and $W_1(k;N)=\Pr(S_N=k|X_0=1)$, for $k=0, 1, \ldots, N$. Conditioning on whether the initial day is dry or wet,

$$W(k;N)=(1-p)W_0(k;N)+pW_1(k;N). \qquad (1)$$

As indicated in the Introduction, Gabriel has derived expressions for $W_0(k;N)$ and $W_1(k;N)$.

3. Recurrence relation

A recurrence relation for $W_0(k;N)$ and $W_1(k;N)$ is obtained by conditioning on X_1. In order for there to be exactly k wet days out of N, either

1) $X_1=0$ and exactly k of the remaining $N-1$ days are wet, or

TABLE 1. Distribution of the number of wet days for State College data for 1–28 February (1950–69).*

N	k	$W_0(k; N)$	$W_1(k; N)$	$W(k; N)$
5	0	0.121	0.098	0.112
	1	0.269	0.245	0.259
	2	0.302	0.304	0.303
	3	0.206	0.227	0.214
	4	0.085	0.103	0.092
	5	0.017	0.023	0.020
10	0	0.015	0.012	0.014
	1	0.064	0.055	0.060
	2	0.141	0.128	0.136
	3	0.207	0.198	0.203
	4	0.220	0.221	0.220
	5	0.176	0.185	0.180
	6	0.108	0.119	0.112
	7	0.050	0.057	0.053
	8	0.016	0.020	0.018
	9	0.004	0.005	0.004
	10	0.000	0.000	0.000

* $P_{01}=0.344$ $P_{11}=0.472$ $p=0.400$.

2) $X_1=1$ and exactly $k-1$ of the remaining $N-1$ days are wet.

Note that the first possibility does not exist if $k=N$, and the second cannot occur if $k=0$.

Suppose that $X_0=0$. Then the first possibility occurs with probability $P_{00}W_0(k; N-1)$ and the second with probability $P_{01}W_1(k-1; N-1)$, so that

$$W_0(k; N)=P_{00}W_0(k; N-1)+P_{01}W_1(k-1; N-1). \quad (2)$$

Similarly,

$$W_1(k; N)=P_{10}W_0(k; N-1)+P_{11}W_1(k-1; N-1). \quad (3)$$

Here $k=0, 1, \ldots, N$, and $N=1, 2, \ldots$, with initial conditions $W_0(0; 0)=W_1(0; 0)=1$ and the conventions that $W_0(N; N-1)=W_1(-1; N-1)=0$. This recurrence relation has been derived by Helgert (1970) and others.

For a given set of transition probabilities (P_{ij}) and initial probability p, Eqs. (2) and (3) can be used to compute recursively $W_0(k; N)$ and $W_1(k; N)$, letting $N=1, 2, \ldots$ and $k=0, 1, \ldots, N$. Also, $W(k; N)$ can be determined from $W_0(k; N)$ and $W_1(k; N)$ using (1).

4. Numerical example

In any practical application, before the distribution of the number of wet days can be calculated, the parameters of the Markov chain model must be estimated. Table 1 includes the usual relative frequency estimates for the transition probabilities (these are the maximum likelihood estimates) and the initial probability, based on State College daily precipitation data for the month of February (omitting 29 February in leap years) during the years 1950 through 1969. Here a wet day is taken to be one on which at least 0.01 inch of precipitation falls. The probability distribution has been calculated with the aid of a computer, and is given in Table 1 for $N=5$ and 10.

Using Gabriel's formula to compute the entire probability distribution, for $N=1, 2, \ldots, 30$, requires roughly five times as much computing time as that for the recurrence relation approach. Further, this comparison of computing times does not take into account the relative difficulty of writing a computer program for Gabriel's complicated formula.

5. Remarks

Exact expressions for probability distributions of interest are often either too complicated to be useful in applications or impossible to derive explicitly. However, the researcher should be aware that many times, when faced with such a situation, alternative methods can be developed. As we illustrate here, for the Markov chain and other related probability models, recurrence relations can be used to compute certain probabilities associated with the model.

Acknowledgments. I wish to thank Prof. Frank Haight for his critical reading of the paper. This research was conducted by the author while supported by a National Science Foundation Traineeship.

REFERENCES

Gabriel, K. R., 1959: The distribution of the number of successes in a sequence of dependent trials. *Biometrika*, **46**, 454–460.

——, and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90–95.

Helgert, H. J., 1970: On sums of random variables defined on a two-state Markov chain. *J. Appl. Prob.*, **7**, 761–765.