# NOTES

## Computing a Probability Distribution for the Start of the Rains from a Markov Chain Model for Precipitation

R. D. STERN

*Department of Applied Statistics, University of Reading, England*

11 July 1981 and 17 November 1981

ABSTRACT

The start of the rains is defined to be the first day, or two-day rain spell, in which greater than a specified total amount of rain occurs. A method of calculating the probability distribution of the start of the rains from a Markov chain model of the daily precipitation is given. The application of this method is illustrated with data from Samaru, Nigeria.

### 1. Introduction

Markov chain models of the daily occurrence of precipitation have been used widely since Gabriel and Neumann (1962). These models have sometimes been combined with gamma-distributed rainfall amounts to provide a model for the probability of precipitation and of precipitation amounts on a daily basis (e.g., Katz, 1977; Buishand, 1977). The models, however, are of comparatively little value by themselves and it is important to be able to calculate, from these models, further probabilities that are of direct practical use. From the model for the occurrence of precipitation, Katz (1974, 1977) showed how a simple recurrence relation could be used to calculate the probability distribution of the total number of rain days and the total precipitation in a given period. Stern (1980b) used similar recurrence relations to calculate the probability of the occurrence of dry spells of any given length. In areas with a distinct dry season, the date at which the rains start is an important agroclimatological variable and the purpose of this note is to illustrate the way in which a probability distribution for this date can be calculated. The event which signals the start of the rains is here defined to be the first day, or two-day rain spell, in which the total precipitation is greater than a given amount $x$. In the examples considered later, $x$ is usually taken as 20 mm (Davy *et al.*, 1976). This event was also used by Virmani (1975) to investigate the probability distribution of dates of pre-season cultivation and planting.

### 2. Markov chain model

The rainfall on day $n$ is described by a pair of random variables, $(J_n, X_n)$, where $J_n = 1$ if precipitation occurs on day $n$ and $J_n = 0$ otherwise. $X_n$ is the precipitation that occurs on day $n$; thus $X_n = 0$ if $J_n = 0$. It is assumed that $J_n$ is a two-state Markov chain with transition probabilities $p_{ij}(n) = \Pr(J_n = j|J_{n-1} = i)$. The conditional distribution functions of the rainfall amounts are given by

$$\left.\begin{aligned} F(x) &= \Pr(X_n < x|J_n = 1) \\ F_2(x) &= \Pr(X_{n-1} + X_n < x|J_{n-1} = J_n = 1) \end{aligned}\right\}. \quad (1)$$

The conditional density function of $X_n$ is assumed to be gamma, i.e., $f(z) = (k/\mu)^k z^{k-1} \exp(-kz/\mu)/\Gamma(k)$, $z > 0$, $k, \mu > 0$, where $F(x) = \int_0^x f(z)dz$. The parameter $\mu$ is the mean rain per rain day and $1/\sqrt{k}$ is the coefficient of variation of the distribution. If $k = 1$, $f(z) = (1/\mu)\exp(-z/\mu)$, the negative exponential distribution.

The start of the rains is the first occurrence of the event

$$X_{i-1} + X_i > x \quad \text{for} \quad i \geq m.$$

It is convenient to be able to choose an earliest starting date $m$ so that very early rainfalls can, if required, be ignored. A random variable $Z_n = 1$ if the event signaling the start of the rains has occurred by day $n$, otherwise $Z_n = 0$. We define

$$\left.\begin{aligned} r_0(n) &= \Pr(Z_n = 0, \quad J_n = 0) \\ r_1(n) &= \Pr(Z_n = 0, \quad J_{n-1} = 0, \quad J_n = 1) \\ r_2(n) &= \Pr(Z_n = 0, \quad J_{n-1} = 1, \quad J_n = 1) \\ r_3(n) &= \Pr(Z_n = 1) \end{aligned}\right\}. \quad (2)$$

If rain occurs on day $n$ we also have to evaluate the probability that it will be sufficient for the event to occur. Hence, the terms $v_i(n)$, $i = 1, 2, 3$, are defined

0021-8952/82/030420-04$05.00

where

$$\left.\begin{aligned} v_i(n) &= 0 \quad \text{for} \quad n < m \\ v_1(m) &= \Pr(X_m > x \mid J_{m-1} = 0, \quad J_m = 1) \\ v_2(m) &= v_3(m) \\ &= \Pr(X_{m-1} + X_m > x \mid J_{m-1} = J_m = 1) \end{aligned}\right\}, \quad (3)$$

and, for $n > m$,

$$\left.\begin{aligned} v_1(n) &= \Pr(X_n > x \mid J_{n-1} = 0, \quad J_n = 1) \\ v_2(n) &= \Pr(X_{n-1} + X_n > x \mid J_{n-1} = J_n = 1, \\ &\qquad J_{n-2} = 0, \quad X_{n-1} < x) \\ v_3(n) &= \Pr(X_{n-1} + X_n > x \mid J_{n-2} = J_{n-1} \\ &\qquad = J_n = 1, X_{n-2} + X_{n-1} < x) \end{aligned}\right\}. \quad (4)$$

The probability that the rains have started by day $n$ is then given by

$$r_3(n) = r_3(n-1) + r_0(n-1)p_{01}(n)v_1(n)$$
$$+ r_1(n-1)p_{11}(n)v_2(n) + r_2(n-1)p_{11}(n)v_3(n). \quad (5)$$

The full recurrence relations for $r(n) = [r_0(n), r_1(n), r_2(n), r_3(n)]$ are given by the transition matrix

$$r(n) = r(n-1) \times \begin{bmatrix} p_{00} & p_{01}(1-v_1) & 0 & p_{01}v_1 \\ p_{01} & 0 & p_{11}(1-v_2) & p_{11}v_2 \\ p_{10} & 0 & p_{11}(1-v_3) & p_{11}v_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6)$$

For simplicity of presentation the suffix $n$ has been omitted from the terms in the transition matrix. In calculating the $v_i(n)$, the additional assumption is made that the successive $X_n$ are conditionally independent given $J_n$. The only terms that are relatively complicated to evaluate are $v_2(n)$ and $v_3(n)$ for $n > m$. They can be written as

$$\left.\begin{aligned} v_2(n) &= [F(x) - F_2(x)]/F(x) \\ v_3(n) &= \left[\int_0^x F(x-z) \right. \\ &\qquad \left. \times (1 - F(x-z))f(z)dz\right] \Big/ F_2(x) \end{aligned}\right\}. \quad (7)$$

The term $v_3(n)$, in general, must be evaluated by numerical integration.

The extension of these recurrence relations for second-order Markov chains is straightforward. It is also important to note that the recurrence relations do not require stationary transition probabilities. They would be of little practical value without this flexibility. The assumption that has been made, of a single distribution of rainfall amounts, is also limiting. At most sites the distribution, particularly the mean rain per rain day, $\mu$, changes through the year. The distribution is also sometimes different for the first and subsequent rain days in a spell. This complicates the calculation of $v_2(n)$ and $v_3(n)$, which both then have to be evaluated by numerical integration. Alternatively, approximate values can be evaluated from the formulas given, using average values of $\mu$ and $k$ for each calculation.

## 3. Examples

To illustrate the recurrence relations, a simple model is first considered, where the probabilities of precipitation are stationary and are given by $p_{01} = 0.25$ and $p_{11} = 0.75$. Rainfall amounts are assumed to be exponentially distributed with mean $\mu = 10$ mm, and it is assumed that 20 mm are required for the rains to start. For the exponential distribution the terms $v_i(n)$ can all be evaluated explicitly. Thus, letting $x' = x/\mu = 2$, in this example,

$$\left.\begin{aligned} v_1(n) &= \exp(-x') && = 0.1353, \quad n \geqslant m \\ v_2(m) &= v_3(m) = (1 + x')\exp(-x') && = 0.4060 \\ v_2(n) &= x'/(\exp(x') - 1) && = 0.3130, \quad n > m \\ v_3(n) &= [x' - 1 + \exp(-x')]/[\exp(x') - 1 - x'] && = 0.2587, \quad n > m \end{aligned}\right\}. \quad (8)$$

We assume that the earliest starting date is such that the system is in equilibrium prior to this date. It is then straightforward to show that

$$r(m-1) = [0.5, 0.125, 0.375, 0], \quad (9)$$

and use of the recurrence relations gives the cumulative probability of the rains beginning within the first four days as

| Day | $m$ | $(m+1)$ | $(m+2)$ | $(m+3)$ |
|---|---|---|---|---|
| Probability | 0.1692 | 0.2547 | 0.3304 | 0.3978 |

The data for a 48-year record at Samaru (1928–75) provide a more realistic example. Samaru is in northern Nigeria (11°11′N, 7°37′E). The mean annual rainfall was 1100 mm, almost all of which occurred between April and October. The method of

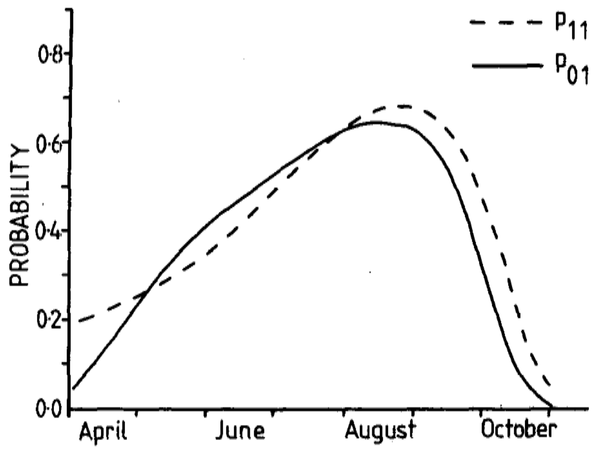


FIG. 1. Estimated probability of rain, Samaru.

fitting a Markov chain model to this set of data is described by Stern (1980a) and Fig. 1 gives the curves fitted to the proportion of rain days. The probabilities $p_{01}(n)$ and $p_{11}(n)$ are not stationary. They were estimated by fitting a linear model to the logit transform of the proportion of rain days.

The equations for the estimated probabilities of rain are given by

$$p_{i1}(n) = \exp[g_i(n)] \times \{1 + \exp[g_i(n)]\}^{-1}, \quad i = 0, 1, \tag{11}$$

where

$$g_0(n) = -3.204 + 9.984 \times 10^{-2}n - 1.394 \times 10^{-3}n^2 + 1.017 \times 10^{-5}n^3 - 2.801 \times 10^{-8}n^4, \tag{12}$$

$$g_1(n) = -1.488 + 1.663 \times 10^{-2}n - 1.897 \times 10^{-4}n^2 + 3.056 \times 10^{-6}n^3 - 1.253 \times 10^{-8}n^4. \tag{13}$$

Values of $n$ from 1–214 correspond to 1 April–31 October. The mean rainfall per rain day was 8.99 mm in April and 13.00 mm in May–September, while the common value of $k$ was estimated as 0.854.

The cumulative probability of the start of the rains is given in Fig. 2. This graph and Table 1 show that, with 1 April as the earliest possible starting date, there is an estimated 20% chance that the event signalling the start of the rain will occur by 17 April. However, the median starting date is not until early May and in one year in five the starting date is in mid-May or later. It is also important to evaluate the risks to crops which are planted early in the season. Using similar recurrence relations (see Stern 1980), we can show that there is estimated to be at least a 50% chance of a dry spell of seven or more days in the next month if a crop is planted before 1 May (Fig. 3). If, for some crops, this risk is too great, it may be worth waiting until later in the year before accepting the event as a start to the rains. Thus Fig. 2 also gives the distribution of the start of the rains if the earliest possible date is 1 May.

As a check of the adequacy of the model that is fitted to the daily data, the distribution of the start

TABLE 1. Percentage points of the distribution of the start of the rains for Samaru calculated directly from the data and from the model.

| Earliest date | | 20% | 50% | 80% |
|---|---|---|---|---|
| 1 April | Data | 20 April | 2 May | 20 May |
| | Model | 17 April | 2 May | 14 May |
| 1 May | Data | 5 May | 13 May | 24 May |
| | Model | 4 May | 10 May | 21 May |

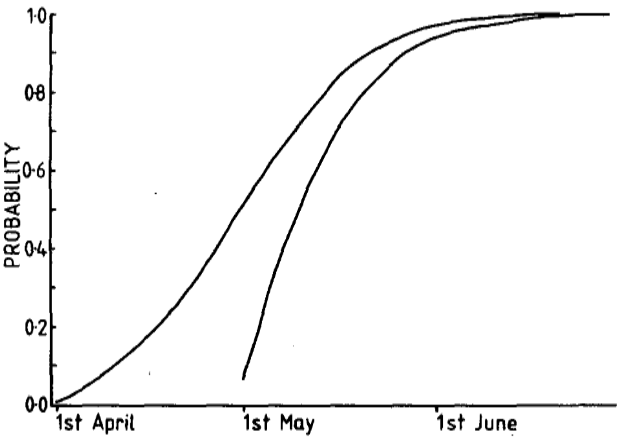


FIG. 2. Cumulative probability of the start of the rains, with the earliest starting dates of (a) 1 April and (b) 1 May. The start is defined as the first day, or two-day spell, in which greater than 20 mm occurs.

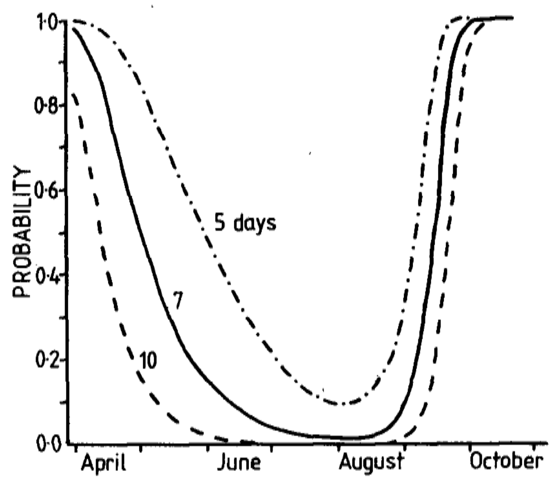


FIG. 3. Probability of a dry spell of greater than 5, 7 and 10 days in the next 30 days after an initial rain day.

of the rains can be calculated by merely noting the date each year on which 20 mm was first exceeded. Table 1 gives the observed percentage points of the start of the rains together with the estimates from the model and shows that, for this characteristic, the model fits the data well.

*Acknowledgments.* I wish to thank the Overseas Development Administration of the British Government for their support of the work of the Tropical Agricultural Meteorology group at Reading University. I also thank colleagues within the group for their assistance and the reviewers for constructive comments on an earlier draft of this paper.

REFERENCES

Buishand, T. A., 1977: *Stochastic Modelling of Daily Rainfall Sequences.* Wageningen, Veenman and Zonen, 211 pp.

Davy, E. G., F. Mattei and S. I. Solomon, 1976: Special environmental report, No. 9. World Meteorological Organization, 289 pp.

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

Katz, R. W., 1974: Computing probabilities associated with the Markov chain model for precipitation. *J. Appl. Meteor.*, **13**, 953–954.

——, 1977: Precipitation as a chain-dependent process. *J. Appl. Meteor.*, **16**, 671–675.

Stern, R. D., 1980a: Analysis of daily rainfall at Samaru, Nigeria, using a simple two-part model. *Arch. Meteor. Geophys. Bioklim.*, **B28**, 123–135.

——, 1980b: The calculation of probability distributions for models of daily precipitation. *Arch. Meteor. Geophys. Bioklim.*, **B28**, 137–147.

Virmani, S. M., 1975: The agricultural climate of the Hyderabad region in relation to crop planning. Internal Report, Farming Systems Program, ICRISAT, Hyderabad, India, 55 pp.