551.5.506.3 : 311.17

# A statistical model suitable for describing weather persistence

T. R. SRINIVASAN

*Meteorological Office, Poona*

(*Received 24 June 1958*)

## 1. Introduction

Persistence is a factor of major importance in meteorological studies. In a system where there is no persistence, the probability $p$ of occurrence of an event and $q$ $(=1-p)$ that of non-occurrence are independent of what has happened previously. It can be easily seen that the cumulative frequencies $F_n$ of occurrences of runs of length $n$ or more are terms of the geometric series $F$, $pF$, $p^2F$ so that the ratio of the successive terms of the actual cumulative frequency distribution provides an estimate of the probability $p$. However, in persistent series such as are usually met with in meteorology the ratios of the cumulative figures are not constant and this feature which is brought about due to the effect of persistence requires to formulate a different model. Newnham (1916) assuming that the probability continuously increases with the length of runs worked the persistence of wet and dry weather at 50 stations. Gold (1929) discussed the problem when runs are restricted by the length of series and studied the frequencies of occurrence of sequences in a series of events of two types with equal chance of occurrence and non-occurrence. Cochran (1936) made a notable contribution on the subject by extending it to unequal chances and applied it for the study of wet and dry sequences. Here with $p$ the probability of the occurrence, $x$ the duration of the spell and $m$ the number of units of time (days) in the restricted series (season) and $n$ the number of series (seasons or years), the frequency of spells of length $x$ units is given by

$$S_x = n[p^x\, q\{2+q(m-x-1)\}]$$

Williams (1952) studied the behaviour of certain weather spells exhibiting tendency to persist by fitting logarithmic series $\alpha p^x/x$ for frequencies of runs of length $x$; the parameters $\alpha$ and $p$ being easily determined from the frequencies and the first moment of the distribution.

In analysing sequence of wet spells and dry spells in respect of some of the Indian stations, it has been found that the geometric series gives a good fit, in some logarithmic series, while in some others either of these fails to give a good fit. It has thus become necessary to search for more suitable models; a generalised model has been described and a method of deriving the parameters has been worked out.

## 2. Generalised models for frequency of runs

If $S_x$ is the frequency of runs of like events of length $x$ we assume the following model for representing $S_x$ so that if we stop at the $r^{\text{th}}$ stage, we have $S_x$ equals

$$A_1 p^x + A_2 p^x x^{-1} + \ldots A_r\, p^r\, x^{-r+1} + \varepsilon_r\, x^{-r} \quad (1)$$

where $\varepsilon_r\, x^{-r}$ is the residual after a representation with $r$ terms having been fitted so that $\Sigma[\epsilon_r\, x^{-r}]^2$ provides a measure of the deviation of the calculated frequencies and the actual frequencies.

Model I

$$S_x = A_1 p^x \quad \ldots \quad \ldots \quad (2)$$

$$\therefore \sum_1^\infty S_x = \sum_1^\infty A_1 p^x = N = \frac{A_1 p}{1-p} \quad (3)$$

$$\text{and } \sum_1^\infty x S_x = \frac{A_1 p}{(1-p)^2}$$

$$= G_1 \text{ the first moment} \quad (4)$$

Solving for $A_1$ and $p$ we get

$$p = \frac{G_1 - N}{G_1} \quad \ldots \quad \ldots \quad (5)$$

$$\text{and } A_1 = \frac{N^2}{G_1 - N} \quad \ldots \quad \ldots \quad (6)$$

Here $S_x/(S_x - 1) = p$, so that when successive frequencies bear a constant ratio, this model is suitable.

Model II

$$S_x = A_1 p^x + A_2 p^x x^{-1} \quad \ldots \quad (7)$$

This is an extended model comprising of both constant probability (when $A_2 = 0$) and the logarithmic model of Williams (when $A_1 = 0$).

If $G_r$ and $L_r$ represent the $r^{\text{th}}$ moment of the geometric series $p^x$ and logarithmic series $p^x x^{-1}$ respectively, the $r^{\text{th}}$ moment of the frequency distribution represented by (7) is

$$T_r = \Sigma x^r S_x = A_1 \Sigma x^r p^x + A_2 \Sigma x^{r-1} p^x$$

Therefore,

$$T_0 = A_1 G_0 + A_2 L_0 \quad (8)$$

$$T_1 = A_1 G_1 + A_2 G_0 \quad (9)$$

$$T_2 = A_1 G_2 + A_2 G_1 \quad (10)$$

$$T_3 = A_1 G_3 + A_2 G_2 \quad (11)$$

and in general

$$T_r = A_1 G_r + A_2 G_{r-1}$$

We know $G_0 = \dfrac{p}{1-p}$

$$G_1 = \frac{p}{(1-p)^2}$$

$$G_2 = \frac{p(1+p)}{(1-p)^3}$$

$$G_3 = \frac{p(1+4p+p^2)}{(1-p)^4}$$

Eliminating $A_2$ from equations (9) and (10) we get

$$A_1 = \frac{(1-p)^2}{p^2} \left\{ T_2(1-p) - T_1 \right\} \quad (12)$$

Eliminating $A_2$ from equations (10) and (11) we get

$$A_1 = \frac{(1-p)^3}{2p^2} (1+p) \left\{ T_3 \frac{1-p}{1+p} - T_2 \right\} \quad (13)$$

Or re-arranging,

$$p^2(T_3 + T_2) - 2p(T_3 - T_2) + T_3 - 3T_2 + 2T_1 = 0 \quad (14)$$

This is a quadratic equation and provides two values for $p$ given by

$$p = \frac{T_3 - T_2 \pm \sqrt{4T_2^2 - 2T_1(T_2 + T_3)}}{T_3 + T_2} \quad (15)$$

In the above equation if $2T_2^2 < T_1(T_2 + T_3)$, $p$ becomes imaginary and hence, we will not consider such cases. If $2T_2^2 > T_1(T_2 + T_3)$ we get two values for $p$.

$$p_1 = \frac{T_3 - T_2 + \sqrt{4T_2^2 - 2T_1(T_2 + T_3)}}{T_3 + T_2}$$

$$p_2 = \frac{T_3 - T_2 - \sqrt{4T_2^2 - 2T_1(T_2 + T_3)}}{T_3 + T_2}$$

We shall have to restrict our attention only to such cases when $p$ is positive.

*i.e.*, $(T_3 - T_2)^2 > 4T_2^2 - 2T_1(T_2 + T_3)$

*i.e.*, $T_3^2 - 3T_2^2 > 2(T_2T_3 - T_1T_2 - T_1T_3)$

The value of $A_1 = \frac{(1-p)^2}{p^2}\left(T_2\,\overline{1-p} - T_1\right)$

If $T_1 > (1-p)T_2$, then $A_1$ becomes negative and hence $S_x$ becomes negative when

$$x > \frac{A_2}{|A_1|}$$

so that in this case the summation of the frequencies is not valid.

As an illustration we shall now apply the model

$$S_x = A_1 p^x + A_2 p^x x^{-1}$$

to the distribution of the rain spells in June at Poona (data represent for the years 1891—1950) which is given in Table 1.

For the above distribution, $T_1 = 726$, $T_2 = 3510$ and $T_3 = 26178$. The values of $p_1$ and $p_2$ are as follows:

$$p_1 = 0{\cdot}847 \quad \text{and} \quad p_2 = 0{\cdot}680$$

The values of the parameters $A_1$ and $A_2$ corresponding to $p_1$ and $p_2$ are given below—

| $p$ | $A_1$ | $A_2$ |
|---|---|---|
| 0·847 | —6·2 | 171·4 |
| 0·680 | 88·0 | 66·7 |

Since $A_1$ is negative for $p = 0{\cdot}847$, this is omitted. We now consider the value of $p = 0{\cdot}680$ and the corresponding values for $A_1$ and $A_2$. Frequencies computed on the basis of the above values for $p$, $A_1$ and $A_2$ are found in column 3 of Table 1.

**TABLE 1**

**Frequency distribution of rain spells in June at Poona during the years 1891—1950**

| Length of rain spell (days) | Frequency | | |
|---|---|---|---|
| | Actual | Calculated* from the method found in this paper | Calculated* from Williams' Log series |
| 1 | 92 | 101·0 | 114·4 |
| 2 | 65 | 53·9 | 48·1 |
| 3 | 32 | 33·4 | 27·0 |
| 4 | 16 | 21·5 | 17·0 |
| 5 | 14 | 14·1 | 11·5 |
| 6 | 12 | 9·4 | 8·0 |
| 7 | 4 | 6·3 | 6·9 |
| 8 | 6 | 4·2 | 5·0 |
| 9 | 2 | 2·9 | 3·7 |
| 10 | 7 | 1·9 | 3·0 |
| 11 | 1 | 1·7 | 2·2 |
| 12 | 1 | 0·9 | 1·7 |
| 13 | 0 | 0·6 | 1·3 |
| 14 | 0 | 0·4 | 1·0 |
| 15 | 1 | 0·3 | 0·8 |

*Zero error adjustment has been made

It can also be seen that the observed and the calculated frequencies agree well.

### 3. Acknowledgements

The author wishes to express his grateful thanks to Sarvashri P. Jagannathan, Meteorologist and K. S. Ramamurty, Assistant Meteorologist for many valuable suggestions and for going through the manuscript and also to the Director of Agricultural Meteorology for the facilities provided in connection with the above study.

*REFERENCES*

| | | |
|---|---|---|
| Cochran, W. G. | 1936 | *J. R. Statist. Soc.*, Suppl. 3, p. 49. |
| Gold, E. | 1929 | *Quart. J. R. met. Soc.*, **55**, p. 307. |
| Newnham E. V. | 1916 | *Ibid.*, **42**, p. 153. |
| Williams, C. B. | 1952 | *Ibid.*, **78**, pp. 91-96. |