Qatar Univ. Sci. J. (1996), 16(2): 231-235

A MARKOV CHAIN MODEL FOR RIVER FLOW

A.J. SALIM* and B.Y. THANOON**
*Department of Mathematics, College of Education, University of Tikrit, Iraq.
**Department of Mathematics, College of Science, University of Mosul, Iraq.

نموذج سلسلة ماركوف لمنسوب النهر

عبد الغفور جاسم سالم و باسل يونس ذنون

تتعامل هذه الورقة مع اجراء تحويل ثلاثي لمنسوب النهر، إذ ان المنسوب المحول ينمذج باعتباره متسلسلة ماركوف ذات ثلاثة حالات ، وتناقش من خلال هذه الورقة مسألتي التقاطع العلوي والتقاطع السفلي سوية . ويتم أشتقاق التوزيع الاحتمالي المستقر للمنسوب المحول اضافة للتوزيعات الاحتمالية لبعض المتغيرات ذوات العلاقه بالمسأله. وأخيرا تطبق النتائج النظريه التي حصلنا عليها في هذه الورقه على نهر (Jökulsa) في ايسلاندا .

ABSTRACT

This paper is concerned with a trinary transformation of river flow. The transformed flow is modeled as a three-states Markov chain and both upcrossing and downcrossing problems are studied simultaneously. The stationary distributions of some related variables are derived. An application to Jokulsa Eystri river flow is presented.

Key Words: Markov Chain; Crossing; Stationary Distribution; Waiting Time.

INTRODUCTION

A class of properties which have special significance in hydrology deals with the instants of which observations from a time series are greater or less than specified values. For flow series those aspects will be of interest in relation to the regulation and abstraction of river flows by reservoir storage. However, as Lawrance and Kottegoda [1] observed, the theory of crossing properties has not found great use in hydrology, perhaps because of the discrete and non-Gaussian nature of most hydrologic series. Battaglia [2] considered a binary transformation of a time series and studied upcrossing problems by means of well-known methods developed for Markov chains. Our main interest in this paper will be in a more general case. We shall consider a trinary transformation that will denote river flow by three states; namely; state -1 (downcrossing), state 0 (steady state) and state 1 (upcrossing).

Let $\{X_t\}$ be a strictly stationary time series observed at regular time intervals. For example, X_t is the daily flow of a river and u, w are some critical levels of flow. The behavior around u and w may be analysed by defining a trinary transformation of $\{X_t\}$ of the form

$$J_t = \begin{cases} -1; \text{if } X_t < w \\ 0; \text{if } w \le X_t \le u \\ 1; \text{if } X_t > u \end{cases}$$

The process $\{J_t\}$ whose dependence upon the past is limited to a finite number of preceding values. Such process is, a Markov chain with three states corresponding to the events $E_1 = \{X_t; X_t < w\}$, $E_2 = \{X_t; w \le X_t \le u\}$ and $E_3 = \{X_t; X_t > u\}$.

Letting

$a = P(J_t = 0/J_{t-1} = -1)$
$b = P(J_t = -1/J_{t-1} = 0)$
$c = P(J_t = 1/J_{t-1} = 0)$
$d = P(J_t = 0/J_{t-1} = 1)$

obviously a, b, c and d are constants independent of time because the process is stationary. These constants are supposed to be strictly positive. The transition matrix P of $\{J_t\}$ has the following form:

$$\begin{array}{cc} \text{states} \rightarrow & \begin{array}{ccc} -1 & 0 & 1 \end{array} \\ \begin{array}{c} \downarrow \\ -1 \\ P = 0 \\ 1 \end{array} & \begin{bmatrix} 1-a & a & 0 \\ b & 1-b-c & c \\ 0 & d & 1-d \end{bmatrix} \end{array}$$

Stationary distribution

It is well-known that if the chain is ergodic then it has a unique stationary distribution (see Cox and Miller [3], p. 108). To check the ergodicity of the chain, we need to show that the transition matrix P is irreduciable and has one of the eigenvalues equal to one and exceed all other eigenvalues in modulus. It is easily shown that P is irreducible. Now by solving the system $|P - \lambda I| = 0$ we have

$$\lambda^3 + A_1\lambda^2 + A_2\lambda + A_3 = 0$$

where

$A_1 = (a+b+c+d) - 3$
$A_2 = 3 - 2(a+b+c+d) + (ac+ad+bd)$
$A_3 = (a+b+c+d) - (1+ac+ad+bd)$

Then

$\lambda_1 = 1$
$\lambda_2 = 0.5\,[\{2 - (a+b+c+d)\} + \{(a+b+c+d)^2 - 4(ac+ad+bd)\}^{1/2}]$
$\lambda_3 = 0.5\,[2 - (a+b+c+d) - \{(a+b+c+d)^2 - 4(ac+ad+bd)\}^{1/2}]$

since $0 < a, b, c, d < 1$, then $|\lambda_2|$ and $|\lambda_3|$ are less than one. Hence, the system is ergodic. Therefore, there is a unique stationary distribution which can be obtained from the solution of the system $(I - P)' X = 0$, where X is column vector and I is the identity matrix. The solution of this system gives the stationary probabilities of states which are $\pi_{-1} = bd/A$, $\pi_0 = ad/A$ and $\pi_1 = ac/A$ where $A = bd+ad+ac$ and $\pi_i = P(\text{state } i)$; $i = -1, 0, 1$.

Distribution of crossing

Crossing occurs when there is a transition from state 0 to state 1 (upcrossing) or when there is a transition from state 0 to state −1 (downcrossing).

Let Q_i and R_i denote the case of upcrossing and downcrossing in state i; respectively, i.e.:

$$Q_i = \begin{cases} 1; \text{if } w \le X_{t-1} \le u \text{ and } X_t > u \\ 0; \text{otherwise} \end{cases}$$

$$R_i = \begin{cases} 1; \text{if } w \le X_{t-1} \le u \text{ and } X_t < w \\ 0; \text{otherwise} \end{cases}$$

Let Y_i denotes the case of crossing (up and down) at the ith step, then $Y_i = Q_i + R_i$.

Let $N_U(0, t)$, $N_D(0, t)$ denote the number of upcrossing and downcrossing in the interval (0, t); respectively. Let $N_Y(0, t)$ denotes the number of crossing (up and down) in the interval (0, t). The distribution of Y_i depends on the previously occupied state or on the state that will be occupied next or on both. Now, let $F_{i,j}(\theta)$ (i, j = -1, 0, 1) be the moment generating function (m.g.f.) of the variable "transition from state i to state j in one step." Let $P(\theta)$ denote the matrix with typical elements $P_{i,j} F_{i,j}$, with $P_{i,j}$ being the transition probabilities from state i to state j. The m.g.f. of $N_Y 0, t)$ is given by (see Cox and Miller [3]).

$$F_{N_Y}(\theta) = (\pi_{-1}, \pi_0, \pi_1)\, P(\theta)^t (1, 1, 1)$$

in our case

$F_{-1,-1}(\theta) = F_{-1,0}(\theta) = F_{-1,1}(\theta) = 1,$
$F_{0,-1}(\theta) = F_{0,1}(\theta) = \bar{e}^{\theta}$ $F_{0,0}(\theta) = 1$ and
$F_{1,-1}(\theta) = F_{1,0}(\theta)$ $F_{1,1}(\theta) = 1.$

Hence

$$P(\theta) = \begin{bmatrix} 1-a & a & 0 \\ b\bar{e}^{\theta} & 1-b-c & c\bar{e}^{\theta} \\ 0 & d & 1-d \end{bmatrix}$$

The eigenvalues $\lambda_1(\theta)$, $\lambda_2(\theta)$ of $P(\theta)$ are the solution of the characteristic equation $|P(\theta) - \lambda(\theta) I| = 0$, i.e.

$\lambda^3(\theta) + B_1\lambda^2(\theta) + B_2\lambda(\theta) + B_3 = 0$, where

$B_1 = (a+b+c+d) - 3,$
$B_2 = 3 - 2(a+b+c+d) + (ab+ac+ad+bd+cd) - (ab+cd)\,\bar{e}^{\theta}$ and
$B_3 = (a+b+c+d) - (ab+ac+ad+bd+cd) + (ab+cd)\,\bar{e}^{\theta}$
$\quad - (acd+abd)\,\bar{e}^{\theta}\,(acd+abd) - 1$

The solution of this characteristic equation gives the following eigenvalues

$$\lambda_1(\theta) = -\frac{1}{3}(2\,p\ \sin r + B_1)$$

$$\lambda_2(\theta) = -\frac{1}{3}(\sqrt{3}\ p\ \cos r - p \sin r + B_1)$$

$$\lambda_3(\theta) = -\frac{1}{3}(\sqrt{3}\ p\ \cos r + p \sin r - B_1)$$

where

$$p = \sqrt{(B_1^2 - 3B_2)}\ ,\ q = 2B_1^3 - 9B_1B_2 + 27B_3 \text{ and}$$

$$r = \frac{1}{3}\sin^{-1}\left(\frac{-q}{2p^3}\right)$$

We note that $\lambda_i(0) = \lambda_i$, $i = 1, 2, 3$ in particular $\lambda_1(0) = 1$. Now since the matrix $P(\theta)$ is regular and one of its eigenvalues exceeds all others in absolute value, we can use the method of diagonal representation to write $\{\Phi(\theta)\}^t$ in the following form (see [3] page 139).

$$\{P(\theta)\}^t = Q(\theta)\begin{bmatrix} \lambda_1^t(\theta) & 0 & 0 \\ 0 & \lambda_2^t(\theta) & 0 \\ 0 & 0 & \lambda_3^t(\theta) \end{bmatrix} Q^{-1}(\theta)$$

where the columns $q_i(\theta)$ of $Q(\theta)$ are solutions of the equations $P(\theta)\, q_i(\theta) = \lambda_i(\theta)\, q_i(\theta)$, $(i = 1, 2, 3)$. The mean of $N_Y(0, t)$ may be obtained by using decomposition $Y_i = Q_i + R_i$. Now Y_i is the number of transitions to state -1 or state 1 at step i, so that

$$\begin{aligned} E\{N_Y(0, t)\} &= E\{N_U(0,t)\} + E\{N_D(0,t)\} \\ &= E(Q_1+Q_2+\ldots+Q_t)+E(R_1+R_2+\ldots+R_t) \\ &= t(\pi_1 + \pi_{-1}) \\ &= t(bd+ac/bd+ad+ac). \end{aligned}$$

And also, since the eigenvalue $\lambda_1(\theta)$ ($\lambda_1(0) = 1$) of $P(\theta)$ will exceed in absolute value all other eigenvalues in a neighbourhood of $\theta = 0$. Hence, as mentioned by (Cox & Miller [3] p. 139) asymptotically.

$$\operatorname{Log} E(e^{-Y_t\, \bar{e}^{\theta}}) \cong t \operatorname{Log}(\lambda_1(\theta)).$$

Thus a symptotically Y_t behaves like a sum of independent random variables.

Some interesting related variables

We define now some interesting variables related to the crossing problem. Let D_i denotes the number of times the process remains in state i. To find the probability distribution of D_i we observe that

$$p(D_{-1} = k) = p(X_j < w,\ 1 < j \le k,\ X_{k+1} \ge w / X_o \ge w,\ X_1 < w)$$

by using the Markovian property:

$$\begin{aligned} &p(D_{-1} = k) = a(1-a)^{k-1};\ k = 1, 2 \ldots \\ &p(D_0 = k) = (b+c)(1-b-c))^{k-1};\ k = 1, 2 \ldots \qquad (1) \\ &p(D_1 = k) = d(1-d)^{k-1};\ k = 1, 2 \ldots \end{aligned}$$

Consider now the variable T_i to be time between two consecutive crossing (up or down), $(i = -1, 1)$. i.e. Let T_1 denote the time between two upcrossing and T_{-1} denote the time between two consecutive downcrossing; T_1 and T_{-1} now are equal to the first return to state 1 or to state -1, respectively. Now $p(T_1 = k) = p(N_U(1, t)) = 0$,

$$\begin{aligned} &N_U(t, t+1) = 1/N_U(0, 1) = 1, \text{ and also} \\ &p(T_1 = k) = p(N_D(1, t) = 0, \\ &N_D(t, t+1) = 1/N_D(0, 1) = 1,\ t = 2, 3, \ldots \end{aligned}$$

given a first crossing at step 1, a second one at time (t + 1) may occur only if the process assume once in (2, t) a value between u and w. Let j denote the step at which this happens; then

$$\begin{aligned} p(T_1 = t) &= \sum_{j=2}^{t} p(X_j > u \text{ for } i < j,\ w \le X_k \le u \text{ for } j \le k \le t, \\ &\qquad X_{t+1} > u / w \le X_0 \le u,\ X_1 > u) \\ &= \sum_{j=2}^{t} p(J_2 = 1, \ldots, J_{j-1} = 1, J_j \ne 1, J_{j+1} \ne 1, \ldots, J_t \ne 1, J_{t+1} = \\ &\qquad 1/J_0 \ne 1, J_1 = 1) \\ &= \sum_{j=2}^{t} p(J_0 \ne 1, J_1 = 1, \ldots, J_{j-1} = 1, J_j \ne 1, J_{j+1} \ne 1, \ldots, J_t \ne 1, \\ &\qquad J_{t+1} = 1)/p(J_0 \ne 1, J_1 = 1) \\ &= \sum_{j=2}^{t} [p(J_{t+1} = 1/J_t \ne 1)\{p(J_t \ne 1/J_{t-1} \ne 1)\}^{t-j} p(J_j \ne 1/J_{j-1} \\ &\qquad = 1)\{p(J_{j-1} = 1/J_{j-2} = 1)\}^{j-2}] \end{aligned}$$

$$= [(acd)/\{d(a+b) - ac][\{(a+b-ac)/(a+b)\}^{t-1} - (1-d)^{t-1}],$$
$$t = 1, 2, 3, \ldots \qquad (2a)$$

Similarly for downcrossing

$$\begin{aligned} p(T_{-1} = t) &= \sum_{j=2}^{t} p(X_j < w \text{ for } i < j,\ w \le X_k \le u \text{ for } j \le k \le t, \\ &\qquad X_{t+1} < w / w \le X_0 \le u,\ X_1 < w) \end{aligned}$$

$$= [(abd)\{a(d+c) - bd\}][\{(d+c-bd)/(d+c)\}^{t-1} - (1-a)^{t-1}],$$
$$t = 1, 2, 3, \ldots \qquad (2b)$$

Consider now the waiting time for crossings (up or down). Let W_U and W_D denote the waiting times for upcrossing and downcrossing respectively. Their probability distribution may be derived as follows:

$p(W_U=k)=p(J_2 \neq 1, J_3 \neq 1, \ldots, J_k \neq 1, J_{k+1} = 1/J_0 = 1, J_1 \neq 1)$,

and also

$p(W_D=k)=p(J_2 \neq -1, J_3 \neq -1, \ldots, J_k \neq -1, J_{k+1} = 1/J_0 = -1, J_1 \neq -1)$.

Then

$p(W_U = k) = ac/(ab)[\{a(1-c)+b\}/(a+b)]^{k-1}$, $k = 1, 2, \ldots$
$p(W_D=k)=bd/(c+d)[\{d(1-b)+c\}/(c+d)]^{k-1}$, $k = 1, 2, \ldots$ (3).

The following table (1) summarizes the probability mass function (p.m.f.), the mean and the variance of each of the previous variables

Table (1)

Variable	**p.m.f.**	**Mean**	**Variance**
D_{-1}	$a(1-a)^{k-1}$, k=1, 2, ...	$1/a$	$(1-a)/a^2$
D_0	$(b+c)(1-b-c)^{k-1}$, $b+c<1$, k=1, 2, ...	$(b+c)^{-1}$	$(1-b-c)/(b+c)^2$
D_1	$d(1-d)^{k-1}$, k=1, 2, ...	$1/d$	$(1-d)/d^2$
T_1	$[(acd)/\{d(a+b) - ac\}][\{(a+b-ac)^{t-1} - (1-d)^{t-1}(a+b)^{t-1}\}/(a+b)^{t-1}]$, t=1, 2, ...	$[d(a+b) + ac]/acd$	$(a+b)^2[d(a+b) - -ac(d+1)]/a^2c^2 [d(a+b) - ac] + [d(a+b) + ac(d-1)]/ d^2 [d(a+b) - ac]$
T_{-1}	$[(abd)/\{a(d+c)- -bd\}][\{(d+c-bd)^{t-1} -(1-a)^{t-1}(d+c)^{t-1}\}/(d+c)^{t-1}]$, t=1, 2, ...	$[a(d+c)+bd]/abd$	$(d+c)^2[a(d+c) - bd(a+1)]/b^2d^2[a(d+c) -bd]+[a(d+c) + bd(a-1)]/a^2[a(d+c) -bd]$
W_U	$(ac)/(a+b)[a(1-c) a(1-c)+b\}/(a+b)]^{k-1}$, k=1, 2, ...	$(a+b)/ab$	$[(a+b)(a+b-2ac) /a^2c^2]$
W_D	$(bd)/(c+d)[\{ d(1-b)+c\}/(c+d)]^{k-1}$, k=1, 2, ...	$(c+d)/bd$	$[(c+d)(c+d-2bd) /b^2d^2]$

APPLICATION

An application of the present method to Jokulsa river flow data for the period from January 1, 1972 to December 31, 1974 is considered. We choose the critical levels $u = 27$, which is approximately equal to the overall mean flow of cold seasons, and $w = 53$, which is the mean flow of non-cold seasons. Then the estimated transition matrix is

$$\begin{array}{c|c} \text{States} \rightarrow & \begin{array}{ccc} -1 & 0 & 1 \end{array} \\ \downarrow & \\ \begin{array}{c} -1 \\ 0 \\ 1 \end{array} & \begin{bmatrix} 0.982 & 0.018 & 0 \\ 0.019 & 0.949 & 0.032 \\ 0 & 0.032 & 0.978 \end{bmatrix} \end{array}$$

So that the estimated stationary probabilities are $\pi_{-1} = 0.3279$, $\pi_0 = 0.3279$ and $\pi_1 = 0.3442$. The mean number of crossings in this period is $E(N_Y(0, t)) = t\,(0.6721) = 735.9$.

Consider now the duration of an excursion over level $u = 53$ and below the level $w = 27$ given that the initial state is J_0. Note that from equation (1)

$p(D_{-1} = k) = 0.0183\,(0.98172)^{k-1}$, $k = 1, 2, \ldots$

$p(D_0 = k) = 0.0512\,(0.9488)^{k-1}$, $k = 1, 2, 3, \ldots$

$p(D_1 = k) = 0.0320\ (0.9680)^{k-1}$, $k = 1, 2, 3, \ldots$
and the probability mass function of the time between two consecutive crossings (up or down) can be obtained from equation (2) to be

$p(T_1 = k) = 0.0305\ [(0.9844)^{k-1}, -(0.96880)^{k-1}]$, $k = 1, 2, \ldots$

$p(T_{-1} = k) = 0.0202\ [(0.9904)^{k-1}, -(0.9817)^{k-1}, k = 1, 2]$

The following table (2) summarizes the mean and the variance of each of the previous variables:

Table (2)

Variables	Mean	Variance
D_{-1}	54.7045	2938.4
D_0	19.5388	362.2
D_1	31.2598	945.9
T_1	95.3636	2941.27
T_{-1}	195.2232	7580.1619
W_U	64.0841	400
W_D	104.2359	13383

Finally the probability mass functions of the waiting times are provided from equation (3) and are given by

$p(W_U = k) = 0.0156(0.9844)^{k-1}$, $k = 1, 2, \ldots$

$p(W_D = k) = 0.0096(0.9904)^{k-1}$, $k = 1, 2, \ldots$

REFERENCES

[1] **Lawrance, A.J.** and **Kottegoda, N.T., 1977**. Stochastic modeling of riverflow time series. J.R. Stat. Soc., Ser. A, 140: 1-47.

[2] **Battaglia, F., 1981**. Upcrossings of discrete hydrologic processes and Markov chain. J. Hydrol., 53: 1-16.

[3] **Cox, D.R.** and **Miller, H. D., 1965**. The Theory of Stochastic Processes, Methuen, London.