

LETTERS TO THE EDITOR

CONDITIONAL PROBABILITIES IN MORAN'S MODEL

JOHN HAIGH* AND
RICHARD P. LEWIS,* *University of Sussex*

Abstract

A result proved by Saunders et al. (1984) is generalised. This generalisation, and its proof, may help to explain some curiosities noted in the original result.

MARKOV CHAINS; CONDITIONAL PROBABILITIES; GENEALOGY

The purpose of this note is to throw light on a remark made by Saunders et al. (1984) on a result (their Theorem 2) that follows from the proposition below and Equation (3). We hope that the generalisation represented by the proposition, and its proof, may help to explain why the curious results noted by Saunders et al. arise.

Consider the discrete-time finite Markov chain $\{\mathbf{A}(n)\}$, where $\mathbf{A}(n) = (A_1(n), A_2(n))$, with state space $\{(u, v) : 1 \leq v \leq u \leq N\}$ and transition matrix P , where $P((u, v), (u, v)) = 1 - c(u)$, $P((u, v), (u - 1, v)) = c(u) - c(v)$, $P((u, v), (u - 1, v - 1)) = c(v)$, and all other entries $P((u, v), (u', v'))$ are 0. In order that the matrix be stochastic, and that $\mathbf{A}(n)$ converge to $(1, 1)$ whatever its initial value, we need the constraints $0 = c(1) < c(2) \leq c(3) \leq \dots \leq c(N) \leq 1$.

It follows that $\{A_1(n)\}$ is also a Markov chain, with state space $\{1, 2, \dots, N\}$, and transition matrix Q , where $Q(u, u) = 1 - c(u)$ and $Q(u, u - 1) = c(u)$. Define $P(\cdot)$ by

$$(1) \quad P(n, i, j, k, l) \equiv \Pr(A_2(n) = l \mid A_1(n) = k, A_1(0) = i, A_2(0) = j)$$

where $1 \leq l \leq j$, $1 \leq k \leq i$ and $1 \leq l \leq k$ for $P(\cdot)$ to be non-zero.

Proposition. Whatever the values of $\{c(k)\}$, the value of $P(\cdot)$:

- (i) is independent of n ,
- (ii) satisfies $P(n, i, j, k, l) = P(n, i, k, j, l)$.

Note. Theorem 2 of Saunders et al. (1984) corresponds to the case when $c(u) = u(u - 1)/N^2$. Similarly, their Theorem 6 corresponds to $c(u) = u(u + \theta - 1)/N(N + \theta)$, but with a relabelling of the states since, if $\theta > 0$, $\mathbf{A}(n)$ converges to $(0, 0)$; thus $0 = c(0) < c(1) \leq c(2) \dots$. The context in which this problem arises is the consideration of the numbers of ancestors, n generations ago, of the samples labelled $\{1, 2, \dots, i\}$ and $\{1, 2, \dots, j\}$ in the current generation, under Moran's model. The exact expression for $P(\cdot)$ given by Saunders et al. can be obtained from (3) and (5) below: we have not found a 'genetic' interpretation of why the roles of j and k are interchangeable.

Proof of proposition. The definition of $P(\cdot)$ shows that

$$P(n, i, j, k, l) = P^n((i, j), (k, l)) / Q^n(i, k).$$

Received 26 November 1984; revision received 11 February 1985.

* Postal address: Mathematics Division, University of Sussex, Falmer, Brighton BN1 9QH, UK.

Since Q only has non-zero entries on and immediately below its main diagonal, the denominator evaluates as $c(i) c(i-1) \cdots c(k+1) S_n(i, k)$, where $S_n(i, k)$ is the sum of all terms of the form $(1 - c(i))^{\alpha_i} \cdots (1 - c(k))^{\alpha_k}$ ($\alpha_m \geq 0$; $\alpha_i + \cdots + \alpha_k = n - i + k$). Thus S_n represents those probabilities associated with ‘waits’ in the states $\{i, i-1, \dots, k\}$ in the chain $A_1(\cdot)$.

But clearly, if $A_1(m+1) = A_1(m)$, then $A_2(m+1) = A_2(m)$. Any path in the chain $\{A(\cdot)\}$ from (i, j) to (k, l) will contain precisely $j-l$ transitions of the form $(u, v) \rightarrow (u-1, v-1)$, and $r \equiv (i-k) - (j-l)$ transitions of the form $(u, v) \rightarrow (u-1, v)$. (Notice that r is also expressed as $(i-j) - (k-l)$.) Thus $P^n((i, j), (k, l)) = c(j) c(j-1) \cdots c(l+1) S_n(i, k) T(i, j, k, l)$, where $T(i, j, k, l)$ is the sum of all terms of the form

$$(2) \quad (c(u_1) - c(v_1))(c(u_2) - c(v_2)) \cdots (c(u_r) - c(v_r))$$

with $u_m - v_m = i - j + 1 - m$ and $i \geq u_1 > u_2 > \cdots > u_r \geq k + 1$. (An example may help here: suppose $i = 17, j = 11, k = 8, l = 5$, so that $r = 3$. Consider the array

$$\begin{array}{ll} c(17) - c(11) & c(16) - c(10) \cdots c(11) - c(5) \\ c(16) - c(11) & c(15) - c(10) \cdots c(10) - c(5) \\ c(15) - c(11) & c(14) - c(10) \cdots c(9) - c(5). \end{array}$$

Then the terms in the expansion of T are those products of one term from each row, with the constraint that the term from the $(m+1)$ th row cannot come from a column to the left of the term in the m th row.) Thus

$$P(n, i, j, k, l) = \frac{c(j) \cdots c(l+1) S_n(i, k) T(i, j, k, l)}{c(i) \cdots c(k+1) S_n(i, k)}$$

i.e.

$$(3) \quad P(n, i, j, k, l) = \frac{C(j) C(k)}{C(l) C(i)} T(i, j, k, l)$$

where $C(m) \equiv c(m) \cdots c(2)$.

Equation (3) shows that $P(\cdot)$ is independent of n , and its derivation shows why this is so: the ‘waits’ in the chains $\{A(n)\}$ and $\{A_1(n)\}$ are the same. It also shows that proving (ii), i.e. that the roles of j and k are interchangeable, is equivalent to proving

$$(4) \quad T(i, j, k, l) = T(i, k, j, l).$$

Suppose w_1, w_2, \dots are integers, and $\varepsilon_1, \varepsilon_2, \dots$ each take one of the values ± 1 ; write $(\mathbf{w}, \boldsymbol{\varepsilon}) \equiv (w_1, w_2, \dots, w_r, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_r)$, and $t(\mathbf{w}, \boldsymbol{\varepsilon}) \equiv \prod_{m=1}^r \varepsilon_m c(w_m)$. The term $t(\mathbf{w}, \boldsymbol{\varepsilon})$ arises in the formal expansion of $T(i, j, k, l)$ from (2) if, and only if, Conditions (a), (b) and (c) hold:

- (a) when $\varepsilon_m = +1$, then $k+1 \leq w_m \leq i$; write $w_m^* = w_m + m - 1$;
- (b) when $\varepsilon_m = -1$, then $l \leq w_m \leq j$; write $w_m^* = w_m + i - j$;
- (c) given (a) and (b), then $w_1^* \geq w_2^* \geq \dots \geq w_r^*$.

Any $(\mathbf{w}, \boldsymbol{\varepsilon})$ satisfying (a), (b), (c) is said to be (i, j, k, l) -admissible (or just admissible when the context is clear), so that (2) can be rewritten

$$(5) \quad T(i, j, k, l) = \sum t(\mathbf{w}, \boldsymbol{\varepsilon})$$

the sum being over all the admissible vectors $\{(\mathbf{w}, \boldsymbol{\varepsilon})\}$.

We may suppose that $j \geq k+1$ (if $j = k$, then (4) is trivial, and if $j < k$, then interchange the roles of j and k). To each admissible $(\mathbf{w}, \boldsymbol{\varepsilon})$, let $B^+(\mathbf{w}, \boldsymbol{\varepsilon})$ be the list of values w_m corresponding to $\varepsilon_m = +1$, and $B^-(\mathbf{w}, \boldsymbol{\varepsilon})$ be the list of values w_m corresponding to $\varepsilon_m = -1$. Note that $B^-(\mathbf{w}, \boldsymbol{\varepsilon})$ may contain repetitions, but that the members of $B^+(\mathbf{w}, \boldsymbol{\varepsilon})$ are distinct.

Definition. Suppose $0 \leq m \leq r$, that B^+ is a list of m distinct integers in the interval $[k+1, r]$ and B^- is a list of $r-m$ integers in the interval $[l, j]$, possibly with repetitions. We say that (B^+, B^-) belongs to (i, j, k, l) .

We have shown how to construct a pair $B^+(\mathbf{w}, \boldsymbol{\varepsilon}), B^-(\mathbf{w}, \boldsymbol{\varepsilon})$ that belongs to (i, j, k, l) when $(\mathbf{w}, \boldsymbol{\varepsilon})$ is admissible. Conversely, suppose (B^+, B^-) belongs to (i, j, k, l) ; it is not hard to show that there is a unique admissible $(\mathbf{w}, \boldsymbol{\varepsilon})$ with $B^+(\mathbf{w}, \boldsymbol{\varepsilon}) = B^+$ and $B^-(\mathbf{w}, \boldsymbol{\varepsilon}) = B^-$. This sets up a bijection between admissible vectors $(\mathbf{w}, \boldsymbol{\varepsilon})$ and pairs (B^+, B^-) that belong to (i, j, k, l) .

Suppose (B^+, B^-) belongs to (i, j, k, l) , and D is the list of distinct integers that belong to either B^+ or B^- . There are two cases to consider:

(i) No value w_m , with $k+1 \leq w_m \leq j$, falls in D . In this case, (B^+, B^-) also belongs to (i, j, j, l) .

(ii) Some value w_m , with $k+1 \leq w_m \leq j$, falls in D . Let w_0 be the largest such value. Define (C^+, C^-) , that will also belong to (i, j, k, l) , by switching one w_0 between B^+ and B^- in the following way: if $w_0 \in B^+$, move it to B^- , but if $w_0 \notin B^+$, move one w_0 from B^- to B^+ . Because B^+ contains at most one term w_0 , this defines an involution. The sum of the terms $t(\mathbf{w}, \boldsymbol{\varepsilon})$ for the two pairs $(\mathbf{w}, \boldsymbol{\varepsilon})$ that correspond to (B^+, B^-) and (C^+, C^-) is clearly zero, since the two terms differ only in precisely one minus sign.

Now every pair (B^+, B^-) that belongs to (i, k, j, l) arises from (i), and the corresponding values of $t(\mathbf{w}, \boldsymbol{\varepsilon})$ are the same in the expansions of both $T(i, k, j, l)$ and $T(i, j, k, l)$. Since Case (ii) shows how the values of $t(\mathbf{w}, \boldsymbol{\varepsilon})$ for (B^+, B^-) belonging to (i, j, k, l) but not to (i, k, j, l) are paired off to cancel out, Equation (5) demonstrates that (4) is correct.

Reference

SAUNDERS, I. W., TAVARÉ, S. AND WATTERSON, G. A. (1984) On the genealogy of nested subsamples from a haploid population. *Adv. Appl. Prob.* **16**, 471–491.