LOW ORDER MOMENTS OF SOME DISTRIBUTIONS ARISING FROM TWO-STATE MARKOFF CHAINS

By P. V. Krishna Iyer and D. Ray
Defence Science Laboratory, New Delhi

Introduction

A sequence of observations $y_1, y_2, \ldots y_n$ can be considered to be random in the characters $A, B, C, \ldots K$ provided every one of the y's take these characters with probabilities $p_1, p_2 \ldots p_k$ respectively subject to the condition $\Sigma p^r = 1$. Now defining a system of scoring for the differences, P. V. K. Iyer and Singh[6] have discussed a number of distributions and their moments for such a sequence by considering the relationship between consecutive observations y_r and y_{r+1}, alternate observations y_r and y_{r+2} and observations like y_r and y_{r+s}, r and s taking values from 1 to $n-s$. The moments of some of these distributions were obtained by using the result[5] that the r-th factorial moment is $r!$ times the expectation for r of the characters considered in the distribution. The purpose of this paper is to obtain the first and second moments of the distributions of AA, AB and AB and BA joins together (*i.e.*, total number of runs of A's and B's *minus* one) for a two-state (A, B) Markoff chain, by considering consecutive observations. For the sake of convenience the expected number of A's and its variance in a Markoff chain has been given in the second section. It may be mentioned that David,[4] Bateman,[2] Bartlett,[1] Patankar[7] and Whittle[9] have obtained the moments for the distribution of the number of transition numbers for a Markoff chain having the matrix (p_{ij}) for large n. We shall be concerned in this paper at present with the derivation of the *exact* expressions for the first two moments of the three distributions mentioned above for a sequence of n observations. A more detailed investigation on the exact distributions and cumulants for the general case of consecutive and alternate observations is in progress and will be published shortly.

2. Expected Number of A's and Its Variance

This can be obtained by evaluating the probability $Pr(A)$ that the r-th observation is A. Now a two-state Markoff chain is defined by the following quantities:

(i) conditional probability that the r-th observation is A or B given $(r-1)$-th observation, is defined by the Matrix

$$\begin{pmatrix} p_1 & p_2 \\ q_1 & q_2 \end{pmatrix}$$

where

$$P_r(A/r-1, A) = p_1, \qquad Q_r(B/r-1, A) = q_1$$
$$P_r(A/r-1, B) = p_2, \qquad Q_r(B/r-1, B) = q_2$$

(ii) the probability that the initial observation is A or B is P and Q respectively.

Let P_r and Q_r be the probabilities that the r-th observation is A and B respectively. It can easily be seen that

$$\left.\begin{aligned} P_r &= p_1P_{r-1} + p_2Q_{r-1} \\ Q_r &= q_1P_{r-1} + q_2Q_{r-1} \end{aligned}\right\} . \tag{1}$$

The solution of the above difference equations is given by

$$P_r = A_1\lambda_1{}^r + A_2\lambda_2{}^r \tag{2}$$

where λ_1 and λ_2 are the roots of the quadratic

$$\begin{vmatrix} \lambda - p_1 & -p_2 \\ -q_1 & \lambda - q_2 \end{vmatrix} = (\lambda - 1)(\lambda - \delta) = 0$$

where

$$\delta = p_1 - p_2 \qquad (p_1 > p_2).$$

Using the initial conditions $P_1 = P$ and $P_2 = Pp_1 + Qp_2$, we get

$$P_r = \frac{p_2}{1-\delta} + \beta\delta^{r-1}$$

$$Q_r = \frac{q_1}{1-\delta} - \beta\delta^{r-1} \text{ where } \beta = \frac{Pq_1 - Qp_2}{1-\delta}.$$

The expected number of A's in a sequence of n observations is equal to:

$$E(W) = \sum_1^n P_r = \frac{np_2}{1-\delta} + \beta\frac{1-\delta^n}{1-\delta}.$$

$$\sim \frac{np_2}{1-\delta}.$$

The second factorial moment is 2 times the expectation for two A's. Now two A's can be formed from:

(i) two consecutive observations,

(ii) two A's separated by, 1, 2, 3, ... $n-2$ observations.

Probability for consecutive A's is $P_r p_1$. Probability for r-th and $(r+s)$-th observations to be A and A is

$$P_r P_{r+s}(A/rA)$$

where

$$P_{r+s}(A/rA)$$

is the conditional probability of $(r+s)$-th observation to be A, when the r-th observation is also A. It can be easily seen that

$$P_{r+s}(A/rA) = \frac{p_2 + q_1 \delta^{(r+s)-r}}{1-\delta}.$$

Thus the expectation for two A's is

$$\frac{\mu'_{[2]}(W)}{2!}$$
$$= \frac{p_2}{1-\delta}\left[\frac{n(n-1)p_2}{2(1-\delta)} + \beta\left\{\frac{n-1}{1-\delta} - \delta \cdot \frac{1-\delta^{n-1}}{(1-\delta)^2}\right\}\right]$$
$$+ \frac{q_1\delta}{(1-\delta)^2}\left[\frac{p_2}{1-\delta}\left\{(n-1) - \delta \cdot \frac{1-\delta^{n-1}}{1-\delta}\right\}\right.$$
$$\left. + \beta\left\{\frac{1-\delta^{n-1}}{1-\delta} - (n-1)\,\delta^{n-1}\right\}\right].$$

Therefore variance

$$V(W) = 2\mu'_{[2]} - \mu'^{2}_1 + \mu'_1$$
$$= \frac{nq_1p_2(1+\delta)}{(1-\delta)^3} + \frac{2p_2q_1\delta}{(1-\delta)^4}(\delta^n - 1)$$
$$- \frac{2p_2\beta}{(1-\delta)^3}(1-\delta^n) + \beta \cdot \frac{1-\delta^n}{1-\delta} - \beta^2\left(\frac{1-\delta^n}{1-\delta}\right)^2$$
$$+ \frac{2q_1\delta\beta}{(1-\delta)^3}(1-\delta^n) - \frac{2n\delta^n\beta(q_1-p_2)}{(1-\delta)^2}$$
$$\sim \frac{np_2q_1(1+\delta)}{(1-\delta)^3}$$

which agrees with that of Bernstein[3] quoted by Uspensky.[8]

The Asymptotic values for $E(W)$ and $V(W)$ have also been given in a different form by Patankar.[7]

3. Expectation and Variance for Two Consecutive A's

Using the results of Section 2, we note that the probability for the r-th and $(r+1)$-th observation to be A is $P_r p_1$ and therefore the expectation of the number of AA's is

$$E(X) = p_1 \sum_{1}^{n-1} P_r$$

$$= p_1 \left\{ \frac{(n-1)\,p_2}{1-\delta} + \beta \cdot \frac{1-\delta^{n-1}}{1-\delta} \right\}$$

$$\sim \frac{(n-1)\,p_1 p_2}{1-\delta}.$$

Two AA's can be obtained in the following ways:

(1) from three consecutive observations like r, $r+1$ and $r+2$, the probability for which is $P_r p_1^2$,

(2) from four consecutive observations like r, $r+1$, $r+2$, $r+3$, the probability for which is $P_r p_1^3$,

(3) from two sets of two consecutive observation and $(r, r+1)$ and $(r+s, r+s+1)$, $s > 2$. The probability for this is $P_r p_1 P_{r+s}$ $(A/r+1, A)\, p_1$. Adding the above probabilities, we get

$$\frac{\mu'_{[2]}(X)}{2!} = p_1^2 \left\{ \frac{(n-2)\,p_2}{1-\delta} + \beta \cdot \frac{1-\delta^{n-2}}{1-\delta} \right\}$$

$$+ \frac{p_2 p_1^2}{1-\delta} \left[\frac{(n-3)(n-2)\,p_2}{2(1-\delta)} \right.$$

$$\left. + \beta \left\{ \frac{n-3}{1-\delta} - \delta \cdot \frac{1-\delta^{n-3}}{(1-\delta)^2} \right\} \right]$$

$$+ \frac{p_1^2 q_1 \delta}{(1-\delta)^2} \left[\frac{p_2}{1-\delta} \left\{ (n-3) - \delta \cdot \frac{1-\delta^{n-3}}{1-\delta} \right\} \right.$$

$$\left. + \beta \left\{ \frac{1-\delta^{n-3}}{1-\delta} - (n-3)\,\delta^{n-3} \right\} \right],$$

Therefore

$$V(X) = \frac{p_1 p_2 (n-1)}{1-\delta} + \frac{2 p_1^2 p_2 (n-2)}{1-\delta}$$

$$\times \beta p_1 \left\{\frac{1-\delta^{n-1}}{1-\delta} + 2p_1 \frac{1-\delta^{n-2}}{1-\delta}\right\}$$

$$\times \frac{p_2{}^2 p_1{}^2}{(1-\delta)^2}(-3n+5) + \frac{2p_2 p_1{}^2 \beta}{(1-\delta)^2}$$

$$\times \left\{(n-3) - \delta \cdot \frac{1-\delta^{n-3}}{1-\delta} - (n-1)(1-\delta^{n-1})\right\}$$

$$- \beta^2 p_1{}^2 \left(\frac{1-\delta^{n-1}}{1-\delta}\right)^2 + \frac{2p_1{}^2 q_1 \delta}{(1-\delta)^2}$$

$$\times \left[\frac{p_2}{1-\delta}\left\{(n-3) - \delta \cdot \frac{1-\delta^{n-3}}{1-\delta}\right\}\right.$$

$$\left.\times \beta \left\{\frac{1-\delta^{n-3}}{1-\delta} - (n-3)\,\delta^{n-3}\right\}\right]$$

$$\sim \frac{np_1p_2}{1-\delta}\left\{1 + 2p_1 - \frac{3p_1p_2}{1-\delta} + \frac{2p_1q_1\delta}{(1-\delta)^2}\right\}.$$

The asymptotic values of $E(X)$ and $V(X)$ have been given by Bartlett[1] and are the same as given above.

4. Expectation and Variance of the Number of AB's for Two Consecutive Observations

The probability that two consecutive observations r and $r+1$ will be A and B is $P_r q_1$. Hence the expectation of the number of AB's is

$$E(Y) = q_1 \sum_{r=1}^{n-1} P_r = q_1 \left\{\frac{(n-1)\,p_2}{1-\delta} + \beta\,\frac{1-\delta^{n-1}}{1-\delta}\right\}$$

$$\sim \frac{(n-1)\,p_2 q_1}{1-\delta}.$$

Two AB joins can be obtained:

(i) from four consecutive observations like r, $r+1$, $r+2$ and $r+3$, the probability for which is $P_r q_1 p_2 q_1$, *i.e.*, $P_r q_1{}^2 p_2$,

(ii) from two pairs of consecutive observations like $(r, r+1)$ and $(r+s, r+s+1)$, $s > 2$.

Probability for the above is

$$P_r q_1 P_{r+s}(A/r+1, B)\, q_1,$$

i.e.,

$$P_r P_{r+s}(A/r+1, B)\, q_1^2$$

where

$$P_{r+s}(A/r+1, B) = \frac{p_2}{1-\delta}\{1 - \delta^{(r+s)-(r+1)}\}$$

using these results it can be shown that

$$\frac{\mu'_{[2]}(Y)}{2!} = \frac{p_2 q_1^2}{1-\delta}\left[\frac{(n-2)(n-3)p_2}{(1-\delta)} - \frac{p_2\delta}{(1-\delta)^2}\left\{(n-3) - \delta\cdot\frac{1-\delta^{n-3}}{1-\delta}\right\} - \frac{\beta\delta}{1-\delta}\left\{\frac{1-\delta^{n-3}}{1-\delta} - (n-3)\,\delta^{n-3}\right\} \times \beta\left\{\frac{n-3}{1-\delta} - \frac{\delta(1-\delta^{n-3})}{(1-\delta)^2}\right\}\right].$$

Therefore

$$V(Y) = \frac{p_2^2 q_1^2}{(1-\delta)^2}\left[(-3n+5) - \frac{2\delta}{(1-\delta)^2} \times \{(n-3) - (n-2)\,\delta + \delta^{n-2}\}\right] - \frac{2p_2 q_1^2\beta}{(1-\delta)^3}[2(1-\delta^{n-1}) + (n-1)(\delta^n - \delta^{n-2})] - \beta^2 q_1^2\left(\frac{1-\delta^{n-1}}{1-\delta}\right)^2 + q_1\left\{\frac{(n-1)p_2}{1-\delta} + \beta\cdot\frac{1-\delta^{n-1}}{1-\delta}\right\}$$

$$\sim \frac{np_2q_1}{1-\delta}\left\{1 - \frac{3q_1p_2}{1-\delta} - \frac{2\delta q_1 p_2}{(1-\delta)^2}\right\}.$$

The asymptotic values for this distribution also have been given by Bartlett[1] and reduce to the expression given above.

5. Expectation and Variance for the Number of *AB* and *BA* Joins from Consecutive Observation

The expectation is equal to the sum of expectations for *AB* and *BA* joins. This can readily seen to be:

$$E(Z) = \frac{2p_2q_1(n-1)}{1-\delta} + \beta(q_1 - p_2)\frac{1-\delta^{n-1}}{1-\delta}$$

$$\sim \frac{2(n-1)p_2q_1}{1-\delta}.$$

Simultaneous occurrence of *AB* and *BA* joins is possible:

(i) from three consecutive observations r, $r+1$, $r+2$, like *ABA* and *BAB*, the probabilities for which are $P_r q_1 p_2$ and $Q_r p_2 q_1$;

(ii) from four consecutive observations r, $r+1$, $r+2$, $r+3$ like *ABAB*, *ABBA*, *BAAB* and *BABA*, the probabilities for whcih are,

$$P_r q_1 p_2 q_1,\ P_r q_1 q_2 p_2,\ Q_r p_2 p_1 q_1 \text{ and } Q_r p_2 q_1 p_2 \text{ respectively};$$

(iii) from two pairs of consecutive observations like $(r, r+1)$ and $(r+s, r+s+1)$, $s > 2$ for all types of joins like *ABAB*, *ABBA*, *BAAB* and *BABA*.

The probabilities for all these cases are

$$P_r q_1 P_{r+s}(A/r+1, B)\, q_1,$$

$$P_r q_1 Q_{r+s}(B/r+1, B)\, p_2,$$

$$Q_r p_2 P_{r+s}(A/r+1, A)\, q_1$$

and

$$Q_r p_2 Q_{r+s}(B/r+1, \text{A})\, p_2$$

respectively where

$$Q_{r+s}(B/r+1, B) = \frac{\{q_1 + p_2{}^{(r+s)-(r+1)}\}}{(1-\delta)}$$

$$Q_{r+s}(B/r+1, A) = \frac{q_1\{1-\delta^{(r+s)-(r+1)}\}}{(1-\delta)}.$$

Adding theses we get

$$\frac{\mu'_{[2]}(Z)}{2!} = (n-2)\, p_2 q_1 + \frac{2p_2{}^2 q_1{}^2}{(1-\delta)^2}\Big[(n-3)(n-2) - \frac{\delta}{1-\delta}$$

$$\times \Big\{(n-3) - \delta \cdot \frac{1-\delta^{n-3}}{1-\delta}\Big\}\Big] - \frac{2p_2 q_1 (q_1 - p_2)\,\beta}{(1-\delta)}$$

$$\times \Big[\frac{\delta}{1-\delta}\Big\{\frac{1-\delta^{n-3}}{1-\delta} - (n-3)\,\delta^{n-3}\Big\}$$

$$-\Big\{\frac{n-3}{1-\delta} - \frac{\delta\,(1-\delta^{n-3})}{(1-\delta)^2}\Big\}\Big] + \frac{q_1 p_2\,(p_2{}^2 + q_1{}^2)\,\delta}{(1-\delta)^2}$$

$$\times \Big[\frac{n-3}{1-\delta} - \frac{\delta}{(1-\delta)^2}(1-\delta^{n-3})\Big]$$

Therefore

$$V(Z) = 2(n-2) q_1 p_2 + \frac{4p_2^2 q_1^2}{(1-\delta)^2} \Big[(-3n+5) - \frac{\delta}{(1-\delta)^2}$$

$$+ \{(n-3) - (n-2)\delta + \delta^{n-2}\} \Big]$$

$$- \frac{4p_2 q_1 (q_1 - p_2) \beta}{(1-\delta)^3} [(n-1)(\delta^n - \delta^{n-2}) + 2(1-\delta^{n-1})]$$

$$+ \frac{2(p_2^2 + q_1^2) p_2 q_1 \delta}{(1-\delta)^4} [(n-3) - (n-2)\delta + \delta^{n-2}]$$

$$- \beta^2 (q_1 - p_2)^2 \cdot \left(\frac{1-\delta^{n-1}}{1-\delta}\right)^2 + 2p_2 q_1 \cdot \frac{n-1}{1-\delta}$$

$$+ \frac{\beta (q_1 - p_2)(1-\delta^{n-1})}{(1-\delta)}$$

$$\sim \frac{4np_2 q_1}{1-\delta} \left\{1 - \frac{3q_1 p_2}{1-\delta} - \frac{2\delta q_1 p_2}{(1-\delta)^2}\right\}.$$

This is equal to four times the variance of AB.

6. Covariance between $X(AA)$ and $Y(AB)$

The simultaneous occurrence of AA and AB joins is possible:

(i) from three consecutive observations r, $r+1$, $r+2$ like AAB, the probability for which is $P_r p_1 q_1$,

(ii) from four consecutive observations r, $r+1$, $r+2$, $r+3$ like $AAAB$, the probability being $P_r p_1^2 q_1$,

(iii) from two pairs of consecutive observations and $(r, r+1)$ and $(r+s, r+s+1)$, $s > 2$.

The probability of this is

$$P_r p_1 P_{r+s} (A/r+1, A) q_1.$$

Hence from above

$$\mu'_{XY} = p_1 q_1 \left\{\frac{p_2(n-2)}{1-\delta} + B\frac{1-\delta^{n-2}}{1-\delta}\right\}$$

$$+ \frac{2p_1 p_2 q_1}{1-\delta} \Big[\frac{(n-3)(n-2) p_2}{2(1-\delta)}$$

$$+ \beta \left\{\frac{n-3}{1-\delta} - \delta \frac{1-\delta^{n-3}}{(1-\delta)^2}\right\}\Bigg]$$

$$+ \frac{p_1 q_1 \delta}{(1-\delta)^2}\left[\frac{p_2}{1-\delta}\left\{(n-3) - \delta \frac{1-\delta^{n-3}}{1-\delta}\right\}\right.$$

$$\left. + \beta \left\{\frac{1-\delta^{n-3}}{1-\delta} - (n-3)\,\delta^{n-3}\right\}\right](q_1 - p_2).$$

Therefore covariance

$$\text{Cov.}\ (X, Y)$$

$$= \mu'_{XY} - \mu'_X \mu'_Y$$

$$= \frac{p_1 q_1 p_2^2}{(1-\delta)^2}(-3n+5) - \beta^2 p_1 q_1 \left(\frac{1-\delta^{n-1}}{1-\delta}\right)^2$$

$$+ p_1 q_1 \left\{\frac{(n-2)\,p_2}{1-\delta} + \beta \cdot \frac{1-\delta^{n-2}}{1-\delta}\right\}$$

$$+ \frac{2 p_2 p_1 q_1 \beta}{(1-\delta)^2}\left\{(n-3) - \frac{\delta(1-\delta^{n-3})}{1-\delta} - (n-1)\,(1-\delta^{n-1})\right\}$$

$$+ \frac{p_2 p_1 q_1 (q_1 - p_2)\,\delta}{(1-\delta)^3}\left\{(n-3) - \frac{(1-\delta^{n-3})\delta}{1-\delta}\right\}$$

$$+ \frac{p_1 q_1 (q_1 - p_2)\,\delta}{(1-\delta)^2}\,\beta \left\{\frac{1-\delta^{n-3}}{1-\delta} - (n-3)\,\delta^{n-3}\right\}$$

$$\sim \frac{n p_1 q_1 p_2}{1-\delta}\left\{1 + \frac{(q_1 - p_2)\,\delta}{(1-\delta)^2} - \frac{3 p_2}{1-\delta}\right\}.$$

This agrees with that of Bartlett.[1]

SUMMARY

This paper gives the first two *exact* moments for the distribution of the number of *AA*, *AB* and *AB* and *BA* joins for a two-state Markoff chain. The covariance for the number of *AA* and *AB* joins has also been given. The asymptotic values given by Bartlett and others for the transition numbers which correspond to *AA* and *AB* joins are the same as given in this paper.

References

1. Bartlett, M. S. .. "The frequency goodness of fit test for probability chain," *Proc. Camb. Phil. Soc.*, 1951, **47**, 86.
2. Bateman, G. .. "On the power function of the longest run as a test for randomness in a sequence of alternatives," *Biometrika*, 1948, **35**, 97.
3. Bernstein, S. .. "Sur l'ex tension du theoreme limite du calcul des probabilities aux sommes de quantites dependantes," *Math. Ann.*, 1927, **97**.
4. David, F. N. .. "A power function for tests of randomness in a sequence of alternatives," *Biometrika*, 1947, **34**, 335.
5. Krishna Iyer, P. V. .. "A theorem on factorial moments and its applications," *Ann. Math. Stat.*, 1958, **29**, 254.
6. ——— and Singh, B. N. .. "On certain probability distributions arising from a sequence of observations and their applications," *J. Ind. Soc. Agri. Stat.*, 1955, **7**, 127.
7. Patankar, V. N. .. "A note on recurrent events," *Proc. Camb. Phil. Soc.*, 1955, **51**, 97.
8. Uspensky, J. V. .. *Introduction to Mathematical Probability*, McGraw-Hill Book Company.
9. Whittle, P. .. "Some distribution and moment formulæ for the Markov chain," *J.R.S.S.*, 1955, **17 B**, 235.