

STOCHASTIC MODELS FOR STRUCTURE OF DAIRY FEMALE POPULATION

BY

J.P. JAIN

I.A.S.R.I. New Delhi-12

(Received : September, 1976)

1. INTRODUCTION

Jain and Narain (1974) adopted Lefkovitch's (1965) extension of Leslie's (1945) deterministic model for studying the pattern of growth of dairy female population grouped in unequal stage-groups. The model assumes both time and age-scale variables as discrete and the number of individuals in various stage-groups as fixed. However, because of the finiteness of the population, the numbers in various stage-groups are always subject to some random fluctuations. Consequently, it would be more appropriate to consider such numbers as stochastic variables following a certain probability distribution. Pollard (1966) formulated the stochastic analogue of Leslie's (1945) equal age-step discrete model wherein the number of births and the number of transfers occurring during any arbitrary interval of time were both assumed to follow binomial law. Since equal age-step grouping may not be relevant in dairy populations where the estimates of rates of mortality, fertility and culling are usually obtained for different stage-groups rather than age-groups of equal duration, it would be desirable to attempt a stochastic analogue of Lefkovitch's (1965) model in the study of such populations. In the present paper two stochastic versions of the same are considered. The number of individuals born during any arbitrary interval of time is assumed to follow a Poisson distribution or a binomial distribution and the number of transfers occurring during the same interval a binomial distribution. Linear matrix recurrence relations are derived for the means, variance and covariance of the random numbers in the various stage-groups. Results about the asymptotic behaviour of the population are also mentioned. Further, the relative suitability of the two stochastic models in the study of growth in dairy populations is compared on the basis of the results of Monte Carlo experiments.

2. THE STOCHASTIC MODELS

2.1. The deterministic model and some basic results

Let us consider the female population at discrete intervals of time $t=0, 1, 2, \dots$ grouped in distinct stage-groups as defined in Jain and Narain (1974). Further, let n_{xt} be the number of females in stage-group x at time t . The numbers in different stage groups at time $t+1$ can be obtained deterministically by taking a linear function of those alive in different stages at time t , i.e.

$$\left. \begin{aligned}
 n_{1, t+1} &= \sum_{x=4}^s n_{xt} a_{1x} \\
 n_{2, t+1} &= n_{1t} a_{21} + \sum_{x=4}^s n_{xt} a_{2x} \\
 n_{3, t+1} &= \sum_{x=1}^2 a_{3x} n_{xt} \\
 n_{4, t+1} &= \sum_{x=1}^4 a_{4x} n_{xt} \\
 n_{p, t+1} &= a_{p, p-1} n_{p-1, t} \quad (p \geq 5)
 \end{aligned} \right\} \dots(1)$$

where the expressions for a_{ix} 's ($i, x=1, 2, \dots, s$) are as obtained by Jain and Narain (1974).

These equations can be written as

$$n_{t+1} = F n_t \quad \dots(2)$$

Hence,
$$n_t = F^t n_0 \quad \dots(3)$$

where n_t and n_{t+1} are column vectors and F is the matrix of a_{ix} 's.

Some of the results which will be used in the sequel are

$$E(M_1) = k_1 \gamma_1 \quad \dots(4)$$

$$E(M_3) = k_3 \gamma_1 \quad \dots(5)$$

$$\text{var}(M_1) = \gamma_1 k_1 + k_1^2 \sigma^2 \quad \dots(6)$$

$$\text{cov}(M_1, M_2) = k_1 k_2 \sigma^2 \quad \dots(7)$$

$$\text{cov}(M_1, N_2) = k_1 k_2 \text{cov}(M, N) \quad \dots(8)$$

$$\text{var}(M_3) = \gamma_1 k_3 (1 - 11 - k_3) + k_3^2 \sigma^2 \quad \dots(9)$$

$$\text{cov}(M_3, N_2) = k_2 k_3 \text{cov}(M, N) \quad \dots(10)$$

$$\text{cov}(M_2, N_1) = k_1 k_2 \text{cov}(M, N) \quad \dots(11)$$

where M and N are discrete random variables taking non-negative integral values with means γ_1 and γ_2 and variances σ^2 and τ^2 respectively. M_1 , M_2 and N_1 are random variables having the conditional Poisson distributions $P(Mk_1)$, $P(Mk_2)$ and $P(Nk_1)$ respectively and M_3 and N_2 having the conditional binomial distributions $B(M, k_3)$ and $B(N, k_2)$ respectively.

2.2. Concepts and notations

For all $i < x$, a_{ix} in matrix F can be viewed as the probability that a female in stage-group x at time t will give birth during the unit interval to a daughter, and that this daughter will be in the herd in stage-group i at time $t+1$ after allowing for mortality, involuntary culling etc. Further, for all $i \geq x$, a_{ix} can be reckoned as the probability that a female from stage-group x will survive to be in stage-group i after a unit interval of time.

Let the number of females in the stage-group x at time t be a random variable n_{xt} with expected value e_{xt} and variance $C^{(t)}x, x$. Further, let the covariance, $\text{Cov}(n_{xt}, n_{x't})$ be denoted by $C^{(t)}x, x'$.

Let $n_{i,t+1}^{(x)}$ for i strictly less than x denote the number of daughters contributed and remaining at time $t+1$ in stage group i by the mothers who were among the females in stage-group x at time t (i.e. by n_{xt} mothers). Further, let $n_{i,t+1}^{(x)}$ for $i \geq x$ denote the number of females in stage-group i at time $t+1$ transferred from among those who were in stage-group x at time t . Obviously for all i 's ($i=1, 2, \dots, s$)

$$n_{i,t+1} = \sum_x n_{i,t+1}^{(x)} \quad \dots(12)$$

2.3. Poisson-binomial model

Consider the n_{xt} females in stage-group x at time t . Each of them has fixed probabilities a_{ix} of contributing or entering stage-groups of $i < x$ and $i \geq x$ as the case may be at time $t+1$. Further, n_{xt} mothers in stage-group x at time t contribute daughters to stage-group i at the mean rate of $n_{xt} a_{ix}$ per unit time. The probability

that any single mother out of n_{xt} contributes a daughter to stage-group i in any infinitesimal interval of time is very small. However, n_{xt} being adequately large, it can be assumed that the number of daughters contributed to stage-group i by n_{xt} follows Poisson law. Hence, for $i < x$, $n^{(x)}_{i,t+1}$ is a Poisson variable $P(n_{xt} a_{ix})$ conditional on n_{xt} . Further, each of the n_{xt} females in stage-group x have equal probability of surviving a unit time interval since they all possess the same breed characteristics and also the survival of any female is independent of the others. Hence, for $i \geq x$, $n^{(x)}_{i,t+1}$ which is the number in stage-group i at time $t+1$ from among those who were in stage-group x at time t , is a binomial variate $B(n_{xt}, a_{ix})$ conditional on n_{xt} . The birth, death and culling processes are assumed independent. In the sequel, $1 - a_{ix}$ is denoted by d_{ix} .

In the light of the foregoing the numbers in different stage-groups at time $t+1$ as given by the equations (1) can be expressed as :

$$\begin{aligned}
 n_{1,t+1} &= \sum_{x=4}^s n^{(x)}_{1,t+1} \\
 &= \sum_{x=4}^s P [a_{1x} n_{xt} / n_{xt}] \\
 n_{2,t+1} &= n^{(1)}_{2,t+1} + n^{(x)}_{2,t+1} \\
 &= B[(n_{1t}, a_{21}) / n_{1t}] + \sum_{x=4}^s P [a_{2x} n_{xt} / n_{xt}] \\
 n_{3,t+1} &= \sum_{x=1}^2 B [(n_{xt}, a_{3x}) / n_{xt}] \\
 \text{Similarly,} \\
 n_{4,t+1} &= \sum_{x=1}^4 B [n_{xt}, a_{4x} / n_{xt}] \\
 n_{p,t+1} &= B [(n_{p-1,t}, a_{p,p-1}) / n_{p-1,t}] \quad (p \geq 5)
 \end{aligned}
 \tag{13}$$

Using (4) and (5), the expectations of the random variables $n_{x, t+1}$ will then be given by

$$\begin{aligned}
 e_{1, t+1} &= \sum_{x=4}^s a_{1x} e_{xt} \\
 e_{2, t+1} &= a_{21} e_{1t} + \sum_{x=4}^s a_{2x} e_{xt} \\
 e_{3, t+1} &= \sum_{x=1}^2 a_{3x} e_{xt} \\
 e_{4, t+1} &= \sum_{x=1}^4 a_{4x} e_{xt} \\
 e_{p, t+1} &= a_{p, p-1} e_{p-1, t} \quad (p \geq 5)
 \end{aligned}
 \tag{14}$$

Using (6) to (11), the variances and covariances become

$$\begin{aligned}
 C_{1,1}^{(t+1)} &= \sum_{x=4}^s (a_{1x}^2 C_{x,x}^{(t)} + a_{1x} e_{xt}) + \sum_{\substack{x \neq x' \\ 4}}^s a_{1x} a_{1x'} C_{x,x'}^{(t)} \\
 C_{2,2}^{(t+1)} &= (a_{21}^2 C_{1,1}^{(t)} + a_{21} d_{21} e_{1t}) + \sum_{x=4}^s (a_{2x}^2 C_{x,x}^{(t)} + a_{2x} e_{xt}) \\
 &\quad + 2a_{21} \sum_{x=4}^s a_{2x} C_{1,x}^{(t)} + \sum_{\substack{x \neq x' \\ 4}}^s a_{2x} a_{2x'} C_{x,x'}^{(t)} \\
 C_{3,3}^{(t+1)} &= \sum_{x=1}^2 (a_{3x} C_{x,x'}^{(t)} + a_{3x} d_{3x} e_{xt}) + 2a_{31} a_{32} C_{1,2}^{(t)} \\
 C_{4,4}^{(t+1)} &= \sum_{x=1}^4 (a_{4x} C_{x,x}^{(t)} + a_{4x} d_{4x} e_{xt}) + \sum_{\substack{x \neq x' \\ 1}}^4 a_{4x} a_{4x'} C_{x,x'}^{(t)} \\
 C_{p,p}^{(t+1)} &= a_{p,p-1}^2 C_{p-1,p-1}^{(t)} + a_{p,p-1} d_{p,p-1} e_{p-1, t} \quad (p \geq 5)
 \end{aligned}
 \tag{15}$$

$$\begin{aligned}
 C_{1,2}^{(t+1)} &= a_{21} \sum_{x=4}^s a_{1x} C_{1,x}^{(t)} + \sum_x \sum_{x'=4}^s a_{1x} a_{2x'} C_{x,x'}^{(t)} \\
 C_{1,3}^{(t+1)} &= \sum_{x'=1}^2 a_{3x'} \sum_{x=4}^s a_{1x} C_{x',x}^{(t)} \\
 C_{1,4}^{(t+1)} &= \sum_{x'=1}^4 a_{4x'} \sum_{x=4}^s a_{1x} C_{x',x}^{(t)} \\
 C_{1,p}^{(t+1)} &= a_{p,p-1} \sum_{x=4}^s a_{1x} C_{p-1,x}^{(t)} \quad (p \geq 5) \\
 C_{2,3}^{(t+1)} &= a_{21} \sum_{x=1}^2 a_{3x} C_{1,x}^{(t)} + \sum_{x'=1}^s a_{3x'} \sum_{x=4}^2 a_{2x} C_{x',x}^{(t)} \\
 C_{2,4}^{(t+1)} &= a_{21} \sum_{x=1}^4 a_{4x} C_{1,x}^{(t)} + \sum_{x'=1}^4 a_{4x'} \sum_{x=4}^s a_{2x} C_{x',x}^{(t)} \\
 C_{2,p}^{(t+1)} &= a_{21} a_{p,p-1} C_{1,p-1}^{(t)} + a_{p,p-1} \sum_{x=4}^s a_{2x} C_{p-1,x}^{(t)} \quad (p \geq 5) \\
 C_{3,4}^{(t+1)} &= \sum_{x'=1}^2 a_{3x'} \sum_{x=1}^4 a_{4x} C_{x',x}^{(t)} \\
 C_{3,p}^{(t+1)} &= a_{p,p-1} \sum_{x=1}^2 a_{3x} C_{x,p-1}^{(t)} \quad (p \geq 5) \\
 C_{4,p}^{(t+1)} &= a_{p,p-1} \sum_{x=1}^4 a_{4x} C_{p-1,x}^{(t)} \quad (P \geq 5) \\
 C_{p,p'}^{(t+1)} &= a_{p,p-1} a_{p',p'-1} C_{p-1,p'-1}^{(t)} \quad (p \neq p' \geq 5)
 \end{aligned}
 \tag{16}$$

Equations (14) to (16) completely define the recurrence relations for the first two moments and can be written in the form

$$\begin{bmatrix} \underline{e}_{t+1} \\ \underline{C}^{(t+1)} \end{bmatrix} = \begin{bmatrix} F & O \\ E & F \times F \end{bmatrix} \begin{bmatrix} \underline{e}_t \\ \underline{C}^{(t)} \end{bmatrix}, \quad \dots \tag{17}$$

where F is the $s \times s$ matrix defined earlier, $F \times F$ is its direct product of dimension $s^2 \times s^2$ and E is $s^2 \times s$ matrix which can be decomposed as follows :

$$E' = [E_1 \ E_2 \ \dots \ E_i \ \dots \ E_s] \quad \dots(18)$$

where each E_i is a $s \times s$ matrix with only i th row comprising not all zero elements, all other row vectors being zero, x th element of i th row being

$$a_{ix} d_{ix} \text{ for } i \geq x$$

and

$$a_{ix} \text{ for } i < x$$

The vector $\underline{C}^{(t)}$ has as its elements the variances and covariances $C_{ij}^{(t)}$ arranged in ascending order of subscripts i and j .

Recurring relation (7) t times, we get

$$\begin{aligned} \begin{bmatrix} \underline{e}_t \\ \underline{C}^{(t)} \end{bmatrix} &= \begin{bmatrix} F & O \\ E & F \times E \end{bmatrix}^t \begin{bmatrix} \underline{e}_0 \\ \underline{C}^{(0)} \end{bmatrix} \\ &= \begin{bmatrix} \underline{e}_t \\ (F \times F)^t \underline{C}^{(0)} + \sum_{j=1}^t (F \times F)^{t-j} E \underline{e}_{j-1} \end{bmatrix} \quad \dots(19) \end{aligned}$$

These relations are precisely of the same form as obtained by Pollard (1966).

2.4 Binomial-binomial model

Following Pollard (1966) in addition to viewing the number of transfers $n_{i,t+1}^{(x)}$ from stage-group x to i ($i \geq x$) during the time interval $(t, t+1)$ as a binomial variable, the number of births taking place during the same interval i.e. $n_{i,t+1}^{(x)}$ ($i < x$) can also be considered as a binomial variate $B(n_{xt}, a_{ix})$ conditional on n_{xt} . With this formulation, the means, variances are still given by the relations (17 and 19) excepting that the x th element of the i th row of matrix E_i in relation (18) takes the value $a_{ix} d_{ix}$ for all x .

3. THE ASYMPTOTIC BEHAVIOUR OF THE POPULATION

To study the asymptotic behaviour of the population we need the dominant latent root of the matrix in equation (17), the characte-

ristic equation of which is

$$\begin{bmatrix} F - \omega I & O \\ E & F \times F - \omega I \end{bmatrix} = 0$$

or

$$|F - \omega I| |F \times F - \omega I| = 0$$

Thus, the matrix has as its latent roots all the roots of F together with all of $F \times F$. If ω_1 is the dominant latent root of F then ω_1^2 is the dominant latent root of $F \times F$. There are thus three cases to be considered *viz.*, (i) $\omega_1 > 1$; (ii) $\omega_1 = 1$; and (iii) $\omega_1 < 1$. Since the matrix in relation (17) is exactly analogous to the one derived by Pollard (1966), the results concerning the asymptotic behaviour of the population in the above three situations are as follows.

Case (i) In this situation ω_1^2 is the dominant latent root of the matrix in equation (17) and the random variable n_t/ω_1^{2t} converges with probability one to a random scalar multiple of a fixed stable stage distribution vector.

Case (ii) In this case there is a pair of dominant latent roots and each is equal to unity. The population will thus attain both stable stage and size distribution.

Case (iii) In this situation F contributes the dominant latent root ω_1 and the population will ultimately become extinct.

4. MONTE CARLO EXPERIMENTS USING THE TWO MODELS

In order to see which of the two models formulated here gives a better fit to the population under consideration, the Monte Carlo experiments were performed on the IBM 1620 computer.

The numerical values of the elements of generation matrix F were computed from the expressions given by Jain and Narain (1974) using the estimates of vital characteristics and other parameters as reported by Amble and Jain (1967) for the cross-bred herd of dairy cattle of Indian military farms. The population was assumed to be divided into seven different stage-groups with the initial population vector as $n'_0 = [50 \ 50 \ 15 \ 65 \ 50 \ 30 \ 20]$. The unit of time taken was one calving interval which for this herd equalled 14.4 months.

The pseudo-Poisson deviates were generated by the subroutine outlined by Knuth (1969). Since no direct method for generating random binomial variable is readily available, a normal approximation to this distribution was used and pseudo-random normal deviates were generated by the subroutine programmed by Pillai (1972). The Monte Carlo experiments and the theoretical mean/

TABLE I
Monte Carlo Experimental Results Using the two Models

Stage-group	Poisson-binomial model				Binomial-binomial model			
	Actual	Expected	Variance	Norm. devn.*	Actual	Expected	Variance	Norm. devn.*
(a) After 20 units of time								
1	178	156	971	0.706	189	156	862	1.124
2	172	142	806	1.057	176	142	729	1.259
3	48	35	74	1.512	35	35	70	0.000
4	248	194	1430	1.428	256	194	1309	1.714
5	122	124	620	-0.081	160	124	572	1.505
6	110	108	476	0.092	118	108	440	0.477
7	98	94	367	0.209	118	94	340	1.304
(b) After 21 units of time								
1	180	165	1088	0.455	206	165	966	1.319
2	171	150	902	0.699	194	150	817	1.539
3	40	37	81	0.333	54	37	76	1.950
4	256	205	1605	1.273	257	205	1470	1.356
5	169	131	694	1.063	174	131	640	1.700
6	113	114	533	-0.043	147	114	492	1.487
7	99	99	410	0.000	107	99	379	0.411

* Standardised deviation of the simulated values from the theoretical mean

variance/covariance calculations were carried out independently after starting with the same initial vector.

Table 1 gives the expected numbers and variances in different stage-group as also the simulated numbers for both Poisson-binomial and binomial-binomial models at two consecutive points of time viz. at 20 and 21 units of time. The expected numbers and their variances are theoretical values for the means and variances obtained using the iterative equation (17) with appropriate values of the elements of matrix *E* under the respective two models. The results of the two experiments are as follows :

TABLE 2
Ratios of Means and Variances Under the two Models

Stage-group	Poisson-binomial model		Binomial-binomial model	
	Means 21/20	Variances 21/20	Means 21/20	Variances 21/20
1	1.0556	1.1198	1.0556	1.1206
2	1.0555	1.1198	1.0555	1.1207
3	1.0555	1.0972	1.0555	1.0857
4	1.0556	1.1229	1.0556	1.1230
5	1.0558	1.1191	1.0558	1.1189
6	1.0559	1.1185	1.0559	1.1182
7	1.0554	1.1172	1.0554	1.1147

(a) The variances of the numbers in different stage-groups under the Poisson-binomial model are more than their counterparts in the binomial-binomial model as expected. On the other hand, the standard normal deviation of the actual numbers from the theoretical means at two consecutive points of time ranged from -0.04 to 1.51 in the case of Poisson-binomial model as against 0 to 2 for the other model which implies that the former model gives a closer fit to the population.

(b) We know that when $\omega_1 > 1$ (case (i) of Section 3)

$$\lim_{t \rightarrow \infty} e_{j,t+1}/e_{j,t} = \omega_1 ;$$

$$\lim_{t \rightarrow \infty} C_{ij}^{(t+1)} / C_{ij}^{(t)} = \omega_1^2$$

These ratios were calculated for $t=20$ and an estimate of 1.0556 was obtained for ω_1 . Further, for $t=20$ the ratios of expectations were seen to be almost independent of j but the ratios of variances still depended on j for both the models as can be seen from *Table 2*. Thus, for either model, a much larger value of t is necessary to be able to calculate ω_1^2 with reasonable accuracy from the ratios of variances. Further, the value of ω_1^2 computed from the ratios of variances deviated less from the actual value ($\omega_1^2 = 1.0556^2 = 1.1141$) in the case of Poisson-binomial model.

Thus, from the foregoing and from the consideration that it is more logical to describe the birth process by Poisson law, it can be concluded that the Poisson-binomial model is more suited than the other model in the study of population growth in dairy herds.

SUMMARY

Two stochastic analogues of the deterministic discrete model (Lefkovitch, 1965; Jain and Narain, 1974) for studying the growth of female dairy population grouped in unequal stage-groups have been presented. The number of individuals born during any arbitrary interval of time is assumed to follow a Poisson/binomial distribution and the number of transfers occurring during the same interval a binomial distribution. Linear matrix recurrence relation is derived which determines precisely the first two moments of the stage-group random variables at each unit of time. Results about the asymptotic behaviour of the population are mentioned. Almost all the results are analogous to the ones obtained earlier by Pollard (1966) while studying the stochastic version of Leslie's (1945) equal age-step deterministic model wherein both births and transfers were assumed to follow a binomial law. Further, Monte Carlo experiments using the two models are described and their results compared.

ACKNOWLEDGEMENTS

The author is grateful to Dr. Kanwar Sen, Head Department of Mathematical Statistics, Delhi University and to Dr. Prem Narain, Senior Professor of Statistics, I.A.S.R.I, New Delhi for their valuable suggestions and to Dr. D. Singh, Director, I.A.S.R.I. for his keen interest in the investigation.

REFERENCES

- [1] Amble, V.N. and Jain, J.P. (1967) : Comparative performance of different grades of crossbred cows on military farms in India. *J. Dairy Sci.* 50 : 1695-1702.
- [2] Jain, J.P. and Narain P. (1974) : The use of population generation matrix in dairy herds. *J. Ind. Soc. Agri. Statist.* 26 : 71-92.
- [3] Knuth, D.E. (1969) : The art of Computer Programming. Semi-numerical Algorithms. Vol. 2 pp. 117-18. Addison-Wesley Publishing Co. U.S.A.
- [4] Lefkovich, L.P. (1965) : The study of population growth in organisms grouped by stages. *Biometrics* 21 : 1-18.
- [5] Leslie, P.H. (1945) : On the use of matrices in certain population mathematics. *Biometrika* 33 : 183-212.
- [6] Pillai, S.S. (1972) : Computer Programmes for Agri. Research. Vol. 1, p. 125. I.A.R.S., New Delhi.
- [7] Pollard, J.H. (1966) : On the direct matrix product in analysing certain stochastic population models. *Biometrika* 53 : 397-415.