USES OF STOCHASTIC MODELS IN THE EVALUATION OF POPULATION POLICIES. II. EXTENSION OF THE RESULTS BY COMPUTER SIMULATION

EDWARD B. PERRIN
UNIVERSITY OF WASHINGTON

1. Introduction

The results obtained by the application of the conventional techniques of the theory of renewal processes and the theory of semi-Markov, or Markov renewal, processes to the general fertility model, as described in [1] and [2], for example, are interesting and quite possibly of some practical value. There remain, however, a number of important respects in which the problem does not yield to the application of these techniques, at least in our present state of knowledge. For example, most of the results that exist thus far for the most general model of human reproduction are asymptotic results (see references in [2]) with respect to time and hence are relations which are at best only approximately true during any short period of observation of the system. It would be of interest to know how well the distribution of the numbers of renewals of a given event, such as a live birth, that occur in a given time period for an individual woman is approximated by the respective asymptotic distribution, given a particular set of parameters. What does the exact family size distribution really look like after short periods of marriage under a given model, and how long must the process be observed before this distribution is reasonably well approximated by the asymptotic results obtained from the results of renewal theory?

As a second area of consideration, there is the very important problem of the generalization of the set of stochastic models by the relaxation of some of the more restrictive assumptions in order to allow the mathematical system to conform more closely to the biological system which it is attempting to describe. For example, it is clearly important to investigate the behavior of the particular model of the reproductive process described in [2] when the fecundability of the individual female is allowed to be a function of factors such as age, parity, and so forth. There are a number of directions in which this and other models could

Supported in part by USPHS Grant GM 1269-01.

be similarly modified, with most of the important modifications yielding systems in which a reasonably complete mathematical analysis is at present impossible. For both of these problems—the question of the goodness of fit of asymptotic results and of modifications of the basic model which prove mathematically intractable—it is natural to turn for further insight to an analysis involving the use of data from a simulation of the system under the conditions in question.

2. The model

The fertility model which will be investigated by simulation is the stochastic version with the three different pregnancy outcomes described in [1] and [2]. The important components of the system are presented in table I. Briefly, a

TABLE I

STATES OF THE REPRODUCTIVE PROCESS

State	Description of State	Probability Density Function of Length of Stay in State	Probability the Next State Visited Will Be S_0	S_1	B_1	B_2	B_3
S_0	nonpregnant, susceptible	$\lambda(t)$	0	1	0	0	0
S_1	pregnant	$f_i(t)$, given B_i, $i = 1, 2, 3$, visited next	0	0	θ_1	θ_2	θ_3
B_1	post live birth, nonsusceptible	$g_1(t)$	1	0	0	0	0
B_2	post still birth, nonsusceptible	$g_2(t)$	1	0	0	0	0
B_3	post early fetal death, nonsusceptible	$g_3(t)$	1	0	0	0	0

married woman is assumed to be in one and only one state with respect to reproduction at any time during her active reproductive period. The states in which a woman can be found are designated S_0, S_1, B_1, B_2, B_3, respectively, and represent the susceptible, nonpregnant state S_0, pregnancy S_1, and the postpartum infecundable state following a live birth B_1, a still birth B_2, and an early fetal loss B_3, in that order. The length of time spent in any state is for each woman assumed to be a random variable, as is the outcome of each pregnancy, with probability density function as designated in table I. The important physical events of the system are, of course, the passage of a woman from S_0 into S_1, that is, a conception, the passage from S_1 to B_1, that is, a live birth, and so forth. It is the relationships between such events, that is, the birth rates, the distribution of birth intervals, and so forth, in a population that we generally seek to investigate and many results of a general nature currently exist for this model (for example, [1], [3], [4]).

Although the choice of time unit for this model is quite arbitrary, the calendar month will be taken to be the standard time unit here, primarily because of its biological significance but also for the added simplicity of dealing only with distributions of integral valued random variables in the simulation procedure.

3. The simulation procedure

Considerable literature exists on the use of Monte Carlo procedures on high speed computers for the solution of specific problems (see, for example, [5]). The direct simulation of a probabilistic problem is the simplest form of the Monte Carlo method and has been used successfully in many areas of research, including in conjunction with biological investigations [6], [7]. As a technique, it has often been useful in supplying background material in understanding qualitatively some of the properties of stochastic models and also in making possible the study of particular problems and extensions of the models not amenable to mathematics. It will be our aim to illustrate both of these uses in connection with the present fertility model.

The actual simulation procedure for this model, written for the IBM-7094 computer, consists, of course, of putting an individual woman through the process for a given number of months, randomly selecting at the proper point and according to the specified distributions both the state to which the individual will pass next and the length of stay in that state. This process is repeated for a specified number of women, with the major programming effort being directed toward keeping track of the proper variables for an entire cohort of women. The output for the simulated cohort of females is designed to provide the usual information of demographic interest, together with some optional features. A typical output includes, for example, the observed distribution of the length of the intervals between live births by birth number; the distribution of number of occurrences of live births, still births and fetal losses, annually and accumulated by five year periods; the mean length of birth interval by birth number, for each completed family size, and so forth. For this particular program there also exists the option, which is seldom used, of obtaining a print out of the path traced by each individual woman through the system. Although the computer time involved in the simulation varies somewhat with the complexity of the distributions specified and the present program is not claimed to be optimal in any sense, the reproductive experience of a cohort of 1000 women can be simulated for a period of 15 years under this model in approximately one minute of 7094 time for an average set of parameters.

4. Results

The results presented here are selected to illustrate as simply as possible the use to which the simulation procedure is being put, to gain some insight into certain facets of the process. The data are specifically directed toward answering the following two questions.

(1) How good are the asymptotic results concerning birth rates, birth intervals

and expected family size for a case where the assumptions of the Markov renewal process hold?

(2) What is the effect of a simple form of age dependence on the birth rates, birth intervals and expected family size in an otherwise homogeneous population?

In table II are specified the various component distributions for what will be

TABLE II

ASSUMPTIONS FOR THE "CONSTANT" CASE

(1) Monthly probability of conception: $\lambda(t)$, a geometric distribution, parameter $\rho = .20$, that is, constant monthly probability .20 of conception.
(2) Probabilities associated with various pregnancy outcomes: $\theta_1 = .8$, $\theta_2 = .02$, $\theta_3 = .18$.
(3) Distributions of length of stay in pregnancy and post pregnancy states:

No. of months (t)	$f_1(t)$	$f_2(t)$	$f_3(t)$	$g_1(t)$	$g_2(t)$	$g_3(t)$
1	—	—	.05	—	—	.25
2	—	—	.15	—	.10	.50
3	—	—	.30	—	.25	.25
4	—	—	.30	—	.30	—
5	—	—	.10	.05	.25	—
6	—	—	.10	.10	.10	—
7	.05	.25	—	.20	—	—
8	.10	.25	—	.30	—	—
9	.60	.25	—	.20	—	—
10	.25	.25	—	.10	—	—
11	—	—	—	.05	—	—
Mean	9.05	8.50	3.55	8.00	4.00	2.00
Variance	.55	1.25	1.65	2.10	1.30	0.50

(4) Length of follow up: 1000 women followed for 15 years (180 months) each.

designated the "constant" case, that is, a case for which the parameters are not time dependent in other than the restricted sense of a Markov renewal process. Whereas this particular selection of probabilities cannot necessarily be defended as the best or the most natural, it is typical of the distributions observed for these particular variables in many noncontracepting populations around the world. An average post live birth infecundable period of eight months is not unusual, for instance, in a society practicing breast feeding extensively [8]. Likewise, a pregnancy wastage rate of 20 per cent is probably often close to fact [9], and so forth.

Tables III, IV, and V give the results of a single computer run in which the reproductive experience of 1000 women was simulated for 15 years under the assumption that each woman was subject to the probability distributions given in table II and that these probabilities remained constant over the period. Table III gives the average length of the interval from the $(n-1)$st to the nth live birth. The average length of the interval for all except the first live birth would be predicted to be

$$\mu_{11} = \frac{1}{\theta_1}\left[\frac{1-\rho}{\rho} + \sum_{i=1}^{3} \theta_i(\mu_i + \nu_i)\right], \tag{4.1}$$

where μ_i and ν_i are the means of f_i and g_i, respectively.

For this set of parameters, $\mu_{11} = 23.61$. Likewise, the average length of the interval to the first live birth would be $\mu_{01} = \mu_{11} - \nu_1$. Here $\mu_{01} = 23.61 - 8.00 = 15.61$.

Each of these tables illustrates the kind of insight into the process that can be provided by a simulation procedure. Table III shows, for example, that

TABLE III

AVERAGE INTERVALS FROM $(n - 1)$ST TO nTH LIVE BIRTH
"CONSTANT" CASE, 1000 FAMILIES, 15 YEARS

n	Number Completing Interval	Average Length of Interval Observed
1	1000	15.55
2	1000	23.23
3	1000	23.78
4	1000	23.60
5	999	23.19
6	980	23.96
7	853	22.85
8	521	21.78
9	130	19.38
10	6	18.50

whereas both the predicted average interval length of 15.61 months from marriage until the first live birth and the predicted average interval of 23.61 months between subsequent live births approximate the average observed (that is, simulated) intervals well for the lower birth orders, there is a decided departure from the predicted interval length in the higher order births. It is significant that this decrease in average time between births of successive parity number comes about without any change in the birth parameters during the period of observation. This decrease is, of course, directly related to the fact that the population was observed cross sectionally after a given period of time (15 years), in effect, truncating the process abruptly so that the longer birth intervals did not have a chance to become completed. This effect would also be present in a collection of completed families under the assumption that the monthly probability of conception ρ becomes zero for each female at the end of a given period. The fact that the average interval between births does decrease sizably for higher birth orders even with no change in birth parameters should serve as somewhat of a caution regarding the inferences which can be drawn when changes in these measurements are observed in the population.

Table IV is an examination of the agreement between the birth rates and family size observed in the simulation process and the essentially asymptotic

TABLE IV

BIRTH RATES
"CONSTANT" CASE, 1000 FAMILIES, 15 YEARS

Year No.	Observed Annual Birth Rate	Observed Average Family Size (end of year)	Predicted Average Family Size	
			Uncorrected	Corrected
1	.408	.408	.508	.403
2	.488	.896	1.017	.912
3	.581	1.477	1.525	1.420
4	.445	1.922	2.033	1.928
5	.547	2.469	2.541	2.436
6	.467	2.936	3.049	2.944
7	.550	3.486	3.558	3.456
8	.469	3.955	4.066	3.961
9	.542	4.497	4.574	4.469
10	.508	5.005	5.082	4.977
11	.480	5.485	5.591	5.487
12	.512	5.997	6.099	5.994
13	.499	6.496	6.607	6.502
14	.514	7.010	7.116	7.011
15	.500	7.510	7.624	7.519

results derived from the application of renewal theory to the model. The asymptotic annual fertility rate, that is birth rate, predicted for the process is $12/\mu_{11} = 12/23.61 = 0.508$, where μ_{11} is the expected waiting time between births. The predicted average family size (uncorrected) is $t/\mu_{11} = 0.0424t$, where t is in months. The predicted average family size (corrected) is $t/\mu_{11} + \mu_{11}^{(2)}/2\mu_{11}^2 - \mu_{01}/\mu_{11} = 0.0424t - 0.105$, where μ_{01} is the expected waiting time from marriage to the first birth.

Whereas the annual birth rate shows a good deal of cyclic fluctuation in the early years of the cohort it settles down reasonably well in 15 years to the asymptotic annual birth rate predicted (that is, 0.508). The predicted average family size agrees rather impressively with that observed, especially when the predicted value is corrected for the fact that all 1000 women began in the nonpregnant susceptible state S_0. Although this is a particularly simple case, it is interesting that the average family size at the end of 15 years of simulation (7.510 children per family) actually does not differ greatly from that observed in noncontracepting populations, for example the Hutterites of central Canada [10].

Table V is a summary of the average interval length between successive births by completed family size generated in the simulation run. It shows what you might expect, namely, that as completed family size increases the average interval between births decreases, since the births must occur more closely together for the larger families in order to accomplish the eventual larger size. It is enlightening to see the relatively large variation in completed family size and length of the respective birth intervals, keeping in mind that each woman is

TABLE V

AVERAGE LIVE BIRTH INTERVAL BY COMPLETED FAMILY SIZE
"CONSTANT" CASE, 1000 FAMILIES, 15 YEARS

Number of Families	Completed Family Size	1	2	3	4	5	6	7	8	9	10
0	1	—									
0	2	—	—								
0	3	—	—	—							
1	4	14.0	37.0	26.0	45.0						
19	5	19.8	34.3	34.8	35.2	30.4					
127	6	20.1	28.2	29.3	27.5	26.7	28.7				
332	7	17.0	24.2	24.5	24.1	24.2	25.4	24.7			
391	8	13.9	21.5	22.2	22.4	21.9	22.5	22.3	22.4		
124	9	11.8	19.4	19.6	20.2	20.3	20.1	19.8	20.1	19.4	
6	10	10.7	18.2	18.0	19.5	18.2	18.7	16.8	18.2	18.3	18.5
1000											

subjected to the same probabilistic phenomenon and that any differences observed are due to random variation alone, and are not due to differing fecundability or other parameters.

The next set of results illustrate the use of simulation procedures to investigate

TABLE VI

AVERAGE INTERVALS FROM $(n - 1)$ST TO nTH LIVE BIRTH
"AGE DEPENDENT" CASE, 1000 FAMILIES, 15 YEARS

n	Number Completing Interval	Average Length of Interval Observed
1	1000	15.60
2	1000	23.73
3	1000	23.73
4	1000	24.21
5	988	24.04
6	911	24.69
7	659	23.62
8	302	23.13
9	61	22.30

a simple time dependent extension of the basic model. This version, called the age dependent case, employs the same assumptions as the constant case, except that the monthly probability of conception is assumed to remain constant until month 120 and then to decrease linearly to zero by month 180. This is a simple example of a case where, because of the dependence of a parameter on time, the asymptotic results from renewal theory are not valid and little is known of methods to predict the behavior of the model.

Table VI illustrates the interesting and perhaps obvious finding that the

TABLE VII

BIRTH RATES
"AGE DEPENDENT" CASE, 1000 FAMILIES, 15 YEARS

Year No.	Observed Annual Birth Rate	Observed Average Family Size End of Year	Predicted Average Family Size (Corrected), "Constant" Case
1	.413	.413	.403
2	.457	.870	.912
3	.571	1.441	1.420
4	.426	1.867	1.928
5	.565	2.432	2.436
6	.449	2.881	2.944
7	.534	3.415	3.456
8	.461	3.876	3.961
9	.540	4.416	4.469
10	.496	4.912	4.977
11	.471	5.383	5.487
12	.477	5.860	5.994
13	.425	6.285	6.502
14	.352	6.637	7.011
15	.284	6.921	7.519

average length of the intervals between the middle order births increase as the fecundability of the population decreases. In the absence of age dependence, the expected (birth) interval length to the first and higher order births would be 15.61 and 23.61, respectively. It is seen, however, that the truncation effect noted previously here more than offsets this declining fecundability, resulting in a reduced interval between higher order births, again illustrating the difficulty in making inferences about changes in birth parameters from observed data on field populations.

TABLE VIII

AVERAGE BIRTH INTERVAL BY COMPLETED FAMILY SIZE
"AGE DEPENDENT" CASE, 1000 FAMILIES, 15 YEARS

Number of Families	Completed Family Size	Order of Birth Interval 1	2	3	4	5	6	7	8	9
0	1	—								
0	2	—	—							
0	3	—	—	—						
12	4	38.3	31.1	21.0	39.8					
77	5	19.5	28.7	30.1	29.5	34.9				
252	6	18.5	25.2	25.6	25.9	25.7	30.4			
357	7	14.2	24.0	23.5	24.1	23.2	23.6	25.4		
241	8	13.0	21.1	21.2	21.2	21.3	21.8	21.9	23.9	
61	9	12.3	18.3	20.1	20.1	19.1	19.1	19.9	20.0	22.3

Table VII shows the effect of the age dependent decline in fecundability on the birth rate (which would tend to 0.508 in the absence of age dependence). The effect is that the birth rate declines in the last five years of observation although perhaps not so rapidly as one might expect, with the result that the average family size in 15 years is about 0.6 child less than in the previous case. Table VIII shows the magnitude of the effect of the decline in fecundability on the distribution family size and length of birth interval by completed family size. Again the effect is felt throughout, but most strongly in the late births.

Obviously this latter case can be extended greatly to study a more complex dependence of fecundability on age, the current program being designed to accept any functional dependence which can be approximated by a large number of connecting linear segments. Likewise, for example, a very general distribution may be placed on the time of entry of a woman into the process (that is, age at marriage) and on the time of exit from the process (that is, age at menopause). Results of simulation runs under more general assumptions such as these are currently being studied. As one would expect, of course, the complexity of interpretation increases with the complexity of the model simulated, a fact perhaps not always fully appreciated by the (uninitiated) beginner.

5. Conclusions

It would seem that the use of Monte Carlo techniques, as illustrated in the previous section, to simulate human reproductive patterns for the purpose of understanding the processes more completely both qualitatively and quantitatively and for the purpose of extending existing models beyond present mathematical limitations can be a valuable adjunct to more classical techniques of investigation of population dynamics. The implications for those involved in the task of evaluation of public population policies are obvious and real. It may be only through such simulation techniques, for example, that we can realistically estimate the effect of a changing pattern of age at marriage on the birth rate or the effect of shifts in differential fertility in certain segments of the population on general population growth, or the importance of age and parity in the over all effectiveness of various plans for population control. Already some attempts have been made to estimate the level of effort necessary to achieve a given level of population control in the face of varying public acceptance under a simple model [11]. One would naturally want next to extend this, as is being done, to more complex specific situations by examination of the output of a simulation procedure.

In short, then, despite the fact that the problems of computer programming and output interpretations are real and not to be ignored and in spite of the additional fact that simulation data is by definition specific to a particular choice of parameters, it appears that there is sufficient information and insight to be gained in the output of a simulation process to justify its use in the investigation of population dynamics and in the accompanying problem of evaluation of population policies.

REFERENCES

[1] E. B. PERRIN and M. C. SHEPS, "Human reproduction: a stochastic process," *Biometrics*, Vol. 20 (1964), pp. 28–45.

[2] M. C. SHEPS, "Uses of stochastic models in the evaluation of population policies. I. Theory and approaches to data analysis," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 4, pp. 115–136.

[3] M. C. SHEPS and E. B. PERRIN, "The distribution of birth intervals under a class of stochastic fertility models," *Population Studies*, Vol. 17 (1964), pp. 321–331.

[4] L. HENRY, "Fécondité et famille, modèles mathématiques," *Population*, Vol. 16 (1961), pp. 27–48 and pp. 261–282.

[5] J. M. HAMMERSLEY and D. C. HANDSCOMB, *Monte Carlo Methods*, London, Methuen, 1964.

[6] W. F. BODMER, "Discrete stochastic processes in population genetics," *J. Roy. Statist. Soc. Ser. B*, Vol. 22 (1960), pp. 218–244.

[7] M. S. BARTLETT, "Monte Carlo studies in ecology and epidemiology," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, 1961, Vol. 4, pp. 39–55.

[8] R. G. POTTER, J. E. GORDON, M. PARKER, and J. B. WYON, "A case study of birth interval dynamics," *Population Studies*, Vol. 19 (1965), pp. 81–96.

[9] C. TIETZE, "Pregnancy rates and birth rates," *Population Studies*, Vol. 16 (1962), pp. 31–37.

[10] M. C. SHEPS, "An analysis of reproductive patterns in an American isolate," *Population Studies*, Vol. 19 (1965), pp. 65–80.

[11] M. C. SHEPS and E. B. PERRIN, "Changes in birth rates as a function of contraceptive effectiveness: some applications of a stochastic model," *Amer. J. Public Health*, Vol. 53 (1963) pp. 1031–1046.