A COMPOUND POISSON PROCESS FORMULATION OF THE PARTIY DISTRIBUTION

Lester R. Curtin and Chirayath M. Suchindran, University of North Carolina

1. Introduction

Various attempts have been made to describe the parity distribution as the realization of some type of Poisson process. Dandekar (1955) developes a modified Poisson distribution which is applied to data on the number of children born in a fixed time period. Brass (1958) and Singh (1968) assume that the number of live birth conceptions follow an underlying Poisson process with modifications for non-susceptible periods following a live birth. Further modifications for heterogeneity among women and for conceptions which end in fetal loss (pregnancy wastage) must also be considered. The probability distributions which result from these models are somewhat cumbersome and difficult to apply.

An assumption of an underlying non-homogeneous Poisson process leads to more theoretical models of the parity distribution such as those by Hoem (1969) and Nour (1972). This paper derives a model of the parity distribution which incorporates Nour's concept of conditional fecundability. The resulting model is a realization of a compound Poisson process and is a particular case of Hoem's model. Estimation of the model parameters from U.S. cohort fertility data will be briefly examined.

2. The H_1 and H_2 distributions

Suppose we observe a cohort of women of current age x. Assume that there has been no mortality, that each woman has been susceptible to the risk of a live birth conception for a fixed number n of time units, and that the probability of a live birth conception in a unit time is a constant p, $0 < p < 1$. Under these assumptions, the number of births to a woman aged x is a random variable having a Binomial distribution with parameters n and p.

Actually, the number of time units that a woman is susceptible to the risk of conception can be considered a random variable. That is, n will vary among women due to the influence of such variables as age at first marriage, non-susceptible periods following a live birth conception (the nine months of gestation plus a period of postpartum amenorrhea), and non-susceptible periods associated with pregnancy wastage. We consider two cases. For the first case, we assume that n is a random variable having a Poisson distribution with parameter λ. This gives the compound distribution for the number of births to a woman aged x as a Poisson distribution with mean λp For the second case, we assume that n has a Negative Binomial distribution with parameters K and p'. The resulting compound distribution is then a Negative Binomial distribution with parameters K and pp'.

Heterogeneity amoung women is introduced by considering the parameter p of the Binomial distribution as a random variable having a Beta distribution with parameters a and b. Specifically, let

$$f(p) = \frac{\Gamma(a)\Gamma(b-a)}{\Gamma(b)} p^{a-1}(1-p)^{b-a-1} \qquad (2.1)$$

with $b > a > 0$ and $0 > p > 1$.

This gives the parity distribution conditional on age as either Katti's (1968) H_1-distribution (for case 1) or the H_2-distribution (for case 2). Using Gurland's (1957) notation for compound distribution we have

$$H_1(\lambda,a,b) \sim \mathrm{Bin}(n,p)\hat{}_n\ \mathrm{Poisson}(\lambda)\hat{}_p\ \mathrm{Beta}(a,b)$$

$$H_2(k,a,b,p') \sim \mathrm{Bin}(n,p)\hat{}_n\ \mathrm{Neg.Bin}(k,p')\hat{}_p\ \mathrm{Beta}(a,b)\ .$$

The probability generating functions are given by

$$g_{H_1}(x) = {}_1F_1[a;b;\lambda(s-1)] \qquad (2.2)$$

and

$$g_{H_2}(s) = {}_2F_1[k,a;b;p'(s-1)] \qquad (2.3)$$

where ${}_1F_1[a;b;\lambda(s-1)]$ is the confluent hypergeometric function and ${}_2F_1[k,a;b;p'(s-1)]$ is the hypergeometric function. These are defined (Erdéyli, 1953) as follows.

$${}_1F_1[a;b;\lambda(s-1)] = \sum_{n=1}^{\infty} \frac{(a)n}{(b)n} \frac{\lambda^n(s-1)^n}{n!}$$

$${}_2F_1[k,a;b;p;(s-1)] = \sum_{n=0}^{\infty} \frac{(k)n(a)n}{(b)n} \frac{[p'(s-1)]^n}{n!}$$

and

$$(a)n = \frac{\Gamma(a+n)}{\Gamma(a)} = \begin{cases} 1 & \text{if } n = 0,-1,\dots \\ \prod_{k=0}^{n-1}(a+k) & \text{if } n = 1,2,\dots \end{cases} .$$

Differentiating the probabiltiy generating function gives the probability density functions for the H_1 and H_2 distributions as

$$p_{H_1}(x) = \frac{\lambda^x}{x!} \frac{(a)x}{(b)x}\, {}_1F_1[a+x;\ b+x;\ -\lambda] \qquad (2.4)$$

and

$$P_{H_2}(x) = \frac{(p')^x}{x!} \frac{(k)x(a)x}{(b)x}\, {}_2F_1[k+x;a+x;b+x;-p'] \qquad (2.5)$$

The factorial moments of these distributions are given by simple recurrence relations. For the H_1-distribution

$$\mu'_{(r+1)} = \frac{\lambda(a+r)}{(b+r)} \mu'_{(r)} \quad \text{for } r = 0,1,2,\dots \qquad (2.6)$$

For the H_2-distribution the relation is

$$\mu'_{(r+1)} = \frac{p'(k+r)(a+r)}{(b+r)} \mu'_{(r)} \quad \text{for} \quad r = 0,1,2,...$$

The mean and variance of each distribution are easily desired to be:

$$\mu'_{H_1} = \frac{\lambda a}{b} \quad \text{and} \quad \sigma^2_{H_1} = \frac{\lambda a}{b}\left[1 + \frac{\lambda(b-a)}{b(b+1)}\right]$$

while

$$\mu'_{H_2} = \frac{Kap'}{b} \quad \text{and} \quad \sigma^2_{H_2} = \frac{kap'}{b}\left[1 + \frac{kp'(b-a)}{b(b+1)} + \frac{p'(a+1)}{(b+1)}\right].$$

Setting $\mu_{H_1} = \mu_{H_2}$ it is seen that $\sigma^2_{H_1} < \sigma^2_{H_2}$. Also, both distributions are over-dispersed in the sense that

$$\frac{\sigma^2_{H_1}}{\mu_{H_1}} > 1 \quad \text{and} \quad \frac{\sigma^2_{H_2}}{\mu_{H_2}} > 1 .$$

3. The Compound Poisson Process

A compound Poisson process is defined by Parzen (1962) in the following manner. Consider the stochastic process $\{x(t), t > 0\}$. Let

$$x(t) = \sum_{n=1}^{N(t)} Y_n \qquad (3.1)$$

such that $\{Y_n;\ n = 1,2,...\}$ are independently identically distributed random variables and $\{N(t), t > 0\}$ is a Poisson process with intensity function $\nu(t)$. Then $x(t)$ is said to be a compound Poisson process. Also, we can define

$$m(t) = \int_0^t \nu(\tau)d\tau \qquad (3.2)$$

as the mean value function of the Poisson process $N(t)$.

We now define $x(t)$ to be the number of live births in the interval $(0,t)$ and $N(t)$ is the number of time units that a woman is susceptible to the risk of a live birth conception. As before we can define p as the probability of a live birth conception in a unit time given that the woman is susceptible to the risk of a live birth conception. This corresponds to Nour's definition of conditional fecundability (Nour, 1972). In the context of the compound Poisson process we now have

$$Y_n = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } (1-p) \end{cases} \qquad (3.3)$$

for $n = 1,2,...$.

The unconditional fecundability (the probability of a live birth conception) can now be defined as $p\nu(\tau)\Delta\tau + o(\Delta\tau)$ where the probability of a woman being susceptible to the risk of conception in the interval $(\tau,\tau+\Delta\tau)$ is given by $\gamma(\tau)\Delta\tau + o(\Delta\tau)$. Thus the "force of fertility" is simply

$$\phi(t) = p\nu(t) . \qquad (3.4)$$

Given the above formulation, we can derive the distribution of $x(t)$ for an arbitrary, but fixed, value of t. Let $P_k(t|p)$ be the probability of k births $(k = 0,1,2,...)$ in the interval $(0,t)$ given the value of p. We have

$$P_k(t|p) = \exp\left[-\int_0^t \phi(\tau)d\tau\right] \frac{\left[\int_0^t \phi(\tau)d\tau\right]^k}{k!};\ k = 0,1,2,...$$

This can also be written as

$$P_k(t\ p) = \exp[-pm(t)]\ \frac{[pm(t)]^k}{k!};\ k = 0,1,2,...$$

Assume that p has a denisity function $f(p)$, we have that $P_k(t)$, the unconditional probability of k births in $(0,t)$, is given by

$$P_k(t) = \int_0^1 P_k(t|p)f(p)dp \qquad (3.5)$$

since $0 < p < 1$. An obvious choice of $f(p)$ is the Beta distribution (equation 2.1). Substitution yields:

$$P_k(t) = \frac{[m(t)]^k}{k!}\frac{\Gamma(b-a)\Gamma(a)}{\Gamma(b)}\int_0^1 e^{-pm(t)}p^{k+a-1}(1-p)^{b-a-1}dp$$

or

$$P_k(t) = \frac{[m(t)]^k}{k!}\frac{(a)k}{(b)k}\ {}_1F_1[a+k;b+k;-m(t)] \qquad (3.6)$$

for $k = 0,1,2,...$. This is simply the H_1-distribution.

If $N(t)$ is considered to be a homogeneous Poisson process, then $\nu(t) = \nu$ and $m(t) = \nu t$. This gives

$$P_k(t) = \frac{[\nu t]^k}{k!}\frac{(a)k}{(b)k}\ {}_1F_1[a+k;b+1;-\nu t] \qquad (3.7)$$

and the probability generating function for $X(t)$ is then $g(s) = {}_1F_1[a;b;\nu t]$.

Further heterogeneity amoung women can be introduced by assuming that ν is a random variable having a Gamma distribution with parameters k and β. That is,

$$f(\nu) = \frac{1}{\beta^k\Gamma(k)}\ \nu^{k-1}\exp[-\nu/\beta] . \qquad (3.8)$$

Letting $p' = 1/(B+1)$, the probability generating function for $X(t)$ now becomes

$$g(s) = {}_2F_1[k,a;b;p't] \qquad (3.9)$$

which is the H_2-distribtuion.

4. Estimation of Parameters

The maximum likelihood equations for the estimators of the parameters of the H_1 and H_2 distributions involve finite series and can not be solved explicitly. Iterative procedures, such as those outlined in Kaplan and Elston (1972) can be used to obtain the maximum likelihood estimates. However, for these distributions, the iterative procedures either fail to converge or converge to an arbitrary upper or lower bound. This could be a result of a very flat likelihood surface. Table I shows how similar H_1-distributions can be obtained for quite different values of the parameters a and b.

Minimum chi-square estimates have the same problems as the maximum likelihood estimates. The minimization procedures used so far have been inadequate to give valid estimates. Moment type estimates are also inadequate.

In an attempt to get preliminary estimates, the function

$$\sum_i |O_i - E_i| \qquad (4.1)$$

was minimized. Here O_i refers to the observed frequency and E_i refers to the expected frequency for the desired distribution. These estimates can only be considered as very rough estimates and are used only as the initial step in the examination of the goodness of fit of the H_1 or H_2 distributions.

5. Modified Models

The data used is from Heuser (1976) and consists of the parity distribution by single years for the 1920 birth cohort (while women only). Initial estimates of the parameters of the H_1-distribution, conditional on age, indicate a possible lack of fit of the model. A modified H_1-distribution can be developed by adjusting the zero-parity class.

We let $(1-\alpha)$ be the proportion of the cohort at age x which can be considered as *never* having been susceptible to the risk of conception due to natural sterility or due to never being married. The modified H_1-distribution is then given by

$$P_0^* = (1-\alpha) + \alpha P_0$$

and

$$P_i^* = \alpha P_i \; , \quad i = 1,2,\ldots \qquad (5.1)$$

where the $P_0, P_1, P_2, \ldots,$ are the probability under the H_1-distribution.

The modified H_1-distribution does not provide an adequate fit to the data at the older ages. The major discrepancy lies in the class where parity equals two. The H_2-distribution fails to correct this problem. In an attempt to correct the problem with the second parity class, two mixtures were considered. Namely, a mixture of two H_1-distributions with different values of the parameter λ:

$$(1-\alpha)H_1(a,b,\lambda_1) + \alpha H_1(a,b,\lambda_2) \qquad (5.2)$$

and a mixture of the H_1 and H_2 distributions as

$$(1-\alpha)H_1(a,b,\lambda) + \alpha H_2(k,a,b,p') \qquad (5.3)$$

Examples of estimates of the parameters of the four distributions considered are presented for various ages. The distributions used are the modified H_1-distribution (Table II), the H_2-distribution (Table III), the mixture of two H_1-distributions (Table IV) and the mixture of an H_1 and an H_2-distribution (Table V).

6. Conclusions

Of the four distributions considered, the modified H_1-distribution appears to provide the best approximation to the observed parity distribution except at the oder ages. After age 40, the mixture of two H_1-distributions is a better approximation. This is indicated by Table VI. Again, the difference between the observed distribution and the fitted, or expected, distributions is most apparent at parity two. It seems that a certain proportion of the population terminate their reproduction after their second birth. Perhaps the model can be improved by treating the study population as a mixture of two populations with one group consisting of people who wish to terminate their fertility at two and the other group who does not terminate at two.

It is also obvious that the estimation procedures must be improved. It may yet be possiole to obtain maximum likelihood estimates for the above distribution. These problems will be examined in subsequent reports.

Acknowledgments

This research was in part supported by the National Institute of Child Health and Human Development (Grant HD-00371). The authors would also like to thank Jackie O'Neal for her conscientious typing of this manuscript.

References

1. Brass, W. (1958). "The distribution of births in human populations," *Population Studies* 12, 51-72.

2. Dandekar, V.M. (1955). "Certain modified forms of Binomial and Poisson distributions," *Sankhyā A* 15, 237-250.

3. Erdélyi, A. (1953). *Higher Transcendental Functions, Vol. I.* McGraw-Hill Book Company, Inc., New York.

4. Gurland, J. (1957). "Some interrelations amoung compound and generalized distributions," *Biometrika* 44, 265-268.

5. Heuser, R.L. (1976). Fertility tables for birth cohorts, by color: United States 1917-1973. DHEW Pub. no. (HRA) 76-1152, National Center for Health Statistics, Rockville, MD

6. Hoem, J.M. (1969). "Fertility rates and reproduction rates in a probabilitic setting," *Biométric-Praximetre* 10, 38-66.

7. Kaplan, E. and Elston, R. (1972). "A subroutine package for maximum likelihood estimation (MAXLIK)," Institute of Statistics Mimeo Series No. 823, University of North Carolina.

8. Katti, S.K. (1966). "Interrelations amoung generalized distributions and their components," *Biometrics* 22, 44-52.

9. Nour, El-Sayed (1972). "A stochastic model for the study of human fertiltity," Institute of Statistics Mimeo Series No. 879, University of North Carolina.

10. Parzen, E. (1962). *Stochastic Processes.* Holden-Day, Inc., San Francisco.

11. Philipson, C. (1960). "The theory of confluent hypergeometric functions and its application to compound poisson processes," *Skandinarisk Aktuarietidskrift* 43, 136-162.

12. Singh, S.N. (1968). "Chance mechanisms of the variation in the number of births per couple," *Journal of the American Statistical Association* 63, 209-213.

TABLE I: The H_1-distribution with $\lambda = 2.5$

PARAMETER		PARITY							
A	B	0	1	2	3	4	5	6	≥ 7
4	5	.1489	.2674	.2582	.1743	.0913	.0392	.0143	.0064
8	10	.1423	.2670	.2641	.1768	.0906	.0378	.0133	.0081
24	30	.1376	.2706	.2684	.1790	.0903	.0367	.0125	.0049
40	50	.1367	.2706	.2693	.1796	.0902	.0365	.0123	.0048
56	70	.1363	.2706	.2697	.1798	.0902	.0363	.0122	.0049

TABLE II: Estimates of the Parameters of the Modified H_1-distribution

	PARAMETER			
Age	A	B	λ	α
20	1.708	3.660	0.943	0.526
25	2.385	3.256	1.553	0.839
30	4.011	4.643	2.058	0.964
35	4.531	5.312	2.695	0.971
40	5.456	6.283	2.934	0.968
45	4.953	5.803	2.935	0.978

TABLE III: Estimates of the Parameters of the H_2-distribution

	PARAMETER			
Age	k	A	B	p'
20	1.525	1.198	2.346	0.275
25	2.857	3.174	3.998	0.422
30	4.766	5.188	6.151	0.475
35	5.137	33.156	38.950	0.560
40	6.159	4.130	4.861	0.504
45	5.153	17.332	20.286	0.630

TABLE IV: Estimates of the Parameters of the Mixture $(1-\alpha)H_1(A,B,\lambda_1) + \alpha H_1(A,B,\lambda_2)$

		PARAMETERS			
Age	α	A	B	λ_1	λ_2
25	0.159	1.246	2.426	2.031	0.893
30	0.484	2.758	4.015	2.605	2.402
35	0.209	3.374	4.440	2.812	2.952
40	0.527	3.390	4.473	3.201	3.038
45	0.969	3.437	4.475	3.291	3.109

TABLE V: Estimates of the Parameters of the Mixture $(1-\alpha)H_1(a,b,\lambda) + \alpha H_2(k,a,b,p')$

			PARAMETER			
Age	α	λ	A	B	k	p'
25	0.478	2.360	2.367	4.700	2.031	0.389
30	0.166	5.343	4.612	6.427	2.311	0.491
35	0.090	2.158	4.713	6.061	2.809	0.323
40	0.088	5.831	4.165	4.924	2.688	0.557
45	0.004	5.476	3.480	4.623	3.201	0.601

TABLE VI: The Observed and Expected Parity Distributions

					PARITY					
Age	Parity Distribution	0	1	2	3	4	5	6	≥ 7	$\sum \lvert O_i - E_i \rvert$
35	Observed	.1351	.1929	.2950	.1908	.0962	.0443	.0220	.0238	—
	Modified H_1	.1352	.2243	.2493	.1912	.1128	.0542	.0220	.0110	.1170
	H_2	.1362	.2243	.2212	.1695	.1113	.0657	.0360	.0358	.1802
	Mixture H_1, H_1	.1350	.2466	.2512	.1828	.1046	.0495	.0200	.0103	.1352
	Mixture H_1, H_2	.1300	.2454	.2548	.1859	.1055	.0492	.0196	.0096	.1336
	Observed	.1161	.1685	.2722	.1965	.1148	.0599	.0328	.0452	—
	Modified H_1	.1101	.2015	.2432	.2023	.1293	.0673	.0296	.0167	.1220
	H_2	.1106	.1968	.2101	.1746	.1244	.0806	.0476	.0553	.1680
	Mixture H_1, H_1	.1100	.2197	.2449	.1951	.1223	.0635	.0281	.0164	.0987
	Mixture H_1, H_2	.1095	.2179	.2432	.1949	.1233	.0647	.0291	.0174	.0993