The Distribution by Age of the Frequency of First Marriage in a Female Cohort

A. J. COALE and D. R. MCNEIL*

The schedule recording first marriage frequencies has been shown to take the same basic form in different populations, with differences only in the origin, area, and horizontal scale. It is shown here that a representative schedule is very closely approximated by a simple closed form frequency function, which is the limiting distribution of the convolution of an infinite number of exponentially distributed components. The schedule is approximated equally well by the convolution of a normal distribution (of age of entry into a marriageable state) and as few as three exponentially distributed delays. The latter convolution provides a plausible model of nuptiality, a model that receives surprising empirical support.

1. A STANDARD CURVE OF FIRST MARRIAGE FREQUENCIES FITTED BY A CLOSED FUNCTION

In an earlier article [2], evidence was presented for the existence of a standard age schedule of rates at which women enter first marriage, a schedule taking the same basic form in populations characterized by markedly different mean ages at marriage, and by markedly different proportions remaining celibate. In different populations (or more precisely, among different cohorts of women) the distribution of the ratio of first marriages in each age interval to person-years lived in the interval differs only in origin, total area (the proportion of the cohort ever-marrying by the end of life), and horizontal scale. A "standard" distribution of the frequency of first marriage so defined was constructed by making minor adjustments (to remove evidently particular features) to the schedule of first marriage frequencies recorded in Sweden from 1865 to 1869.

Given the existence of a standard curve, it would be useful to have a mathematical formula which fits it. The proportion of women married by age x (excluding those who never marry) can be treated as a probability distribution function,[1] and it is this function, $F(x)$, say, which is sought. In fitting a probability distribution, one often proceeds by plotting the empirical distribution on log- or semi-log-paper and by this means searching for a straight-line relationship. In this case of the (standardized) distribution of first marriages, this method did not yield a straight-line relationship, but when the same procedure was applied to the risk function (still excluding those who never marry)

$$r(x) = F'(x)/\{1 - F(x)\} \tag{1.1}$$

an extremely close fit of the empirical risk to the "double exponential" function

$$r(x) = 0.174e^{-4.411e^{-0.309x}} \tag{1.2}$$

was obtained. Accordingly, Coale suggested that (1.2) be used as a formula to represent the standardized risk of first marriage.

Unfortunately, neither the distribution function corresponding to (1.2) nor its derivative (the frequency function) is expressible in closed form. Nor are the behavioral implications of such a risk function at all evident. When this material on nuptiality was presented at a conference on mathematical demography held at the East-West Population Center in Honolulu in the summer of 1971, Griffith Feeney suggested that the first marriage distribution curve might be the convolution of a distribution describing the age of entry into the marriage market, and a distribution of delays between entry and actual marriage. He had not had opportunity to make the requisite calculations, and conjectured the possibility that the distribution of delay might be a simple exponential, and the distribution of entry into the marriage market normal.

This suggestion is in fact a special case of an alternative procedure for fitting probability distributions, which is based on the assumption that the distribution function sought is the convolution of a (possibly infinite) number of simple components. In the case when the risk function approaches a constant asymptote, r, say, the convolution contains at least one exponentially distributed component, which may be removed by application of the formula

$$F_1(x) = F(x) + F'(x)/r. \tag{1.3}$$

In equation (1.3), $F_1(x)$ is the distribution which results

* A.J. Coale is director, Office of Population Research and professor, Department of Economics, and D.R. McNeil is faculty associate, Office of Population Research and associate professor, Department of Statistics, both at Princeton University, Princeton, N.J. 08540. The research on which this article is based was made possible by financial support of the general program of the Office of Population Research by the Ford and Rockefeller Foundations.

[1] In a cohort experiencing no mortality, the proportion ever married would be a probability distribution function $F(x)$, and $F'(x)$ would be the density of first marriages. If differential mortality by marital status has a negligible effect on the proportion ever married (as is in general the case), the relation of $F(x)$ to the frequency of first marriage is negligibly different from the relation in the hypothetical absence of mortality.

© Journal of the American Statistical Association
December 1972, Volume 67, Number 340
Applications Section

after the exponential component has been removed. To derive equation (1.3), note that if $F(x)$ is the convolution of $F_1(x)$ and the exponential distribution $1-e^{-rx}$, then by the convolution formula

$$F(x) = \int_0^x F_1(z) re^{-r(x-z)} dz. \tag{1.4}$$

Differentiating both sides of equation (1.4) with respect to x, we find

$$\begin{aligned} F'(x) &= F_1(x)r - r\int_0^x F_1(z)re^{-r(x-z)}dz, \\ &= F_1(x)r - rF(x) \end{aligned}$$

which yields equation (1.3).

It is easy to calculate $F_1(x)$ by first choosing the value of r as the asymptote of the risk function (0.174 in the case of the standardized first marriage distribution) and then numerically differentiating the empirical distribution function. Such calculation reveals that the risk function $r_1(x)$ corresponding to the residual distribution $F_1(x)$ (hypothetically the distribution of entry into the state of readiness for marriage) approaches a constant (0.483) as $x \to \infty$. If the distribution of $F_1(x)$ were normal, as Feeney suggested, then $r_1(x)$ should be asymptotically linear with positive slope.

Repeated application of the formula (1.3) confirms that the distribution corresponding to the risk function (1.2) yields asymptotes $r=0.174$, $r_1=0.174+0.309$, $r_2=0.174+2(0.309)$, $r_3=0.174+3(0.309)$, $\cdots$. The removal of the nth exponential component results in a residual distribution $F_n(x)$, where

$$F_n(x) = F_{n-1}(x) + F'_{n-1}(x)/r_{n-1}. \tag{1.5}$$

It may be noted that when the nth component is taken out, the mean of the residual distribution is reduced by the amount $1/r_{n-1}$, and the variance by $1/r^2_{n-1}$. Since the sum $\sum 1/r_n$, being a harmonic series, diverges (though $\sum 1/r_n^2$ converges) the procedure cannot be repeated indefinitely without reducing the mean of the residual to $-\infty$.

This difficulty is easily remedied by the device of adding an appropriate constant term to the argument of F_n each time another exponential component is removed. This has the effect of leaving the mean of the residual distribution unchanged while reducing the variance. The question then arises: can the distribution corresponding to the "double exponential" risk (1.2) be represented as an infinite convolution of (mean-corrected) exponential distributions, plus a constant term? To answer this question write

$$X_n = a + \textstyle\sum_{j=1}^n \left(Z_j - \dfrac{1}{\alpha + (j-1)\lambda} \right), \tag{1.6}$$

where Z_j is exponentially distributed with mean $1/\{\alpha+(j-1)\lambda\}$, and the Z_j are independent. It is possible to write down the distribution of X_n explicitly, and the relevant calculations are presented in the appendix. It is shown there (Theorem 2) that X_n has a limiting distribution $\bar{G}(x)$ as $n \to \infty$, with frequency function

$$\bar{g}(x) = \frac{\lambda}{\Gamma(\alpha/\lambda)} e^{-\alpha(x-\mu)-e^{-\lambda(x-\mu)}}, \tag{1.7}$$

where Γ indicates the well-known gamma function, $\psi=\Gamma'/\Gamma$ the digamma function, $\mu=a+(1/\lambda)\psi(\alpha/\lambda)$, and a is the mean of $\bar{g}(x)$. (In the case $\alpha=\lambda$, $\bar{g}(x)$ is the frequency function of the Gompertz extreme value distribution (see [4]).)

It may be shown (see appendix, Theorem 5) that the risk function of the distribution (1.7) possesses the expansion

$$\bar{r}(x) = \alpha \left\{ \textstyle\sum_{j=0}^{\infty} \dfrac{\Gamma(1+\alpha/\lambda)}{\Gamma(j+1+\alpha/\lambda)} e^{-\lambda j(x-\mu)} \right\}^{-1}. \tag{1.8}$$

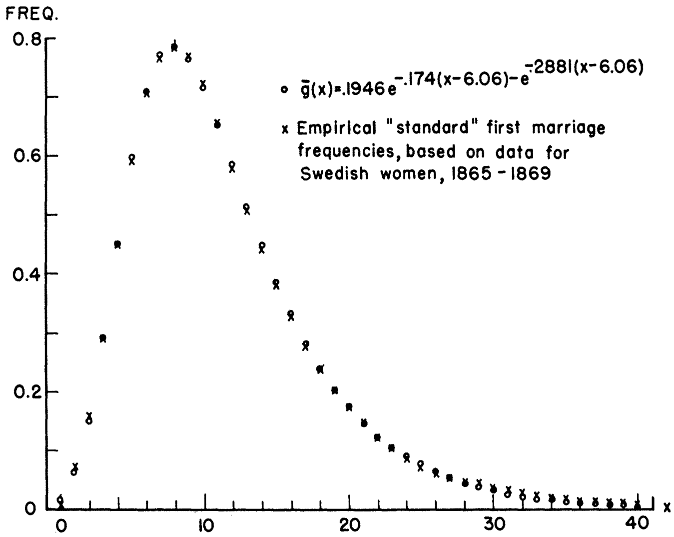
On the other hand, the analogous expansion for the right-hand side of equation (1.2) is, expanding the exponential as a series,

$$r(x) = 0.174 \left\{ \textstyle\sum_{j=0}^{\infty} \dfrac{\Gamma(1)}{\Gamma(j+1)} e^{-0.309j(x-4.803)} \right\}^{-1}. \tag{1.9}$$

Comparing (1.8) and (1.9), we see that the expressions within the brackets are identical, if, and only if, $\alpha=0$, $\lambda=0.309$, and $\mu=4.803$. However the asymptotes of $\bar{r}(x)$ and $r(x)$ as $x\to\infty$, given by taking the terms with $j=0$ in (1.8) and (1.9), are α and 0.174, respectively, so we must have $\alpha=0.174$ if the risk functions are to agree at $x=\infty$. It follows that $\bar{r}(x)$ and $r(x)$ are essentially different, so that the distribution given by (1.2) cannot be represented as an infinite convolution of the form (1.6). Since, however, the procedure of successively removing exponential components does not, in practice, distinguish between the two distributions, the question arises: does the distribution given by equation (1.7) fit the data as well as that given by (1.2)?

It was found that if $\alpha=0.174$, $\lambda=0.288$ and $a=11.36$, the answer is in the affirmative. (These values were chosen as follows: 11.36 is the mean of the empirically based "standard" fertility schedule; 0.174 is the asymptotic value of the empirically calculated risk function; and 0.288 is the value of X that ensures agreement of $\bar{g}(x)$ with the "standard" curve in the neighborhood of the mode.) In fact, the frequency distribution implied by (1.2), and that expressed in (1.7), with these parameter values, differ by less than 0.001 over the whole range, so that there is no visible difference between the two when plotted on a fairly large scale. Thus we may effectively replace the "standard" frequency distribution of first marriage rates by $\bar{g}(x)$, given by equation (1.7) with the above values of α, λ and a. This expression has the advantage of being a simple, closed form, frequency function. The fit of this distribution to the standard first marriage frequencies is shown in Figure A. (The "standard" curve has its origin at

A. COMPARISON OF FIRST MARRIAGE FREQUENCIES, FROM SWEDISH DATA AND CALCULATED FROM EQUATION (1.7)



about the earliest age at which a consequential number of first marriages occur.)

2. THE STANDARD CURVE AS THE CONVOLUTION OF A NORMAL CURVE AND m EXPONENTIALLY DISTRIBUTED DELAYS

We turn now to possible behavioral implications of this model. The representation (1.6), when taken to the limit as $n \to \infty$, may be written as

$$\begin{aligned} \overline{X} &= a + \sum_{j=1}^{\infty}\left(Z_j - \frac{1}{\alpha + (j-1)\lambda}\right) \\ &= \sum_{j=1}^{m} Z_j + Y_m \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} Y_m = a &- \sum_{j=1}^{m} \frac{1}{\alpha + (j-1)\lambda} \\ &+ \sum_{j=m+1}^{\infty}\left(Z_j - \frac{1}{\alpha + (j-1)\lambda}\right). \end{aligned} \tag{2.2}$$

In other words, $\overline{X}$ is the convolution of m exponentially distributed components together with an additional component Y_m. Denoting the frequency distributions of Y_m and $\sum_1^m Z_j$ by $g_m(x)$ and $h_m(x)$, respectively, it may be shown (see appendix, Theorems 1 and 3) that

$$g_m(x) = \frac{\lambda}{\Gamma(m + \alpha/\lambda)} e^{-(\alpha+m\lambda)(x-\mu)-e^{-\lambda(x-\mu)}} \tag{2.3}$$

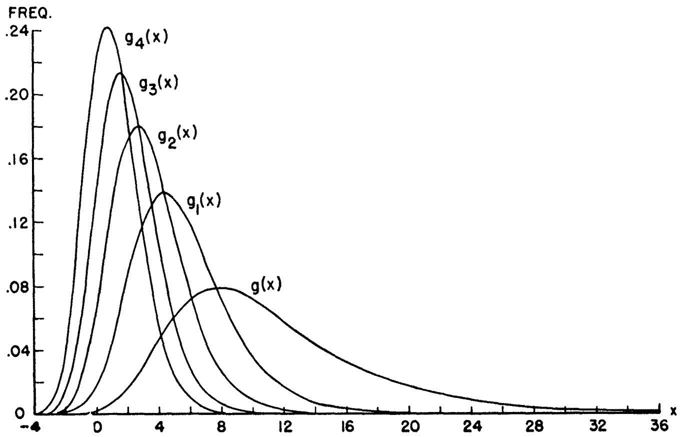
$$h_m(x) = \frac{\lambda\Gamma(m + \alpha/\lambda)}{\Gamma(\alpha/\lambda)(m-1)!}\left(1 - e^{-\lambda x}\right)^{m-1} e^{-\alpha x}. \tag{2.4}$$

The quantity Y_m may be regarded as the residual term after the removal of m exponentially distributed components. When its frequency distribution $g_m(x)$ is plotted for increasing values of m (see Figure B) an interesting pattern emerges: the residual distribution becomes less and less skewed as each component is removed, and seems to be approaching the shape of a normal distribution. This apparent convergence may indeed be proved mathematically. In the appendix (Theorem 4) it is shown that

$$\sqrt{m}\left\{Y_m - \left(a - \sum_{j=1}^{m} \frac{1}{\alpha + (j-1)\lambda}\right)\right\}$$

has a limiting distribution as $m \to \infty$ which is normal with mean zero and variance $1/\lambda^2$.

B. THE RESIDUAL FUNCTION $g_m(x)$



This fact suggests that the theoretical distribution $\overline{G}(x)$ (and thus, in view of the closeness of fit as evidenced by Figure A, the empirical first marriage distribution) could be approximated closely by the convolution of (a) a normal distribution and (b) a moderate number of exponential components with mean values in harmonic progression. In Table 1, $\bar{g}(x)$, calculated from equation (1.7) is compared with the convolution of a normal distribution (with appropriate mean and variance) and m exponentially distributed delays, for $m=1$, 2 and 3. The values of α, λ and a in each case are the same as those calculated in the previous fitting procedure, i.e., $\alpha=0.174$, $\lambda=0.288$, $a=11.36$. The fit with three delays is quite close; in fact, the fit of this convolution to the standard schedule of first marriages is essentially as good as the fit of $\bar{g}(x)$ itself. The absolute value of the area between the standard curve and $\bar{g}(x)$ is only 1.6 percent; between the standard curve and the convolution containing three exponentials, only 1.9 percent.

The similarity of these curves can be anticipated from comparison of the cumulants of $\bar{g}(x)$ with those of the convolution of $h_3(x)$ and a normal curve (with the same mean and variance as $g_3(x)$). The cumulants of a normal curve beyond the second are zero; in addition, the cumulant (of any order) of a convolution is the sum of the corresponding cumulants of the components. Since the rth cumulant of an exponential distribution se^{-sx} is $(r-1)!(1/s)^r$, it follows that the rth cumulant of

1. FIRST MARRIAGE FREQUENCIES, OBSERVED IN SWEDEN AND FITTED BY FOUR FUNCTIONS

x	Conv. of normal and n exponentials: n = 1	n = 2	n = 3	$\bar{g}(x)$	Standard schedule
0	.0055	.0034	.0026	.0018	.0
1	.0102	.0078	.0064	.0064	.0073
2	.0173	.0157	.0153	.0158	.0159
3	.0270	.0274	.0280	.0296	.0291
4	.0388	.0419	.0436	.0456	.0447
5	.0515	.0568	.0589	.0603	.0591
6	.0633	.0692	.0709	.0711	.0706
7	.0724	.0771	.0779	.0771	.0765
8	.0776	.0798	.0796	.0784	.0783
9	.0784	.0780	.0772	.0760	.0767
10	.0753	.0730	.0721	.0711	.0720
11	.0693	.0663	.0654	.0648	.0653
12	.0618	.0588	.0582	.0578	.0576
13	.0539	.0514	.0510	.0508	.0505
14	.0462	.0444	.0442	.0442	.0439
15	.0392	.0381	.0380	.0381	.0379
16	.0331	.0325	.0325	.0326	.0327
17	.0279	.0277	.0277	.0278	.0278
18	.0234	.0234	.0235	.0236	.0237
19	.0197	.0198	.0199	.0200	.0204
20	.0166	.0167	.0168	.0169	.0173
21	.0139	.0140	.0142	.0143	.0147
22	.0117	.0119	.0119	.0120	.0124
23	.0098	.0100	.0101	.0101	.0104
24	.0083	.0084	.0085	.0085	.0085
25	.0069	.0071	.0071	.0072	.0071
26	.0058	.0060	.0060	.0060	.0061
27	.0049	.0050	.0050	.0051	.0053
28	.0041	.0042	.0042	.0043	.0047
29	.0035	.0035	.0036	.0036	.0043
30	.0029	.0030	.0030	.0030	.0037

$h_m(x)$ is $(r-1)!\sum_1^m \{\alpha+(j-1)\lambda\}^{-r}$. The rth cumulant of $\bar{g}(x)$ is $(r-1)!\sum_1^\infty \{\alpha+(j-1)\lambda\}^{-r}$. Thus the convolution of $h_3(x)$ and a normal curve (with the same mean and variance as $g_3(x)$) has exactly the same mean and variance as $\bar{g}(x)$, but the third and higher cumulants are $(r-1)!\sum_1^3 \{\alpha+(j-1)\lambda\}^{-r}$ instead of $(r-1)! \cdot \sum_1^\infty \{\alpha+(j-1)\lambda\}^{-r}$, for $r=3, 4, \cdots$. It turns out that the third cumulants of $h_2(x)$ and $h_3(x)$ are 399.02 and 404.65, while that of $\bar{g}(x)$ is 408.96; for higher cumulants the relative differences are even less. Hence replacing $h_3(x)$ by a normal curve with the same mean and variance yields a distribution not significantly different from $\bar{g}(x)$.

It may be noted that if $h_1(x)$ were sufficiently close to a normal distribution, it would be possible to replace $\bar{g}(x)$ by the convolution of a normal distribution and just one exponential delay—the model suggested by Feeney. The normal component in this more exact model may still be regarded as the time to reach a marriageable age, while the exponential terms may be interpreted as the further delays before the state of marriage is finally reached.

We come, then, to an interpretation of the "standard" schedule of first marriage frequencies as the combination of (1) a normal distribution of attainment of marriageable age with a mean of a_0 (the origin of the "standard" curve or of $G(x)$) plus 2.12 years, and a standard deviation of 2.0 years, and (2) three exponentially distributed delays with average durations of 1.33 years, 2.16 years and 5.75 years (1/(0.174+0.576), 1/(0.174+0.288), and 1/(0.174)), respectively; the mean of the normal curve occurs at 11.36 less the sum of these three "delays," and the variance of the normal curve is the variance of $\bar{g}(x)$ less the sum of the variances of the three delays).

3. IMPLICATIONS FOR THE INTERPRETATION OF NUPTIALITY IN A SPECIFIC POPULATION

In any population (or more strictly speaking, any cohort) other than Swedish females of the 1860's, the distribution of first marriages is also approximated by the convolution of a normal distribution of the age at which women become marriageable and three exponentially distributed delays. The mean of the normal distribution differs from population to population, and the standard deviation of this distribution and the mean length of each of the three delays are the same multiple of the "standard" values given previously. In a cohort characterized by more rapid entry into marriage once the earliest age of marriage is attained, the standard deviation and the three delays are reduced. In the United States in the 1960's we have estimated a mean age of entry into marriageability of 15.6 years, and standard deviation of 1.52 years, and delays of average duration of 10.5, 17, and 45.5 months—a total of 73 months.

Since the distribution of first marriages is closely approximated by the convolution of a normal curve and three exponential distributions, it is natural to ask whether there is an identifiable action or event that corresponds to the normally distributed age of entry into a state of marriageability, and whether there are observable counterparts to the three stages that supposedly intervene between such entry and marriage itself. If so, despite the very similar form of the distribution of first marriage frequencies in different societies, the specific actions or events corresponding to becoming marriageable and to the subsequent prenuptial stages must not be the same from population to population because the arrangements preceding marriage are far from uniform. In contemporary populations of Western European origin, in which marriage is typically a ceremony that unites a couple who select each other on the basis of mutual preference, we may conjecture that the age of becoming marriageable is the age at which serious dating, or going steady begins; that the longest delay is the time between becoming marriageable and meeting (or starting to keep frequent company with) the eventual husband; and that the two shorter delays are the period between beginning to date the future husband and engagement, and between engagement and marriage.

The identifications of events prior to marriage with the components of $\bar{g}(x)$ are of course wholly conjectural until tested by empirical data. We have been able to make one such test using information on "The Choice of Spouse" from a sample survey of married couples in France conducted in 1959 [3]. One of the questions asked of the sample of currently married couples was how long before marriage the couple had known each other and gone together. The distribution of the duration of acquaintance is tabulated [3, p. 113] at intervals of less than six months to one year, one to two years, three to six years, and more than six years.

According to our model of nuptiality, this distribution of duration of acquaintance should be the convolution of two exponential distributions, $(\alpha+\lambda)e^{-(\alpha+\lambda)x}$ and $(\alpha+2\lambda)e^{-(\alpha+2\lambda)x}$. The convolution of these exponentials leads to the distribution

$$F(x) = 1 + \left(\frac{\alpha+\lambda}{\lambda}\right)e^{-(\alpha+2\lambda)x} - \left(\frac{\alpha+2\lambda}{\lambda}\right)e^{-(\alpha+\lambda)x}.$$

When $\bar{g}(x)$ is fitted to the "standard" schedule of first marriage frequencies based on data from Sweden, 1865–69, the value of α is 0.174, and the value of λ is 0.2881. For any other nuptiality experience α and λ must each be adjusted by a factor expressing the pace at which first marriages occur in the given population relative to the pace in Sweden in the 1860's. The age of the bride is tabulated by single years of age for the couples interviewed in the French survey [3, p. 52]. First marriage frequency is defined as the ratio of the number of first marriages in an interval relative to the number of woman-years lived in the interval. We calculated the schedule of frequencies by dividing the number of marriages reported at age x to $x+1$ by the number of women in the sample at age x and above. The standard deviation of this distribution is 4.60 years; the standard deviation of $\bar{g}(x)$ fitted to the Swedish data is 6.82 years; and the ratio is 0.698. In other words, nuptiality experience occurring in one year in Sweden of the 1860's occurred in about 0.7 years for couples covered by the French survey. Thus $\alpha=0.174/0.698$ or 0.249, and $\lambda=0.2881/0.698$, or 0.413, and the convolution of $(\alpha+\lambda)e^{-(\alpha+\lambda)x}$ and $(\alpha+2\lambda)e^{-(\alpha+2\lambda)x}$ is

$$F(x) = 1 + 1.6042e^{-1.0748x} - 2.6042e^{-0.6619x}.$$

Table 2 shows the recorded values of the percent of couples with a duration of acquaintance berfore marriage of less than six months, one year, two years, three years, and six years, together with the values calculated for $F(x)$ at the same durations. The only substantial disagreement is in the residual (longer than six years); and

2. PERCENT OF COUPLES WHO MET NO MORE THAN X YEARS BEFORE MARRYING

x	Observed percent	Calculated percent
0.5	6	6.7
1.0	21	20.4
2.0	50	49.4
3.0	69	70.6
6.0	90	95.3

Source: [3, p. 113]

in the survey those beyond six years consist of couples who had "always" known each other, a response that makes the genuinely relevant period of going together before marriage hard to infer.

The agreement of the two columns in Table 2 can only be viewed as remarkable, given the wholly independent and largely theoretical basis of the second column. The figures in this column were calculated without any reference whatever to the responses given with regard to length of acquaintance. The rationale for the figures in the second column is:

a. the existence of a common pattern of first marriage frequencies in a wide variety of populations;
b. the construction of a "standard" schedule of first marriage frequencies by minor smoothing of data for Sweden, 1865–69;
c. the fitting of a curve $\bar{g}(x)$ to the standard schedule, and the recognition that $\bar{g}(x)$ differs trivially from the convolution of a normal curve and three exponentially distributed delays;
d. the estimation of the time scale of nuptiality for the French sample by calculation of the standard deviation of French first marriage frequencies; and
e. the convolution of the two shorter of the three postulated delays employing the parameters derived from Swedish data modified by the estimated scale factor.

The surprising agreement of theoretical and observed values gives some reason for supposing that there is, in fact, a behavioral basis for considering $\bar{g}(x)$ the convolution of a normal curve and a few exponentially distributed delays. We may calculate the mean age at first marriage of the French couples covered by the survey as 23.1 years and the standard deviation of age at first marriage as 4.60 years. If we accept the theoretical model, at least for this population, we may further infer that the mean interval from acquaintance to marriage was 2.45 years (with two component intervals of 1.52 and 0.93 years), that the mean age of entry into a state of marriageability was 16.6 years (standard deviation 1.89 years), and that the mean interval between attaining marriageability and meeting the future husband was 4.02 years.

APPENDIX

Theorem 1: If Z_j, $j=1, 2, \cdots, n$ are independent random variables, exponentially distributed with

$$E[Z_j] = \{\alpha + \lambda(j-1)\}^{-1},$$

the frequency distribution of the convolution $\sum_{j=1}^{n} Z_j$ is

$$h_n(x) = \frac{\lambda\Gamma(n+\alpha/\lambda)}{\Gamma(\alpha/\lambda)(n-1)!}(1-e^{-\lambda x})^{n-1}e^{-\alpha x}.$$

Proof: (Induction). The theorem is obviously true for $n=1$, for using the identity $\Gamma(1+\alpha/\lambda)=(\alpha/\lambda)\Gamma(\alpha/\lambda)$,

$$h_1(x) = \alpha e^{-\alpha x}.$$

Suppose now that the theorem is true for some particular value of n, say p. Then by the convolution formula

$$h_{p+1}(x) = \int_0^x h_p(z)(\alpha+\lambda p)e^{-(\alpha+\lambda p)(x-z)}dz,$$

$$= e^{-(\alpha+\lambda p)x}\int_0^x \frac{\lambda\Gamma(p+\alpha/\lambda)(\alpha+\lambda p)}{\Gamma(\alpha/\lambda)(p-1)!}(1-e^{-\lambda z})^{p-1}e^{\lambda pz}dz.$$

Since $\Gamma(p+\alpha/\lambda)(p+\alpha/\lambda)=\Gamma(p+1+\alpha/\lambda)$, this becomes

$$\begin{aligned} h_{p+1}(x) &= e^{-(\alpha+\lambda p)x}\frac{\lambda^2\Gamma(p+1+\alpha/\lambda)}{\Gamma(\alpha/\lambda)(p-1)!}\int_0^x e^{\lambda z}(e^{\lambda z}-1)^{p-1}dz,\\ &= e^{-(\alpha+\lambda p)x}\frac{\lambda^2\Gamma(p+1+\alpha/\lambda)}{\Gamma(\alpha/\lambda)(p-1)!}\frac{1}{\lambda p}(e^{\lambda x}-1)^p,\\ &= \frac{\lambda\Gamma(p+1+\alpha/\lambda)}{\Gamma(\alpha/\lambda)p!}(1-e^{-\lambda x})^p e^{-\alpha x},\\ &= h_{p+1}(x). \end{aligned}$$

Theorem 2: The limiting distribution of

$$X_n = a + \sum_{j=1}^n\left\{Z_j - \frac{1}{\alpha+\lambda(j-1)}\right\}$$

as $n\to\infty$ has frequency function

$$\bar{g}(x) = \frac{\lambda}{\Gamma(\alpha/\lambda)}e^{-\alpha(x-\mu)-e^{-\lambda(x-\mu)}}$$

where $\mu=a+(1/\lambda)\psi(\alpha/\lambda)$.

Proof: Using Theorem 1, we have

$$\begin{aligned} \Pr\left[\sum_{j=1}^n Z_j < x\right] &= 1-\int_x^\infty h_n(z)dz\\ &= 1-\frac{\lambda\Gamma(n+\alpha/\lambda)}{\Gamma(\alpha/\lambda)(n-1)!}\int_x^\infty(1-e^{-\lambda z})^{n-1}e^{-\alpha z}dz\\ &= 1-\frac{\Gamma(n+\alpha/\lambda)}{\Gamma(\alpha/\lambda)(n-1)!}\\ &\quad\cdot\int_0^{\exp(-\lambda x)}(1-v)^{n-1}v^{(\alpha/\lambda)-1}dv, \end{aligned}$$

making the change of variable $v=\exp(-\lambda z)$. It follows that

$$\begin{aligned} \Pr[X_n<x] &= \Pr\left[\sum_{j=1}^n Z_j < x + \sum_{j=1}^n\frac{1}{\alpha+\lambda(j-1)} - a\right]\\ &= 1-\frac{\Gamma(n+\alpha/\lambda)}{\Gamma(\alpha/\lambda)(n-1)!}\int_0^{\zeta(x)}(1-v)^{n-1}v^{(\alpha/\lambda)-1}dv, \end{aligned}$$

where $\zeta(x)=\exp[-\lambda(x-a)-\sum(j-1+\alpha/\lambda)^{-1}]$. The following asymptotic formulas for the gamma and digamma function are now needed (see, [1 pp. 257–9]):

$$\begin{gathered} \Gamma(n+z)/(n-1)! = n^z + 0(n^{-1});\\ \sum_{j=1}^n(j-1+z)^{-1} = \psi(n+z)-\psi(z);\\ \psi(n+z)\sim\ln n + 0(n^{-1}). \end{gathered}$$

Using these formulas, the expression for $\Pr[X_n<x]$ becomes, asymptotically

$$\Pr[X_n<x]\to 1-\frac{n^{\alpha/\lambda}}{\Gamma(\alpha/\lambda)}\int_0^{(1/n)e^{-\lambda(x-\mu)}}(1-v)^{n-1}v^{(\alpha/\lambda)-1}dv,$$

where $\mu=a+(1/\lambda)\psi(\alpha/\lambda)$. Making the change of variable $z=nv$, and using the result

$$(1-z/n)^{n-1}\to e^{-z},$$

we derive

$$\Pr[X_n<x]\to 1-\frac{1}{\Gamma(\alpha/\lambda)}\int_0^{e^{-\lambda(x-\mu)}}z^{(\alpha/\lambda)-1}e^{-z}dz.$$

The limiting frequency function of X_n is now obtained, by differentiation, as

$$\bar{g}(x) = \frac{\lambda}{\Gamma(\alpha/\lambda)}e^{-\alpha(x-\mu)-e^{-\lambda(x-\mu)}}.$$

Theorem 3: The frequency distribution $\bar{g}(x)$ may be written as the convolution of $g_m(x)$ and $h_m(x)$, where

$$g_m(x) = \frac{\lambda}{\Gamma(m+\alpha/\lambda)}e^{-(\alpha+m\lambda)(x-\mu)-e^{-\lambda(x-\mu)}}.$$

Proof: Using formula (1.3), if an exponentially distributed component with mean α is removed, the residual frequency is

$$\begin{aligned} g_1(x) &= \bar{g}(x)+\bar{g}'(x)/\alpha\\ &= \frac{\lambda}{\Gamma(\alpha/\lambda)}\left\{1-\left(1-\frac{\lambda}{\alpha}e^{-\lambda(x-\mu)}\right)\right\}e^{-\alpha(x-\mu)-e^{-\lambda(x-\mu)}}\\ &= \frac{\lambda}{\Gamma(1+\alpha/\lambda)}e^{-(\alpha+\lambda)(x-\mu)-e^{-\lambda(x-\mu)}}. \end{aligned}$$

Thus the theorem is true when $n=1$, and by induction, in general.

Theorem 4: Suppose $\bar{X}$ is a random variable with frequency distribution given by $\bar{g}(x)$. Then $\bar{X}$ may be written as the convolution of X_m and W_m, where X_m is defined in equation (1.6) and $\sqrt{m}\,W_m$ is asymptotically normally distributed as $m\to\infty$ with mean 0 and variance $1/\lambda^2$.

Proof: We may write W_m as the infinite convolution

$$W_m = \sum_{j=m+1}^\infty\left\{Z_j-\frac{1}{\alpha+(j-1)\lambda}\right\}.$$

The cumulant generating function of $\sqrt{m}\,W_m$ is thus

$$\begin{aligned} \log E[e^{i\theta\sqrt{m}\,W_m}] &= \log E[e^{i\theta\sqrt{m}\sum_{j=m+1}^\infty\{Z_j-[\alpha+(j-1)\lambda]^{-1}\}}]\\ &= \log\prod_{j=m+1}^\infty E[e^{i\theta\sqrt{m}\{Z_j-[\alpha+(j-1)\lambda]^{-1}\}}]\\ &= \sum_{j=m+1}^\infty\left\{\log\left\{\frac{1}{1-i\theta\sqrt{m}[\alpha+(j-1)\lambda]^{-1}}\right\}\right.\\ &\quad\left.-\frac{i\theta\sqrt{m}}{\alpha+(j-1)\lambda}\right\}, \end{aligned}$$

since if Z is exponentially distributed with mean μ, $E[e^{i\theta Z}]=1/(1-i\theta\mu)$. Expanding the logarithm in a Taylor series we find

$$\begin{aligned} &\log E[e^{i\theta\sqrt{m}\,W_m}]\\ &= \sum_{j=m+1}^\infty\left[\frac{i\theta\sqrt{m}}{\alpha+(j-1)\lambda}-\frac{\theta^2 m}{2\{\alpha+(j-1)\lambda\}^2}\right.\\ &\quad\left.-\frac{i\theta^3m^{3/2}}{3\{\alpha+(j-1)\lambda\}^3}+\cdots-\frac{i\theta\sqrt{m}}{\alpha+(j-1)\lambda}\right]\\ &= -\frac{1}{2}\theta^2\left\{m\sum_{j=m}^\infty\frac{1}{(\alpha+j\lambda)^2}+\frac{2i\theta}{3}m^{3/2}\sum_{j=m}^\infty\frac{1}{(\alpha+j\lambda)^3}+\cdots\right\}. \end{aligned}$$

Now

$$\sum_{j=m}^\infty\frac{1}{(\alpha+j\lambda)^k}\sim\frac{1}{k-1}\left(\frac{1}{\lambda}\right)^k m^{1-k}\quad\text{as } m\to\infty,$$

so

$$\log E[e^{i\theta\sqrt{m}\,W_m}]\sim-\tfrac{1}{2}\theta^2\lambda^{-2}+0(m^{-\frac{1}{2}})\to-\tfrac{1}{2}\theta^2\lambda^{-2}\quad\text{as } m\to\infty.$$

It follows that $\sqrt{m}\,W_m$ is asymptotically normal with mean zero and variance λ^{-2}.

Theorem 5: The risk function associated with the frequency function $\bar{g}(x)$ possesses the expansion

$$\bar{r}(x) = \alpha\left\{\sum_{j=0}^\infty\frac{\Gamma(\alpha/\lambda)}{\Gamma(j+\alpha/\lambda)}e^{-\lambda j(x-\mu)}\right\}^{-1}.$$

Proof: By definition

$$\begin{aligned} 1/\bar{r}(x) &= \int_x^\infty\bar{g}(z)dz/\bar{g}(x),\\ &= \int_{x-\mu}^\infty e^{-\alpha z-e^{-\lambda z}}dz/e^{-\alpha(x-\mu)-e^{-\lambda(x-\mu)}} \end{aligned}$$

using Theorem 2. Integrating by parts in the numerator, this becomes

$$\begin{aligned} 1/\bar{r}(x) &= \frac{-\frac{1}{\alpha}e^{-\alpha z-e^{-\lambda z}}\Big|_{x-\mu}^\infty+\frac{\lambda}{\alpha}\int_{x-\mu}^\infty e^{-(\alpha+\lambda)z-e^{-\lambda z}}dz}{e^{-\alpha(x-\mu)-e^{-\lambda(x-\mu)}}}\\ &= \frac{1}{\alpha}\{1+\lambda e^{-\lambda(x-\mu)}/\bar{r}_1(x)\}, \end{aligned}$$

where $\bar{r}_1(x)$ is the risk function of the frequency distribution $h_1(x)$, defined in Theorem 3. Similarly

$$1/\bar{r}_1(x) = \frac{1}{\alpha+\lambda}\{1 + \lambda e^{-\lambda(x-\mu)}/\bar{r}_2(x)\},$$

where $\bar{r}_2(x)$ is the risk function corresponding to $h_2(x)$, and in general

$$1/\bar{r}_n(x) = \frac{1}{\alpha+n\lambda}\{1 + \lambda e^{-\lambda(x-\mu)}/\bar{r}_{n+1}(x)\},$$

where $\bar{r}_n(x)$ goes with $h_n(x)$. Thus we obtain the expansion

$$1/\bar{r}_n(x) = \frac{1}{\alpha}\left\{1 + \frac{\lambda}{\alpha+\lambda}e^{-\lambda(x-\mu)} + \frac{\lambda^2}{(\alpha+\lambda)(\alpha+2\lambda)}e^{-2\lambda(x-\mu)} + \frac{\lambda^3}{(\alpha+\lambda)(\alpha+2\lambda)(\alpha+3\lambda)}e^{-3\lambda(x-\mu)} + \cdots\right\}$$

$$= \frac{1}{\alpha}\left\{\sum_{j=0}^{\infty}\frac{\Gamma(1+\alpha/\lambda)}{\Gamma(j+1+\alpha/\lambda)}e^{-\lambda j(x-\mu)}\right\}.$$

[*Received December 1971. Revised March 1972.*]

REFERENCES

[1] Abramowitz, M. and Stegun, I.A., eds., *Handbook of Mathematical Functions*, U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., 1964.

[2] Coale, A.J., "Age Patterns of Marriage," *Population Studies*, 25, No. 2 (July 1971), 193–214.

[3] Girard, A., "Le Choix du Conjoint," Institut National d'Etudes Demographiques, Travaux et Documents, Presses Universitaires de France, 1964.

[4] Kimball, B.F., "Sufficient Statistical Estimation Functions for the Parameters of the Distribution of Maximum Values," *Annals of Mathematical Statistics*, 17 (1946), 299–317.