15

The Demographic Theory of Kinship

Figure 15.1. Expected number of female kin alive when Ego (hatched circle) is aged 40, based on birth and death rates of the United States, 1965.

This chapter will extract information on kinship numbers from the age-specific rates of birth and death of a population. A fixed set of age-specific rates implies the probability that a girl aged a has a living mother and great-grandmother, as well as her expected number of daughters, sisters, aunts, nieces, and cousins. Certain assumptions are required to draw the implications, some stronger than others. The formulae of this chapter in effect set up a genealogical table, giving not the names of incumbents in the several positions but the expected number of incumbents. Those of Figure 15.1 are based on birth and death rates of the United States in 1965, whose net reproduction rate R_0 was 1.395 and $\overset{o}{e}_0$ was 73.829, all for females. They offer a different kind of knowledge from what would be provided by a kinship census.

Like earlier chapters, this one supposes a population generated by birth and death with overlapping generations. (Generations do not overlap in annual plants, where all the parents have disappeared before the children come to life. This circumstance requires population models that account for processes both within and between years; see MPM Section 13.2.) The considerable longevity of human beings, as well as other large mammals and long-lived birds, after the birth of their offspring produces simultaneously living kin of many kinds—not only parents and children, but also grandparents, nephews, and cousins. Human beings produce most of their offspring in births of discrete individuals, but this is not recognized in the present analysis, which supposes $m(x)\,dx$ of a daughter in each dx of maternal age. For certain kin this introduces serious qualifications, specified in Section 15.3. Development of this field is due to Lotka (1931), Burch
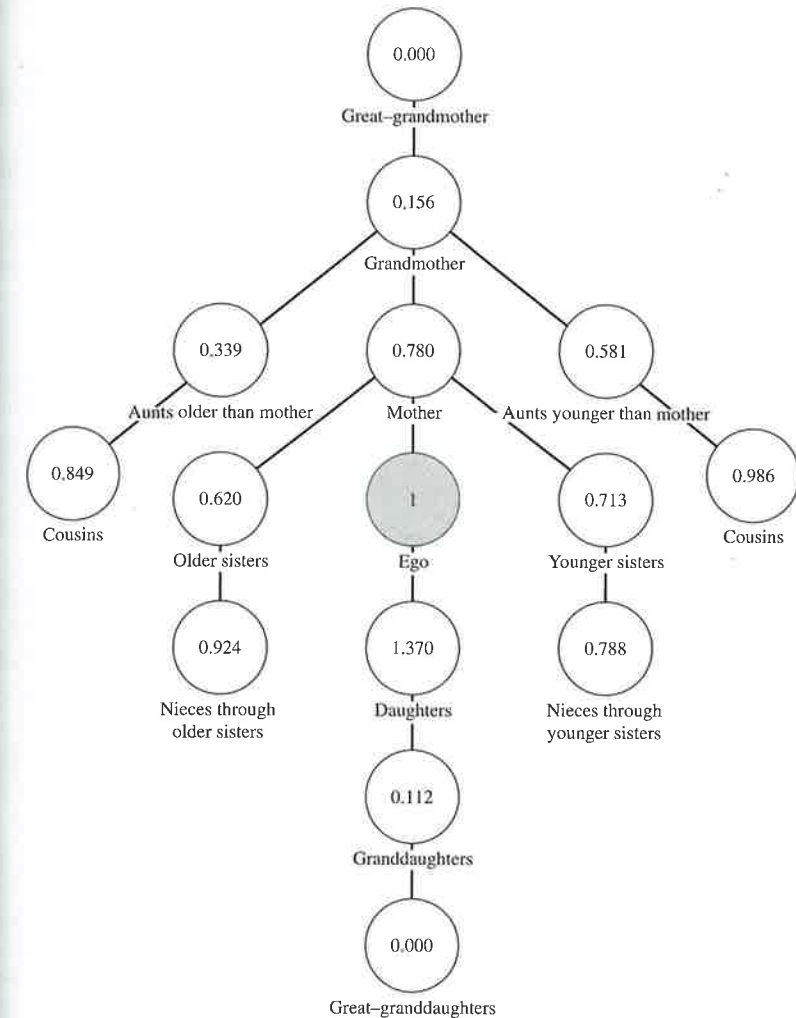
(1970), Coale (1965), Goodman, Keyfitz, and Pullum (1974) and Le Bras (1973).

Explicit recognition of the several degrees of living and dead kin varies from one culture to another, and indeed from one family to another. We disregard here cultural, social, and psychological definitions and deal with numerical relations among average numbers of biological kin as they are determined by birth and death rates. To avoid undue complication, all of the following discussion recognizes female kin only.

15.1   Probability of Living Ancestors

Deterministic models concern both population numbers and probabilities. The two perspectives are at least on the surface distinct.

Counting Method. A large population can be seen as developing according to given rules, and in effect we can make counts of the number of individuals having the kin relations of interest. This is an extension of the notion that l_0 is the number of births occurring at one moment and $l(x)$ is the number of those surviving x years later, the cohort conception of the life table $l(x)$ referred to in Section 2.1. (But we still keep $l_0 = 1$.)

Probability Method. We can start by thinking of an individual and work out probabilities and expected values for his various kin. This is an extension of the interpretation of the life table $l(x)$ as the probability that a child just born will survive for x years.

15.1.1   Living Mother by the Counting Method

Our first approach to finding the probability that a girl aged a has a living mother is to see how a population would have developed starting from B girl children born at a moment $a + x$ years ago. At age x of this maternal generation cohort the survivors were $Bl(x)$, and during the interval x to $x + dx$ they could be expected to give birth to $Bl(x)m(x)l(a)\,dx$ daughters who would live to age a. Of the mothers who gave birth at age x a fraction $l(x + a)/l(x)$ would survive over an additional a years; hence the number of living mothers must be $Bl(x)m(x)\,dx[l(x + a)/l(x)]l(a)$. A woman is counted once for each birth that survives.

All this concerns one cohort of the mother generation. But we seek the probability that a girl aged a, standing before us, has a living mother, without any knowledge of which cohort her mother belonged to, or indeed any knowledge other than the regime of mortality and fertility supposed to apply to all generations and at all times. If births as a function of time are $B(t)$, and the present is time t, girls born $x + a$ years ago numbered $B(t - a - x)$, and the number of living mothers (counted once for each daughter) of all cohorts who gave birth to girls now alive and a years of age is the integral

$$\int_\alpha^\beta B(t - a - x)l(x)m(x)\frac{l(x + a)}{l(x)}l(a)\,dx. \qquad (15.1.1)$$

For the same birth function $B(t)$ the number of daughters, that is, girls born a years ago and surviving to the present, is

$$B(t - a)l(a). \qquad (15.1.2)$$

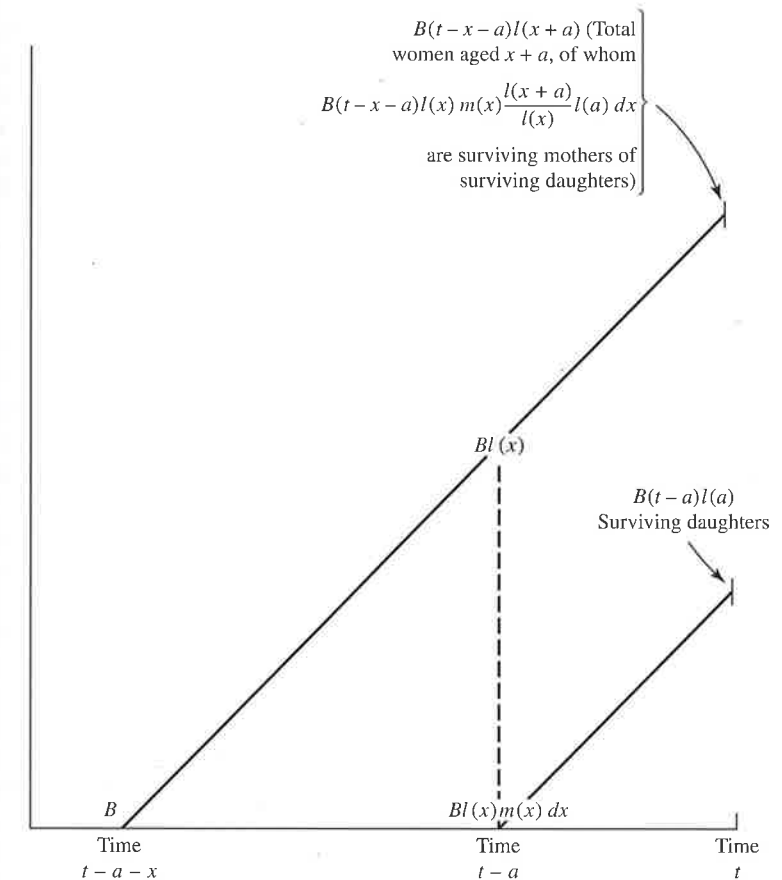Hence the average number of mothers per daughter at time t is the ratio of (15.1.1) to (15.1.2).

Figure 15.2. Cohort of mothers giving birth at age x, and of daughters born at age x of mother.

Now, if the age-specific rates of birth and death have been in effect for a long period of time, the births will be growing exponentially, say according to the curve $B_0 e^{rt}$, where r is determined by (6.1.2). Entering this for the birth function in each of (15.1.1) and (15.1.2), taking the ratio of the first to the second, and canceling $l(a)B_0 e^{r(t-a)}$ from numerator and denominator gives

$$M_1(a) = \int_\alpha^\beta \frac{l(x + a)}{l(x)}e^{-rx}l(x)m(x)\,dx \qquad (15.1.3)$$

for the probability that a girl aged a has a living mother under the given regime of mortality and fertility (Lotka 1931).

All of the problems of this chapter can be solved in this way, by following cohorts through time, then in effect taking a census at a certain moment, and finding the ratio of one census aggregate to another. This does not of

course constitute a census of the real population, which would show the result of changing death rates over time, but is at best a simplified abstract argument devised by analogy to counting population by means of a census and births by means of registrations. Made possible by the deterministic assumption of Section 2.1, it requires no explicit considerations of probability, yet it is both unnecessarily complicated and unnecessarily restricted.

15.1.2 Living Mother by Conditional Probability

Alternatively, the life table $l(x)$ column is taken, not as a cohort, but as the probability of living to age x for a child just born, on the regime of mortality assumed, with $l_0 = 1$. The corresponding approach will provide a result for $M_1(a)$ identical with (15.1.3).

In seeking the probability that a girl chosen at random out of a population with birth rates $m(x)$ and survivorship $l(x)$ has a living mother, we first note that the conditional case is easily solved. If the mother was a known x years old when she gave birth to the girl, then the chance that the mother is alive a years later must be $l(x+a)/l(x)$. This probability, conditional on the mother's age at bearing the girl having been x, is the first part of the solution.

It remains to remove the condition, which is not part of the problem; we do not care about the age x of the mother at childbearing. To eliminate the condition we average over all ages x, giving each x a weight proportional to the number of births occurring at that age of mother under the regime in question. The number of the stable population of ages x to $x + dx$ per current birth is $e^{-rx}l(x)\,dx$, from Euler's argument of Section 5.1, and the births to this fraction are $e^{-rx}l(x)m(x)\,dx$, still taken per one current birth. The last expression is the fraction of births occurring to women aged x to $x + dx$; once again, its total over all x is unity by the (6.1.2) defining r. Hence the unconditional probability that the mother of the girl in question is still alive is obtained by multiplying $e^{-rx}l(x)m(x)\,dx$ by the survivorship $l(x+a)/l(x)$ and totaling over all x.

This is the same expression, derived more compactly, that we obtained as (15.1.3). Because of its compactness the probability method will be preferred in what follows. Note that (15.1.3) is more general than appeared in our interpretation. The derivation did not require the fact that the girl born a years ago is still alive, but consisted in finding the probability of survival of the mother a years after a random birth. Whether the girl born survived does not affect the value of $M_1(a)$, given independent regimes.

15.1.3 Probability of Living Grandmother

Now think of the grandmother of the girl aged a; we once again provisionally take the latter as having been born at age x of her mother. The grandmother

has to live $x + a$ years after the birth of her daughter (the mother of our girl aged a) to be alive now. The chance of a woman living $x + a$ years after the birth of her daughter as calculated above is $M_1(x + a)$. This is now to be seen as the (conditional) probability that the grandmother of the girl aged a is alive, given that the girl aged a was born when her mother was x years old. To obtain the unconditional probability we again multiply by $e^{-rx}l(x)m(x)$ and integrate out x:

$$M_2(a) = \int_\alpha^\beta M_1(x + a)e^{-rx}l(x)m(x)\,dx. \tag{15.1.4}$$

Note that (15.1.4) again does not require the survival of the mother of the girl aged a. It is merely the probability that the grandmother of a randomly selected girl birth of a years ago is now alive.

The procedure supposes not only that the regime is fixed, but also that successive generations are independently subject to the given mortality and fertility. Independence of fertility between generations means, for example, that there is no tendency for daughters born to young mothers themselves to give birth at a young age, and it also disregards social class and other differences in mortality and fertility within heterogeneous populations. This is in addition to requiring birth and death to be independent, both in the same and in different generations; it excludes the possibility that some families have both high birth rates and high death rates. Finally, the chance of a woman having a child at ages x to $x + dx$ is taken as $m(x)\,dx$, whether or not she had a child immediately before this. These several unrealistic features of the model seem to have only a small effect on the probability of living ancestors and descendants, though they are important for the number of sisters, aunts, and nieces.

Once we have $M_2(a)$, the chance that a girl aged a has a living grandmother, we can similarly use it to find the chance of a living great-grandmother. Again suppose that the mother of the girl aged a was x years old at the time of childbearing; the question whether a girl aged a has a living great-grandmother is the probability that the grandmother of the mother, born $x + a$ years ago, is still alive. That the grandmother survives $x + a$ years after her granddaughter was born has probability $M_2(x + a)$, so the unconditional probability that a girl aged a has a living great-grandmother is

$$M_3(a) = \int_\alpha^\beta M_2(x + a)e^{-rx}l(x)m(x)\,dx. \tag{15.1.5}$$

Once more, this does not require the mother or grandmother of the girl aged a to be alive. The argument may be readily extended to even more remote progenitors (Goodman et al. 1974).

15.1.4 Numerical Examples

These and the other formulae of this chapter have been programmed by Tom Pullum, and Table 15.1 shows his results for three countries—strictly speaking, for three regimes of mortality and fertility—the United States, 1967; Venezuela, 1965; and Madagascar, 1966. The first two resemble each other in mortality and the last two in fertility, as the following standardized rates per thousand population having the United States, 1960, age distribution show:

Country and year	Birth rate	Death rate
United States, 1967	16.71	9.12
Venezuela, 1965	41.82	10.97
Madagascar, 1966	44.48	29.10

Thus Venezuela and Madagascar both have about 2.5 times the fertility of the United States, and Madagascar has about 3 times the mortality of the United States and Venezuela. We will later seek a more precise way of connecting the input mortality and fertility with the output kinship probabilities, but the present comparison is suggestive.

Table 15.1 shows, for example, that the chance that a woman aged 20 has a living mother is about 0.96 for the United States, 0.93 for Venezuela, and 0.71 for Madagascar. The complements of these numbers, 0.04, 0.07, and 0.29, are the probability of orphanhood on the mother's side. It was in an effort to see how serious was the problem of orphanhood that Lotka (1931) first developed (15.1.3). The greater difference between Venezuela and Madagascar than between the United States and Venezuela is to be expected; the chance of having living ancestors depends much more on mortality than on fertility rates. Insofar as fertility affects orphanhood, it is through the age of childbearing rather than through the number of children born, as will appear in Section 15.6.

15.1.5 Stable Results Versus a Kinship Census

These formulae and the numbers of Table 15.1 have been worked out for the specified regimes of mortality and fertility, taken as fixed through time and the same in all generations. They are meant to answer the question: what probability of having a living mother, grandmother, and so on does the given schedule of birth and death rates imply?

The fraction of women aged 20 in the United States having living mothers as ascertained by a survey or census would disagree with the result of calculation by (15.1.3) for several reasons: changing mortality and fertility over the preceding years, presence of immigrants from countries with different regimes, misstatement of age in the survey and in the vital statistics on which our calculations are based, or failure of the various independence

Table 15.1. Probability of living mother, grandmother, great-grandmother, and great-great-grandmother, for a female aged $a = 0, 20, 40, 60$, based on mortality and fertility regimes of the United States, Venezuela, and Madagascar

Ancestor and country	Age a			
	0	20	40	60
Living mother $M_1(a)$				
United States, 1967	1.000	0.959	0.785	0.298
Venezuela, 1965	1.000	0.932	0.707	0.223
Madagascar, 1966	1.000	0.713	0.386	0.061
Living grandmother $M_2(a)$				
United States, 1967	0.919	0.653	0.165	0.000
Venezuela, 1965	0.867	0.553	0.112	0.000
Madagascar, 1966	0.600	0.256	0.032	0.000
Living great-grandmother $M_3(a)$				
United States, 1967	0.507	0.090	0.000	0.000
Venezuela, 1965	0.397	0.058	0.000	0.000
Madagascar, 1966	0.164	0.017	0.000	0.000
Living great-great-grandmother $M_4(a)$				
United States, 1967	0.049	0.000	0.000	0.000
Venezuela, 1965	0.030	0.000	0.000	0.000
Madagascar, 1966	0.009	0.000	0.000	0.000

assumptions. Probability of living grandmother $M_2(a)$ requires independence of two generations, and $M_3(a)$ of three generations. We know that longevity runs in families, as well as being different for social classes, and experimenting would be required to find the effect of the mortality correlation between generations. All the formulae would become much more complicated if they took account of such departures from the assumptions of fixity and independence of the vital rates. One can only repeat that the kinship implication of fixed and independent rates constitutes a different kind of knowledge from a kinship census.

15.1.6 An Approximation

Insofar as the net maternity function is concentrated close to the mean age of childbearing κ, the quantity $l(\kappa + a)/l(\kappa)$ ought to be an approximation to our $M_1(a)$. For United States, 1967, females with $a = 20$ this would be $l(46.281)/l(26.281)$. The life table shows $l(x + 20)/l(x)$ for $x = 25$ as 0.96724, and for $x = 30$ as 0.95122; straight-line interpolation between these values gives $l(46.281)/l(26.281) = 0.963$. This compares with the more precisely calculated $M_1(20) = 0.959$ of Table 15.1; the approximation $l(\kappa + 20)/l(\kappa)$ is slightly high, because the curve of $l(x)$ is at this point concave below.

In the same way we would expect that the chance of a living grandmother would be something like $l(2\kappa + a)/l(\kappa)$. For United States women aged 20 the value is

$$\frac{l(2\kappa + 20)}{l(\kappa)} = \frac{l(72.562)}{l(26.281)} = \frac{0.65431}{0.96734} = 0.676,$$

against the $M_2(20) = 0.653$ in Table 15.1, or about 4 percent high. The greater error for grandmothers than for mothers is due to greater variation in age for the former.

Most of the difference between the crude and the correct estimate is accounted for by the concavity of the survivorship curve, along with the variance of ages of childbearing. To establish this, expand the ratio $l(x + a)/l(x)$ in $M_1(a)$ of (15.1.3) around κ by Taylor's theorem, writing the derivative of $l(x + a)/l(x)$ at $x = \kappa$ as $[l(\kappa + a)/l(\kappa)]'$, and so on:

$$\frac{l(x + a)}{l(x)} = \frac{l(\kappa + a)}{l(\kappa)} + (x - \kappa)\left[\frac{l(\kappa + a)}{l(\kappa)}\right]'$$
$$+ \frac{(x - \kappa)^2}{2!}\left[\frac{l(\kappa + a)}{k(\kappa)}\right]'' + \cdots;$$

(15.1.6)

integrate over x after multiplying by $e^{-rx}l(x)m(x)$, so that the term in $x - \kappa$ vanishes; call the variance of ages of mothers σ^2; and then factor out $l(\kappa + a)/l(\kappa)$ to obtain

$$M_1(a) \approx \frac{l(\kappa + a)}{l(\kappa)}\left[1 + \frac{\dfrac{\sigma^2}{2}\left(\dfrac{l(\kappa + a)}{l(\kappa)}\right)''}{\dfrac{l(\kappa + a)}{l(\kappa)}}\right].$$

(15.1.7)

The correction in square brackets equals 0.9957 for United States females of 1967, and produces $(0.9631)(0.9957) = 0.9590$ for $M_1(20)$ versus 0.9594 from the printout on which Table 15.1 was based.

For grandmothers the square bracket of (15.1.7) for the correction holds approximately, but with 2κ in place of κ in each numerator. The correction then is 0.963, and multiplying by $l(2\kappa + 20)/l(\kappa)$ gives 0.651, versus the true 0.653.

Presenting $M_1(a)$ in the form of (15.1.7) serves to show what feature of the net maternity function mainly determines the probability of a living grandmother: the mean age of childbearing κ acting through the factor $l(\kappa + a)/l(\kappa)$. The total of the net maternity function is absent, so to the (fairly close) approximation provided by (15.1.7) the level of fertility has little effect on probability of living ancestors. The variance of ages of childbearing has a small effect, whose amount depends mostly on the second derivative (i.e., the curvature) of $l(\kappa + a)/l(\kappa)$. The $l(x)$ curve is for the most part concave downward up to about age 65, so that its second derivative is negative. That the survival of mothers is negatively related to the variance

of ages of childbearing wherever $l(\kappa + a)/l(\kappa)$ is concave downward is the conclusion from (15.1.7), but the effect is small.

Further conclusions from (15.1.7) are drawn in Section 15.6.

15.2 Descendants

To illustrate how this chapter is an extension of standard demographic techniques the familiar net reproduction rate R_0 will be put into a form suitable for the counting of descendants.

If a cohort of girl births numbers B, born at time zero, the number of survivors to age x will be $Bl(x)$ on the deterministic model, and in the interval x to $x + dx$ these will bear $Bl(x)m(x)\,dx$ girl babies. The total number of daughters to which the cohort will give birth during its existence will be the integral of this last expression over x, and the average number of daughters will be this integral divided by the girl births B:

$$R_0 = \frac{\displaystyle\int_\alpha^\beta Bl(x)m(x)\,dx}{B}$$
$$= \int_\alpha^\beta l(x)m(x)\,dx.$$

(15.2.1)

This, the expected number of girl children to which a girl child will give birth under the regime $l(x)m(x)$, may be regarded as the ratio of one generation to the preceding (cf. Section 11.3.4).

For the number of granddaughters we use (15.2.1) in relation to each of the daughters. Thus, if the average number of girl babies at age x to the mother cohort is expected to be $Bl(x)m(x)\,dx$, and if each of these is expected to have R_0 births, we multiply by R_0 and again integrate up to age β, now to find R_0^2 granddaughters. Similarly the average number of great-granddaughters expected by a girl child will be R_0^3.

For incomplete generations the multiple integrals are formed in the same way, but β is no longer the upper limit. If we want only girl children that will have been born to a female by the time she is age a, where $\beta \geqslant a \geqslant \alpha$, the argument gives an expected

$$\int_\alpha^a l(x)m(x)\,dx$$

(15.2.2)

girl children. The expected number of granddaughters by the time the original cohort is aged $a \geqslant 2\alpha$ can be obtained by noting that $l(x)m(x)\,dx$ daughters would be expected to have been born when the woman aged a was $x \geqslant \alpha$ years of age; since each of these has up to age $a - x$ in which to

bear children, each would be expected to bear

$$\int_{\alpha}^{a-x} l(y)m(y)\,dy$$

daughters in turn. Thus the total number of granddaughters will be the product of this and (15.2.2) added through the possible x:

$$\int_{\alpha}^{a} l(x)m(x) \int_{\alpha}^{a-x} l(y)m(y)\,dy\,dx. \qquad (15.2.3)$$

All the above concern prospective descendants. Now we want to find the average number of girl children that have already been born to women aged a. The same cohort of B births has been followed down from time zero, and by the time it reaches age a there are $Bl(a)$ survivors. The total number of children that have been born up to this time is the integral $\int_{\alpha}^{a} Bl(x)m(x)\,dx$, but we do not want to include all of them in our average— we are concerned only with those born to mothers that lived at least to age a. The fraction of mothers that lived from age x to age a is $l(a)/l(x)$, and this is the fraction that we will take of the daughters born at age x of the mothers. Hence we have, for the total daughters expected to be born to the mothers that survived to age a,

$$\int_{\alpha}^{a} \frac{l(a)}{l(x)} Bl(x)m(x)\,dx = \int_{\alpha}^{a} Bl(a)m(x)\,dx,$$

and on dividing this by $Bl(a)$ mothers living at age a we have the average number of such daughters:

$$B_1(a) = \int_{\alpha}^{a} m(x)\,dx. \qquad (15.2.4)$$

This result is obvious when we consider that any woman alive at age $a > \alpha$ was also alive at age $x \leqslant a$, and that her probability of bearing a daughter in the interval x to $x+dx$ was $m(x)\,dx$. Her total female births to age a must therefore be given by (15.2.4), in which the $l(x)$ function does not enter. One minor difficulty is our assumption, unavoidable if existing fertility tables are to be used, that the $m(x)$ function is the same for women who survived to age a as for all women.

How many of the $\int_{\alpha}^{a} m(x)\,dx$ daughters will still be alive by the time the mother cohort is aged a? The fraction of daughters born at age x of their mothers that survive to age a of their mothers, or $a - x$ years, must be $l(a - x)$. Hence the number of daughters still alive of women aged a must on the average be

$$BL_1(a) = \int_{\alpha}^{a} m(x)l(a-x)\,dx. \qquad (15.2.5)$$

We can build on these results to find expected granddaughters already born for a woman aged a. Consider a daughter born at age x of the original

cohort. By the time the original cohort is aged a, the daughter herself will have averaged $\int_{\alpha}^{a-x} l(y)m(y)\,dy$ daughters as in (15.2.3). Integrating over all daughters born to the original cohort gives us the double integral

$$B_2(a) = \int_{\alpha}^{a} m(x) \int_{\alpha}^{a-x} l(y)m(y)\,dy\,dx \qquad (15.2.6)$$

for the average number of granddaughters so far born to women aged a.

To find the number of such granddaughters who are still alive we must multiply within the inner integral by the chance of survival through the years to the time when the original cohort is aged a (i.e., $a - x - y$ years), that is, by the factor $l(a - x - y)$. Great-granddaughters and further direct descendants raise no new problem.

15.3   Sisters and Aunts

To find the number of older sisters that a girl now aged a is expected to have, we again set the provisional condition that she was born at age x of her mother, when according to (15.2.4) her mother would be expected to have had $\int_{\alpha}^{x} m(y)\,dy$ children. The condition on x is removed as before by multiplying by $e^{-rx}l(x)m(x)\,dx$ and then integrating over x. Thus the expected number of older sisters, say S^{old}, still alive or not, is

$$S^{\text{old}} = \int_{\alpha}^{\beta} \int_{\alpha}^{x} m(y)e^{-rx}l(x)m(x)\,dy\,dx. \qquad (15.3.1)$$

Because a mother aged x was necessarily alive at ages y younger than x, we need no allowance for survivorship of the mother. Note that S^{old} does not depend on the age a; a girl can hardly acquire additional older sisters as she ages, and older sisters once born cannot decrease. Some of them, however, are no longer living; to find older sisters now alive we need only include the factor $l(a + x - y)$ in the inner integral of (15.3.1), and this makes the double integral a function of a, the age of the girl with whom the calculation starts.

Younger sisters ever born require an allowance for survivorship beyond the birth of the girl aged a, and their number must depend on a. If the girl aged a was born when her mother was x years old, the chance that the mother lived on to age $x + u$ and then bore a child is $[l(x + u)/l(x)]m(x + u)\,du$; integrating this over the possible values of u, then multiplying by $e^{-rx}l(x)m(x)\,dx$, and again integrating gives

$$S_a^{\text{young}} = \int_{\alpha}^{\beta} \int_{0}^{a} \left[\frac{l(x+u)}{l(x)} \right] m(x+u)\,du\,e^{-rx}l(x)m(x)\,dx \qquad (15.3.2)$$

for the number of younger sisters. Again this result may be interpreted more generally as the number of girls expected to be born to the mother of a random birth in the a years following.

15.3.1   A Paradox: The Average Girl Seems to Have Too Many Sisters

We find that the average number of sisters ever born according to the United States regime of mortality and fertility was 1.251 for a randomly chosen woman aged $a = 60$ (Table 15.2), while the average number of girls in the completed sisterhood or sorority was 1.26, the gross reproduction rate. The two numbers are very close, yet no allowance has been made for the woman aged a herself—one would have thought that the gross reproduction rate would equal $S^{old} + S_a^{young} + 1$. Have we inadvertently included the woman as her own sister? An examination of the argument leading to (15.3.1) and (15.3.2) will satisfy the reader that we have not. Nevertheless, how can the sisters of a randomly selected girl number 1.251, so that with her the sisterhood numbers 2.251, whereas the average number of girls obtained as the gross reproduction rate is only 1.26? (Goodman, Keyfitz, and Pullum 1975).

The answer lies in the manner of selection: the number of daughters of a randomly selected mother is decidedly smaller than the number of sisters plus one of a randomly selected girl. Consider the following hypothetical distribution of total number of daughters (i.e., of completed sororities):

Number of Daughters	Proportion of Cases
0	0.79
6	0.21

This would give the same gross reproduction rate of 1.26, but now all girls have five sisters, as needs no calculation to establish. The mean size of sororities when a *girl* is chosen at random is 6, very much larger than the average of 1.26 when a *family* is chosen at random. Whenever there is variation among mothers in childbearing, the estimate of the size of sorority from a sample of daughters will be larger than the estimate from a sample of families.

The difference between these two means can be expressed in terms of the variance. Suppose that of completed sisterhoods fraction f_0 is 0, f_1 is 1, and so on in the following scheme:

Table 15.2. Older and younger sisters ever born of a female aged a, birth and death rates of the United States, 1967

Sisters	a			
	0	20	40	60
Older S^{old}	0.610	0.610	0.610	0.610
Younger S_a^{young}	0.000	0.625	0.641	0.641
Total $S^{old} + S_a^{young}$	0.610	1.235	1.251	1.251

Number of daughters	Relative frequency
0	f_0
1	f_1
2	f_2
$\vdots$	$\vdots$

where $f_0 + f_1 + f_2 + \cdots = 1$. Then the gross reproduction rate G is

$$G = 0f_0 + 1f_1 + 2f_2 + \cdots,$$

the mean of the distribution, and the variance of the distribution is σ^2, where

$$\sigma^2 = 1^2 f_1 + 2^2 f_2 + \cdots - G^2.$$

The probability that a randomly chosen girl is a member of a sorority with zero members is 0, that she is a member of a sorority with one member is proportional to f_1, that she is a member of a sorority with two members is proportional to $2f_2$, and so on. This distribution is very different from the preceding one:

Size of sorority	Relative frequency of girls
0	0
1	$1f_1$
2	$2f_2$
3	$3f_3$
$\vdots$	$\vdots$

Now the total frequency is $0 + 1f_1 + 2f_2 + \cdots = G$, and the mean is

$$\frac{0 + (1)(1f_1) + (2)(2f_2) + (3)(3f_3) + \cdots}{0 + 1f_1 + 2f_2 + 3f_3 + \cdots} = \frac{\sigma^2 + G^2}{G}$$

$$= \frac{\sigma^2}{G} + G, \qquad (15.3.3)$$

without approximation. The number of sisters of a randomly selected girl equals this minus 1; i.e., $S = (\sigma^2/G) + G - 1$.

If the random selection were of families, the sorority would average G. The fact that the selection is of girls adds the term σ^2/G, the variance of the distribution of girls in families divided by the mean. Random selection of a daughter will always give a larger sorority than random selection of a family, as long as $\sigma^2 > 0$, that is, as long as there is any variation in family size. In the numerical calculation based on the formulae of this chapter the variance of the distribution is nearly equal to its mean, a relation characteristic of the Poisson distribution.

This does not arise from data but from the model. Recall that we made the probability of birth at any moment independent of births at all other moments. No heterogeneity was allowed for among women. That is why we ended with a Poisson distribution. In real populations some women are sterile and others have many children. Such heterogeneity among women negates the independence assumption and tends to make the variance greater than the mean. On the other hand, insofar as a two-child family is popular in birth-controlling populations, the variance is reduced.

15.3.2   Age Incidence of Childbearing Conditional on Birth of One Child

Within any homogeneous group expressions 15.3.1 and 15.3.2 are exact if $m(y)$ in the first and $m(x+u)$ in the second are conditional on the birth of a girl aged a at age x of the mother. Lacking data showing birth rates at the several ages for women who have had a birth at each age x, one is tempted to use overall values of $m(y)$ and $m(x+u)$. Insofar as the chance of another birth is zero in the months after a birth, a further impropriety is thereby added to the one discussed in detail above. The notch in the curve allowing for pregnancy and postpartum sterility would be compensated for in other parts of the range of mother's childbearing ages by the conditional $m(y)$ being higher than the average $m(y)$ for the entire population. Moreover the selection of mothers implicit in the fact that the ones we are concerned with are of proven fertility would probably add further to the conditional $m(y)$ in those ages y where it is nonzero.

The points raised above for sisters apply also to aunts, cousins, nieces, and other kin that are related through sisters.

15.3.3   Aunts

Sisters are aunts when seen from the viewpoint of the daughter of one of them. For the number of aunts that are older sisters of the mother the matter is simple; since S^{old} in (15.3.1) does not depend on the age of the girl, it must be invariant when taken in relation to the daughter of the girl. Thus S^{old} of (15.3.1) is also the expected number of aunts of a girl aged a who are older sisters of the mother. Under a fixed regime of birth and death, the same in both generations, a girl has the same expected number of older sisters and of maternal aunts older than her mother. This of course is not true of surviving sisters or aunts.

Aunts who are younger sisters of the mother, say A_a^{young}, are again first obtained conditionally on the mother having been of age x at the birth of girl now aged a (Figure 15.3). A mother who was then aged x must now be aged $x + a$, if she is alive. The number of younger sisters expected for a woman aged $x + a$ is S_{x+a}^{young}, entering the argument $x + a$ in (15.3.2). As
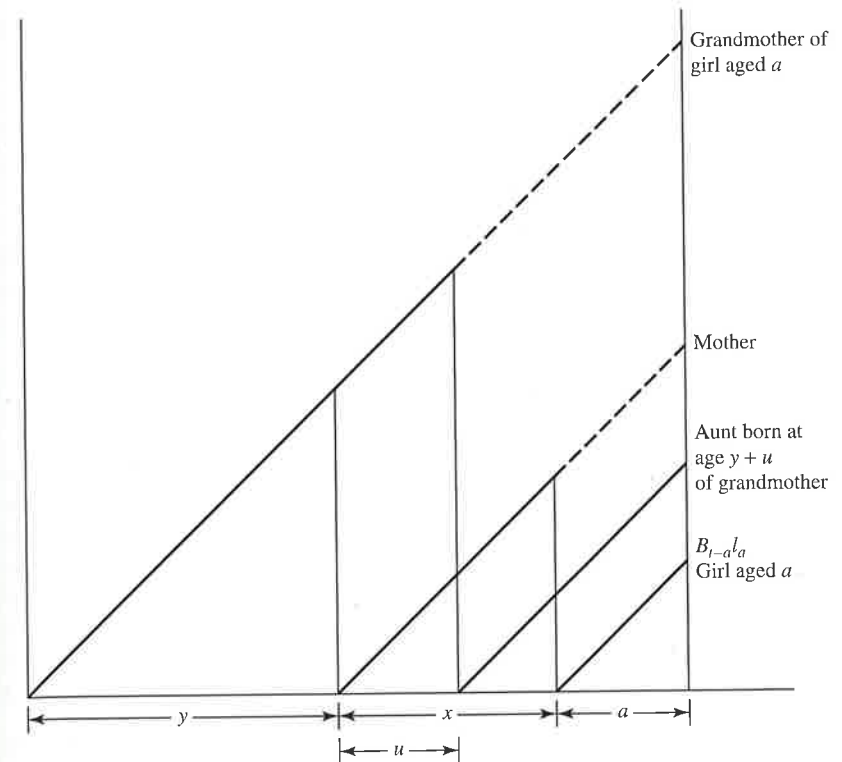
Figure 15.3. Lexis diagram for aunt born after mother.

earlier, we can use this result whether or not the mother is now alive. All that remains is to integrate out the condition that the girl aged a was born when her mother was aged x:

$$A_a^{young} = \int_\alpha^\beta S_{x+a}^{young} e^{-rx} l(x) m(x) \, dx,$$

or written out in full so as to accord with Figure 15.3,

$$A_a^{young} = \int_\alpha^\beta \int_\alpha^\beta \int_0^{x+a} \left(\frac{l(y+u)}{l(y)} m(y+u) \right) \left(e^{-ry} l(y) m(y) \right)$$
$$\times \left(e^{-rx} l(x) m(x) \right) du \, dy \, dx.$$

The application of the same principles to nieces and first cousins (Figure 15.4) is found in Goodman, Keyfitz, and Pullum (1974). [The reader can show how to go on to second- and higher-order cousins, as well as great aunts and other distant relatives.]
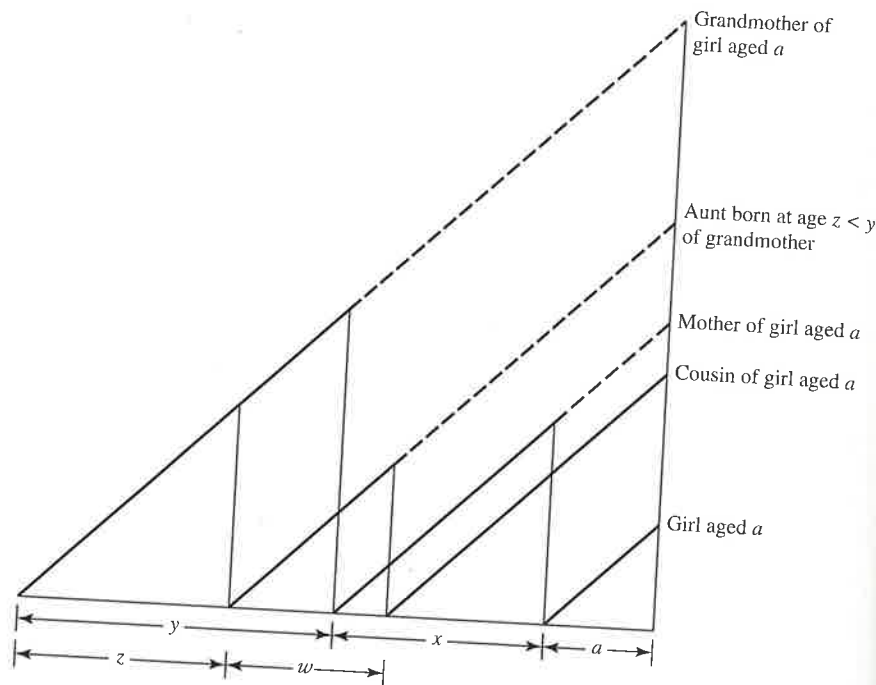
Figure 15.4. Lexis diagram for cousin of a girl aged a through mother's elder sister.

15.4 Mean and Variance of Ages

For each kin whose expected number can be calculated, so can the mean age. Consider the descendants, for example, of daughters already born to a woman aged $a > \alpha$. These number $\int_\alpha^a m(x)\,dx$, and their mean age must be

$$\frac{\int_\alpha^a (a-x)m(x)\,dx}{\int_\alpha^a m(x)\,dx} = a - \bar{x}_a, \qquad (15.4.1)$$

where $\bar{x}_a$ is the mean age of the childbearing function $m(x)$ up to age a. The mean age of children still alive to the same woman aged a is

$$\frac{\int_\alpha^a (a-x)m(x)l(a-x)\,dx}{\int_\alpha^a m(x)l(a-x)\,dx}. \qquad (15.4.2)$$

The mean age of granddaughters of women aged $a > 2\alpha$ is similarly

$$\frac{\int_\alpha^a l(x)m(x)\int_\alpha^{a-x}(a-x-y)l(y)m(y)\,dy\,dx}{\int_\alpha^a l(x)m(x)\int_\alpha^{a-x}l(y)m(y)\,dy\,dx}. \qquad (15.4.3)$$

One can go to the variance of ages of descendants in the successive generations. Thus for the variance of ages of living daughters of a woman aged a we would have

$$\sigma_a^2 = \frac{\int_\alpha^a (x-\bar{\bar{x}})^2 m(x)l(a-x)\,dx}{\int_\alpha^a m(x)l(a-x)\,dx}, \qquad (15.4.4)$$

where $\bar{\bar{x}}$ is the mean age of women up to age a at the birth of their children, weighted by the survival function $l(a-x)$. Here, as elsewhere, no account is taken of heterogeneity in ages and rates of childbearing, or of the spacing of children imposed by the sterile period of pregnancy and afterward. The simplification has negligible consequences for the expected number of daughters, or for the probability of a living grandmother, but does matter for expected sisters, aunts, and more distant collateral relatives, as well as for variances in all kin.

15.4.1 Ascertainment

Such results illustrate the concept of ascertainment, the way information has been obtained, used in genetics and applied in demography by Mindel C. Sheps. Consider the expected children of a given person. When the person is a child just born, her expected future daughters are $\int_\alpha^\beta l(x)m(x)\,dx$ prospectively; once she has passed age β, they are $\int_\alpha^\beta m(x)\,dx$ retrospectively; the corresponding mean ages are $\int_\alpha^\beta xl(x)m(x)\,dx/R_0$ and $\int_\alpha^\beta xm(x)\,dx/G_0$, the latter always being greater. The expected grandchildren calculated retrospectively for a woman of 85 differ from the prospective number of grandchildren of a child just born. In a cohort, individuals may be ascertained by the occurrence of a "signal" event at some point in their lives—the signal event may be having a second child, or being caught in a survey at time t. (Sheps and Menken 1973, p. 341.) Expressions for extended kin provide further illustrations.

15.5 Changing Rates of Birth and Death

This chapter has been restricted to the stable case, in which we suppose a fixed regime of mortality and fertility to be in force over a long past period

and continuing into the present. Yet the theory can be extended to cover certain kinds of change in the regime.

For an example of how changing rates would be accommodated, let us reconsider the probability that a woman aged a has a living mother, the expression $M_1(a)$ of (15.1.3). The conditional probability that the mother is alive, $l(x+a)/l(x)$, must now be determined by the chance of survival appropriate to the changing death rates actually experienced by the cohort aged $x+a$ at time t, that is, born at time $t-x-a$. The ratio $l(x+a)/l(x)$ in the formula would have to be taken from the appropriate cohort life tables, a different table for each value of x. This is certainly possible, though awkward enough that no one is likely to do it.

In addition the distribution of x, the age at childbearing, is affected by the instability; if, for instance, the actual age distribution is younger than the stable one, (15.1.3) has to be modified to allow a greater weight to $l(x+a)/l(x)$ for younger x, thereby increasing the probability that the mother is still alive. Thus the factor $e^{-rx}l(x)m(x)$ in $M_1(a)$ would have to be replaced by numbers proportional to the actual ages of mothers prevailing a years earlier, say $w(x|t-a)$. Result 15.1.3 would thus be replaced by

$$M_1'(a) = \int_\alpha^\beta \frac{l(x+a)}{l(x)} w(x|t-a)\, dx,$$

where $w(x|t-a)$ is the age distribution of women bearing children a years ago, or at time $t-a$.

Analogous considerations permit a straightforward rewriting and reinterpreting of all the formulae of this chapter in a way that dispenses with the stable assumption insofar as it affects earlier age distributions. Goodman, Keyfitz, and Pullum (1974) provide these more general formulae. Interpretation for fixed rates is simple: the given schedules $l(x)$ and $m(x)$ imply certain mean numbers of kin. The corresponding statement for changing rates is unavoidably more complicated.

15.6   Sensitivity Analysis

A main use of the kinship formulae here developed is to ascertain the effect of changes in the demographic variables on kinship. How much does a younger age of marriage and of childbearing reduce the number of orphans? What is the effect of a fall in the birth rate on the number of grandchildren of a person of a given age chosen at random? What does a uniform improvement in mortality at all ages do to the number of living aunts of a girl of given age?

Merely looking at the formulae does not tell much more than we know without them. Intuition suggests that the fraction of girls aged a who have living mothers must depend primarily on death rates (specifically those

between the time of childbearing and a years afterward) and secondarily on birth rates—most of all, on whether children are born at young or older ages of mothers. In a smaller way yet it ought to depend on the overall rate of increase, because with given death rates and with birth rates in a given proportional distribution a faster growing population has somewhat younger mothers. But a quick look at (15.1.3) reveals only that $M_1(a)$ is a function of birth and death rates, without clearly suggesting the amount or even the direction of the relation.

15.6.1   Decomposition of $M_1(a)$, the Probability of a Living Mother

Once a computer program for an expression such as (15.1.3) for $M_1(a)$ is available, it is possible to make small variations in any part of the input— add 10 percent to the birth rates at certain ages while leaving the life table intact, for instance—and see the effect on the probability of a living mother. Here we will try to see how such variations operate theoretically, by using the approximation to $M_1(a)$ developed as (15.1.7) or, written slightly differently,

$$M_1(a) \approx \frac{l(\kappa + a)}{l(\kappa)} + \frac{\sigma^2}{2}\left[\frac{l(\kappa + a)}{l(\kappa)}\right]''. \qquad (15.6.1)$$

The main effect of raising the mean age of childbearing is to replace mortality in an interval at the original κ with mortality around $a + \kappa$, as is evident from application of (1.6.5):

$$\frac{l(\kappa + a)}{l(\kappa)} = \exp\left[-\int_\kappa^{\kappa + a} \mu(t)\, dt\right]. \qquad (15.6.2)$$

If the death rate is nearly constant with age, or if a is small, $M_1(a)$ depends little on the value of κ. The second term on the right-hand side of (15.6.1) is negative through most of the life table and is considerably smaller than the first, unless a is a very old age. The rate of increase of the population enters only through κ, which for a given life table is younger the higher the rate of increase.

If mortality $\mu(x)$ between κ and $\kappa + a$ increases by an amount k at every age, the survivorship $l(\kappa + a)/l(\kappa)$ will diminish in the ratio e^{-ka}, and this is the only effect of a constant mortality addition on $M_1(a)$. Hence the new $\overline{M}_1(a)$ is equal to $e^{-ka}M_1(a)$. The reason that the weighting factor $e^{-rx}l(x)m(x)$ in (15.1.3) remains unaffected is that a uniform change in mortality causes a change in r that offsets the change in $l(x)$ as far as age distribution is concerned, a matter discussed in Section 10.1.

15.6.2   Other Progenitors

If mortality at all ages is increased in the uniform amount k, the probability of a living grandmother will change to

$$\overline{M}_2(a) = \int_\alpha^\beta e^{-k(a+x)} M_1(a+x) e^{-rx} l(x) m(x) \, dx, \qquad (15.6.3)$$

which cannot be simplified without approximation. But let x be replaced in $e^{-k(a+x)}$ by κ', the mean age at childbearing for the mothers still alive. Then, taking the exponential outside the integral, we have approximately

$$\overline{M}_2(a) \approx e^{-k(a+\kappa')} M_2(a).$$

For great grandmothers

$$\overline{M}_3(a) \approx e^{-k(a+\kappa'+\kappa'')} M_3(a),$$

where κ'' is the mean age at childbearing for grandmothers still alive. In practice we do not have data on κ' or κ'' and would suppose them to be close to κ; hence the outcome in general is

$$\overline{M}_i(a) \approx e^{-k[a+(i-1)\kappa]} M_i(a). \qquad (15.6.4)$$

If k is small, so that e^{-ka} is nearly $1 - ka$, we obtain the following finite approximations for the difference in the several $M_i(a)$ on adding k to the force of mortality:

$$\Delta M_1(a) = -ka M_1(a)$$

$$\Delta M_2(a) = -k(a + \kappa) M_2(a)$$

$$\Delta M_3(a) = -k(a + 2\kappa) M_3(a).$$

With these approximations, if one of two populations has mortality higher at every age by 0.003, and if κ is 27.5, for women aged $a = 20$ the chance of having a living mother is 0.94 as high as in the other population; of a living grandmother, 0.86; and of a living great-grandmother, 0.78. On the more precise (15.6.4) the last three numbers become 0.942, 0.867, and 0.799, respectively.

Venezuela has somewhat higher mortality than the United States; we note from Table 15.1 that for a Venezuelan girl aged 20 the probability of having a living mother is in the ratio $0.932/0.959 = 0.97$; of a living grandmother, in the ratio 0.85; of a living great-grandmother, in the ratio 0.64, all ratios to the United States. The gradient as one advances to more remote ancestors is steeper than that of (15.6.4) based on fixed differences of $\mu(x)$.

A more complete analysis would decompose the difference between the two countries in, say, the probability of a living grandmother into two components: (1) that due to mortality differences, and (2) that due to differences in the pattern of births. This is readily accomplished arithmetically,

once an appropriate computer program is available, simply by permuting the input data, as was done in Table 10.2.

15.6.3   Effect of Birth Pattern on Living Progenitors

The main variation in $M_1(a)$ as far as births are concerned occurs through the mean age of childbearing κ. We found that

$$M_1(a) \approx \frac{l(\kappa + a)}{l(\kappa)};$$

therefore taking logarithms of both sides and differentiating gives

$$\frac{1}{M_1(a)} \frac{dM_1(a)}{d\kappa} = -\left[\mu(\kappa + a) - \mu(\kappa)\right],$$

where $\mu(\kappa)$ is the force of mortality at age κ. In finite terms

$$\frac{\Delta M_1(a)}{M_1(a)} \approx -\left[\mu(\kappa + a) - \mu(\kappa)\right] \Delta\kappa;$$

that is, the proportionate change in the chance of a living mother is minus the difference in death rates over an a-year interval times the absolute change in κ. With the death schedule of Madagascar, 1966 (Keyfitz and Flieger 1971), and its κ of about 27.5, using for $\mu(27.5)$ the approximation $_5M_{25} = 0.01740$, and for $\mu(\kappa + a) = \mu(27.5 + 20) = \mu(47.5)$ the rate $_5M_{45} = 0.02189$, we have

$$\frac{\Delta M_1(a)}{M_1(a)} \approx -(0.02189 - 0.01740)\Delta\kappa = -0.0045\Delta\kappa.$$

For each year later of average childbearing the chance of a woman of 20 having a living mother is lower by 0.45 percent.

The change with a is found in the same way to be

$$\frac{\Delta M_1(a)}{M_1(a)} = -\mu(\kappa + a)\Delta a,$$

and for $a = 20$, $\kappa = 27.5$, this is

$$-\mu(47.5)\Delta a = -0.02189\Delta a.$$

For a 5-year interval the proportionate decrease in $M_1(20)$ ought to be 5 times as great or 0.109. In fact Table IIIa of Goodman, Keyfitz, and Pullum (1974) shows

$$\frac{M_1(25) - M_1(15)}{2} = -\frac{0.7817 - 0.6421}{2} = -0.0698;$$

and as a proportion of $M_1(20) = 0.7126$, this is $-0.0698/0.7126 = -0.098$, about as close to -0.109 as we can expect with the crude approximations used.

Table 15.3. Effect of changed birth rate on probability of ancestor being alive, Madagascar females, 1966

Age of woman a	Difference in Probability of Having a Living:		
	Mother $\overline{M}_1(a) - M_1(a)$	Grandmother $\overline{M}_2(a) - M_2(a)$	Great-grandmother $\overline{M}_3(a) - M_3(a)$
	After Lowering the Birth Rate for Women 20–24 by 0.01		
0	0.	−0.00268	−0.00529
20	−0.00025	−0.00423	−0.00109
40	−0.00228	−0.00103	−0.00000
60	−0.00067	−0.00000	−0.00000
	After Lowering the Birth Rate for Women 40–44 by 0.01		
0	0.	+0.00395	+0.00404
20	+0.00109	+0.00345	+0.00057
40	+0.00199	+0.00051	0.00000
60	+0.00036	0.00000	0.00000

To find the effect of a change in fertility at particular ages one can run the program twice, once with the observed regime of mortality and fertility, and once with the specific birth rate for age 20–24 lowered by 0.01. Differences for progenitors are shown in Table 15.3. A drop in fertility at age 40–44 lowers the average age of childbearing and hence raises the chance of a living grandmother. Other items can be similarly interpreted.

15.6.4   Comparison of Effect of Birth and Death Rates

Robert Sembiring (1978) has experimented to determine whether the numbers of particular kin are affected more by birth or by death rates. As an example of the procedure we consider the number of female cousins that a girl aged a would be expected to have through her mother's sister and use the technique of permuting the input data, as in Section 10.1.

To separate the effects of mortality from those of fertility nine calculations were made of the curve of expected cousins by age. Three levels of fertility were used, those of Costa Rica for 1960, the United States for 1959–61, and Sweden for 1958–62, all pertaining to 1960 or thereabout. The gross reproduction rates of the three countries were 3.891, 1.801, and 1.080, respectively. Each of these levels of fertility was paired with each of three levels of mortality taken from the Coale and Demeny West series model tables, with $\overset{\circ}{e}_0$ values of 70, 55, and 40. The numbers chosen represent approximately the range of mortality and fertility among human populations. The resulting nine curves for the average number of living cousins in the female line (i.e., daughters of maternal aunts) are shown in Figure 15.5.
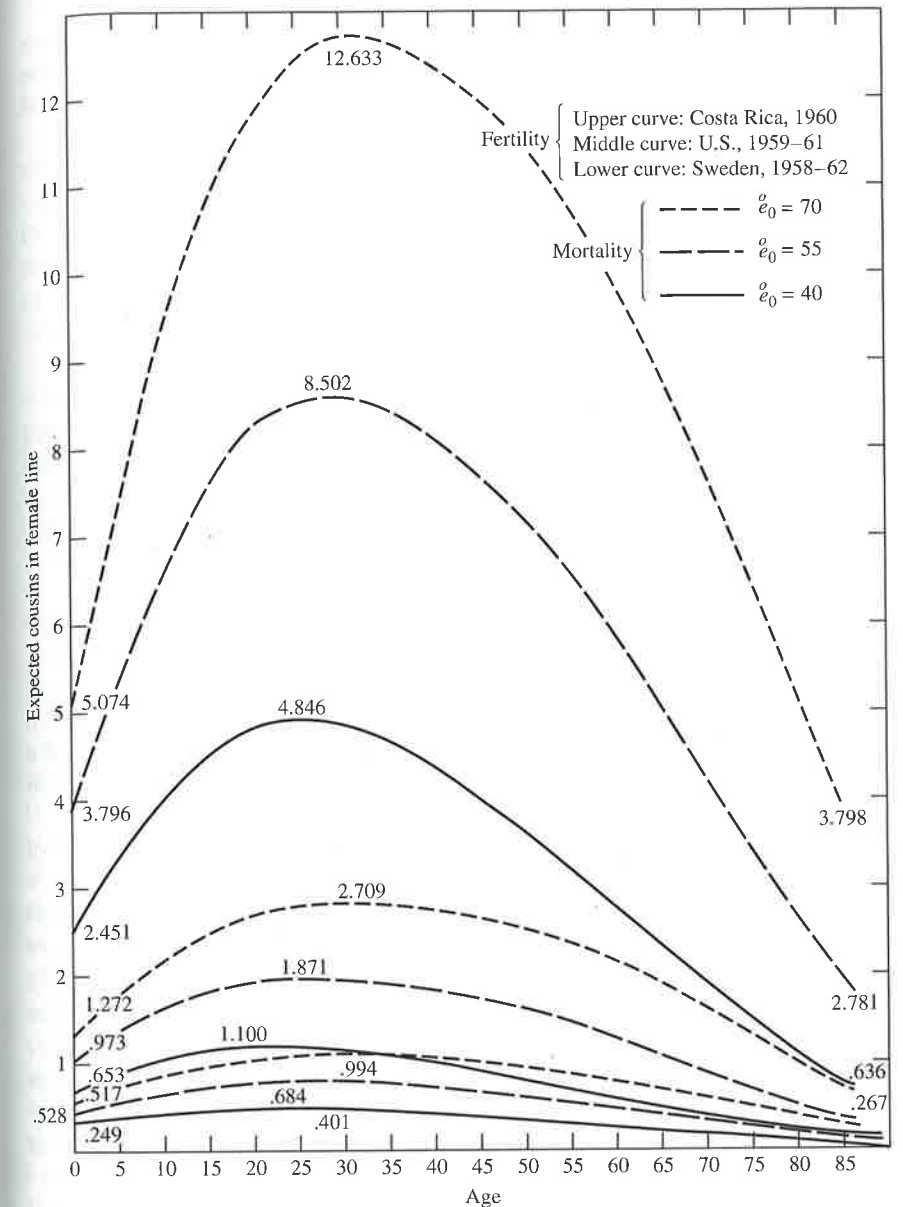
Figure 15.5. Expected number of cousins still alive.

The three curves for high (Costa Rican) fertility are above those for intermediate fertility, and these again are mostly above those for low fertility. Apparently fertility has more effect on the number of living cousins than does mortality. With high fertility the number of cousins reaches a sharp

Table 15.4. Mean number of living first cousins in the female line of a woman aged 20, for artificial populations constructed by fertility of Costa Rica, the United States, and Sweden, about 1960, in all combinations with mortality of Coale–Demeny model West tables having $\overset{o}{e}_0$ values of 40, 55, and 70 years

| | Fertility | | |
Mortality	Costa Rica $G_0 = 3.891$	United States $G_0 = 1.801$	Sweden $G_0 = 1.080$
$\overset{o}{e}_0 = 40$	4.7292	1.0998	0.4009
$\overset{o}{e}_0 = 55$	7.9888	1.8502	0.6789
$\overset{o}{e}_0 = 70$	11.3518	2.6006	0.9657

peak at ages 25 to 45; with lower fertility and lower mortality the curve peaks less sharply. Note that no actual population combines a gross reproduction of 3.891 with an expectation of life of 70 years; therefore the peak of 12.633 female maternal parallel cousins is purely hypothetical. The combination of United States or Swedish fertility with this high expectation of life represent possible real situations.

All of the expressions in this chapter apply to male as well as to female kin, but with one difference. This difference arises out of the fact that we know the mother was alive at the birth of her child, but we know only that the father was alive 9 months before the birth. To apply to males, the formulae would have to be adjusted for the three-quarters of a year of additional mortality, which could be appreciable for certain kin in populations subject to high death rates.

A rough approximation to the total number of first cousins (i.e., of both sexes) would be obtained by multiplying the material parallel female cousins here given by 8; this would be improved by making the corresponding calculation for the male line and for mixed lines. An approximation to the number of cousins implied by other schedules of mortality and fertility would be obtained by two-way quadratic interpolation, using $\overset{o}{e}_0$ and the gross reproduction rate as indices. This is especially feasible for mortality; note that the number of cousins for $\overset{o}{e}_0 = 55$ is almost exactly the mean of the numbers for $\overset{o}{e}_0 = 40$ and $\overset{o}{e}_0 = 70$ (Table 15.4). Interpolation may be useful even with an available program because of the large amount of computer time required for the exact calculation.

15.7 The Inverse Problem: Deriving Rates from Genealogies

The inverse problem is of practical interest to those who must make inferences regarding birth and death rates for areas or times for which

registration systems are not in existence, or if in existence are grossly incomplete. For a given genealogy, the problem now is to find the regime of mortality and fertility.

If one-half of women aged 40 have living mothers in a certain population, what is the expectation of life? According to Table 15.1, Venezuela showed a probability of 0.707 and Madagascar of 0.386, and these had $\overset{o}{e}_0$ values of 67.7 and 38.5 years, respectively. By straight-line interpolation, supposing ages at childbearing to be sufficiently similar among all three populations to leave the answer unaffected, we find $\overset{o}{e}_0 = 48.9$ corresponding to our $M_1(40) = 0.5$.

Looked at formally, what were relatively simple integrals to evaluate when the regime of mortality and fertility was given become difficult—in most instances unsolvable—integral equations when the mean number of kin is known but the regime is unknown. If we observed mean numbers of the various living kin, we would have a set of equations, most of them containing multiple integrals, and they would have to be solved as a simultaneous set. Thus all of the expressions for different kin might be equated to observations and solved simultaneously for the unknown rates of birth and death.

Yet for many practical purposes we can avoid most of the difficulties just mentioned by supposing that all life tables can be laid out in a straight line, indexed by the expectation of life at age 0 or age 10, and correspondingly that schedules of childbearing can be arranged according to the gross reproduction rate G_0. All variations in mortality and fertility beyond these two dimensions will be neglected in this simple method.

If the number of cousins to women aged 20 is represented as a height above the $(G_0, \overset{o}{e}_0)$ plane, so that the collection of such information regarding women aged 20 is a quadratic surface over the plane, an observed number of cousins can be represented by a plane parallel to the $(G_0, \overset{o}{e}_0)$ plane and cutting the quadric surface in a second-degree curve, which may now be projected down onto the $(G_0, \overset{o}{e}_0)$ plane. To make the regime entirely determinate we need some other fact about kin. Continuing to confine ourselves to age 20 for the sake of this example (though we need not stay with the same age), we note the fraction of women aged 20 having a living mother. This also can be represented as a quadric surface over the same $(G_0, \overset{o}{e}_0)$ plane, and the given observation as a plane again cutting the quadric surface in a quadratic. The intersection of the two quadratics gives the regime of mortality and fertility. The object is to choose kin that provide curves intersecting as nearly as possible at right angles to each another. Examples are probability of living mother (or grandmother) and number of daughters (or granddaughters) ever born.

In practice any one pair of observations will be unacceptable as neglecting most of the data. Indeed, errors are so pervasive that we will do none too well using all of the information available. A large number of data pairs will

each provide a point of intersection, and with given accuracy of enumeration the precision of any point will be greater the closer the lines defining it come to making a right angle with each other. The several estimates obtained from pairs of kin can be weighted by the sine or other suitable function of the angle that the lines make with each other.

15.8 Incest Taboo and Rate of Increase

An incest taboo has the advantage for the group that adheres to it of compelling biological and social mixing, and of stirring individual initiative in the search for a mate. In addition, it avoids the confusion that would result if one's father were also one's uncle, these being very different roles in most societies. It promotes political alliances among families, and it avoids intrafamilial conflict over women. Most such advantages are greater the wider the degree of incest prohibition: a taboo reaching as far as second cousins will compel more mixing than a taboo against brother–sister matings only.

But a price has to be paid for the advantages—the wider the taboo, the more individuals will fail to find mates, especially in sparsely settled populations. From the viewpoint of reproduction the incest taboo is a luxury, and the question is how much of it a group can afford. The ideal approach would be an analysis of trade-offs: find the point at which the advantages of increased mixing are exactly offset by the lesser reproduction. Unfortunately the elements of this equation are incommensurable, and no theory seems to exist that will provide a quantitative measure of the net advantage of mixing.

However, it is possible to deal with one side of the problem: the cost in rate of increase of various degrees of incest taboo. Though determinate, this is mathematically difficult, as are all questions of population increase in which the rate of reproduction depends on the size of the group and the random number of possible mates. Having little hope of finding a closed solution, Hammel (1972, Hammel et al. 1976, 1979) and his coworkers addressed the problem by simulation.

They used mortality rates assembled from the Maghreb and ancient Rome, and fertility rates from the Cocos-Keeling Islands, reduced by 20 percent to be slightly below stationarity without any incest taboo. A group of 65 individuals with a kinship structure of three generations of genealogical depth was derived from the previous evolution of the model. Five runs were then made with no incest prohibition, five with a prohibition of one degree (sibling and parent–child), five with a prohibition of two degrees (up to first cousin), and five with a prohibition of three degrees (effectively up to second cousin). All runs covered 100 years.

The outcome, as anticipated, was a more rapid rate of population decrease the broader the taboo. With no taboo the mean of the five runs was

a rate of $r = -0.001$; with sibling and parent–child exclusion the mean was -0.006; with exclusion up to first cousin the mean rate was -0.018; with exclusion up to second cousin the rate was -0.036. The one-degree prohibition apparently costs 0.005, the two-degree an additional 0.012, and the three-degree a further 0.018 lowering of r, on this rough model.

Insofar as the rates used were realistic, a breeding group of the order of 65 individuals could not afford any exclusion, not even siblings. However, as mentioned above, the authors had reduced the Cocos-Keeling fertility by 20 percent. A group that would increase at 1 percent per year with no taboo at all could tolerate the sibling and parent–child taboo, which would reduce it to $0.01 - 0.005 = 0.005$; but it could not afford to go as far as the first cousin taboo, which would bring it down to $0.005 - 0.012 = -0.007$, or a half-life of a century. The calculation suggests that the incest taboo, aside from its other functions, is capable of holding down the rate of increase in small dispersed populations.

15.9 The Bias Imposed by Age Difference on Cross-Cousin Marriage

Among social groups practicing cross-cousin marriage, more instances have been observed of Ego, a male, marrying his mother's brother's daughter (MBD) than his father's sister's daughter (FSD). It is also usual for men to marry women younger than themselves. The question is whether the age bias by itself would lead to the bias toward MBDs. A realistic model would be complicated, but the effect of brides being younger than grooms can be shown with some simplified arithmetic (Hammel, 1972).

Consider a population in which men marry 5 years older than women, and children are born to the couple when the husband is age 25 and the wife age 20. Brothers and sisters are all the same age. Then, if Ego (male) is age E, his father will be age $E + 25$, and his father's sister also $E + 25$. His FSD will be $E + 25 - 20 = E + 5$. On the other hand, his mother will be $E + 20$, his mother's brother also $E + 20$, and his MBD $E + 20 - 25 = E - 5$. Thus, of the two kinds of cross-cousins, the FSD is 5 years older than Ego. If he is seeking a bride 5 years younger, he will find the MBD the right age.

A similar calculation can be made for parallel cousins. Ego's father being $E + 25$ years of age and his father's brother also $E + 25$, his FBD is $E + 25 - 25 = E$, and similarly for the other parallel cousin, designated as MSD. Parallel cousins are the same age as Ego on this simple model.

These purely demographic (some would say, merely logical) considerations mean that the tendency for men to marry kinswomen younger than themselves leads to the MBD marriage. Of the four kinds of cousins only MBDs are the right age to permit men to be older than their brides generation after generation.

Hammel was the first to point out that age preferences for *either* older or younger wives would have equivalent effects, and indeed that any heritable property would work as well as age. The contribution of Hammel and Wachter was to show by simulation that the effect remains considerable even in the face of all the obvious sources of randomness and to study the dependence of the effect on the size of the age gap. Simulation has been useful here and in other instances where analytic solutions are out of reach. Kunstadter et al. (1963) used it to find the fraction of individuals who would have an MBD cousin to marry in a tribe when that was preferred.

The approach in this chapter, via stable population theory, takes a deterministic approach appropriate for large populations. We should, however, point out two other important approaches, both of which take more account of individuals and their properties. First, we might recognize that the vital rates apply as probabilities to discrete individuals. If we suppose that they do so independently, we are led to stochastic branching process models (a simple branching process model appears in Section 16.4; see also Chapter 15 of MPM). These models have been used by Pullum (1982, Pullum and Wolf 1991) to derive entire probability distributions of the numbers of kin of various kinds, but without taking the age-specificity of the vital rates into account.

If we take this approach to its limit, we would want to keep track of each individual, with all of his or her *i*-state variables (age, marital status, health, employment, etc.) and relationships to other individuals. We would then apply to each individual the probabilities of birth, death, marriage, and any other demographic transitions of interest. Doing so repeatedly would project the population forward in time subject to those rates. Repeating that exercise many times would produce the probability distribution of population trajectories (including all the information on all the individuals) implied by the vital rates. Such models are called *i*-state configuration models (Caswell and John 1992) or individual-based models (DeAngelis and Gross 1992, Grimm et al. 1999) in the ecological literature, and microsimulation models in the human demographic literature (e.g., Wachter et al. 1997, Wolf 2001). They have been applied to problems of kinship by, e.g., Hammel et al. (1979), Ruggles (1993), Wachter et al. (1997), Wachter (1997), and the chapters in Bongaarts et al. (1987).

The approaches of the 15 chapters through this one may be called *macrodemography*, following a usage going back through sociology, economics, and physics, ultimately to a source in Greek metaphysics. Microsimulation methods are an example of *microdemography*, in which properties of individuals and their random variation are recognized as the source of change in population aggregates. Chapter 16 introduces some aspects of microdemography.

16
Microdemography

Physics accounts for heat by the motion of molecules, medicine accounts for disease by the action of germs, and economics accounts for aggregate prices and production by the activities of individuals seeking to maximize their utility. The fact that nontrivial problems of aggregation arise, and that the microelements often turn out on closer examination to be unrealistic constructions, does not deprive them of explanatory and predictive value.

Microdemography helps us to understand such macrophenomena as birth rates of regions and nations. How much reduction of the birth rate results from couples substituting 99 percent efficient contraception for methods 95 percent efficient? If the number of abortions was equal to the number of births in a country, could we conclude that suppression of abortion would double the number of births? How can the probability of conception be measured? What difference does it make to the increase of a population if parents aim for three children rather than two? These and other questions of microdemography are the subject of the present chapter.

16.1   Births Averted by Contraception

The theory of birth as a Markov renewal process have been developed by Sheps (1964, Sheps and Perrin 1963, Sheps and Menken 1973), and Potter (1970) has shown how this theory can be applied to calculate births averted. Tietze (1962) was a pioneer in this field and Lee and Isbister (1966) made important early suggestions.