The Demography of Kinship

Formal Demography
Stanford Summer Short Course
James Holland Jones, Instructor

August 18, 2006

Outline

1. Stable Population Theory
   (a) Counting Method
   (b) Probability Method
   (c) Mothers and Orphans
   (d) Daughters
   (e) Sisters
   (f) Two-Sex Extensions
2. Pullam's Probability Generating Functions
3. Demographic Microsimulation

Counting Kin

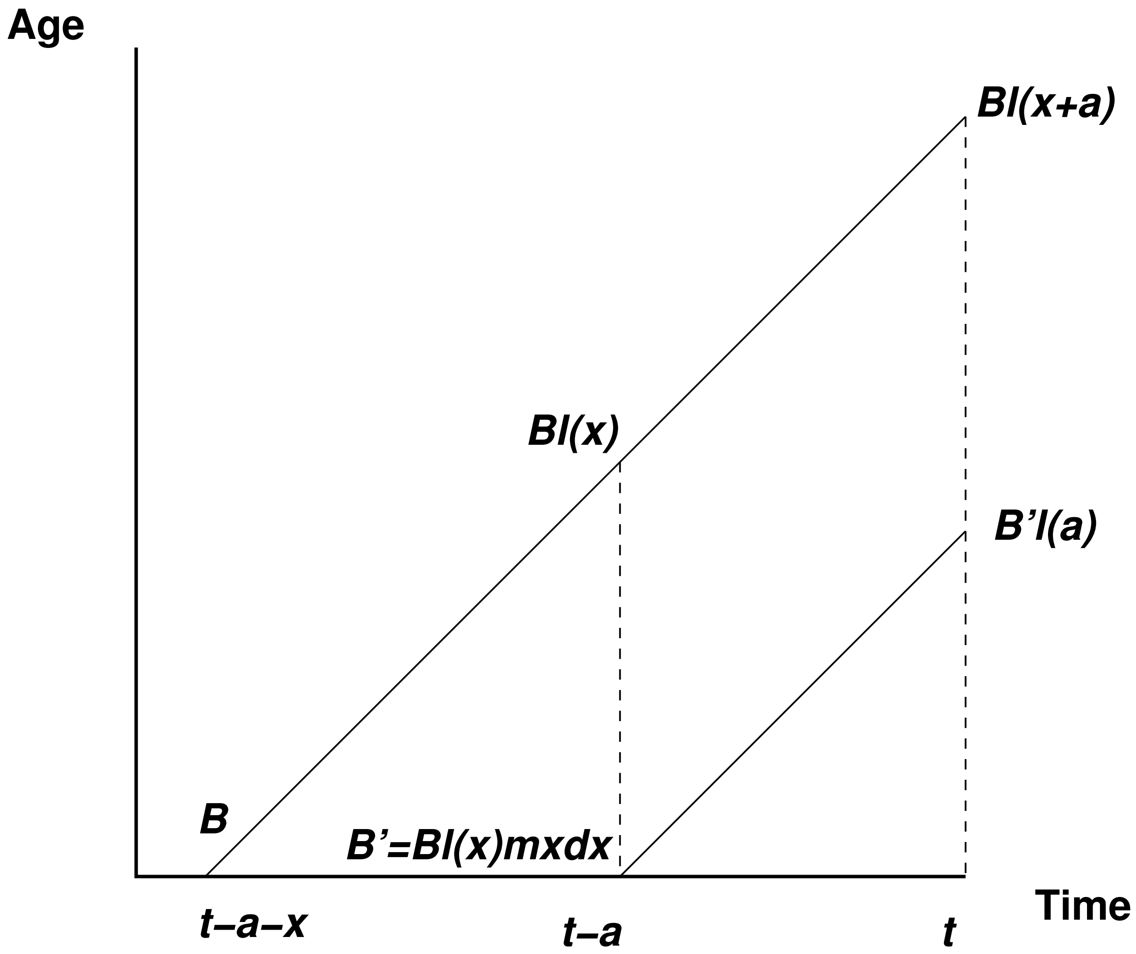
There are basically two approaches to the demographic theory of kinship

1. Counting Method
2. Probability Method

We will first focus on the counting method, because it is intuitive (if a little involved...)

The Problem: What is the probability that a girl, age a has a living mother?

Counting Method



Now in Words...

B female children are born $a + x$ years ago

At age x, a fraction $Bl(x)$ will remain

During the interval $x + dx$, the will give birth to

$$Bl(x)m(x)l(a)dx$$

daughters who survive to age a

- Remember, that if $m(x)$ is the age-specific fertility $rate$, the number of births in a small interval dx is simply $m(x)dx$

Of all the mothers who gave birth, a fraction $l(x+a)/l(x)$ will survive the additional a years of the a-year-old girl in question

The number of living mothers is thus:

$$Bl(x)m(x)dx\frac{l(x+a)}{l(x)}l(a) \tag{1}$$

Now, this is for a cohort of women born $a + x$ years ago

But we don't know what cohort a girl's mother belonged to

Generalize the number of births $B \to B(t)$

The present time is t, so girls born $x + a$ years ago number $B(t - a - x)$

Substitute this birth function into equation 1 and integrate across a ages x for which a woman can give birth

$$\int_{\alpha}^{\beta} B(t-a-x)l(x)m(x)dx\frac{l(x+a)}{l(x)}l(a) \tag{2}$$

Note that this is the number of living mothers

However, We want the average number of mothers per daughter age a

Assuming the same birth function $B(t)$ applies, the number of daughters born a years ago alive now will be

$$B(t-a)l(a)$$

The ratio of equation 2 to this will give us the value we seek

So, what's $B(t)$?

In deriving the renewal equation, we made the assumption (based on the baseline assumptions that birth and death rates have been constant for a long time) that births were increasing exponentially as $B_0 e^{rt}$

Substituting this and doing some algebra, we get:

$$M_1(a) = \int_{\alpha}^{\beta} l(x+a)e^{-rx}m(x)dx \tag{3}$$

Lotka (1931) was the first to derive this relationship

The Probability Method

The conditional probability that a woman age x when she gave birth survives an additional a years is simply

$$l(x + a)/l(x)$$

We need to integrate across all the possible ages of mothers

This is key But, mothers are not uniformly distributed across the age interval $[\alpha \; \beta]$

We therefore need to weight by the age distribution

Stable population theory tells us that the fraction of women age x in the stable population is

$$c(x) = be^{-rx}l(x) \tag{4}$$

We want our weight to be the number in the population at age x **per current birth** b

The number of births born to this fraction is $e^{-rx}l(x)m(x)dx$

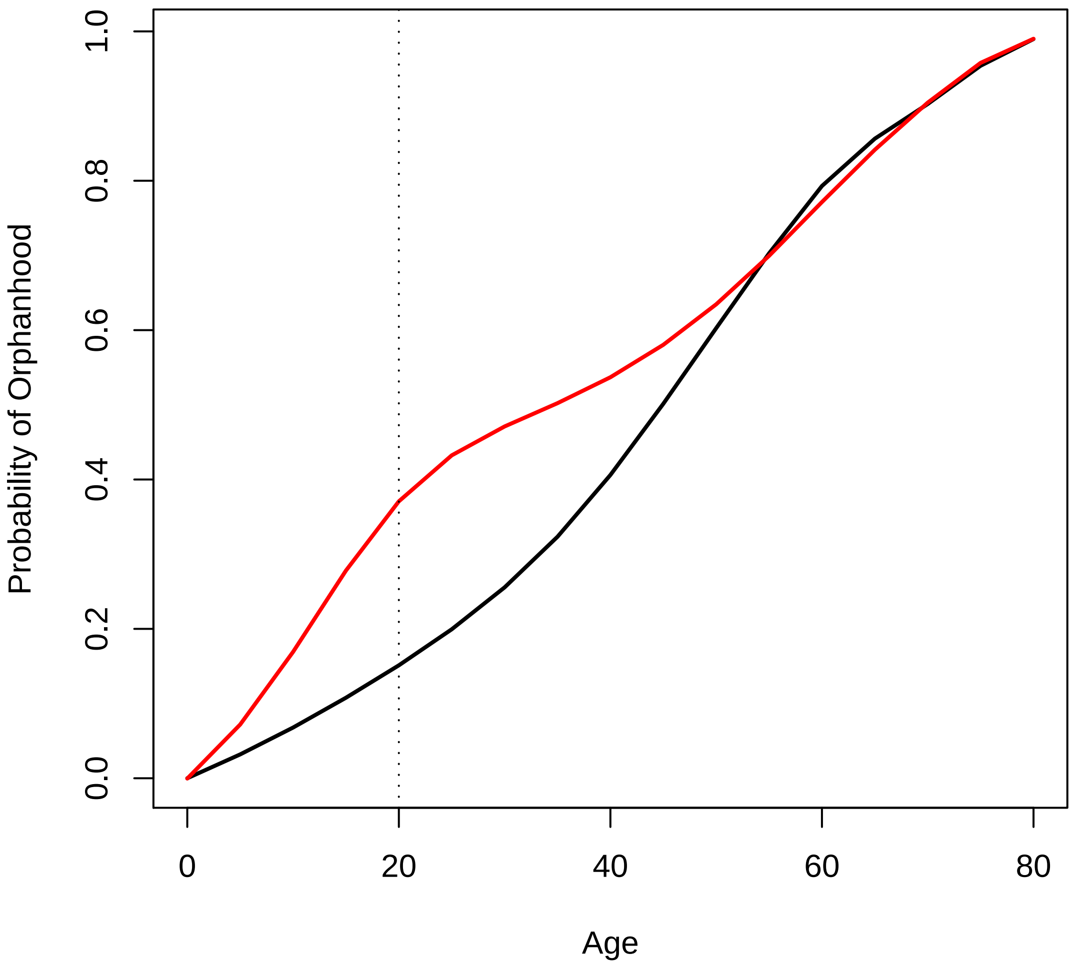
Integrate across ages and multiply by $l(x+a)/l(x)$ to get

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

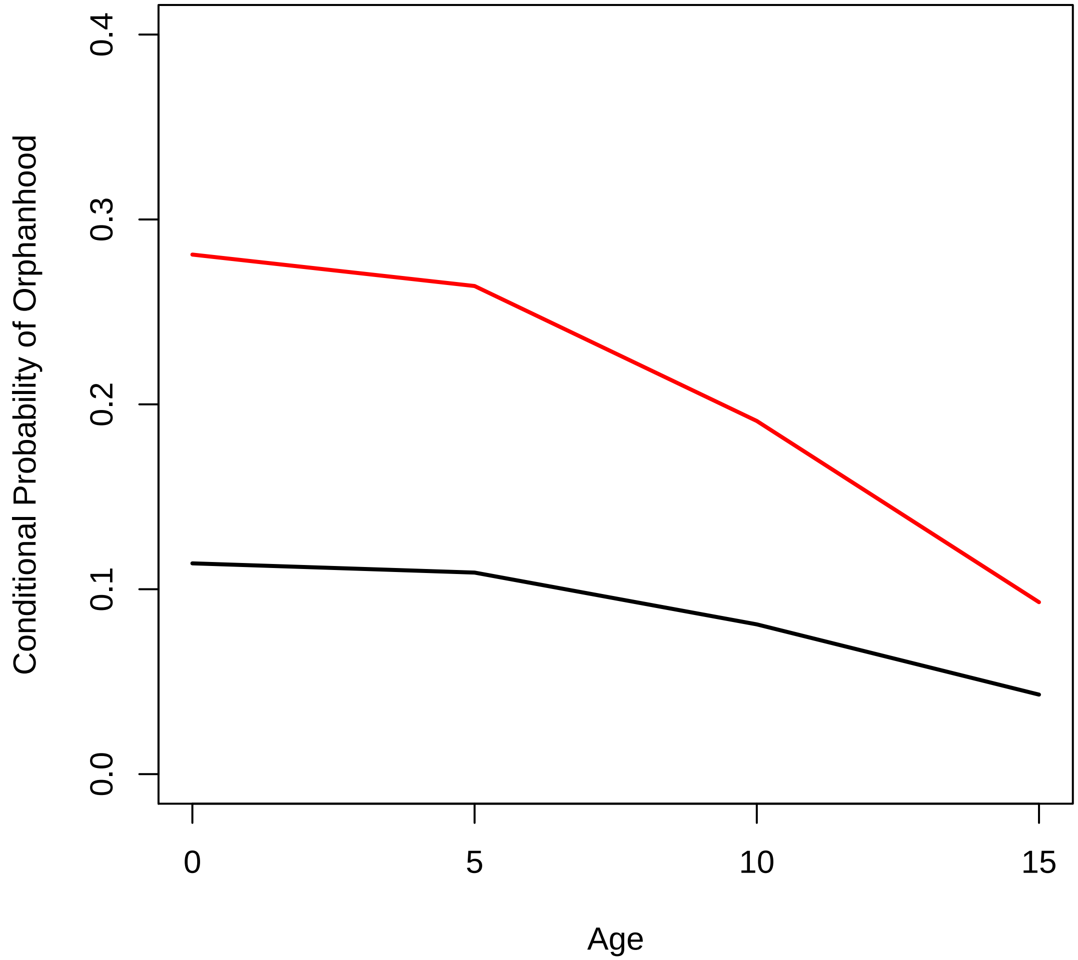
(Again)

We can use this formula for indirect estimation of adult mortality when it is coupled with an orphanhood survey (Henry 1960)

Maternal Orphanhood and the HIV/AIDS Epidemic



Conditional Orphanhood and the HIV/AIDS Epidemic



Works For Grandmothers Too...

The grandmother of a girl age a must live $x + a$ years after the birth of her daughter (who gave birth herself at age x) to be alive at the time of our hypothetical census

This chance is calculated simply as $M_1(x + a)$ using equation 3 above

This is the conditional probability (i.e., conditional on her having a daughter exactly $x + a$ years ago)

To calculate the unconditional probability, we follow the same logic as for mothers:

Multiply by $e^{-rx}l(x)m(x)$ and integrate across all possible x

$$M_2(a) = \int_{\alpha}^{\beta} M_1(x + a)e^{-rx}l(x)m(x)dx \tag{5}$$

Some Things to Consider

We have assumed that we have fixed regime mortality and fertility

We have also assumed for equation 5 that vital rates are independent across generations

- No correlated mortality
- Daughters of young mothers do not themselves go on to reproduce early
- Birth rates and death rates not correlated

In practice, these problems have a small effect on the probabilities of living ancestors and descendants, but they can seriously throw off similar calculations for sisters, aunts, and nieces

Relaxing Stable Population Assumptions

Consider again the probability that a girl age a has a living mother

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

The term $e^{-rx}l(x)m(x)dx$ represents a weighting factor, where the weights are the number of mothers giving birth at age x

This value comes from the characterisitc equation

$$1 = \int_{\alpha}^{\beta} e^{-rx}l(x)m(x)dx$$

which is Lotka's solution to the renewal equation where we assume that $l(x)$ and $m(x)$ have been constant for a relatively long time

We can substitute in any arbitrary weighting function $W(x)$

$$M_1(a) = \int_{\alpha}^{\beta} = \frac{l(x+a)}{l(x)} W(x|t-a)dx$$

A clear choice for $W(x|t-a)$ would be the observed age distribution of mothers age x, $t-a$ years ago

Variable-r

In the absence of good information on the age of mothers $t - a$ years ago, we can use variable-r extensions to stable theory to correct for changing vital rates

In deriving equation 3 for the probability that a girl age a has a living mother, we inegrated the term $l(x+a)/l(x)$

The age structure in the stable population is given by equation 4:

$$c(x) = be^{-rx}l(x)$$

To get the age distribution of births, we multiply by the number of births at age x $m(x)dx$ and divide by the birth rate b to make it the distribution of maternal ages per birth:

$$\psi(x) = e^{-rx}l(x)m(x)$$

If the vital rates have not been constant, the age distribution of births will be different

The variable-r extension to classical stable theory shows that the age structure for a population subject to changing vital rates is

$$c(x) = be^{-\int_0^x r(s)ds} l(x)$$

Multiply by the number of births, divide by the birth rate to get the ages of mothers at birth

We now use this to weight the $l(x+a)/l(x)$ values, yielding:

$$M_1(a) = \int_\alpha^\beta l(x+a) e^{-\int_0^x r(s)ds} m(x)dx$$

Does this help us? A research question!

An Approximation

Keyfitz (1977) notes that a reasonable approximation of $M_1(a)$ will be the probability of surviving from the mean age of childbearing κ to $\kappa + a$

Note that the mean age of childbearing is:

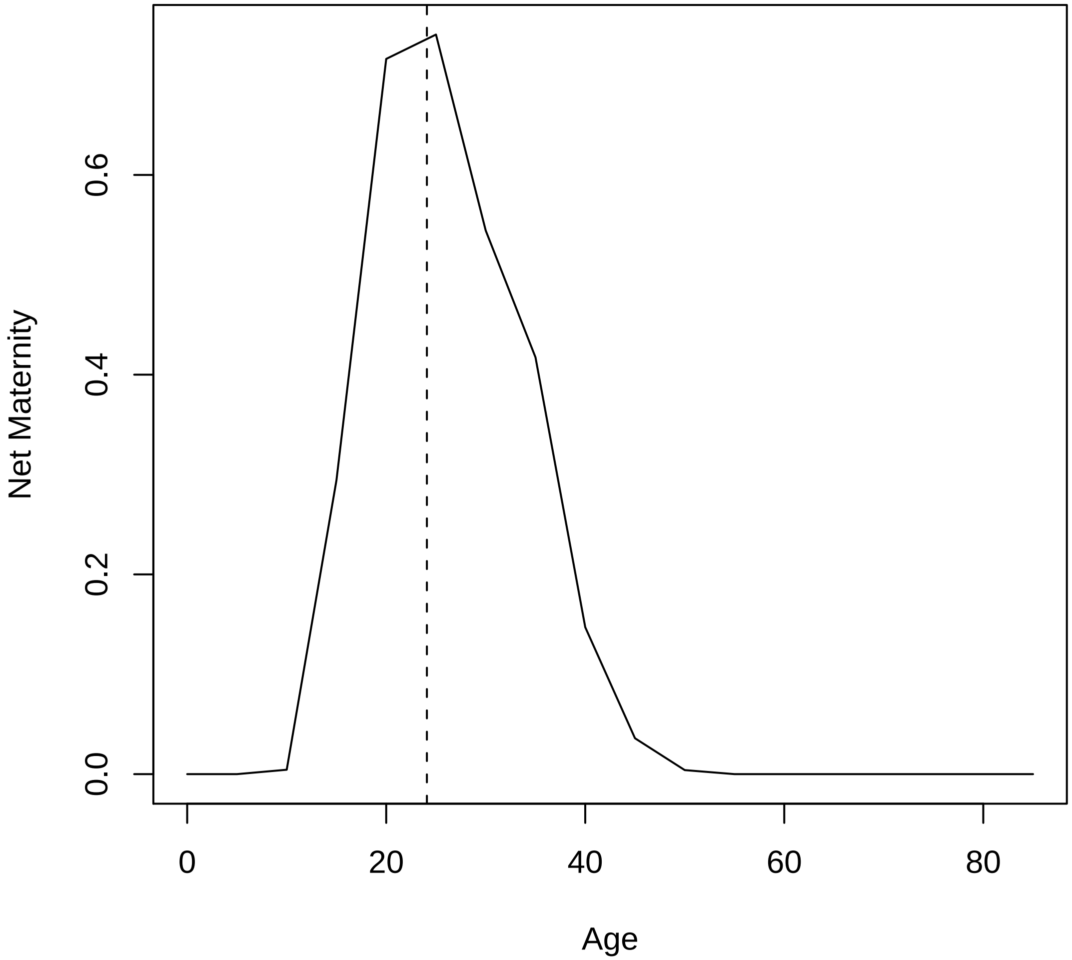
$$\kappa = \int_{\alpha}^{\beta} a\, e^{-ra} l(a) m(a) da$$

This approximation results from the fact that net maternity function is concentrated around κ

The adequacy of this approximation will depend on exactly how concentrated reproduction is around this age

For Uganda (using the WHO model life table) for a 10 year-old girl, this will be $\approx l(36)/l(26)$ which is about 0.84

Mean Age at Birth



Descendants

We use the same logic as we used for calculating the number of living mothers

The number of survivors to age x of a cohort of B girls is $Bl(x)$

in the interval $x + dx$ these girls will bear $B\,l(x)m(x)dx$ daughters

To get the total number of daughters to which the cohort will give birth, we integrate over all possible ages x.

The average number of daughters will be this divided by the number of births B

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

$$= \int_\alpha^\beta l(x)m(x)dx$$

This, of course, is the expected number of daughters of a girl born into a fixed $l(x)m(x)$ regime

It is also the ratio of the current generation to the previous one

Average Number of Daughters Born to a Woman age a

Same cohort of B girls

$Bl(a)$ survivors at age a

Total number of daughters born up to a is

$$\int_{\alpha}^{a} Bl(x)m(x)dx$$

But we don't want all the girls born – only those born to mothers who survived to age a

The fraction of mothers who live from age x to age a is $l(a)/l(x)$

Put it all together

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

Divide by the $Bl(a)$ mothers to get:

$$B_1(a) = \int_{\alpha}^{a} m(x)dx$$

Which is pretty much what we should have expected it to be...

This result can be used for indirect estimation of childhood mortality (Brass 1953)

Older Sisters

When we move laterally in a genealogy, we need to separate the calculations for older and younger siblings

The number of older sisters that a woman has is simply the sum of the number of daughters her mother had prior to her birth

Note that because these daughters are necessarily older than ego, we don't need to condition on her mother's survival

The expected number of daughters born to a woman before age x is

$$B_1(x) = \int_{\alpha}^{x} m(s)ds$$

This is for a girl whose mother was age x when the girl was born

To get the average across all cohorts, we integrate across the age distribution of mothers:

$$S^{(+)}(a) = \int_{\alpha}^{\beta} \left[\int_{\alpha}^{x} m(y)dy \right] e^{-rx} l(x)m(x)dx$$

In order to only count currently living sisters, we need to multiply the inner integral by the probability that the daughter survives to the present $l(a + x - y)$

$$S^{(+)}(a) = \int_{\alpha}^{\beta} \left[\int_{\alpha}^{x} m(y)l(a + x - y)dy \right] e^{-rx} l(x)m(x)dx$$

Younger Sisters

The number of younger sister a girl age a has must incorporate of ego's mother beyond the birth of ego as well as on the age of ego a

If the girl is age a and was born when her mother was age x, the probability that mother lived to age $x + u$ and had another daughter is

$$\frac{l(x+u)}{l(x)} m(x+u)$$

Again, we need to calculate this for all living cohorts by integrating across the age distribution of mothers:

$$S_a^{(-)} = \int_{\alpha}^{\beta} \left[\int_0^a \frac{l(x+u)}{l(x)} m(x+u) du \right] e^{-rx} l(x) m(x) dx$$

Aunts Are Sisters When Seen From Ego's Mother's Perspective

The expected number of maternal aunts who are older than ego's mother is simply the average number of older sisters ego's mother has

$$A_1^{(+)}(a) = \int_\alpha^\beta = S^{(+)}(a+x)W(x|t-a)dx$$

And the expected number of maternal aunts who are younger than ego's mother is simply the average number of younger sisters ego's mother has

$$A_1^{(-)}(a) = \int_\alpha^\beta = S^{(-)}(a+x)W(x|t-a)dx$$

Where if the population is stable, we replace $W(x|t-a)$ by the Lotka weights $e^{-rx}l(x)m(x)dx$

Again if we care about the number of $surviving$ aunts, we would need to multiply the inner integral (i.e., the function $S^{(+)}(a+x)$) by an appropriate survival probability (See the Goodman et al (1974) paper for these)

Rethinking Levi-Strauss

Is the *function* of MBD marriage to reticulate unilineal kindreds and thereby concentrate political power and wealth?

Claude Levi-Strauss argues this in his classic *The Elementary Structures of Kinship*

Hammel (1972) considered the hypothesis that age preferences for marriage partners could bias marriages toward cross-cousin marriage

Consider this extremely simplified model:

- Men marry women 5 years younger than themselves
- All children are born to a couple when the man is 25 and the woman is 20
- All siblings are the same age

Ego is a man age A

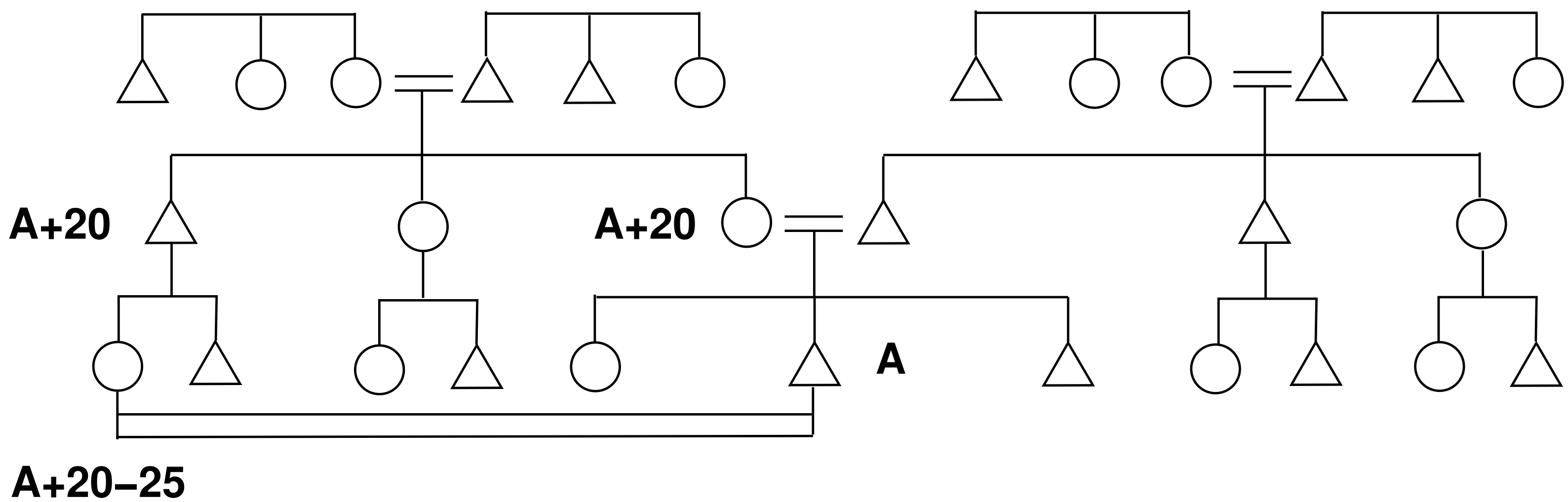
his father is $A + 25$
his mother is $A + 20$

his father's sister is also $A + 25$

her daughter will be $A + 25 - 20 = A + 5$

in contrast, his MBD will be $A + 20 - 25 = A - 5$

Note that both *parallel cousins* (MSD & FBD) will be of the same age as ego


A+20
A+20
A
A+20–25

Applying Hammel's Logic to Orphanhood

Consider the case of orphan adoption

Who is the best relative to be culturally favored foster parent for orphaned children?

One criterion for "best" foster parent is one that is likely to be alive

Identify the child as Ego, age A:

Ego's father's sister is $A + 25$
Ego's mother's mother is $A + 40$
Ego's father's mother is $A + 45$
Ego's mother's sister is $A + 20$

How Does Orphanhood Change with Mean Age of Childbearing?

We have seen from equation 3 that the probability that a girl age a has a living mother is:

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

An approximation of this function is that conditional probability that a mother survives from age of mean childbearing κ and $\kappa + a$:

$$M_1(a) \approx l(\kappa + a)/l(\kappa)$$

Take logarithms of both sides and differentiate with respect to meean age

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

Assume unit time, rearrange

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

Both Parents

Define the probability of a girl age a having a living mother $M_1(a)$ and the probability of having a living father $F_1(a)$

If female and male mortality is independent, then the probability that a girl age a has both parents living is

$$P_1(a) = M_1(a)\, F_1(a) \tag{6}$$

referring back to equation 3, we see that the probability of a girl having a living parent when she is age a is approximately

$$\frac{l(\kappa + a)}{l(\kappa)}$$

where κ is the mean age of childbearing

The probability that a child age a has living parents who were ages $x^{(m)}$ and $x^{(f)}$ respectively when she was born (denote as $p_1(a)$) can be written as a function of the joint age of her parents

$$p_1(a) \approx f(x^{(m)} + a, x^{(f)} + a)$$

Aside: Taylor Series Expansion of a Function of Two Variables

Note that a Taylor series expansion of a the function of two variables $f(x_1, x_2)$ around the point $(\bar{x}_1, \bar{x}_2)$ is:

$$\begin{aligned} f(x_1, x_2) &\approx f(\bar{x_1}, \bar{x}_2) + \left[(x_1 - \bar{x}_1)\frac{\partial f}{\partial x_1} + (x_2 - \bar{x}_2)\frac{\partial f}{\partial x_2}\right] \\ &+ \frac{1}{2!}\left[(x_1 - \bar{x}_1)^2\frac{\partial^2 f}{\partial x_1^2} + 2(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\frac{\partial^2 f}{\partial x_1 \partial x_2} + (x_2 - \bar{x}_2)^2\frac{\partial^2 f}{\partial x_2^2}\right] + \ldots \end{aligned}$$

Taylor Series Expansion of Biparental Survival Function

Expand $p_1(a)$ around the joint mean age of childbearing (κ_m, κ_f) and take expectation (i.e., integrate across all ages)

$$P_1(a) = f(\kappa^{(m)}, \kappa^{(f)}) + \frac{1}{2}\left[\mathrm{Var}(x_m)\frac{\partial^2 f}{\partial x_m^2} + 2\mathrm{Cov}(x_m, x_f)\frac{\partial^2 f}{\partial x_m \partial x_f} + \mathrm{Var}(x_f)\frac{\partial^2 f}{\partial x_f^2}\right]$$

where

$$f(\kappa^{(m)}, \kappa^{(f)}) = \frac{l(\kappa^{(m)} + a)}{l(\kappa^{(m)})} \cdot \frac{l(\kappa^{(f)} + a)}{l(\kappa^{(f)})}$$

What Does This Expansion Tell Us?

As the mean age of childbearing – for either women or men – increases, $P_1(a)$ will decrease

If the second derivatives are positive, variance in either women's or men's ages at childbearing will increase $P_1(a)$

But $l(x)$ is a concave function, so it's second derivatives will be less than zero and variance therefore decreases $P_1(a)$

Negative covariance between the ages of mothers and fathers will increase $P_1(a)$

Positive covariance between the ages of mothers and fathers will decrease $P_1(a)$