Population Projections

We now review in some detail the cohort component method of population projection. We assume that we have population counts by age at time t and we want to project the population to time $t+n$ given a life table and a set of fertility rates. For simplicity we take the length of the projection to be the same as the width of the age groups. (We work with five-year intervals, but we could work with single years just as well.) We assume initially that the population is closed to migration.

Application to Sweden in 1993

The basic inputs are the initial population, the life table person-years lived, and the maternity rates. Box 6.1 in the textbook has the basic inputs for Sweden in 1993, and these are reproduced below and in the companion computing log. The data are given in terms of fertility rates, but I converted to maternity rates dividing by 2.05.

Age	1993 ${}_5N_x^F$	${}_5L_x^F$	${}_5F_x^F$	Survival Ratio[1]	Maternity Rate[2]	1998 ${}_5N_x^F$
0	293,395	497,487	0	0.9950[1]	0	293,574
5	248,369	497,138	0	0.9993	0	293,189
10	240,012	496,901	0	0.9995	0.0029	248,251
15	261,346	496,531	0.0059	0.9993	0.0250	239,833
20	285,209	495,902	0.0443	0.9987	0.0587	261,015
25	314,388	495,168	0.0731	0.9985	0.0639	284,787
30	281,290	494,213	0.0549	0.9981	0.0382	313,782
35	286,923	492,760	0.0215	0.9971	0.0126	280,463
40	304,108	490,447	0.0036	0.9953	0.0019	285,576
45	324,946	486,613	0.0001	0.9922	0.0001	301,731
50	247,613	480,665	0	0.9878	0	320,974
55	211,351	471,786	0	0.9815	0	243,039
60	215,140	457,852	0	0.9705	0	205,109
65	221,764	436,153	0	0.9526	0	204,944
70	223,506	402,775	0	0.9235	0	204,793
75	183,654	350,358	0	0.8699	0	194,419
80	141,990	271,512	0	0.7750	0	142,324
85+	112,424	291,707	0	*0.5179*[1]	0	131,768

[1]Entries are ${}_5L^F{}_x/\ {}_5L^F{}_{x-5}$ except for the first and last rows, which are ${}_5L^F{}_0/5l_0$ and $T^F{}_{85}/T^F{}_{80}$.

[2]Entries are average of ${}_5F^F{}_x$ and ${}_5F^F{}_{x+5}\ {}_5L^F{}_{x+5}/\ {}_5L^F{}_x$. See the text for details

Projecting the Existing Population

Projecting the population above age 5 at time $t+5$ is relatively easy because in a closed population they are just the survivors of the population at time t. All we need is the probability of surviving from age $(x, x+5)$ to age $(x+5, x+10)$, which is ${}_5L_{x+5}/\,{}_5L_x$ from the life table, so

$$ {}_5P_{x+5}^{t+5} = {}_5P_x^t \frac{{}_5L_{x+5}}{{}_5L_x} $$

All rates and survival probabilities are for the female population. I omit the superscript F to avoid clutter.

The only slight complication concerns the open-ended age group, say 85+. This group consists of two subgroups, those 80-84 at baseline who would then be 85-89, and those 85+ at baseline who would be 90+ five years later. The survival probabilities are ${}_5L_{85}/\,{}_5L_{80}$ for the first group and T_{90}/T_{85} for the second. If the last age group is 85+, however, we do not have T_{90}. It is then customary to combine the last two age groups at baseline and project them together (so 80-84 and 85+ are treated as 80+, who then become 85+) using the survival ratio T_{85}/T_{80}, where $T_{80} = {}_5L_{80} + T_{85}$, so

$$ {}_\infty P_{85}^{t+5} = \left({}_5P_{80}^t + {}_\infty P_{85}^t \right) \frac{T_{85}}{{}_5L_{80} + T_{85}} $$

(The textbook describes on page 121 the procedure to be followed when T_{90} is available, but uses the combined projection in Box 6.1, so we'll stick to that.)

The resulting survivals ratios are shown in the first column of the second panel in the table. (Ignore the first entry for now.) This is all you need to project the population above age 5 in 1998.

Projecting Births

The population under age 5 at $t+5$ consists of births during the period from t to $t+5$ who survive to the end of the projection period. Female births, in turn, are obtained by applying the maternity rates to the women exposed to the risk of having a child between t and $t+5$. Thus, the number of girls under 5 at the end of the projection period depends on the initial population of women in the reproductive ages, their survival probabilities, the maternity rates, and the survival probabilities for female births.

There are essentially two ways to proceed, and both yield exactly the same answer. Here I will focus on women and average the rates that apply to them over the projection interval. (The textbook describes an alternative approach that focuses on the rates and averages the women exposed to them.)

Consider, then, women age 15-19 at the start of the projection, who will become 20-24 if they survive. This cohort is exposed to the rates at 15-19 and to the rates at 20-24, with the later discounted by the probability of surviving to be 20-24. The relevant maternity rates are then ${}_5F_{15}$ and ${}_5F_{20}\ {}_5L_{20}/\,{}_5L_{15}$, which we average and multiply by 5, the width of the period. The resulting births have to survive from birth to age 0-4, which occurs with probability ${}_5L_0/5l_0$. (This ratio is shown in the first row of the table.) Putting all of this together, the contribution from women age $(x, x+5)$ to girls under 5 at $t+5$ is

$$ {}_5P_x \frac{5}{2}\left({}_5F_x + {}_5F_{x+5}\frac{{}_5L_{x+5}}{{}_5L_x}\right)\frac{{}_5L_0}{5 l_0} $$

Note that the 5's cancel out; but I left them in for clarity. To obtain the total number of girls under 5 we sum these contributions over all reproductive ages. The results are shown in the last two columns of the table. You now have a complete projection for 1998.

The Leslie Matrix

The calculations required can be laid out conveniently as the result of multiplying a projection matrix by a population vector:

$$ \boldsymbol{p}_{t+5} = \boldsymbol{L}\,\boldsymbol{p}_t $$

Here $\boldsymbol{p}_t$is a column vector with the population at time t in each of the k age groups, and $\boldsymbol{L}$ is a k by k projection matrix known as the *Leslie* matrix, with entries given by the coefficients we have just derived. I will not try to write the entire matrix in symbols (the textbook does so in equation 6.10, although it treats the open-ended group differently by using one more value of T_x than we have here). Instead, you can see it in full glory in the companion computing logs.

The Leslie matrix can be used to project the population repeatedly. If you do that you will find that the population will continue to grow (or decline until it becomes extinct), but the age distribution eventually will stop changing. What's more interesting, the final age distribution depends on the Leslie matrix (and hence on the fertility and mortality rates) but not on the initial age distribution. This result is the jewel in the crown of mathematical demography.

Stable Populations

Stability means that the population at time $t + 5$ is just proportional to the population at time t. In symbols, for sufficiently large t

$$ \boldsymbol{p}_{t+5} = \boldsymbol{L}\,\boldsymbol{p}_t = \lambda \boldsymbol{p}_t $$

If we write $\boldsymbol{p}$ for the population vector in a stable population, the equation becomes simply

$$ (\boldsymbol{L} - \lambda \boldsymbol{I})\boldsymbol{p} = \boldsymbol{0} $$

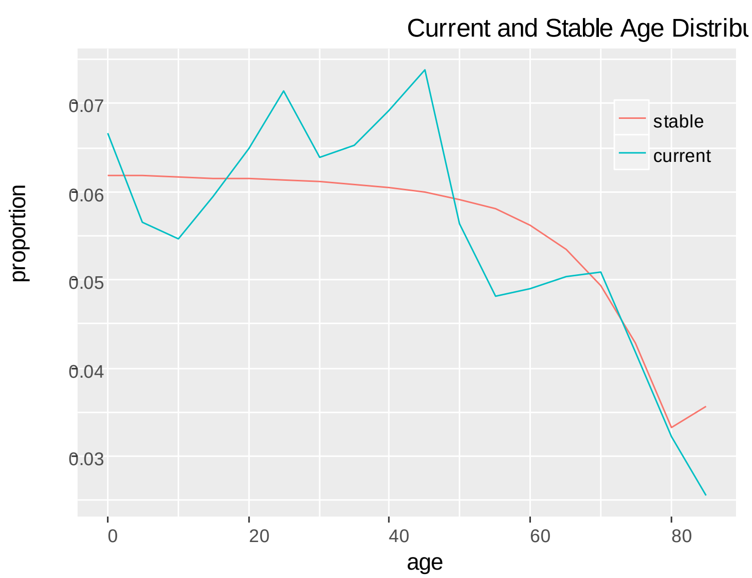
where $\boldsymbol{I}$ is the identity matrix. This is a well-known equation in matrix algebra. It has a solution only if the determinant of $\boldsymbol{L} - \lambda\boldsymbol{I}$ is zero, that is, if

$$ |\boldsymbol{L} - \lambda \boldsymbol{I}| = 0 $$

This equation is called the *characteristic equation* of the matrix $\boldsymbol{L}$. For a real matrix $\boldsymbol{L}$ the equation has n roots λ_i called the *eigenvalues* of $\boldsymbol{L}$. The corresponding vectors $\boldsymbol{p}_i$ are called the *eigenvectors* of $\boldsymbol{L}$.

In demography we are mostly interested in the first eigenvalue and eigenvector. For most (reasonable) Leslie matrices the first eigenvalue and the corresponding eigenvector are real. The eigenvalue is one if the population is stationary; values above and below one indicate growth and decline respectively. The implied instantaneous growth rate is called the *intrinsic growth rate* or *Lotka's* r. The corresponding eigenvector is proportional to the stable age distribution.

Using a matrix algebra package one can compute eigenvalues and eigenvectors directly. For the Leslie matrix in our example the first eigenvalue is λ =1.00111253, which over 5 years is equivalent to a growth rate of $r = \ln(\lambda)/5$ =0.000222383; so at 1993 rates Sweden would end up growing by 0.02% per year. The first eigenvector, scaled so the entries add to one, is shown on the sidebar and the graph below. If you project the Swedish population for about 100 years the age distribution will look very similar to this vector. (And that's about as close as we'll get to proving any of these results.)



	Current	Stable
0	0.0667	0.0619
5	0.0565	0.0618
10	0.0546	0.0617
15	0.0594	0.0616
20	0.0649	0.0614
25	0.0715	0.0613
30	0.0640	0.0611
35	0.0652	0.0608
40	0.0692	0.0605
45	0.0739	0.0600
50	0.0563	0.0592
55	0.0481	0.0580
60	0.0489	0.0562
65	0.0504	0.0535
70	0.0508	0.0494
75	0.0418	0.0429
80	0.0323	0.0332
85+	0.0256	0.0356

When we return to stable population theory we will learn simple yet accurate methods for estimating Lotka's r and the stable age distribution from standard demographic data.

Open Populations

Migration complicates the picture. Out-migration can be handled by introducing death and emigration as competing risks. In-migration is harder to model (particularly if you assume that immigrants may be subject to different fertility and mortality schedules, at least initially) and is usually handled by making assumptions about absolute numbers rather than rates. See Section 6.3.3 in the textbook for a description of how the closed-population procedures can be adapted to handle immigration.