

last revised: 15 February 2009

### 3 Differential Reproduction

#### 3.1 Social mobility revisited

We first discussed intergenerational social mobility in Section 1.4. For simplicity, the model assumed that every parent had exactly one child. Of course, in reality, some parents have multiple children while other parents have none. Perhaps more crucially, the average number of children may vary by social class. That is, there may be *differential reproduction* across social classes.

To address this issue formally, let  $\mathbf{x}_t(i)$  denote the number of parents in social class  $i$  in period  $t$ , let  $r(i)$  denote the average number of children for parents in social class  $i$ , and let  $P(i, j)$  denote the probability that each child of a parent in class  $i$  will occupy class  $j$ . Population dynamics are now given by the system of equations

$$\mathbf{x}_{t+1}(i) = \sum_j \mathbf{x}_t(j)r(j)P(j, i) \quad \text{for all } i \in \{1, \dots, n\}.$$

Intuitively, in period  $t$ , each parent in class  $j$  has (on average)  $r(j)$  children, and each of these children will occupy class  $i$  with probability  $P(j, i)$ . Summing over all classes  $j$ , we thus obtain the total number of individuals in class  $i$  in period  $t + 1$ . Constructing a diagonal “reproduction” matrix

$$R = \begin{bmatrix} r(1) & 0 & \dots & 0 \\ 0 & r(2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r(n) \end{bmatrix},$$

the population dynamics can be rewritten in matrix form as

$$\mathbf{x}_{t+1} = \mathbf{x}_t RP.$$

Note that the reproduction matrix  $R$  is assumed to be fixed over time, just like the transition matrix  $P$ . If we let  $\mathbf{x}_0$  denote the initial (frequency) distribution of the population across social classes, subsequent distributions are given by

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_0 RP \\ \mathbf{x}_2 &= \mathbf{x}_1 RP = \mathbf{x}_0 (RP)^2 \\ \mathbf{x}_3 &= \mathbf{x}_2 RP = \mathbf{x}_0 (RP)^3 \end{aligned}$$

and thus, by induction, we obtain

$$\mathbf{x}_t = \mathbf{x}_0 (RP)^t.$$

In the special case where every parent has exactly one child,  $R$  is the identity matrix, and the population dynamics simplify to those given in section 1.7. However, it is important to recognize that we are now moving beyond Markov chain models. In the general case where  $R$  is not the identity matrix, rows of the matrix  $RP$  are not probability vectors. Consequently,  $\mathbf{x}_t$  is now interpreted as a *frequency* distribution rather than a *probability* distribution. Even if we “normalize” the initial size of the population to 1, so that  $\mathbf{x}_0$  is a probability vector, population size in subsequent periods may be larger or smaller than 1 depending on the reproduction matrix. Furthermore, given our specification of the reproduction process, we are focusing exclusively on deterministic population dynamics, following our “macro” interpretation of the Markov chain model. We might set  $r(i) = 1.5$  to indicate that the average number of children is 1.5 for parents in class  $i$ . But obviously, no particular parent in class  $i$  has 1.5 children, and keeping track of the distribution of family size across parents would require a more elaborate model.<sup>1</sup>

To illustrate, let’s return to the social mobility example from Section 1.7, and consider the effect of differential reproduction.

```
>> x = [.2 .3 .5] % initial distribution of population by social class
```

```
x =
    0.2000    0.3000    0.5000
```

```
>> R = diag([1.5 1.2 .9]) % reproduction matrix
```

```
R =
    1.5000         0         0
         0    1.2000         0
         0         0    0.9000
```

```
>> P = [.6 .4 0; .3 .4 .3; 0 .7 .3] % transition matrix
```

```
P =
    0.6000    0.4000         0
    0.3000    0.4000    0.3000
         0    0.7000    0.3000
```

```
>> x * R % number of children by parent’s social class
```

```
ans =
    0.3000    0.3600    0.4500
```

```
>> x * R * P % number of children by child’s class
```

```
ans =
```

---

<sup>1</sup>While we are moving beyond the Markov chain framework, note that the model remains *linear* in the sense discussed in the (not yet written) Introduction. That is, the model can be written in the form  $\mathbf{x}_{t+1} = \mathbf{x}_t A$  where  $A$  is matrix of coefficients that is constant over time.

```

    0.2880    0.5790    0.2430
>> sum(ans)    % size of population after 1 generation
ans =
    1.1100
>> x*(R*P)^2    % frequency distribution after 2 generations
ans =
    0.4676    0.6038    0.2741
>> sum(ans)    % population size after 2 generations
ans =
    1.3455

```

This example assumes that lower-class parents average  $r(1) = 1.5$  children, middle-class parents average  $r(2) = 1.2$  children, and upper-class parents average  $r(3) = 0.9$  children. Having normalized the initial distribution so that population size is 1, we see that the population grows over the next 2 periods from 1.0 to 1.11 to 1.34.

To see what happens over a longer time span, we can compute population distributions for the first 15 generations using a for loop.

```

>> y = [ ]; for t = 0:15; y = [y; [.2 .3 .5]*(R*P)^t]; end
>> y    % distribution over classes for 15 generations
y =
    0.2000    0.3000    0.5000
    0.2880    0.5790    0.2430
    0.4676    0.6038    0.2741
    0.6382    0.7431    0.2914
    0.8419    0.9232    0.3462
    1.0901    1.1664    0.4258
    1.4010    1.4822    0.5349
    1.7944    1.8890    0.6780
    2.2950    2.4105    0.8631
    2.9333    3.0778    1.1008
    3.7480    3.9309    1.4052
    4.7883    5.0209    1.7945
    6.1170    6.4136    2.2920
    7.8142    8.1927    2.9277
    9.9821   10.4655    3.7399
   12.7515   13.3688    4.7773
>> s = sum(y,2)    % population size in each generation
s =
    1.0000

```

```

1.1100
1.3455
1.6727
2.1113
2.6823
3.4180
4.3614
5.5686
7.1120
9.0841
11.6037
14.8226
18.9346
24.1875
30.8977

```

From these computations, we merely see that the population is growing. However, a more interesting pattern emerges once we compute the growth factor for each period (dividing population size in period  $t$  by the population size in period  $t - 1$ ), and normalize each frequency distribution so that it becomes a probability vector.

```
>> s(2:16)./s(1:15)    % growth factor for each period
```

```
ans =
1.1100
1.2122
1.2432
1.2622
1.2704
1.2743
1.2760
1.2768
1.2771
1.2773
1.2774
1.2774
1.2774
1.2774
1.2774
1.2774

```

```
>> z = [y(:,1)./s y(:,2)./s y(:,3)./s]    % proportion of population in each class
```

```
z =
0.2000    0.3000    0.5000
0.2595    0.5216    0.2189
0.3476    0.4488    0.2037
0.3816    0.4442    0.1742
0.3988    0.4373    0.1640
0.4064    0.4348    0.1588
0.4099    0.4336    0.1565

```

0.4114	0.4331	0.1555
0.4121	0.4329	0.1550
0.4124	0.4328	0.1548
0.4126	0.4327	0.1547
0.4127	0.4327	0.1547
0.4127	0.4327	0.1546
0.4127	0.4327	0.1546
0.4127	0.4327	0.1546
0.4127	0.4327	0.1546

Thus, in the long run, the population growth rate converges to 27.74%, while the probability distribution over social classes is converging to an outcome with 41.27% lower class, 43.27% middle class, and 15.46% lower class.

### 3.2 Eigenvalues and eigenvectors

The preceding computations illustrate a general result that we will discuss in Section 3.4. But it may be useful first to review some concepts from matrix algebra. Given a square matrix  $A$ , a column vector  $\mathbf{v}$ , and a scalar  $\lambda$  which satisfy the equation

$$A\mathbf{v} = \lambda\mathbf{v},$$

we say that  $\lambda$  is an *eigenvalue* of  $A$ , and that  $\mathbf{v}$  is the corresponding *eigenvector*. It is apparent that the scaling of eigenvectors is arbitrary, since doubling every element of  $\mathbf{v}$  would simply double both sides of the equation. Alternatively, we could multiply every element of  $\mathbf{v}$  by  $-1$ , or we could normalize  $\mathbf{v}$  so that it becomes a probability vector. Nevertheless, for any eigenvalue of  $A$ , the associated eigenvector is unique up to the scaling factor.<sup>2</sup>

If  $A$  is an  $n \times n$  matrix, then it has  $n$  eigenvalues (not necessarily distinct), each associated with an eigenvector (arbitrarily scaled). Arranging the  $n$  eigenvectors as the columns of the matrix  $V$ , and placing the  $n$  eigenvalues along the main diagonal of the diagonal matrix  $L$ , we can thus write

$$AV = VL.$$

Eigenvalues and eigenvectors are obtained in Matlab using the `eig` function. To illustrate, the eigenvalues and eigenvectors of the matrix  $A = (RP)'$  are given below. (The rationale for using this matrix will become apparent in the next section.)

```
>> A = (R*P)' % transpose of the RP matrix

A =
    0.9000    0.3600         0
    0.6000    0.4800    0.6300
```

---

<sup>2</sup>Matlab scales eigenvectors so that the inner product  $\mathbf{v}'\mathbf{v} = 1$ .

```

0    0.3600    0.2700

>> [V,L] = eig(A) % columns of V are eigenvectors; entries of L are eigenvalues

V =
-0.6682    0.5766    0.2515
-0.7006   -0.5252   -0.7681
-0.2503   -0.6259    0.5889

L =
 1.2774         0         0
      0    0.5721         0
      0         0   -0.1995

```

To verify these Matlab results, we can begin with the first eigenvalue and eigenvector, checking whether  $A\mathbf{v} = \lambda\mathbf{v}$  for this pair  $(\lambda, \mathbf{v})$ . We can then consider all eigenvalues and eigenvectors simultaneously (checking whether  $AV = VL$ ).

```

>> A * V(:,1)

ans =
-0.8536
-0.8949
-0.3198

>> L(1,1) * V(:,1)

ans =
-0.8536
-0.8949
-0.3198

>> A*V

ans =
-0.8536    0.3298   -0.0502
-0.8949   -0.3005    0.1532
-0.3198   -0.3581   -0.1175

>> V*L

ans =
-0.8536    0.3298   -0.0502
-0.8949   -0.3005    0.1532
-0.3198   -0.3581   -0.1175

```

Alert readers will have noticed that the largest eigenvalue of the  $A$  matrix ( $= 1.2774$ ) is equal to the long-run growth factor computed in the preceding section. We'll see why in Section 3.4. But first, we need one more result from matrix algebra.

### 3.3 The Perron-Frobenius Theorem

In general, both eigenvalues and the elements of eigenvectors may be real (positive, zero, or negative) or complex (having both real and imaginary components). However, for the special case where the  $A$  matrix is primitive, we have the following result.

**Perron-Frobenius Theorem.** *If  $A$  is a non-negative, primitive matrix, then (i) one of its eigenvalues is positive and is greater (in absolute value)<sup>3</sup> than all the other eigenvalues, and (ii) there is a positive eigenvector (i.e., an eigenvector with positive elements) corresponding to that eigenvalue.*

The particular eigenvalue described in this theorem is often called the *dominant* eigenvalue, for reasons that will become apparent in the next section.

For the present example, it is easy to see that the  $A$  matrix is primitive by computing the  $A^2$  matrix.

```
>> A^2  
  
ans =  
    1.0260    0.4968    0.2268  
    0.8280    0.6732    0.4725  
    0.2160    0.2700    0.2997
```

Thus, as guaranteed by the Perron-Frobenius Theorem, we see that dominant eigenvalue of  $A$  is positive ( $= 1.2774$ ) and has a greater absolute value than the other eigenvalues.<sup>4</sup> Further, we see that the associated eigenvalue can be rescaled so that all elements are positive. In the Matlab computations above, all of the elements of the associated eigenvector are negative. Thus, we multiply by  $-1$  to obtain a positive eigenvector.

```
>> V(:,1) * -1    % to obtain a positive eigenvector  
  
ans =  
    0.6682  
    0.7006  
    0.2503
```

---

<sup>3</sup>For complex eigenvalues, which can be written as  $\lambda = a + bi$  where  $i = \sqrt{-1}$ , the absolute value of  $\lambda$  equals  $\sqrt{a^2 + b^2}$ . While all eigenvalues are real in the present example, we will return to this issue in Ch xx.

<sup>4</sup>For this example, Matlab has fortuitously listed the eigenvalues and eigenvectors from largest to smallest, beginning with the dominant eigenvalue. More generally, Matlab does not always list the dominant eigenvalue first. So you should pay close attention to the output of the `eig` function when searching for the dominant eigenvalue and the associated eigenvector.

### 3.4 The stable-growth equilibrium

For our present purposes, the Perron-Frobenius Theorem is important because it underpins the following result.

**Theorem 2.** *Suppose that population dynamics are given by  $\mathbf{x}_t = \mathbf{x}_0(RP)^t$  where  $RP$  is a primitive matrix. As  $t$  becomes large, the population reaches a “stable growth equilibrium” determined by  $\mathbf{x}\lambda = \mathbf{x}RP$  where  $\lambda$  is the dominant eigenvalue of  $(RP)'$  and  $\mathbf{x}$  is the associated eigenvector. The dominant eigenvalue  $\lambda$  is the population growth factor. The associated eigenvector  $\mathbf{x}$ , normalized as a probability vector, is the limiting distribution of the population across social classes.*

Before attempting to explain this result, let's first illustrate it. For the present example, we have already seen that the population growth factor converges to the dominant eigenvalue of  $A$  (equal to 1.2774). To show that the limiting distribution is given by the corresponding eigenvector, we simply normalize so that this eigenvector becomes a probability vector.

```
>> V(:,1)'/sum(V(:,1)) % the limiting distribution
```

```
ans =  
    0.4127    0.4327    0.1546
```

Thus, as we have already seen, the population will eventually reach a stable growth equilibrium in which 41.27% are lower class, 43.27% are middle class, and 15.46% are lower class.

To illustrate further, let's return to the original social mobility example from Chapter 1, which assumed that every parent has one child (and hence  $R = I$ ). Having originally obtained the limiting distribution through iterated multiplication of the  $P$  matrix, Theorem 2 tells us that we can also obtain this distribution by computing the eigenvectors and eigenvalues of the  $P'$  matrix.

```
>> [V,L] = eig(P')
```

```
V =  
   -0.5676    0.8090    0.3090  
   -0.7568   -0.3090   -0.8090  
   -0.3243   -0.5000    0.5000
```

```
L =  
    1.0000         0         0  
         0    0.4854         0  
         0         0   -0.1854
```

```
>> V(:,1)'/sum(V(:,1)) % the limiting distribution
```

```
ans =  
    0.3443    0.4590    0.1967
```



Note that the dominant eigenvalue of  $P'$  is equal to 1. Adopting the “macro” interpretation of the Markov chain, this makes intuitive sense: population size is constant because each parent has exactly one child. Indeed, it can be shown formally that the dominant eigenvalue of  $P'$  is 1 for any probability transition matrix  $P$ .<sup>5</sup> Consequently, the stable growth condition  $\mathbf{x}\lambda = \mathbf{x}RP$  becomes  $\mathbf{x} = \mathbf{x}P$ , we can see that Theorem 1 (from Chapter 1) is simply a special case of Theorem 2.

Having offered some illustrations, we can now examine why Theorem 2 holds. We’ll start by considering why it involves the *transpose* of the  $RP$  matrix. In Theorem 2, the stable-growth condition is written as  $\mathbf{x}\lambda = \mathbf{x}RP$  where  $\mathbf{x}$  is a *row* vector. In contrast, the eigenvalue equation from the preceding section was written as  $A\mathbf{v} = \lambda\mathbf{v}$  where  $\mathbf{v}$  is a *column* vector. Recall from matrix algebra that  $(FG)' = G'F'$  for any two matrices  $F$  and  $G$  of the proper sizes. Thus, the stable-growth condition can be rewritten as

$$(RP)'\mathbf{x}' = \lambda\mathbf{x}'$$

where  $\lambda$  is an eigenvalue of  $(RP)'$  and the column vector  $\mathbf{x}'$  is the associated eigenvector.<sup>6</sup> This explains why we used the  $(RP)'$  matrix rather than the  $RP$  matrix to illustrate eigenvalues and eigenvectors in the preceding section. Relatedly, for the remainder of this section, it will be convenient to write the population dynamics in the form

$$\mathbf{x}_t = A^t\mathbf{x}_0$$

where  $A = (RP)'$  and  $\mathbf{x}_t$  is now specified as a *column* vector.

Next, we can rewrite the population dynamics equation to make explicit the role played by the eigenvalues and eigenvectors of the  $A$  matrix. The “trick” is to write the initial condition  $\mathbf{x}_0$  as a weighted sum of the eigenvectors of  $A$ . That is, we need to solve for constants  $c_1, c_2, \dots, c_n$  such that

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$$

where the column vector  $\mathbf{v}_i$  is the eigenvector corresponding to the  $i$ th largest eigenvalue of  $A$ . Organizing the eigenvectors into the matrix  $V$ , and the constants into a column vector  $\mathbf{c}$ , this equation may be rewritten as

$$\mathbf{x}_0 = V\mathbf{c}$$

and hence the vector of constants is determined by the equation

$$\mathbf{c} = V^{-1}\mathbf{x}_0.$$

Given this specification of the initial condition, we now obtain

$$\mathbf{x}_1 = A\mathbf{x}_0 = AV\mathbf{c} = VL\mathbf{c}$$

---

<sup>5</sup>Upper and lower bounds on the dominant eigenvalue of a non-negative matrix  $A$  can be derived from the row and column sums of  $A$ . See Farina and Rinaldi (2000, p 36, Theorem 11). In particular, if every column sum (or every row sum) of  $A$  equals 1, then the dominant eigenvalue of  $A$  is 1.

<sup>6</sup>Perhaps this explains why demographers specify transition matrices “backwards.” See Chap 9.

where the last substitution makes use of the equation  $AV = VL$ . The distributions for subsequent periods are given by

$$\mathbf{x}_2 = A\mathbf{x}_1 = AVL\mathbf{c} = VL^2\mathbf{c}$$

$$\mathbf{x}_3 = A\mathbf{x}_2 = AVL^2\mathbf{c} = VL^3\mathbf{c}$$

and thus by induction we obtain

$$\mathbf{x}_t = VL^t\mathbf{c}.$$

To verify that this specification of the population dynamics generates the same results as our initial specification  $\mathbf{x}_t = \mathbf{x}_0(RP)^t$ , consider again the numerical example from Section 3.1. We'll solve first for the vector of constants and then compute frequency distributions for several periods.

```
>> A = (R*P)'; % A is the transpose of the (RP) matrix
>> [V,L] = eig(A); % displayed previously
>> x0 = [.2 .3 .5]' % initial condition (as column vector)

x0 =
    0.2000
    0.3000
    0.5000

>> c = inv(V) * x0 % solving for constants
c =
   -0.4848
   -0.3385
    0.2831

>> V * c % to verify solution
ans =
    0.2000
    0.3000
    0.5000

>> V * L * c % distribution in first generation

ans =
    0.2880
    0.5790
    0.2430

>> V * L^2 * c % distribution in second generation

ans =
    0.4676
```

```

0.6038
0.2741

>> V * L^3 * c    % distribution in third generation

ans =
0.6382
0.7431
0.2914

```

The frequency distributions are now specified as column vectors rather than row vectors. But otherwise, these results are identical to those computed in Section 3.1.

Finally, using this alternative specification of population dynamics, it becomes possible to see why the dominant eigenvalue and the corresponding eigenvector play such an important role in Theorem 2. Recognizing that  $L$  is a diagonal matrix,

$$L^t = \begin{bmatrix} \lambda_1^t & 0 & \dots & 0 \\ 0 & \lambda_2^t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^t \end{bmatrix} = \lambda_1^t \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & (\lambda_2/\lambda_1)^t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\lambda_n/\lambda_1)^t \end{bmatrix}.$$

Recall that the eigenvalues have been arranged in order of their absolute value, with  $\lambda_1$  being the dominant eigenvalue guaranteed by the Perron-Frobenius Theorem. Because  $\lambda_1$  is greater in absolute value than the other eigenvalues,

$$(\lambda_i/\lambda_1)^t \rightarrow 0 \quad \text{as } t \rightarrow \infty \quad \text{for all } i \neq 1.$$

Consequently, for  $t$  sufficiently large,

$$\mathbf{x}_t = VL^t\mathbf{c} \approx \lambda^t V \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \mathbf{c} = \lambda^t V \begin{bmatrix} c_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \lambda_1^t c_1 \mathbf{v}_1.$$

Thus, while all of the eigenvalues of  $A$  affect population dynamics in the short run, the largest eigenvalue  $\lambda_1$  becomes “dominant” in the long run. For all practical purposes, the approximation is exact for  $t$  sufficiently large. Consequently, to show that the dominant eigenvalue is the long-run growth factor, we can write

$$\mathbf{x}_{t+1} = \lambda_1^{t+1} c_1 \mathbf{v}_1 = \lambda_1 (\lambda_1^t c_1 \mathbf{v}_1) = \lambda_1 \mathbf{x}_t.$$

From the Perron-Frobenius Theorem, we also know that all of the elements of the associated eigenvector  $\mathbf{v}_1$  will be positive. Recognizing that the constant  $c_1$  merely rescales this eigenvector, the frequency distribution in each period (and hence the limiting distribution) must be proportional to  $\mathbf{v}_1$ .

### 3.5 The intergenerational dynamics of IQ scores

Our social mobility examples have assumed social classes based upon occupation or income. However, many other applications of the differential reproduction model are possible. For instance, we might partition the population into classes based on intelligence (as measured by IQ score). Given the average number of children born to parents in each IQ class, as well the (probabilistic) link between parent's IQ class and child's IQ class, we might then consider how the distribution of the population across IQ classes would change over time.

While the data are not in a form that we can use directly, Table 3.1 gives some

**Table 3.1 Heritability of IQ scores**

**Mean of child's IQ by parent's IQ class**

		<i>mother's IQ class</i>						
		<i>0-74</i>	<i>75-84</i>	<i>85-94</i>	<i>95-105</i>	<i>106-15</i>	<i>116-25</i>	<i>126+</i>
<i>father's</i> <i>IQ class</i>	<i>0-74</i>	74.4	90.3	103.3	101.1	97.5	97.8	113.6
	<i>75-84</i>	82.4	89.9	102.2	101.3	102.5	113.9	109.0
	<i>85-94</i>	91.1	99.6	102.4	106.0	107.8	108.3	109.4
	<i>95-105</i>	93.8	99.2	106.0	108.2	108.9	111.6	113.6
	<i>106-15</i>	104.0	102.0	104.1	109.1	112.9	113.4	116.1
	<i>116-25</i>	98.3	113.9	101.9	110.9	114.2	108.9	115.8
	<i>126+</i>	113.6	107.4	113.0	113.1	114.9	115.2	123.5

**Standard deviation of child's IQ by parent's IQ class**

		<i>mother's IQ class</i>						
		<i>0-74</i>	<i>75-84</i>	<i>85-94</i>	<i>95-105</i>	<i>106-15</i>	<i>116-25</i>	<i>126+</i>
<i>father's</i> <i>IQ class</i>	<i>0-74</i>	19.7	17.6	13.0	16.0	13.0	22.7	10.8
	<i>75-84</i>	15.6	12.0	10.0	16.6	12.0	9.5	3.0
	<i>85-94</i>	16.1	10.1	16.7	10.2	10.5	12.3	4.8
	<i>95-105</i>	16.7	11.2	11.2	11.4	11.0	11.1	6.3
	<i>106-15</i>	13.0	12.0	18.7	12.0	10.8	12.4	13.8
	<i>116-25</i>	12.0	9.5	13.2	11.5	9.2	22.2	10.7
	<i>126+</i>	10.8	17.2	9.0	9.6	11.0	13.3	13.0

Source: Samuel H Preston and Cameron Campbell (1993) "Differential Fertility and the Distribution of Traits: The Case of IQ," *American Journal of Sociology* 98:997-1043. See Tables 2 and 3, p 1002. From data originally collected by Reed and Reed (1965).

indication of the empirical relationship between parents' IQ and child's IQ. From the top panel, we see that the mean of the child's IQ tends to rise as either the mother or father is drawn from a higher IQ class, and that the impact of the father and the mother is roughly symmetric. The bottom panel gives some indication of the level of intergenerational "mobility" between IQ classes.<sup>7</sup> We cannot recover from this table the precise transition probabilities into each IQ class that the child might occupy (although we will derive some estimates below). Nevertheless, we can see that there is considerable intergenerational mobility. In particular, even if both parents belong to the same IQ class, the child could belong to a higher or lower IQ class.

The data in Table 3.1 are interesting in their own right. For instance, we might examine whether intergenerational mobility across IQ classes is closer to the pole of perfect mobility or zero mobility. However, following Preston and Campbell (1993), we'll shift our focus to the population level, addressing some long-standing concerns regarding the dynamics of the IQ distribution when there is differential reproduction. In many countries (including the US), parents from lower IQ classes tend to have more children than parents from higher IQ classes. This has prompted some commentators – both inside and outside academia – to worry that the average IQ of the population will necessarily fall over time. Implicitly, their intuitions seem to follow from the zero-mobility case where each child always belongs to the same IQ class as the parent. In this simple case, it is straightforward to show that the percentage of the population in the IQ class with the highest reproduction rate will indeed grow over time, converging to 100% in the long run.<sup>8</sup>

Should we worry about a downward trend in IQ scores? Obviously, real-world population processes are more complex than the simple models we've considered. Nevertheless, our differential reproduction model offers some assurance that fears of this worst-case scenario are unfounded. Even though we cannot recover the precise intergenerational transition probabilities, the data in Table 3.1 strongly suggest that the transition matrix is primitive. Consequently, Theorem 2 implies that the population will reach a stable growth equilibrium, and the Perron-Frobenius Theorem guarantees that all elements of the limiting distribution are positive. Thus, in the long-run, not everyone in the population will belong to the lowest class.

Attempting to illustrate this result using the data in Table 3.1, two obstacles arise immediately. First, while child's IQ depends on both mother's and father's IQ, our present model maintains the fiction that each child has only one parent. We will return to this issue in Chapter xx where we develop a non-linear "two-sex" version of the differential reproduction model. But for now, we'll adopt a simpler approach. If we assume that marriage always occurs between individuals within the same class – that marriage is completely *endogamous* – then the two-sex model reduces to a

---

<sup>7</sup>Preston and Campbell (1993, p 1002, Table 3) indicate that the entries in this table are variances. But from the mobility tables presented in Reed and Reed (1965), it seems clear to me that the entries must be standard deviations.

<sup>8</sup>I'll leave this simple exercise for the reader.

one-sex model. Essentially, we are restricting attention to the main diagonals of the two panels in Table 3.1.

Second, as already indicated, the data are not presented in a form that we can use directly. For instance, while we can see that parents in the 0-74 IQ class have children whose IQ has mean 74.4 and standard deviation 19.7, we cannot recover the precise intergenerational transition probabilities from the 0-74 class to each other class. However, if we are willing to assume that these IQ scores are normally distributed, it is straightforward to estimate transition probabilities from each class. Let  $\mu(i)$  and  $\sigma(i)$  denote the mean and standard deviation of the child's IQ given parents in class  $i$ . Further let  $b(i)$  denote the lower bound on IQ in class  $i$ .<sup>9</sup> Estimated intergenerational transition probabilities can then be written as

$$P(i, j) = F\left(\frac{b(j+1) - \mu(i)}{\sigma(i)}\right) - F\left(\frac{b(j) - \mu(i)}{\sigma(i)}\right)$$

where  $F$  denotes the cumulative distribution function (cdf) for the standard normal distribution. Using Matlab, we thus obtain the estimated transition matrix below.

```
>> mu = [74.4 89.9 102.4 108.2 112.9 108.9 123.5];
>> sigma = [19.7 12.0 16.7 11.4 10.8 22.2 13.0];
>> b = [-inf 75 85 95 106 116 126 inf];
>> F = inline('0.5 * erfc(-x/sqrt(2))') % standard normal cdf
F =
    Inline function:
    F(x) = 0.5 * erfc(-x/sqrt(2))
>> P = []; for i = 1:7; for j = 1:7; ...
P(i,j) = F((b(j+1)-mu(i))/sigma(i)) - F((b(j)-mu(i))/sigma(i)); ...
end; end; P
P =
    0.5121    0.1926    0.1474    0.0935    0.0370    0.0130    0.0044
    0.1072    0.2343    0.3231    0.2456    0.0750    0.0135    0.0013
    0.0504    0.0983    0.1801    0.2565    0.2069    0.1289    0.0788
    0.0018    0.0191    0.1025    0.3000    0.3296    0.1877    0.0592
    0.0002    0.0047    0.0438    0.2127    0.3515    0.2745    0.1126
    0.0634    0.0775    0.1248    0.1824    0.1774    0.1540    0.2206
    0.0001    0.0014    0.0126    0.0749    0.1929    0.2943    0.4238
```

It is immediately apparent that this matrix is primitive because every element is positive. Substantively, while large “jumps” may occur only rarely, it is possible for parents in any IQ class  $i$  to have a child in any IQ class  $j$ . More crucially, small

---

<sup>9</sup>Because the normal distribution has support  $[-\infty, \infty]$ , we set  $b(1) = -\infty$  and  $b(n+1) = \infty$ .

jumps – transitions between adjacent IQ classes – are quite common. Thus, even if large jumps never occurred, the matrix would still be primitive because transitions from any class  $i$  to any class  $j$  could occur over multiple generations.

Following the original analysis by Preston and Campbell (1993), let's now specify some hypothetical reproduction matrices and examine the consequences for the limiting distribution. In particular, we will examine one case where

$$r(i) = 0.5i \quad \text{for } i \in \{1, \dots, 7\}$$

so that family size is positively related to parent's IQ, and a second case where

$$r(i) = 4 - 0.5i \quad \text{for } i \in \{1, \dots, 7\}$$

so that family size is negatively related to parent's IQ. In reality, the negative relationship between these variables is much weaker than assumed in the second case. However, these extreme examples will be useful for illustrating the theoretical argument.

```
>> r = .5*(1:7); R = diag(r)    % positive relationship between reproduction and IQ
```

```
R =
    0.5000         0         0         0         0         0         0
         0    1.0000         0         0         0         0         0
         0         0    1.5000         0         0         0         0
         0         0         0    2.0000         0         0         0
         0         0         0         0    2.5000         0         0
         0         0         0         0         0    3.0000         0
         0         0         0         0         0         0    3.5000
```

```
>> A = (R*P)'; [V,L] = eig(A)
```

```
V =
 -0.0509  -0.0450  -0.3881   0.2480  -0.2832   0.1224   0.2643
 -0.0728  -0.0889  -0.4444  -0.3003  -0.0018   0.4885   0.3074
 -0.1644  -0.2462  -0.5344  -0.6021  -0.3831  -0.6937   0.3066
 -0.3950  -0.5730  -0.0804  -0.2690  -0.5369   0.4531   0.2835
 -0.5449  -0.5111   0.4858   0.5023   0.4997  -0.1267  -0.1175
 -0.5070  -0.0260   0.2771   0.3434   0.4298  -0.1720  -0.7485
 -0.5050   0.5825  -0.2166  -0.2076  -0.2243   0.1194   0.2954
```

```
L =
 2.5828         0         0         0         0         0         0
         0    1.0398         0         0         0         0         0
         0         0    0.3474         0         0         0         0
         0         0         0    0.2029         0         0         0
         0         0         0         0    0.0766         0         0
         0         0         0         0         0    0.0009         0
         0         0         0         0         0         0   -0.0660
```

```

>> V(:,1)'/sum(V(:,1)) % limiting distribution

ans =
    0.0227    0.0325    0.0734    0.1763    0.2433    0.2263    0.2254

>> r = 4 - .5*(1:7); R = diag(r) % negative relationship between reproduction and IQ

R =
    3.5000         0         0         0         0         0         0
         0    3.0000         0         0         0         0         0
         0         0    2.5000         0         0         0         0
         0         0         0    2.0000         0         0         0
         0         0         0         0    1.5000         0         0
         0         0         0         0         0    1.0000         0
         0         0         0         0         0         0    0.5000

>> A = (R*P)'; [V,L] = eig(A)

V =
 -0.4355  -0.1332  -0.0521  -0.0101  -0.0071  -0.0146  -0.0148
 -0.2966   0.0202   0.2041   0.0006   0.0530  -0.1380  -0.0545
 -0.3969   0.1751   0.2660  -0.0704  -0.2010   0.3594  -0.0053
 -0.5179   0.5117  -0.1030  -0.1844  -0.2156  -0.3892  -0.1337
 -0.4459   0.6389  -0.5687   0.0129   0.5390   0.1846   0.0797
 -0.2771   0.4582  -0.5703   0.3899   0.3323   0.4181   0.7167
 -0.1448   0.2667  -0.4750   0.8993  -0.7136  -0.7009  -0.6774

L =
    2.1714         0         0         0         0         0         0
         0    1.3447         0         0         0         0         0
         0         0    0.5942         0         0         0         0
         0         0         0    0.2700         0         0         0
         0         0         0         0    0.0728         0         0
         0         0         0         0         0    0.0017         0
         0         0         0         0         0         0   -0.0158

>> V(:,1)'/sum(V(:,1)) % limiting distribution

ans =
    0.1732    0.1180    0.1578    0.2060    0.1773    0.1102    0.0576

```

Obviously, the reproduction matrix does affect the limiting distribution. In the first case, where IQ is positively related to reproduction, the limiting distribution is skewed towards the higher IQ classes. Conversely, in the second case, where IQ and reproduction are negatively related, the limiting distribution is skewed toward the lower classes. Nevertheless, even with these extreme assumptions on the reproduction matrix, we see that every IQ class has a positive share in the limiting distribution. Given more realistic assumptions on the reproduction matrix, we would obtain a limiting distribution between these extremes. Thus, given some understanding of



the Perron-Frobenius Theorem, we might tentatively conclude that fears about the dynamics of the IQ distribution are overblown.<sup>10</sup>

### 3.6 Further reading

See any matrix algebra textbook for further discussion of eigenvectors and eigenvalues. Back in the day, I learned matrix algebra from Strang (Academic Press, 1980), which remains one of my standard references. My statement of the Perron-Frobenius Theorem is drawn from the online supplement to Leon (Linear Algebra with Applications, 5th edition), posted at

[www.prenhall.com/divisions/esm/app/ph-linear/leon/html/perron.html](http://www.prenhall.com/divisions/esm/app/ph-linear/leon/html/perron.html)

Some of my analysis of the limiting distribution follows Allman and Rhodes (Cambridge, 2004, Chapter 2.3). More generally, Farina and Rinaldi (Positive Linear Systems, Wiley, 2000) provide a rigorous (but still readable) introduction to linear systems, as well as applications including Markov chains and demography.

My discussion of the intergenerational dynamics of IQ scores draws heavily on the original paper by Preston and Campbell (AJS 1993). The heritability data was originally presented by Higgins, et al (Social Biology 1962) and Reed and Reed (1965). For one empirical analysis reporting a negative correlation between parent's IQ and family size, see Retherford and Sewell (Social Biology 1988). For one commentator concerned about the effect of differential reproduction on the IQ distribution, see Herrnstein (Atlantic Monthly 1989). Lam (AJS 1993) and Coleman (AJS 1993) offer some sophisticated critiques of the Preston and Campbell (AJS 1993) paper.

---

<sup>10</sup>As indicated above, we'll return to this topic in Chapter xx, and examine whether a two-sex version of the model could lead to different conclusions.