3.1 Population dynamics with age distribution

When the mathematical model for population dynamics is applying to a human population, more often than not the age of the individuals in a population matters, for example one’s social behavior (in sociology), one’s health conditions (in public health studies), or one’s risk for a car accident (in insurance management), etc. In all these situations, one is interested in the populations of individuals in each and every age group, and how this distribution changes with time.

Let \( N_k(n) \) be the population size, e.g., the number of individuals, with age \( k \) in the \( n^{th} \) year. Then clearly without any birth and death, there is a “shift” of age from a year to the next year. Combining this with death in each sub-population with probability of death per capita per year \( d_k \), one has a dynamic equation

\[
N_k(n + 1) = N_{k-1}(n) - d_k N_k(n), \quad n \geq 0,
\]

with the new borns which are considered as age 0:

\[
N_0(n + 1) = \sum_{\ell=1}^{\infty} b_\ell N_\ell(n),
\]

where \( b_\ell \) is per year number of births of each individual with age \( \ell \). We see that Eqs. (3.1) and (3.2) together can be expressed in terms of a matrix equation

\[
\begin{bmatrix}
N_0 \\
N_1 \\
N_2 \\
\vdots \\
N_K
\end{bmatrix}(n + 1) =
\begin{bmatrix}
0 & b_1 & b_2 & \cdots & b_K \\
1 & -d_1 & 0 & \cdots & 0 \\
0 & 1 & -d_2 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \ddots & 0 \\
0 & 0 & \cdots & 1 & -d_K
\end{bmatrix}
\begin{bmatrix}
N_0 \\
N_1 \\
N_2 \\
\vdots \\
N_K
\end{bmatrix}(n)
\]

where we have assumed that the oldest age is \( K \).

We note that the percentage of death per year, \( d_k \), is related to the percentage survival per year \( s_k = 1 - d_k \). Then another way to express the population dynamics with
The matrix in (3.4) is known as a Leslie matrix in mathematical ecology.

### 3.1.1 Eigenvalues and their dependence on a model parameter

Let \( A \) denote the \((K + 1) \times (K + 1)\) matrix in Eq. 3.4, which can be then written as

\[
\vec{N}(t + 1) = A \vec{N}(t).
\]

Therefore \( \vec{N}(t) = A^n \vec{N}(0) \), where \( t = 0, 1, \ldots \).

From linear algebra, we know that a matrix like \( A \) has \( K + 1 \) eigenvalues \( \lambda_0 > \lambda_1 > \cdots > \lambda_K \), and corresponding to eigenvalue \( \lambda_\ell \), a left eigenvector

\[
v^{(\ell)} = \begin{pmatrix} v^{(\ell)}_0 \\ v^{(\ell)}_1 \\ \vdots \\ v^{(\ell)}_K \end{pmatrix},
\]

and a right eigenvector

\[
w^{(\ell)} = \begin{pmatrix} w^{(\ell)}_0 \\ w^{(\ell)}_1 \\ \vdots \\ w^{(\ell)}_K \end{pmatrix}.
\]

The left eigenvectors and right eigenvectors satisfy an orthonormal relation

\[
v^{(\ell)} \cdot w^{(m)} = \delta_{\ell m}, \quad (3.5)
\]

where \( \delta_{\ell m} = 1 \) when \( \ell = m \), and = 0 when \( \ell \neq m \). Then, \( A \) can be written as

\[
A = \begin{pmatrix} w^{(0)}_0 & w^{(1)}_0 & \cdots & w^{(K)}_0 \\ w^{(0)}_1 & w^{(1)}_1 & \cdots & w^{(K)}_1 \\ \vdots & \vdots & \ddots & \vdots \\ w^{(0)}_K & w^{(1)}_K & \cdots & w^{(K)}_K \end{pmatrix} \begin{pmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{pmatrix} \begin{pmatrix} v^{(0)}_0 \\ v^{(0)}_1 \\ \vdots \\ v^{(0)}_K \end{pmatrix} = \begin{pmatrix} w^{(0)}_0 & v^{(0)}_0 \\ w^{(0)}_1 & v^{(0)}_1 \\ \vdots & \vdots \\ w^{(0)}_K & v^{(0)}_K \end{pmatrix} \begin{pmatrix} v^{(0)}_0 \\ v^{(0)}_1 \\ \vdots \\ v^{(0)}_K \end{pmatrix}, \quad (3.6)
\]

and

\[
A' = \begin{pmatrix} w^{(1)}_0 & w^{(1)}_0 & \cdots & w^{(1)}_0 \\ w^{(1)}_1 & w^{(1)}_1 & \cdots & w^{(1)}_1 \\ \vdots & \vdots & \ddots & \vdots \\ w^{(1)}_K & w^{(1)}_K & \cdots & w^{(1)}_K \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{pmatrix} \begin{pmatrix} v^{(1)}_0 \\ v^{(1)}_1 \\ \vdots \\ v^{(1)}_K \end{pmatrix} = \begin{pmatrix} w^{(1)}_0 & v^{(1)}_0 \\ w^{(1)}_1 & v^{(1)}_1 \\ \vdots & \vdots \\ w^{(1)}_K & v^{(1)}_K \end{pmatrix} \begin{pmatrix} v^{(1)}_0 \\ v^{(1)}_1 \\ \vdots \\ v^{(1)}_K \end{pmatrix}, \quad (3.7)
\]
Therefore, for a very large \( t \), in “the long time limit”, we have

\[
\vec{N}(t) \simeq \alpha \lambda_0 t \begin{pmatrix} w_0(0) \\ w_1(0) \\ \vdots \\ w_K(0) \end{pmatrix},
\]

where \( \alpha = \sum_{i=0}^{K} v_i(0) N_i(0) \). (3.8)

\( \lambda_0 \) is the annual growth rate of the population, and the corresponding eigenvector \( w(0) \) gives the stationary age distribution within the long time population. According to Perron-Frobenius theorem, the elements of \( w(0) \) associated with the largest eigenvalue \( \lambda_0 \) are all non-negative when all entries of \( A \) are non-negative.

We now show a very important result on the sensitivity of the eigenvalue \( \lambda_\ell \) with respect to a change in the entry \( a_{ij} \), keeping all other entries fixed:

\[
\frac{\partial \lambda_\ell}{\partial a_{ij}} = v_\ell^i w_\ell^j. 
\] (3.9)

To prove this, we note

\[
\lambda_\ell = \sum_{h,k} v_\ell^h a_{hk} w_\ell^k \quad \text{and} \quad 1 = \sum_k v_\ell^k w_\ell^k. 
\] (3.10)

Then,

\[
\frac{\partial \lambda_\ell}{\partial a_{ij}} = \left. \frac{\partial}{\partial a_{ij}} \lambda_\ell \right|_{a_{ij}=a_{ij}} = \left. \frac{\partial}{\partial a_{ij}} \left( \sum_{h,k} v_\ell^h a_{hk} w_\ell^k \right) \right|_{a_{ij}=a_{ij}} = \left. \frac{\partial}{\partial a_{ij}} \left( \sum_{h,k} v_\ell^h w_\ell^k \right) \right|_{a_{ij}=a_{ij}} = \left. v_\ell^i w_\ell^j \lambda_\ell \right|_{a_{ij}=a_{ij}} = \left. v_\ell^i w_\ell^j \right|_{a_{ij}=a_{ij}} = v_\ell^i w_\ell^j. 
\]

### 3.2 Population dynamics in a school district

While growing old year by year is obligatory, the grade distribution among the student population in a school district is not. In this case, a fraction (hopefully very small!) of the students will remaining in the same grade due to unacceptable performance, while others will move into a grade higher. In this case, we have

\[
\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix} (n + 1) = \] (3.11)
in which $X_\ell$ is the number of students in grade $\ell$, $\mu_\ell$ is the percentage of the students in grade $\ell$ with satisfactory performance, and the inhomogeneous term $N_\ell$ represents the new students entering grade $\ell$ from outside. More compactly one writes:

$$X_\ell(n+1) = (1 - \mu_{\ell-1})X_{\ell-1}(n) - \mu_\ell X_\ell(n) + N_\ell,$$

(3.12) 

$$0 \leq \ell \leq K.$$

The total student population,

$$X_T(n) = \sum_{\ell=0}^{K} X_\ell(n),$$

(3.13)

satisfies

$$X_T(n+1) = X_T(n) + \sum_{\ell=0}^{K} N_\ell - \mu_K X_K(n).$$

(3.14)

### 3.3 Networks and scaling concept (Ch. 2)

We shall now develop a mechanistic model for the “degree distribution” of a social network. As we shall see, the mathematics is nearly identical to that in the previous sections. But first let us introduce the notion of a graph, a mathematical object that describes a “network”, which has nodes and edges. For a graph of $n$ nodes, the maximal number of edges is $\frac{n(n-1)}{2}$. We call a node with $k$ edges attached having a “degree $k$”. Such a graph can be mathematically represented by a square matrix called “incidence matrix” which tallies all the nodes and connections: $G = \{g_{ij} \mid g_{ij} = 1 \text{ if there is an edge between } i \text{ and } j; g_{ij} = 0 \text{ if there is no connection}\}$. It is clear that $g_{ij} = g_{ji}$, $G$ is a symmetric matrix. We shall set the diagonal elements to zero.

Dynamic models like that in Eqs. (3.1) and (3.12) have wide applications beyond just population of biological species. Let us now consider an problem from network science, discussed in Chapter 2. The so called preferential attachment model. Let us use the language of the citation network, with $N_k(n)$ be the number of papers, among a “pool of total $n$ papers”, that are cited $k$ times per paper. So

$$n \sum_{k=0}^{n} N_k(n) = n.$$  

(3.15)

Now when a “new paper” is added into the pool, we have a distribution $N_k(n+1)$, which is related to the $N_k(n)$ in terms of the following mechanism: The new paper
cites a particular paper according to the popularity of that paper, determined by its current number of citations. More precisely, the total number of citation among all the \( n \) papers is

\[
\sum_{k=0}^{n} kN_k(n) = nm, \quad \text{where} \quad m = \frac{1}{n} \sum_{k=1}^{n} kN_k(n)
\]

(3.16)
is the mean number of citations of a paper. We shall assume it is always a constant for a new paper. Among the total \( nm \) citations, those on the papers in \( N_k \) is \( km \). So we shall let the probability of the new paper citing a paper with \( k \) citations be

\[
\mu_k = \frac{(k+1)m}{n(m+1)}.
\]

(3.17)
The extra 1 in the numerator and denominator takes care of the following situation: the probability of citing a paper with \( k \) citations is proportional to \( k+1 \), not \( k \) since we want paper with zero citation to have at least a non-zero probability to get started.

Note the \( \mu_k \) in (3.17) is not only a function of \( k \), but also inversely proportional to \( n \); this is a new feature: It yields

\[
\sum_{k=0}^{n} \mu_k(n)N_k(n) = \sum_{k=0}^{n} \left( \frac{(k+1)m}{n(m+1)} \right) N_k(n) = m.
\]

(3.18)

With the probability \( \mu_k \) in (3.17), we have

\[
N_k(n+1) = \left( 1 - \mu_k(n) \right) N_k(n) + \mu_{k-1}(n)N_{k-1}(n).
\]

(3.19)

We note that while \( N_k(n+1) \) always changes with increasing \( n \), the total number of papers, the proportion

\[
p_k(n) \equiv \frac{N_k(n)}{n}
\]

(3.20)

actually will converge to a distribution that eventually independent of increasing \( n \). Let us denote

\[
\lim_{n \to \infty} p_k(n) = p_k^*,
\]

(3.21)

then we obtain

\[
p_k^* = -\mu_k np_k^* + \mu_{k-1}np_{k-1}^*
\]

\[= -\frac{(k+1)m}{m+1} p_k^* + \frac{km}{m+1} p_{k-1}^*.
\]

(3.22)

Therefore,

\[
\frac{p_k^*}{p_{k-1}^*} = \frac{k}{k+2+1/m} = 1 - \frac{2+1/m}{k+2+1/m}.
\]

(3.23)

For large \( k \), \( p_k^*/p_{k-1}^* \simeq 1 - (2+1/m)/k \). This expression is actually consistent with a power law distribution \( q_k \propto k^\lambda \) with large \( k \):

\[
\frac{q_k}{q_{k-1}} = \frac{k^\lambda}{(k-1)^\lambda} = \left( 1 - \frac{1}{k} \right)^{-\lambda} \simeq 1 + \frac{\lambda}{k}.
\]

(3.24)
We identify that the power $\lambda = -(2+1/m)$. Therefore, the degree distribution given in Eq. 3.23 is

$$p_k^* = p_1^* k^{-2-1/m},$$

in which the $p_1^*$ is determined from

$$\sum_{k=1}^{\infty} p_k^* = 1 \implies p_1^* = \left( \sum_{k=1}^{\infty} k^{-2-1/m} \right)^{-1}.$$ (3.26)

What is the mathematical relation between the incidence matrix $G = \{g_{ij}\}$ and the degree distribution of the graph? Actually, we have

$$N_k = \sum_{\ell=0}^{n} \delta \left( k, \sum_{j=1}^{n} g_{\ell j} \right),$$

in which the $\delta$-function is $\delta(i,j) = 1$ if $i = j$ and $\delta(i,j) = 0$ if $i \neq j$. The sum $\sum_{j=1}^{n} g_{\ell j}$ is the degree of the $\ell^{th}$ node. So we see that the incidence matrix is a much more informative description of a graph (network), the degree distribution is only one of the many aspects of the graph.

Now, the dynamics of the evolving network can be represented with full details by the following $n$-dependent matrix

$$G(n) = \begin{bmatrix} G(n-1) & \vdots \\ \vdots & \vdots \\ \vdots & \vdots & \cdots & 0 \end{bmatrix}$$

in which $G(n)$ is a $n \times n$ symmetric matrix and $G(n+1)$ is a $(n+1) \times (n+1)$ matrix. When a new node is added, there is an additional row and similarly an additional column, which represents the attachements of the new, $(n+1)^{th}$ node to the other $n$ existing nodes.

Not everythings can also be represented in a graph. For example, a chemical reaction

$$A + B \rightleftharpoons C + D$$

suggests that there are two species at the end of an edge, if one uses a node to represent a chemical species. One can not represent a system of chemical reactions use a simple graph.