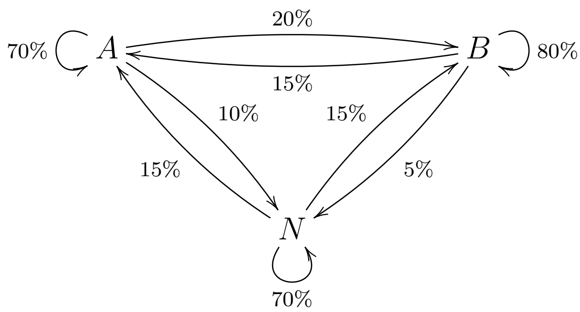
Practice 3
Stochastic matrices and Markov chains

Contents

1 Introductory example

Two companies offer cable TV to a city with 100000 homes. The annual change in the subscription is given by the diagram below. Assuming that the company A has 15000 subscribers, the company B has 20000 and there are 65000 homes without subscribers (N), we will calculate how many subscribers will have each company after 10, 20 and 30 years.



Each home corresponds to one of the following "states": A, B and N. The **vector of initial states** $\vec{x}_0$, that stores the initial proportion of subscribers in each state, is the following one:

$$\vec{x}_0 = \begin{bmatrix} 0,15 \\ 0,20 \\ 0,65 \end{bmatrix}.$$

(Observe that the components of the vector $\vec{x}_0$ have been computed dividing the number of subscribers in each state by the total number of subscribers, that is 100.000).

If $\vec{x}_1 = (a, b, n)$ denotes the **vector of states** that provides the proportions of subscribers (in A, B and N) after 1 year, from the above diagram one can easily deduce the following:

$$a = 0.70 \cdot 0.15 + 0.15 \cdot 0.20 + 0.15 \cdot 0.65,$$

$$b = 0.20 \cdot 0.15 + 0.80 \cdot 0.20 + 0.15 \cdot 0.65,$$

$$n = 0.10 \cdot 0.15 + 0.05 \cdot 0.20 + 0.70 \cdot 0.65.$$

That is,

$$\underbrace{\begin{bmatrix} a \\ b \\ n \end{bmatrix}}_{\vec{x}_1} = \underbrace{\begin{bmatrix} 0.70 & 0.15 & 0.15 \\ 0.20 & 0.80 & 0.15 \\ 0.10 & 0.05 & 0.70 \end{bmatrix}}_{\mathsf{P}} \underbrace{\begin{bmatrix} 0,15 \\ 0,20 \\ 0,65 \end{bmatrix}}_{\vec{x}_0}.$$

The matrix P is called **transition matrix**. Reasoning in a similar way one has that, if $\vec{x}_k$ denotes **vector of states** after k years:

$$\vec{x}_2 = \mathsf{P}\vec{x}_1 = \mathsf{PP}\vec{x}_0 = \mathsf{P}^2\vec{x}_0$$

and, in general,

$$\vec{x}_k = \mathsf{P}^k\vec{x}_0 \text{ for all } k \geq 1.$$

Let's compute, with Scilab, the percentages of subscribers in each state after 10, 20 and 30 years:

```
-->x0=[0.15000; 0.20; 0.65];
-->P=[0.70, 0.15, 0.15; 0.20, 0.80, 0.15;0.10, 0.05, 0.70];
-->x10=P^10*x0
 x10  =
    0.33286896
    0.4714703
    0.19566074
-->x20=P^20*x0
 x20  =
    0.33333216
    0.47612439
    0.19054345
-->x30=P^30*x0
 x30  =
    0.33333333
    0.47618958
    0.19047709
```

Observe that, although $\vec{x}_{10}$ is quite different from $\vec{x}_0$, the difference between $\vec{x}_{10}$ and $\vec{x}_{20}$ are small, and it is even smaller between $\vec{x}_{20}$ and $\vec{x}_{30}$. So, we can say that the distribution of subscribers «tends to stability» and that, after 10 years, the situation is «almost stable».

Computing more vectors $\vec{x}_{40}, \vec{x}_{50}, \dots, \vec{x}_{100}, \dots$ we would see that, as the years pass, the proportion of subscribers in each one of the three states tends to sabilize around a certain value. Now, we clarify the meaning of this imprecise words:

- «Tends to stability» means that the sequence of vectors $\vec{x}_n$ is **convergent** to a certain vector $\vec{v}$, the «final state» or «limit state». That is:
$$\vec{v} = \lim_{n\to\infty} \mathsf{P}^n \vec{x}_0.$$
- The expression «after 10 years, the situation is almost stable» means that the difference between $\vec{x}_{10}$ and the successive vectors $\vec{x}_k$, $k > 10$, is reasonably small.

We have seen, in this example, a process that tends to stability. The question, now, is the following: under what conditions similar processes tend to be stable? In the rest of this paper will answer this question.

2 Stochastic matrices and Markov chains

The approach of the previous example works in general. Suppose we consider a similar process on proportions of individuals belonging to n states and that we have an $n \times n$ matrix, $\mathsf{P} = [p_{ij}]$, called **transition matrix**, such that every entry p_{ij} indicates the probability that a member of the population change from the state j to the state i. Taking into account Probability Theory, $0 \leq p_{ij} \leq 1$ and the sum of entries in each column is 1. Notice that $p_{ij} = 0$ means that no individual with state j will change to state i, while $p_{ij} = 1$ means that any member in state j will move to state i. Note also that p_{ii} represents the probability of remaining in the same state i.

The **initial vector of states**, $\vec{x}_0$, is a vector of $\mathbb{R}^n$ whose ith component is the initial proportion of individuals in the state i, for every $i \in \{1, \dots, n\}$. The **kth vector of states**, $\vec{x}_k$, is a vector of $\mathbb{R}^n$ whose ith component is the proportion of individuals in the state i after k «steps» (usually, unities of time), for $i \in \{1, \dots, n\}$. Observe that the sum of the components of each vector of states is equal to 1.

As in the previous example, one can obtain successive vectors of states $\vec{x}_0, \vec{x}_1, \vec{x}_2, \dots$, from a transition matrix P:

$$\vec{x}_1 = \mathsf{P}\vec{x}_0, \quad \vec{x}_2 = \mathsf{P}\vec{x}_1, \quad \vec{x}_k = \mathsf{P}\vec{x}_{k-1}, \dots$$

or, also,

$$\vec{x}_k = \mathsf{P}^k \vec{x}_0 \text{ for all } k \geq 1.$$

Definition 1. A **probability vector** is a vector whose components are non-negative and such that their sum is 1. A **stochastic matrix** is a square matrix whose columns are probability vectors.

Notice that the transition matrices P that we have considered before are stochastic matrices. Also, the vectors of states $\vec{x}_k$ are probability vectors. The sequences of vectors of states $\vec{x}_0, \vec{x}_1, \ldots$ that have been obtained from transition matrices P are called **Markov chains**, as we precise in the following definition:

Definition 2. A **Markov chain** is a **sequence of probability vectors** $\vec{x}_0$, $\vec{x}_1, \ldots$ such that there exists a stochastic matrix P (**transition matrix**) satisfying the following:

$$\vec{x}_1 = \mathsf{P}\vec{x}_0, \quad \vec{x}_2 = \mathsf{P}\vec{x}_1, \quad \vec{x}_k = \mathsf{P}\vec{x}_{k-1}, \ldots$$

The usual problem associated with a Markov chains is to know whether the evolution of the process over time tends to "stability" and, if so, to know the "final state". In other words, one wants to decide about the existence of a **limit vector of states**

$$\vec{v} = \lim_{k\to\infty} \vec{x}_k.$$

If such a vector exists, we say that the Markov chain is **convergent**. Moreover, taking limits when $k \to \infty$ at both sides of the equality $\vec{x}_{k+1} = \mathsf{P}\vec{x}_k$ we obtain $\vec{v} = \mathsf{P}\vec{v}$. That is, **the vector** $\vec{v}$ **does not change when it is multiplied (on the left) by the matrix** P. In other words, $\vec{v}$ is a **stationary vector** for P, following the terminology that is precised in the definition:

Definition 3. Given a square matrix P, we will say that a non-zero vector $\vec{v}$ is **stationary** for P if $\mathsf{P}\vec{v} = \vec{v}$.

Then, we have the following property:

Proposition 1. Let $\vec{x}_0, \vec{x}_1, \ldots$ be a **convergent** Markov chain whith transition matrix P. If $\vec{v} = \lim_{k\to\infty} x_k$ then $\vec{v}$ is a stationary probability vector for P.

Example 1. Let's check this property with the introductory example.

```
-->x0=[0.15; 0.20; 0.65];
-->P=[0.70, 0.15, 0.15; 0.20, 0.80, 0.15;0.10, 0.05, 0.70];
```

Computing $\vec{x}_k$, with k big enough, we obtain a good aproximation of the limit vector:

```
-->v=P^100*x0
 v  =
    0.33333333
    0.47619048
    0.19047619
```

Let's check that it is a stationary vector:

```
-->clean(P*v-v)
 ans  =
    0.
    0.
    0.
```

Remark. Notice that a non-zero multiple of a stationary vector is, again, a stationary vector. Indeed, if $\vec{v}$ is a stationary vector of a square matrix A and λ is a non-zero scalar, using properties of matrices we have that:

$$\mathsf{A}(\lambda\vec{v}) = \lambda \underbrace{\mathsf{A}\vec{v}}_{\vec{v}} = \lambda\vec{v}.$$

Moreover the stochastic matrices satisfy the following properties (that we do not prove):

1. The product of stochastic matrices is a stochastic matrix.
2. Any stochastic matrix has stationary vectors.

Next, we will see that the matrix P of the introductory example **has a unique stationary <u>probability</u> vector**:

Example 2. Let's compute, first, all the stationary vectors for the matrix P given in the introductory example.

Observe that a satationary vector for P is any non-zero vector $\vec{x}$ such that $\mathsf{P}\vec{x} = \vec{x}$. But this is equivalent to the equality $\mathsf{P}\vec{x} = I\vec{x}$ (where I is the identity matrix). Applying the distributive property we obtain the following equivalent condition:

$$(\mathsf{P} - I)\vec{x} = \vec{0}.$$

This is the matrix expression of a homogeneous system of linear equations. Therefore, the stationary vectors of P are, exactly, the non-zero vectors of the **kernel of the matrix** $\mathsf{P} - I$. Let's compute it with Scilab:

```
-->A=[0.7 0.15 0.15; 0.2 0.8 0.15; 0.1 0.05 0.7]
 A  =

    0.7     0.15    0.15
    0.2     0.8     0.15
    0.1     0.05    0.7

-->kernel(A-eye(3,3))
 ans  =

    0.5449493
    0.7784989
    0.3113996
```

Therefore the set of stationary vectors for the matrix P is

$$\left\{ \lambda \begin{bmatrix} 0.5449493 \\ 0.7784989 \\ 0.3113996 \end{bmatrix} \mid \lambda \in \mathbb{R} \setminus \{0\} \right\}$$

Dividing all the vectors in this set by the sum of their components we will obtain all the stationary probability vectors:

$$\left\{ \frac{1}{\lambda(0.5449493+0.7784989+0.3113996)}\lambda \begin{bmatrix}0.5449493\\0.7784989\\0.3113996\end{bmatrix} \mid \lambda \in \mathbb{R}\setminus\{0\} \right\} = \{ \begin{bmatrix}0.3333333\\0.4761905\\0.1904762\end{bmatrix} \}$$

Evidently, dividing we obtain always the same vector. Therefore **there is only one stationary probability vector**. This vector, by Proposition 1, must coincide with the limit vector of the Markov chain given in Section 1 (compare it with the approximation to the limit vector obtained in Section 1). Furthermore, since there is only one stationary probability vector, this must be the limit of **any** Markov chain with transition matrix P, **independently of the initial vector of states $\vec{x}_0$!**

Not all transition matrices P associated with a Markov chain have a unique stationary probability vector as in the above example. This fact causes that different Markov chains with the same transition matrix P might converge to different vectors. The following example shows this:

Example 3. Let us consider the following stochastic matrix:

$$\mathsf{P} = \begin{bmatrix}1/2 & 0 & 0\\ 0 & 1 & 0\\ 1/2 & 0 & 1\end{bmatrix}$$

and the Markov chain with transition matrix P and initial vector of states $\vec{x}_0 = (1/3, 2/3, 0)$. To study the convergence of this process we will write, in Scilab, the following commands:

```
-->P=[1/2 0 0; 0 1 0; 1/2 0 1];x0=[1/3; 2/3; 0];P^10*x0
 ans  =
    0.0003255
    0.6666667
    0.3330078
-->P^30*x0
 ans  =
    3.104D-10
    0.6666667
    0.3333333
-->P^50*x0
 ans  =
    2.961D-16
    0.6666667
    0.3333333
-->clean(ans)
```

```
 ans  =
    0.
    0.6666667
    0.3333333
```

We see clearly that this process is convergent to the probability vector $(0, 2/3, 1/3)$. However, what happens if we choose a different initial vector of states? Take, for example, the vector $\vec{x}_0 = (1/3, 1/3, 1/2)$:

```
-->x0=[1/3; 1/3; 1/3];P^10*x0
 ans  =
    0.0003255
    0.3333333
    0.6663411
-->P^30*x0
 ans  =
    3.104D-10
    0.3333333
    0.6666667
-->P^50*x0
 ans  =
    2.961D-16
    0.3333333
    0.6666667
-->clean(ans)
 ans  =
    0.
    0.3333333
    0.6666667
```

We see that the process is again convergent, but **the limit is different**: $(0, 1/3, 2/3)$. Both vectors are two different stationary probability vectors.

We can get all the stationary vectors for the matrix P computing the kernel of P $- I$:

```
-->P=[1/2 0 0; 0 1 0; 1/2 0 1]
 P  =

    0.5    0.     0.
    0.     1.     0.
    0.5    0.     1.

-->kernel(P-eye(3,3))
 ans  =
```

```
0.      0.
0.      1.
1.      0.
```

We can deduce, then, that the set of stationary vectors for P are $\lambda_1(0,1,0)+\lambda_2(0,0,1) = (0,\lambda_1,\lambda_2)$, with $\lambda_1,\lambda_2 \in \mathbb{R}$ and some $\lambda_i \neq 0$. Dividing by the sum of their components (that is, $\lambda_1+\lambda_2$) we obtain all the stationary probability vectors:

$$\frac{1}{\lambda_1+\lambda_2}(0,\lambda_1,\lambda_2), \quad \lambda_1,\lambda_2 \in \mathbb{R}.$$

Notice that there are infinitely many of them.

Now we will provide a sufficient condition under which a stochastic matrix P satisfies the "good" conditions of the introductory example, that is, it guarantees the existence of a **unique** stationary probability vector that is, in addition, the limit vector of any Markov chain with transition matrix P, independently of the initial vector of states.

Definition 4. A stochastic matrix P is **regular** if there eixists a natural number k such that all the entries of the matrix P^k are strictly posiive (P^k has not zero entries).

Theorem 1. If P is a regular stochastic matrix, there exists a **unique** stationary probability vector $\vec{v}$ for P. Moreover, if $\vec{x}_0$ is any probability vector and $\vec{x}_k = P\vec{x}_{k-1}$ for all $k \geq 1$, then the Markov chain $\{\vec{x}_k\}$ converges to $\vec{v}$.

Remark. Notice that a stochastic matrix without zero entries is always regular (see the matrix P of the introductory example). However, if a matrix is not regular then it has, necessarily, many zero entries (Example 3). Are the matrices

$$\begin{bmatrix} 1/2 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

regular?

Remark. There are a lot of practical applications (for example the calculation of Google's PageRank) in which it is interesting to calculate a probability vector for a given stochastic matrix P. In almost all these applications the matrix P is huge and, what is done in practice is, *essentially*, instead of computing the kernel of the matrix $P - I$ (as in our example), we start with an initial vector $\vec{x} = \vec{x}_0$ and repeat the iterative process $\vec{x} \leftarrow P\vec{x}$ (storing only the vector $\vec{x}$ that results in each iteration) with the idea of stopping the process when the difference between two consecueive vectors $\vec{x}$ is less than a certain small amount ϵ. The vector obtained in this way is an approximation to the limit vector of a Markov chain and, thus, it will also be a stationary probability vector for the transition matrix P (by Proposition 1) [1]. If the matrix P satisfies the conditions of Theorem 1, it will be guaranteed the existence of a unique stationary probability vector; this is essential, for example, in the Google's PageRank algorithm, as you will see in other practice.

[1] This iterative method is known as the "Power Method" and is commonly used in many problems.