Master of Data Science Online Programme
Course: Linear Algebra
SGA #2: Week 6

Student: Andrei Batyrov (Fall 2022 / Winter 2023)
`arbatyrov@edu.hse.ru`

Higher School of Economics (HSE)
National Research University
Faculty of Computer Science

February 16, 2023

Contents

Problem 4 **1**

Solution 1

Further Considerations 4

Answer 5

List of Figures

1 Random walk graph 1
2 Vector g of probabilities of having various health states in infinite random walk . . 4

Problem 4

A dangerous virus spreads on an island with a population of $10,000$. Every day, island authorities collect statistics and want to understand if they have enough health system resources. Doctors report that every day 15% of healthy people become infected and have a mild illness (which does not require hospitalization), 12% of healthy people become infected and have a difficult illness (they need to get to the hospital). At the same time, 12% of people with a mild form of the disease recover completely, and 15% go into the category of seriously ill patients. In the category of seriously ill patients, the situation is as follows: 20% go into the category of patients with a mild form of the disease and 10% completely recover. Recovered patients may become infected again. At the initial time on the island, 500 patients were identified in a mild form of the disease and 100 patients in a severe form. Luckily, the virus is not lethal. How will the number of patients behave with increasing time? From a mathematical point of view, find the limits of the number of patients in mild and severe forms, if they exist. Please do not forget that with real viruses everything is not so simple.

Solution

This problem can be solved with the Random walk on graphs approach [4]. First let's represent our system as a directed graph. Let Health states $= \{Good, Mild, Ill\}$ be a set of patients health every day, $|\text{Health states}| = 3$. Let Impacts be a set of possible impact on the patients health by the virus disease. Then the random walk graph can be described as follows: $\mathcal{G}\langle V(\mathcal{G}) = \text{Health states}, E(\mathcal{G}) = \text{Impacts}\rangle$. The random graph is shown in Figure 1. Now we need to construct

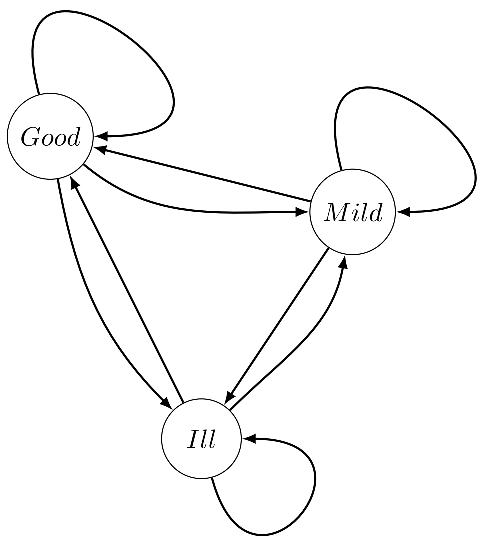

Figure 1: Random walk graph

the Markov transition matrix P with probabilities of walking from one health state to another, where columns will be the health states, and the sum of all elements along any column is equal to 1, since the patient can be in either Good, Mild, or Ill state. Let's list all possible directed edges with their probabilities:

1. $P(Good \to Mild) = 15\% = 0.15$
2. $P(Good \to Ill) = 12\% = 0.12$
3. $P(Good \to Good) = 100\% - P(Good \to Mild) - P(Good \to Ill) = 100\% - 15\% - 12\% = 73\% = 0.73$
4. $P(Mild \to Good) = 12\% = 0.12$
5. $P(Mild \to Ill) = 15\% = 0.15$

6. $P(Mild \to Mild) = 100\% - P(Mild \to Good) - P(Mild \to Ill) = 100\% - 12\% - 15\% = 73\% = 0.73$

7. $P(Ill \to Mild) = 20\% = 0.20$

8. $P(Ill \to Good) = 10\% = 0.10$

9. $P(Ill \to Ill) = 100\% - P(Ill \to Mild) - P(Ill \to Good) = 100\% - 20\% - 10\% = 70\% = 0.70$.

So, our Markov transition matrix is

$$P = \begin{pmatrix} 0.73 & 0.12 & 0.10 \\ 0.15 & 0.73 & 0.20 \\ 0.12 & 0.15 & 0.70 \end{pmatrix}. \tag{1}$$

The elements of the matrix p_{ij} are the probabilities of walking from health state j to health state i. For example, $p_{21} = P(Good \to Mild) = 0.15$. The Markov process allows to find the current state of the system by knowing the previous state. We need to find the stationary state of the system, in this case, the eigenvector g of probabilities of having each health state of the random walk graph if the virus disease never disappears, i.e. it stays infinitely long, such that

$$g = \lim_{k \to \infty} P_\alpha^k X_0, \tag{2}$$

where

- $P_\alpha = (1-\alpha)P + \alpha Q$,
- P is our transition matrix (1),
- α is the factor of "falling out" of the Markov process, i.e. it is the probability that instead of walking along the health state graph, a patient can transit to some random health state with the equal probability of 1/3, since |Health states| = 3,
- Q is a matrix of size of P with all elements equal to the equal probability of 1/3,
- k is the number (index) of the discrete state of the system (number of the day), and
- X_0 is the vector of the initial state of the system. In our case the components of the initial vector: the fraction of Mild patients which is $500/10,000 = 0.05$, the fraction of Ill patients which is $100/10,000 = 0.01$, and the fraction of Good patients which is $1 - 0.05 - 0.01 = 0.94$, i.e. $X_0 = \begin{pmatrix} 0.94 \\ 0.05 \\ 0.01 \end{pmatrix}$.

Let's assume that due to specificity of the viral disease, a patient cannot get from one state to another with the probability of 1/3, for example, an Ill patient cannot get Good with the probability of 1/3 – that would be a miracle! So, we have a pure Markov process, $\alpha = 0$, $P_\alpha = P$, and the matrix Q is not needed. As an example, let's find the state X_1, i.e. the distribution of patients on the next day from the start of the disease. As per (2), $X_1 = \lim_{k \to 1} P^k X_0 = PX_0 = \begin{pmatrix} 0.73 & 0.12 & 0.10 \\ 0.15 & 0.73 & 0.20 \\ 0.12 & 0.15 & 0.70 \end{pmatrix} \begin{pmatrix} 0.94 \\ 0.05 \\ 0.01 \end{pmatrix} = \begin{pmatrix} 0.6932 \\ 0.1795 \\ 0.1273 \end{pmatrix}$. Multiplying this vector by the total population of 10,000, we get 6,932 Good patients, 1,795 Mild patients, and 1,273 Ill patients on the next day from the initial state. Eventually, we need to find (2) for a theoretically infinite number of walks along the edges (virus impacts). To find $\lim_{k \to \infty} P^k$, as per [3], we need to decompose P as a product of three matrices:

$$P^k = TD^kT^{-1}, \tag{3}$$

where T is the transition matrix of eigenvectors of P, and D is the diagonal matrix in the basis of eigenvectors of P. Let's find this decomposition with the following algorithm.

1. Find eigenvalues λ_i of (1). As per [1], eigenvalues of a matrix are the roots of the characteristic polynomial, i.e.

$$det(P - \lambda I) = 0, \tag{4}$$

where det is a function of columns of a square matrix. So, our characteristic equation is

$$\begin{aligned}
det\left[\begin{pmatrix}0.73 & 0.12 & 0.10\\ 0.15 & 0.73 & 0.20\\ 0.12 & 0.15 & 0.70\end{pmatrix} - \lambda\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\right] &= 0,\\
det\left[\begin{pmatrix}0.73 & 0.12 & 0.10\\ 0.15 & 0.73 & 0.20\\ 0.12 & 0.15 & 0.70\end{pmatrix} - \begin{pmatrix}\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{pmatrix}\right] &= 0,\\
det\left[\begin{pmatrix}0.73-\lambda & 0.12 & 0.10\\ 0.15 & 0.73-\lambda & 0.20\\ 0.12 & 0.15 & 0.70-\lambda\end{pmatrix}\right] &= 0,\\
(1-\lambda)(10000\lambda^2 - 11600\lambda + 3349) &= 0.
\end{aligned} \tag{5}$$

The equation (5) has three real roots: $\lambda_1 = 1, \lambda_2 = \frac{58-\sqrt{15}}{100}, \lambda_3 = \frac{58+\sqrt{15}}{100}$. These are the eigenvalues of (1). And the matrix D is

$$D = \begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\ 0 & \frac{58-\sqrt{15}}{100} & 0\\ 0 & 0 & \frac{58+\sqrt{15}}{100}\end{pmatrix}. \tag{6}$$

2. Find eigenvectors of (1), As per [2], each eigenvalue has a corresponding eigenvector, which is a solution vector of the linear system $Pv_i = \lambda_i v_i \Rightarrow (P - \lambda_i I)v_i = 0$. Since we have four eigenvalues, there are four eigenvectors, each of can be obtained by solving its linear system. Let's find v_1 corresponding to λ_1:

$$\left[\begin{pmatrix}0.73 & 0.12 & 0.10\\ 0.15 & 0.73 & 0.20\\ 0.12 & 0.15 & 0.70\end{pmatrix} - \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\right] v_1 = \begin{pmatrix}-0.27 & 0.12 & 0.10\\ 0.15 & -0.27 & 0.20\\ 0.12 & 0.15 & -0.3\end{pmatrix} v_1 = 0. \tag{7}$$

The solution vector is $v_1 = x_3\begin{pmatrix}\frac{170}{183}\\ \frac{230}{183}\\ 1\end{pmatrix}$. By choosing $x_3 = 1, v_1 = \begin{pmatrix}\frac{170}{183}\\ \frac{230}{183}\\ 1\end{pmatrix}$. By constructing and solving a similar system for $\lambda_2 = \frac{58-\sqrt{15}}{100}, \lambda_3 = \frac{58+\sqrt{15}}{100}$, we get $v_2 = \begin{pmatrix}\frac{-3+\sqrt{15}}{3}\\ -\frac{\sqrt{15}}{3}\\ 1\end{pmatrix}, v_3 = \begin{pmatrix}\frac{-3-\sqrt{15}}{3}\\ \frac{\sqrt{15}}{3}\\ 1\end{pmatrix}$. And the matrix T is

$$T = \begin{pmatrix}\dfrac{170}{183} & \dfrac{-3+\sqrt{15}}{3} & \dfrac{-3-\sqrt{15}}{3}\\[2ex] \dfrac{230}{183} & -\dfrac{\sqrt{15}}{3} & \dfrac{\sqrt{15}}{3}\\[2ex] 1 & 1 & 1\end{pmatrix}. \tag{8}$$

Now by applying (2) and (3), we get $g = \lim_{k\to\infty} P^k X_0 = \lim_{k\to\infty} TD^kT^{-1}X_0 = T\left[\lim_{k\to\infty} D^k\right]T^{-1}X_0$. It is easily seen that $\lim_{k\to\infty} D^k$ will have zeros for elements $|d_{ij}| < 1$, so $D^k = \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$. Finally,

$$
\begin{aligned}
g = TD^kT^{-1}X_0 &= \begin{pmatrix} \frac{170}{183} & \frac{-3+\sqrt{15}}{3} & \frac{-3-\sqrt{15}}{3} \\ \frac{230}{183} & -\frac{\sqrt{15}}{3} & \frac{\sqrt{15}}{3} \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{170}{183} & \frac{-3+\sqrt{15}}{3} & \frac{-3-\sqrt{15}}{3} \\ \frac{230}{183} & -\frac{\sqrt{15}}{3} & \frac{\sqrt{15}}{3} \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0.94 \\ 0.05 \\ 0.01 \end{pmatrix} \\
&= \begin{pmatrix} \frac{170}{183} & 0 & 0 \\ \frac{230}{183} & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{170}{183} & \frac{-3+\sqrt{15}}{3} & \frac{-3-\sqrt{15}}{3} \\ \frac{230}{183} & -\frac{\sqrt{15}}{3} & \frac{\sqrt{15}}{3} \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0.94 \\ 0.05 \\ 0.01 \end{pmatrix} \\
&= \begin{pmatrix} \frac{170}{183} & 0 & 0 \\ \frac{230}{183} & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.3139 & 0.3139 & 0.3139 \\ -0.0042 & -0.3915 & 0.4958 \\ -0.3097 & 0.0776 & 0.1903 \end{pmatrix} \begin{pmatrix} 0.94 \\ 0.05 \\ 0.01 \end{pmatrix} \\
&= \begin{pmatrix} 0.3139 & 0.3139 & 0.3139 \\ -0.0042 & -0.3915 & 0.4958 \\ -0.3097 & 0.0776 & 0.1903 \end{pmatrix} \begin{pmatrix} 0.94 \\ 0.05 \\ 0.01 \end{pmatrix} \\
&= \begin{pmatrix} 0.2916 \\ 0.3945 \\ 0.3139 \end{pmatrix}.
\end{aligned}
\tag{9}
$$

Multiplying this vector by the total population of $10,000$, we get $2,916$ Good patients, $3,945$ Mild patients, and $3,139$ Ill patients as the stationary state of the virus disease spread.

Further Considerations

We have studied the case of the stationary state of the system. Let's find the number of days to reach this state. Since, it is the same Markov process of random walk employed in the PageRank algorithm, we can use a similar Python code to facilitate computation. A sample code can be found, for example, on Wikipedia: https://en.wikipedia.org/wiki/PageRank#Python.

In Figure 2 below there is a plot of the vector g versus number of days since the virus disease started. It is clearly seen that within two weeks the system reaches its stationary state. So, the

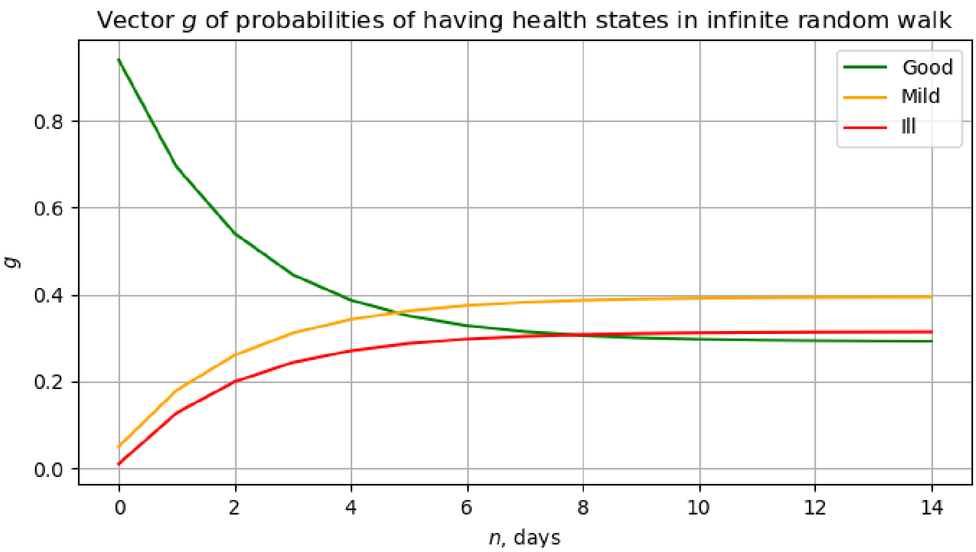

Figure 2: Vector g of probabilities of having various health states in infinite random walk

"Island Healthcare Department" has two weeks to enhance the healthcare system: prepare the hospitals, get more personnel, etc.

Answer

The number of patients in mild form is $3,945$, in severe form is $3,139$, and $2,916$ are healthy.

References

[1] Vsevolod Chernyshev. *Characteristic polynomial*. Faculty of Computer Science, Higher School of Economics. URL: https://smartedu.hse.ru/mod/page/0/701025.

[2] Vsevolod Chernyshev. *Eigenvectors and eigenvalues*. Faculty of Computer Science, Higher School of Economics. URL: https://smartedu.hse.ru/mod/page/0/701022.

[3] Vsevolod Chernyshev. *Matrix diagonalization*. Faculty of Computer Science, Higher School of Economics. URL: https://smartedu.hse.ru/mod/page/0/701028.

[4] Vsevolod Chernyshev. *Random walk on a graph*. Faculty of Computer Science, Higher School of Economics. URL: https://smartedu.hse.ru/mod/page/0/701030.