Iterative and recursive version of simulating epidemic phenomena

Karolina Kęsik
Institute of Mathematics
Silesian University of Technology
Kaszubska 23, 44-100 Gliwice, Poland
Email: karola.ksk@gmail.com

***Abstract*—Forecasting various phenomena can help us analyze the behavior of certain phenomena. Special objects for this type of activity are the weather forecast or the spread of the disease. In the case of an epidemic, there is a risk that the lack of the invention of a vaccine can be very problematic – if a given bacterium or virus have a high probability of being infected. In this paper, we focus on presenting two approaches to the problem of modeling the outbreak epidemic – iterative and recursive. As part of the comparison of these solutions, appropriate simulations were performed in order to indicate the advantages and disadvantages of these solutions. The results were presented and discussed.**

I. Introduction

Forecasting future events is based on several input data that return the most likely changes. In fact, it is not so easy because there may always be some unexpected element that will change predictions. This is particularly evident in the analysis of epidemic phenomena. By analyzing any strain, it is hard to take into account all existing conditions that can lead to mutations. However, simplified models allow us to obtain numerical values over time. An example of this is the anticipated time when a given vaccine should be ready.

Forecasting can take place using different techniques. One of the most important branches of scientific research is artificial intelligence. The development of this technique gives many possibilities. An example of these techniques are artificial neural networks, whose greatest drawback is very long training time. In [1], [2], the authors presented the idea of parallelization. Neural networks can be used not only for classification but also for predictions as shown in [3], [4], [5], [6], [7]. Similarly in [8], where prediction techniques for dynamic seismic slope have been described using these networks and heuristics. An important study is also the prediction of wind speed [9], [10] which can be useful on roads or ski jumps. Another important area of application of different forecasting techniques is medicine, where using selected data, possible occurrences of diseases or their progression can be predicted [11].

In this paper, we compare two approaches to programming the simulation of the epidemic spreading phenomenon and analyze their advantages and disadvantages in relation to differences.

II. Markov chains

Markov chain is one of the stochastic models, which is used to describe probability of evolution in time random process of changing states. Every new state is dependent on events in past. We consider a set S as a space of states, where the number of states is finite (or at most countable) and $X_n = i$ means that the state i is reached at time n.

In our model we assume that $S = \{0, 1, 2\}$ map each state to the appropriate stage of the epidemic – susceptible, infected or recovered. In order for some sequence of variables to be called a Markov chain, the Markov property must be fulfilled. That requires for each $n \in \{0, 1, \dots\}$ and for every state $i_0, \dots, i_{n+1} \in S$ to be occured following equation

$$\begin{aligned} P\{X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0\} \\ = P\{X_{n+1} = i_{n+1} | X_n = i_n\}. \end{aligned} \tag{1}$$

In other words Markov property is related to the memoryless property, which is characteristic for some stochastic processes. If the process has this feature, it means that we only need to know the current state. Knowledge of any of the states in the past does not give the opportunity to predict states that could appear in the future.

Definition of the Markov chain is connected with the matrix of probabilities of transition between states and the probability distribution vector for the variable X_0. First of them, the matrix $\mathbb{P} = (p_{ij})$ is composed of probabilieties of changing the state from i to another state j during one step – which is always . Its dimension is equal $k \times k$, where k is the power of the set S. According to this, we obtain square matrix

$$\mathbb{P} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix}, \tag{2}$$

where $p_{ij} = P\{X_{n+1} = j | X_n = i\}$. It is important to remember about two essensial properties of probability – for each $i \in S$, $\sum_{j=1}^{|S|} p_{ij} = 1$ and also for every i and j $p_{ij} \geq 0$.

The other element is an initial vector of distribution for variable X_0. It can be defined for every $i \in S$ as following

$$\pi = [p_0, p_1, \dots], \tag{3}$$

where each $p_i = P\{X_0 = i\}$ is the probabilities, therefore, the properties mentioned above are also fulfilled.

For probabilities, that are independent of n, we can introduce each element in specified position (i, j) of matrix $\mathbb{P}^m$ as

$$p_{ij}(m) = P(X_{n+m} = j|X_n = i) = P(X_m = j|X_0 = i) \tag{4}$$

and that determines the probability that the Markov chain, which state at period n is i, after m periods will reach the state j.

III. Model of infection spread

Let there be a phenomenon that can be divided into stages. In each period, some changes are made and probability of those changes can be determined. In the proposed model of spreading infection, we assume that we have three stages. Each of them represents the health condition of the population. We can assign every individual to one of the following group

- susceptible $sus(\cdot)$,
- infected $inf(\cdot)$,
- recovered $rec(\cdot)$.

Obviously, there are many elements that affect changes between stages in the model. To simplify it, we have made some assumptions. In our proposition, we assume that

- we do not consider that elements such as climate, demographic changes, social status, age or sex can have any effect on the probability in model,
- only susceptible individual may be infencted,
- leaving the infected group is possible only throughout becoming recovered,
- individual, who is recovered, has acquired immunity.

We have considered two ideas of presenting this model - by using recursion and with Markov chains.

A. Markov chains as a tool for prediction the evolution of infection process

As it was said earlier, we assume that the space of states S is composed of 3 elements. Each of states $\{1, 2, 3\}$ corresponds to one of groups – for example 1 is corresponding to susceptibles.

According to those assumptions, the transition matrix, which consists of probabilities p of every particular transitions during one period, can be defined as

$$\begin{pmatrix} p_{sus,sus} & p_{sus,inf} & p_{sus,rec} \\ p_{inf,sus} & p_{inf,inf} & p_{inf,rec} \\ p_{rec,sus} & p_{rec,inf} & p_{rec,rec} \end{pmatrix} \tag{5}$$

It is easy to notice that the probability of changing state from infected to susceptibles is equal to zero, because the only way to go through infection is to become recovered. In the same way we can see that it is impossible to became susceptibles after being recovered. To predict how the process of spreading the infection in a given population will take place, we need to know the initial state of the population. To present this, we introduce the initial vector of the population, which can be represented as

$$\mathbb{P}_0 = \begin{pmatrix} n_{sus} & n_{inf} & n_{rec} \end{pmatrix}. \tag{6}$$

Each n corresponds to the number of people, who were assigned to the specific group – susceptibles, infected or recovered.

B. Another tool for prediction the evolution of infection process - recursion

The other idea for prediction how the infection would evolve is recursion. We propose that the calculation of the next steps in each stage of infection was based on the state of the population in the previous step. This idea can be defined by following equations

$$sus(t) = sus(t-1) - p_{inf} \cdot inf(t-1), \tag{7}$$

$$inf(t) = inf(t-1) + p_{inf} \cdot inf(t-1) - p_{rec} \cdot inf(t-1), \tag{8}$$

$$rec(t) = rec(t-1) + p_{rec} \cdot inf(t-1), \tag{9}$$

where t and $t-1$ are current and previous step.

In the beginning, we need to specify the number of individuals belonging to each of the states. Initial conditions can be presented as in Eq.(6), where

$$n_{sus} = sus(0), \tag{10}$$

$$n_{inf} = inf(0), \tag{11}$$

$$n_{rec} = rec(0). \tag{12}$$

IV. Iterative and recursive approach

Data: number of healthy n_{sus}, number of infected n_{inf}, number of recovered n_{rec}, the probability of being infected p_{inf} and recovery p_{rec}, stop condition
Result: The number of individuals in each group after a certain time.
Start;
Create of a stochastic matrix in according to Eq. (5);
Create of the initial vector using Eq. (6);
Create a temporary value $\mathbb{K}$ and $t := 0$;
while *stop condition is not met* **do**
  if $t == 0$ **then**
    Calculate $\mathbb{K} = \mathbb{P}_0 \cdot \mathbb{P}$;
  else
    $\mathbb{K} = \mathbb{K} \cdot \mathbb{P}$;
  end
  $t++$;
end
Return $\mathbb{K}$;
Stop;

Algorithm 1: Iterative approach to modeling of the epidemic phenomenon.

Data: number of healthy n_{sus}, number of infected n_{inf}, number of recovered n_{rec}, the probability of being infected p_{inf} and recovery p_{rec}, stop condition
Result: The number of individuals in each group after a certain time.
Start;
Create of the initial conditions using Eq. (10), (11) and (12);
Create a temporary value $\mathbb{K}$ and $t := 1$;
while *stop condition is not met* **do**
  Calculate number of sus according to Eq. (7);
  Calculate number of inf using Eq. (8);
  Calculate number of rec according to Eq. (9);
  Save calculated values as vector $\mathbb{K}$;
  $t++$;
end
Return $\mathbb{K}$;
Stop;

Algorithm 2: Recursive approach to modeling of the epidemic phenomenon.

V. Experiments

Two methods of infection spread prediction were implemented and tested. Simulations were conducted to compare both methods. In each simulation, we use the same probabilies of becoming infected and recovered for iterative and recursive approach. We also tested it for two populations to see how results will differ.

We present a first initial steps of each method to show how calculation are made. In every step we round up the values to one, because we consider individuals. Let us assume that the population is made up of a one hundred people, where 95 people are healthy and only 5 are infected. Hence, at the beginning, the initial vector is

$$\mathbb{P}_0 = \begin{pmatrix} 95 & 5 & 0 \end{pmatrix}. \tag{13}$$

For iterative approach, we define stochastic matrix as

$$\mathbb{P} = \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0 & 0.55 & 0.45 \\ 0 & 0 & 1 \end{pmatrix}. \tag{14}$$

We do the first iteration and we obtain

$$\begin{aligned} \mathbb{P}_0\mathbb{P} &= \begin{pmatrix} 95 & 5 & 0 \end{pmatrix} \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0 & 0.55 & 0.45 \\ 0 & 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} 38 & 60 & 2 \end{pmatrix}. \end{aligned} \tag{15}$$

For recursive approach, we calculate number of individuals after first step in following way

$$sus(1) = 95 - 0.6 \cdot 5 = 92, \tag{16}$$

$$inf(1) = 5 + 0.6 \cdot 5 - 0.45 \cdot 5 = 6, \tag{17}$$

$$rec(1) = 0 + 0.45 \cdot 5 = 2. \tag{18}$$

	Statistic	p-value
Anderson-darling	1.174801	0.003818
Kolmogorov-Smirnov	0.347438	0.046128
Kuiper	0.463567	0.009629
Pearson ξ^2	9.	0.011109
Shapiro-Wilk	0.628608	0.000965
Watson U^2	0.192294	0.002967

Table I: Statistical tests for iterative susceptible table.

	Statistic	p-value
Anderson-darling	0.391341	0.378907
Kolmogorov-Smirnov	0.171128	0.676132
Kuiper	0.199223	0.617289
Pearson ξ^2	1.555556	0.459426
Shapiro-Wilk	0.901546	0.259454
Watson U^2	0.047075	0.615696

Table II: Statistical tests for recursive susceptible table.

	Statistic	p-value
Anderson-darling	0.392656	0.372902
Kolmogorov-Smirnov	0.229084	0.515858
Kuiper	0.241794	0.456394
Pearson ξ^2	0.666667	0.716531
Shapiro-Wilk	0.893534	0.336583
Watson U^2	0.057898	0.464991

Table III: Statistical tests for iterative infected table.

	Statistic	p-value
Anderson-Darling	0.321762	0.546843
Kolmogorov-Smirnov	0.170226	0.684102
Kuiper	0.270848	0.732914
Pearson ξ^2	1.555556	0.459426
Shapiro-Wilk	0.928101	0.463175
Watson U^2	0.043625	0.684242

Table IV: Statistical tests for recursion infected table.

	Statistic	p-value
Anderson-darling	0.461812	0.251231
Kolmogorov-Smirnov	0.233851	0.482461
Kuiper	0.292183	0.250857
Pearson ξ^2	2.333333	0.311403
Shapiro-Wilk	0.864369	0.203849
Watson U^2	0.065498	0.351484

Table V: Statistical tests for iterative recovered table.

	Statistic	p-value
Anderson-darling	0.391341	0.378907
Kolmogorov-Smirnov	0.171128	0.676132
Kuiper	0.199223	0.617289
Pearson ξ^2	1.555556	0.459426
Shapiro-Wilk	0.901546	0.259454
Watson U^2	0.047075	0.615696

Table VI: Statistical tests for recursive recovered table.

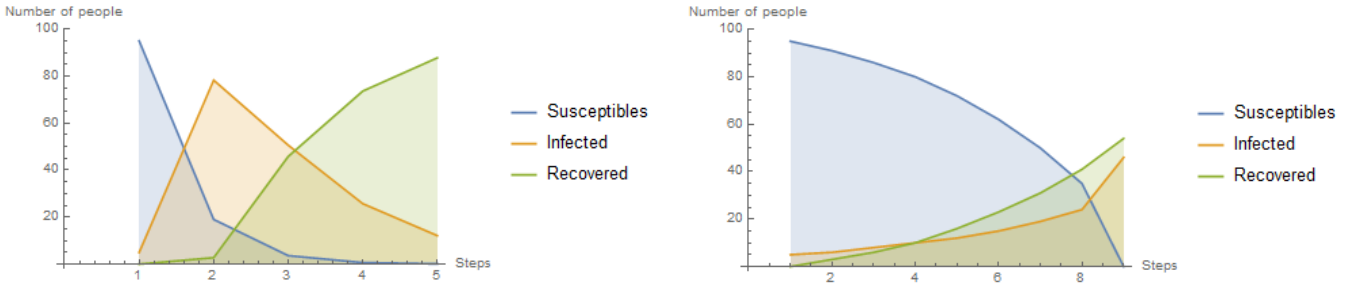

Figure 1: Measurements for the $p_{inf} = 0.8, p_{rec} = 0.55$ and population $\{95, 5, 0\}$.

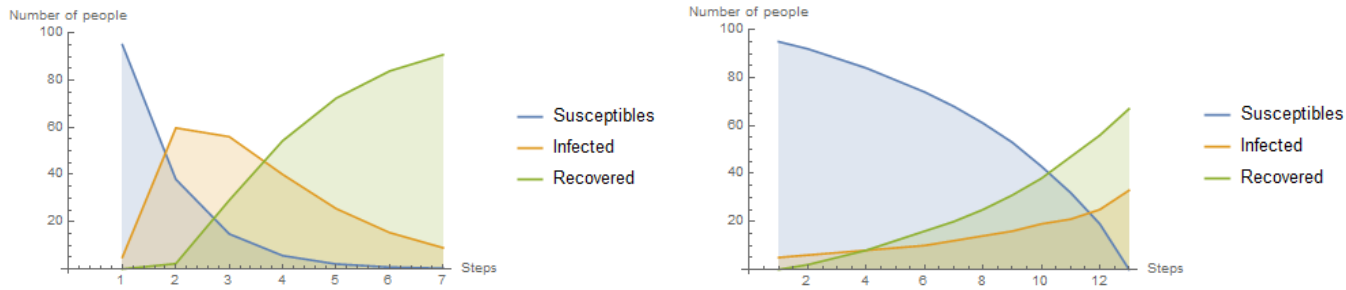

Figure 2: Measurements for the $p_{inf} = 0.6, p_{rec} = 0.45$ and population $\{95, 5, 0\}$.

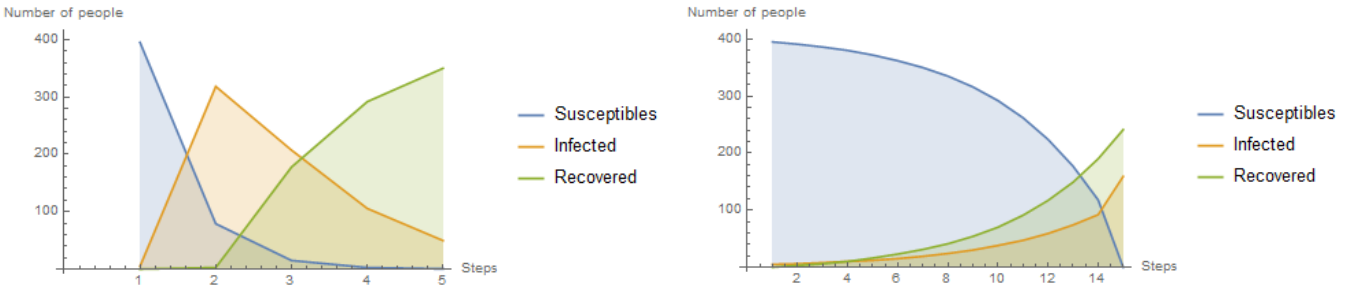

Figure 3: Measurements for the $p_{inf} = 0.8, p_{rec} = 0.55$ and population $\{395, 5, 0\}$.

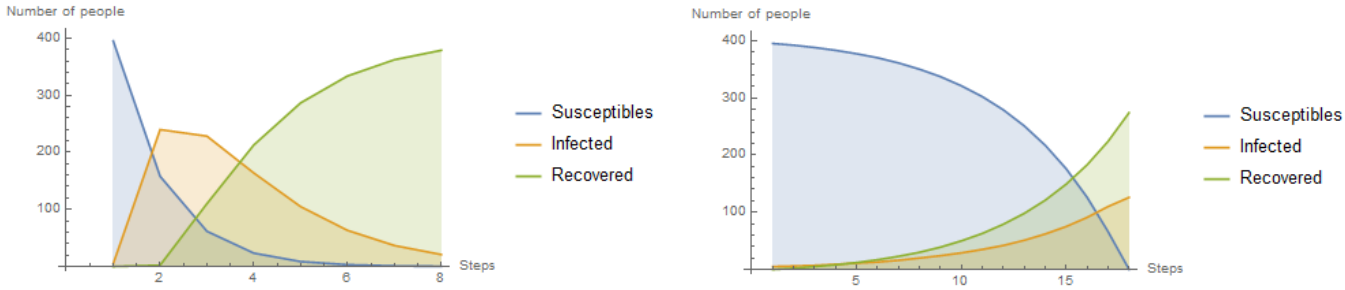

Figure 4: Measurements for the $p_{inf} = 0.6, p_{rec} = 0.45$ and population $\{395, 5, 0\}$.

We have made statistical tests to check out whether the received data is distributed according to the normal distribution. For iterative table of infected and recovered individuals, we obtained a positive result – distribution of data and the normal distribution are compatible. Only for susceptibles, we reject the hypothesis that the data is distributed in accordance with normal distribution. In the case of recursive approach, the null hypothesis is not rejected (at the 5 percent level – the same as for the iteration version). It means that we can assume that the data is distributed in accordance with the normal distribution.

VI. Conclusion

To create a model of prediction of the infection spread we proposed two algoritms. First of them uses Markov chain and the second one – recursion. Both are useful and each of them

has some advantages. Iterative model is is fast and effective even for large populations. Recursive model is more accurate and a bit more reliable, but has one big disadvantage – has a high computational complexity. That makes calculations last for a very long time.

The analysis of the test results allowed us to conclude that the use of Markov chains is better and more optimal solution for prediction. Using this method we obtained rewarding forecast of the spread of infection with specific probabilities of being infected and recovered in the given population.

REFERENCES

[1] D. Połap, M. Woźniak, W. Wei, and R. Damaševičius, "Multi-threaded learning control mechanism for neural networks," *Future Generation Computer Systems*, vol. 87, pp. 16–34, 2018.

[2] D. Połap, "Human-machine interaction in intelligent technologies using the augmented reality," *Information Technology And Control*, vol. 47, no. 4, pp. 691–703, 2018.

[3] D. Lopez, G. Manogaran *et al.*, "Modelling the h1n1 influenza using mathematical and neural network approaches," 2017.

[4] M. H. Esfe, M. R. H. Ahangar, M. Rejvani, D. Toghraie, and M. H. Hajmohammad, "Designing an artificial neural network to predict dynamic viscosity of aqueous nanofluid of tio2 using experimental data," *International Communications in Heat and Mass Transfer*, vol. 75, pp. 192–196, 2016.

[5] G. Capizzi, G. L. Sciuto, C. Napoli, and E. Tramontana, "Advanced and adaptive dispatch for smart grids by means of predictive models," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6684–6691, 2017.

[6] S. Brusca, G. Capizzi, G. Lo Sciuto, and G. Susi, "A new design methodology to predict wind farm energy production by means of a spiking neural network–based system," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 32, no. 4, p. e2267, 2019.

[7] S. Coco, A. Laudani, F. R. Fulginei, and A. Salvini, "A new neural predictor for elf magnetic field strength," *IEEE Transactions on Magnetics*, vol. 50, no. 2, pp. 69–72, 2014.

[8] B. Gordan, D. J. Armaghani, M. Hajihassani, and M. Monjezi, "Prediction of seismic slope stability through combination of particle swarm optimization and neural network," *Engineering with Computers*, vol. 32, no. 1, pp. 85–97, 2016.

[9] B. Doucoure, K. Agbossou, and A. Cardenas, "Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data," *Renewable Energy*, vol. 92, pp. 202–211, 2016.

[10] A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "A new fuzzy-based combined prediction interval for wind power forecasting," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 18–26, 2016.

[11] S. Vijayarani, S. Dhayanand, and M. Phil, "Kidney disease prediction using svm and ann algorithms," *International Journal of Computing and Business Research (IJCBR)*, vol. 6, no. 2, 2015.