Stochastic Processes TMA4265 - Project 1

Candidates: 10067, 10051

November 13, 2020

Problem 1: Modelling the outbreak of measles

1 a)

Since we may observe the random variable X_n on each day $n \in \mathbb{N}_0$ it follows that $\{X_n\}_{n\in\mathbb{N}_0}$ is a discrete time stochastic process. Let P_{ij} denote the probability that an individual transitions from state i on day n to state j on day $n+1$. Then we have the following transition probabilities,

$$P_{01} = \beta, \quad P_{12} = \gamma, \quad P_{20} = \alpha \tag{1}$$

The process may also remain in the same state between subsequent time steps with probabilities

$$P_{00} = 1-\beta, \quad P_{11} = 1-\gamma, \quad P_{22} = 1-\alpha \tag{2}$$

The process $\{X_n\}_{n\in\mathbb{N}_0}$ has well defined transition probabilities which do not depend on the time n. Additionally, the probability of transitioning from one state to another depends only on the current state of the process, and not the previously inhabited states. Hence, $\{X_n\}_{n\in\mathbb{N}_0}$ satisfies the definition of a discrete time Markov chain. Using this fact, we can conclude that there exists a common transition matrix such that

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix}$$

where P_{ij} is the stationary one-step transition probability from state i to state j. Inserting probabilities from (1) and (2) results in the transition matrix

$$\mathbf{P} = \begin{bmatrix} 1-\beta & \beta & 0 \\ 0 & 1-\gamma & \gamma \\ \alpha & 0 & 1-\alpha \end{bmatrix}.$$

where we assume that β, α and γ are real constants in the interval $(0,1)$. Note that $\mathbf{P}$ satisfies $\sum_j P_{ij} = 1$.

1 b)

- A Markov chain is irreducible if each state communicates with every other state. In this example we can observe from $\mathbf{P}$ that there exists non-zero positive probabilities for the one-step transitions P_{01}, P_{12} and P_{20}. Hence, all states are accessible from each other and the Markov chain is said to be irreducible.

- An equivalence class is defined as a class of states where all states in the same class communicate with each other in form of being reflexive, transitive and symmetric. Thus, if any given state communicates with all other states in some equivalence class, then the state belongs to the equivalence class. Since we have shown that all states communicate there can be only one equivalence class consisting of the entire state space $S = \{0,1,2\}$.
  Let Γ_i be defined as
  $$\Gamma_i = \{\nu \in \mathbb{N} : Pr\{X_{n+\nu} = i | X_n = i\} > 0\}$$
  for any $n \in \mathbb{N}_0$. Then the period of state i is defined as the greatest common divisor of all elements in Γ_i. For any state $i \in S$ there is a non-zero positive probability of transitioning back into i at the next time step. Then for any $\nu \in \mathbb{N}$ we must have that $Pr\{X_{n+\nu} = i | X_n = i\} > 0$ so that $\Gamma_i = \mathbb{N}$. Hence, the greatest common divisor of Γ_i is 1 $\forall i$. Therefore, the period of every state is 1, meaning that they are aperiodic.

1 c)

- Let $\{X_n\}$ be one particular sample function of the stochastic process. Let $t_{ij}(n)$ be defined as the smallest integer $t > 0$ for which $X_{n+t} = j$ given that $X_n = i$. Let $T_{ij}(n) = E[t_{ij}(n)]$. Referring to the definition of $t_{ij}(n)$ we see that the Markov property ensures that $E[t_{ij}(n)]$ is constant for any n. Hence, we may replace $T_{ij}(n)$ with T_{ij}. The first step has two possible outcomes:

  1) $X_1 = 1$ s.t. $t_{01}(0) = 1$ with $P_{01} = \beta$
  2) $X_1 = 0$ s.t. $t_{01}(0) = 1 + t_{01}(1)$ with $P_{00} = 1-\beta$

  Hence, $Pr\{t_{01}(0) > 0\} = 1$ and $Pr\{t_{01}(0) > 1\} = 1-\beta$ s.t.
  $$\begin{aligned} T_{01} &= P_{01} \cdot 1 + P_{00} \cdot (1 + T_{01}) \\ &= 1 + P_{00} \cdot T_{01}, \end{aligned}$$

using the property that $P_{00} + P_{01} = 1$. Substituting P_{00} with $1 - \beta$ we get the recurrence

$$T_{01} = 1 \; + \; (1-\beta)T_{01}$$

with solution

$$T_{01} = \frac{1}{\beta} = 20 \tag{3}$$

- In order to go from state 0 to 2 the process must first pass through state 1. Following the same procedure as before we have

$$\begin{aligned} T_{02} &= P_{00} \cdot (1 + T_{02}) + P_{01} \cdot (1 + T_{12}) \\ T_{12} &= P_{11} \cdot (1 + T_{12}) + P_{12} \cdot 1 \end{aligned}$$

  which can be rearranged into

$$\begin{aligned} T_{02} &= 1 + P_{00}T_{02} + P_{01}T_{12} \\ T_{12} &= 1 + P_{11}T_{12} \end{aligned}$$

  Solving the system of equations and substituting probabilities yields

$$T_{02} = \frac{1}{\beta} + \frac{1}{\gamma} = 30 \tag{4}$$

  which is the expected time until the process first enters state 2.

- Since the process must go through state 2 to complete a cycle we may invoke the law of total expectation and state that $T_{00} = T_{02} + T_{20}$. Employing the same procedure as before we get that $T_{20} = \frac{1}{\alpha}$. Therefore

$$T_{00} = \frac{1}{\beta} + \frac{1}{\gamma} + \frac{1}{\alpha} = 130 \tag{5}$$

1 d)

Starting with the initial state of individual being susceptible at state 0, we can simulate 20 years of the Markov chain and we then end up with the results

$$T_{01} \approx 20.01, \quad T_{02} \approx 29.95 \quad \text{and} \quad T_{00} \approx 130.67.$$

which agrees with the analytical result given in (3), (4) and (5).

1 e)

One important property of a Markov chain is that the next state of the process must depend only on the current state. We will show that $\{I_n\}_{n\in\mathbb{N}}$ fails to satisfy this property. In one time step I_n can only increase as a result of susceptible individuals becoming infected. Therefore if we let $S_n = 0$ then $Pr\{I_{n+1} > I_n\} = 0$. If instead we let $S_n > 0$ then $Pr\{I_{n+1} > I_n\} > 0$, since susceptible individuals can become infected. Hence, I_{n+1} depends not only on I_n but also on S_n, and therefore $\{I_n\}_{n\in\mathbb{N}}$ is not a Markov chain.

Z_{n+1} depends on Z_n and R_n. However, given $Z_n = (S_n, I_n)$ at some time n and fixed population N, we can deduce that $R_n = N - S_n - I_n$. Hence, Z_{n+1} depends only on Z_n, and therefore $\{Z_n\}_{n\in\mathbb{N}}$ is a Markov chain.

1 f)

The results of the simulation is shown in figure 1. During the early phase of the outbreak the state of the process is displaced far from the stationary distribution of the Markov chain. As the process continues the Markov chain converges toward the stationary distribution and the population evolves slower.

1 g)

The simulation shows that the long run distribution converges near

$$Y_n/N = (1/N)(S_n, I_n, R_n) = (0.203, 0.073, 0.725).$$

The long run distribution was calculated as $E[Y]$ for the sample function.

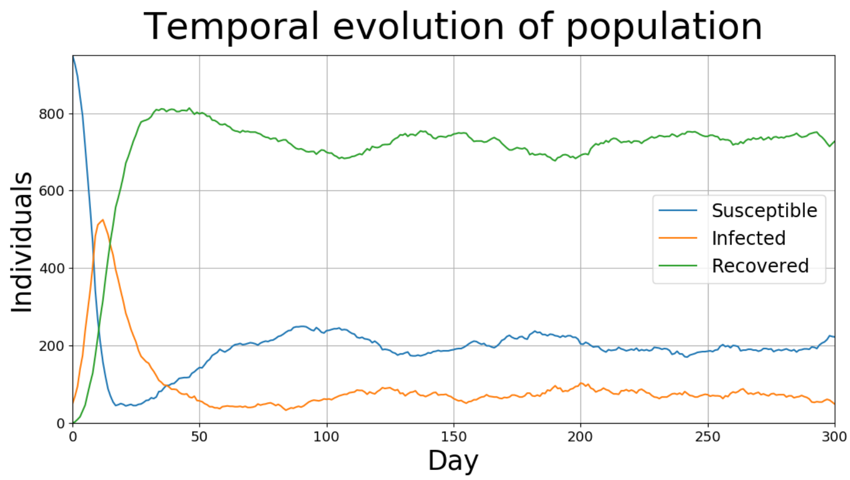

Figure 1: A plot of a single simulation of the temporal evolution of a population. The total population $N = 1000$ is constant through the process, and the starting population is $Y = (950, 50, 0)$.

1 h)

Based on 1000 simulations the estimated expected maximum number of infected individuals is $I_{\max} = 525$. The estimated expected time until reaching the first peak in infections is $t_{\text{peak}} = 11.93$. One may consider an outbreak to be severe if the peak in infected individuals is high and the peak occurs late. This corresponds to a situation in which a large amount of individuals remain infected and symptomatic for a long period of time. The ideal outbreak is short-lived with a small peak.

Problem 2: Insurance claims

2 a)

Assuming that $\lambda = 1.5$, we can compute the probability of more than 100 claims occurring before a total of 59 days. Thus,

$$P(X(59) > 100) = 1 - e^{-59\lambda} \sum_{i=0}^{100} \frac{(59\lambda)^i}{i!} \approx 0.1028. \tag{6}$$

Simulating the process from $t = 0$ to $t = 59$ 1000 times yields the result
$Pr\{X > 100\} \approx 0.106$. Figure 2 shows ten simulation of $\{X(t)\}$ in the time interval from $0 \leq t \leq 59$.

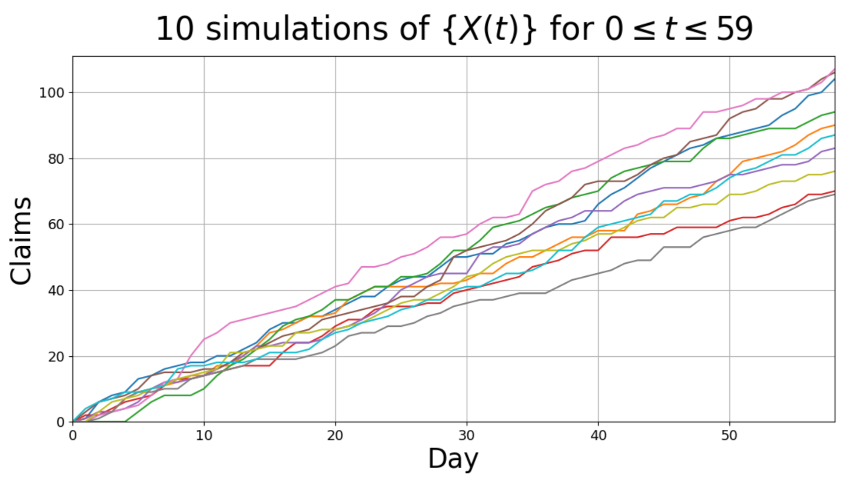

Figure 2: A plot of 10 simulated realizations of $\{X(t)\}$ in the time interval from $0 \leq t \leq 59$, where t is measured in days.

2 b)

Using the properties of a Poisson distribution, $\mathrm{E}[X(t)] = \mathrm{Var}[X(t)] = \lambda t$, we can compute analytically the expected value of $X(t)$.

$$
\begin{aligned}
\mathrm{E}[Z(t)] &= \mathrm{E}\left[\sum_{i=1}^{X(t)} C_i\right] = \mathrm{E}[X(t)C_i] \\
&= \mathrm{E}[C]\ \mathrm{E}[X(t)] = \frac{\lambda t}{\gamma} \approx 8.85
\end{aligned}
$$

The variance may be calculated through the law of total variance,

$$
\begin{aligned}
\mathrm{Var}[Z] &= E\left[\mathrm{Var}\left[Z|X\right]\right] + \mathrm{Var}\left[\mathrm{E}\left[Z|X\right]\right] \\
&= \mathrm{E}\left[X \cdot \mathrm{Var}\left[C\right]\right] + \mathrm{Var}\left[X \cdot \mathrm{E}\left[C\right]\right] \\
&= \mathrm{E}\left[X\right] \cdot \mathrm{Var}\left[C\right] + \mathrm{E}\left[C\right]^2 \cdot \mathrm{Var}\left[X\right] \\
&= \frac{2\lambda t}{\gamma^2} \\
&= 9.735.
\end{aligned}
$$

By 10^5 simulations of the total claims with $t = 59$ we arrive at the approximations

$$
\mathrm{E}[Z] \approx 8.853 \quad \text{and} \quad \mathrm{Var}[Z] \approx 9.759,
$$

which agrees well with the analytically computed results.