Stochastic Modelling - Project 1

Helge Bergo
Lars Sæle

September 2019

Problem 1: Modelling the outbreak of measles

a)

Definition 0.1 *A Markov process X_t is a stochastic process with the property that, given the value of X_t the values of X_S for $s > t$ are not influenced by the values of X_n for $u < t$.*

X_n is a Markov chain because X_n can only be in one state at time n. If a person is susceptible at a given day $n-1$, will not alter our prediction that the person still is susceptible or transitions to infected state at day n. This comes from the assumption that the probability for an individual to shift states remains constant. Also, in this case, we know that if a person is susceptible at time n, the person must have been susceptible all previous days, which makes information about this redundant.

The same argument holds for the infected state as well. If a person is infected at time n makes previous information about the person being susceptible or infected useless, as it does not alter the prediction whether or not the person will stay infected, or shift state to recovered. When the person is recovered, we will still have no use for the previous information about the individual being susceptible or infected, as a recovered person can not change state. By definition 0.1, X_n will have Markov property.

The transition matrix of X_n is $\mathbf{P}$ because all individuals start in state $\mathbf{P}_{11}$, susceptible (S), with the probability $1-\beta$ to stay uninfected the next day. To reach state (I), infected, the probability is therefore β. When a person is infected, he or she can only stay infected with probability $1-\gamma$, or recover and become immune (R), with the probability γ.

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

b)

The Markov chain is reducible, because you have to go through every state in order, and cannot go back to the previous state. A person has to be infected before they can be recovered.

The equivalence states are $\{\mathrm{S}, \mathrm{I}\}$ and $\{\mathrm{I}, \mathrm{R}\}$. Because the chain is reducible, S and I are transient states. R is the absorbing state, and will also be the recurrent state, as the probability for returning to R will always be 1.

For each state, there can only be one transition before the system returns to starting state. This comes from the fact that in each state, you can only stay in your current state, or move to the next, until you reach state R. For that reason, the period of all states will be $d = \gcd(1) = 1$.

c)

We want to find $u = \Pr(X_T = 1 \mid X_0 = 0)$ and $v = E[X_T \mid X_0 = 0]$. We can see from the task at hand, that you can only stay in your current state, or move to the next. Therefore, the probability for transition to state I from state S is $u = \beta = 0.05$. This implicates that $v_I = \frac{1}{\beta} = 20$. We can also see that state R will absorb state I. By the same calculations, the expected time until an infected individual becomes recovered is $v_R = \frac{1}{\gamma} = 5$.

d)

We ran our simulation 1000 times, and calculated the mean values for v_I and v_R

$$v_S = 20.775 \approx 20$$
$$v_R = 4.949 \approx 5$$

which is close to the exact values given in **c)**.

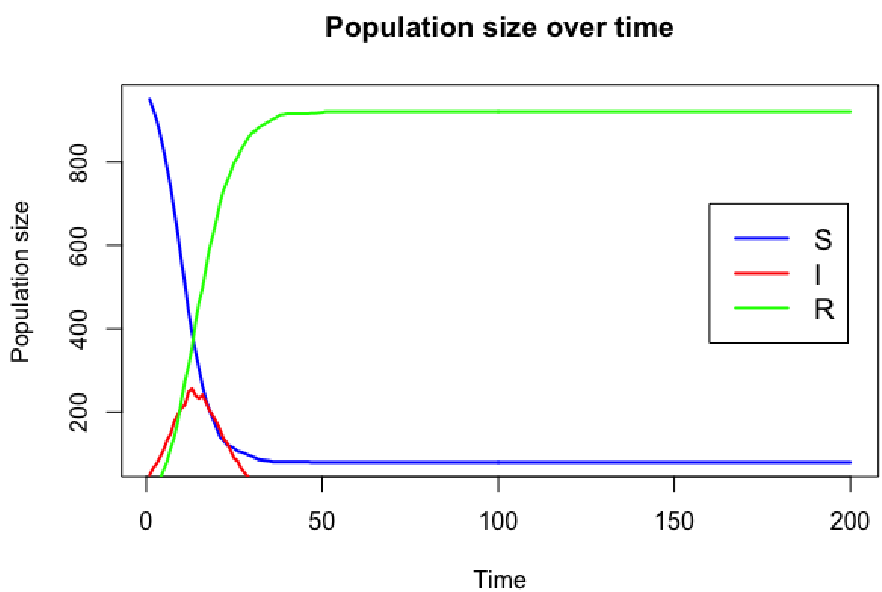

Figure 1: Population size plotted over time, with the different lines showing the states S, I and R over time.

e)

From Figure 1 we can confirm that the I_n peeks at the time when S_n and R_n crosses paths. We see that the number of people in state $I_n \to 0$ as $n \to \infty$. It will then follow that the number of people in state S_n and R_n will converge, as β_n will decrease as I_n decrease.

f)

The simulation was repeated 1000 times for time steps 0 to 200, and the expected maximum number of infected individuals E[max$\{I_0, I_1, ..., I_{200}\}$]=275.237 and the expected time, E[min$\{\arg\max_{n \leq 200}\{I_n\}\}$] $=$ 12.4 days. Both of these values fit well with what can be seen in Figure 1 for the I state (red line).

Problem 2: Insurance claims

a)

For a Poisson process with rate λ, the probability for k events at time t, is given by

$$\Pr\{X(t) \leq k\} = \sum_{x=0}^{k} \frac{e^{-\lambda t}(\lambda t)^x}{x!}.$$

$$\Pr\{X(t) > k\} = 1 - \sum_{x=0}^{k} p(x)$$

In our case, the probability for more than a 100 insurance claims at time 59 will be the following:

$$\begin{aligned}\Pr\{X(59) > 100\} &= 1 - \sum_{x=0}^{100} \frac{e^{-1.5\cdot 59}(1.5\cdot 59)^x}{x!} \\ &= 0.102822,\end{aligned}$$

which gives us that the likelihood for more than 100 claims is relatively small.

We can also see from the simulation that this seems to be true, as the 1000 realisations from the Poisson process with intensity λ and taking the mean of all the realisations. This gave the result of 0.114000, quite close to the theoretical value, but still imprecise.

The simulation was repeated with 10 000 000 realisations getting a final value of 0.102811, much closer to the theoretical value, implying that only 1000 simulations might not be sufficient for more precise measurements.

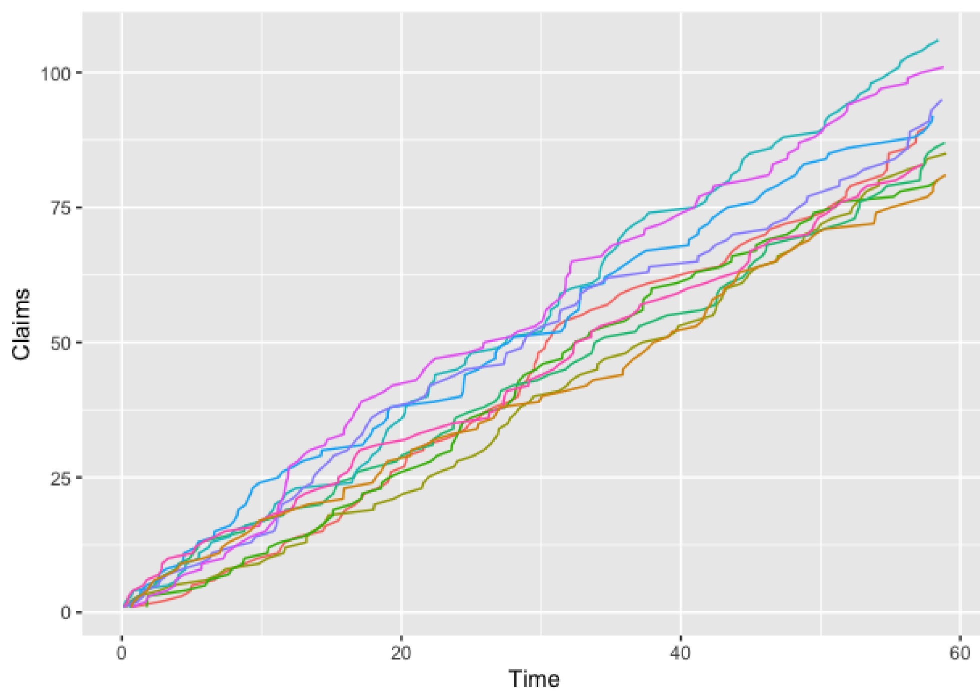

Figure 2: Plot showing 10 realisations of $X(t)$ from time 0 to 59.

Figure 2 shows 10 realisations of the process $X(t)$. In this simulation, two of the realisations ended up with over 100 claims, while most others were close to $\mu = \lambda t = 88.5$, which is the expected number of claims in the long run.

b)

Given that the claim amount (in mill NOK) $C_i \sim \text{Exp}(\beta), \quad i = 1, 2, ...$ with $\beta = 10$, and the total claim amount at time t to be $Z(t) = \sum X(t)C_i$, we can calculate the expectation and variance by the law of total expectation and total variance. The law total expectation is as follows:

$$E[Z(t)] = E[E[Z(t) \mid X(t)]]$$

In our case, with $t = 59$, the expected claim amount is calculated as follows:

$$\begin{aligned}
E[Z(t)] &= E[E[Z(t) \mid X(t)]] \\
&= E\left[E\left[\sum_{i=1}^{X(t)} C_i \mid X(t)\right]\right] \\
&= E\left[\sum_{i=1}^{X(t)} E[C_i]\right] \\
&= E\left[\frac{X(t)}{\beta}\right] \\
&= \frac{\lambda t}{\beta} = \frac{1.5 \cdot 59}{10} \\
&= 8.85
\end{aligned}$$

With the law of total variance being

$$\text{Var}[Z(t)] = E[\text{Var}[Z(t)] \mid X(t)]] + \text{Var}[E[Z(t) \mid X(t)]],$$

the calculations in our case with $t = 59$ will be as follows:

$$\begin{aligned}
\text{Var}[Z(t)] &= E[\text{Var}[Z(t)] \mid X(t)]] + \text{Var}[E[Z(t) \mid X(t)]] \\
&= E\left[\text{Var}\left[\sum_{i=1}^{X(t)} C_i \mid X(t)\right]\right] + \text{Var}\left[E\left[\sum_{i=1}^{X(t)} C_i \mid X(t)\right]\right] \\
&= E\left[\sum_{i=1}^{X(t)} \text{Var}[C_i]\right] + \text{Var}\left[\frac{X(t)}{\beta}\right] \\
&= E\left[\frac{X(t)}{\beta^2}\right] + \frac{1}{\beta^2}\text{Var}[X(t)] \\
&= \frac{\lambda t}{\beta^2} + \frac{\lambda t}{\beta^2} \\
&= \frac{2\lambda t}{\beta^2} = \frac{2 \cdot 1.5 \cdot 59}{100} \\
&= 1.77
\end{aligned}$$

By simulating the Poisson process 1000 times, the estimated expectation $E[Z(t)] = 8.833600$ and variance $\text{Var}[Z(t)] = 1.766720$, which both are close to the theoretical values.