Norwegian University of Science
and Technology
Institutt for matematiske fag

Stochastic modelling
Fall 2020

Project 2

Modelling the common cold

a) The exponential distribution is memoryless, meaning that its distribution is the same regardless of how much time has passed. As the sojourn times, S_S, S_{I_L} and S_{I_H}, are independent of each other and exponentially distributed, the expected time an individual will spend in a given state does not depend on how much time they already spent there or which states they visited earlier only the state they currently are in. Thus $\{X(t) : t \geq 0\}$ satisfies the Markov property

$$\Pr\{X(t+s) = j | X(s) = i, X(u), 0 \leq u \leq s\} = \Pr\{X(t+s) = j | X(s) = j\}$$

for $i, j = 0, 1, ...$ and for all $s \geq 0$ and $t > 0$.

The jump probabilities are

$$\Pr\{S \to I_L\} = 1 - \alpha = 0.9 \quad \Pr\{S \to I_H\} = \alpha = 0.1$$

$$\Pr\{I_L \to S\} = \Pr\{I_H \to S\} = 1 \quad \Pr\{I_L \to I_H\} = \Pr\{I_H \to I_L\} = 0,$$

since these are the probabilities to transition from one state to another after the sojourn time has ended.

The transition rates can be found using the formula $q_{ij} = \Pr\{i \to j\} q_i$, where $q_i = \sum_{j \neq i} q_{ij}$, which results in

$$q_{SI_L} = (1-\alpha)\lambda = 0.009 \quad q_{SI_H} = \alpha\lambda = 0.001 \quad q_{I_L S} = \mu_L = \frac{1}{7} \quad q_{I_H S} = \mu_H = 0.05$$

$$q_{I_H I_L} = q_{I_L I_H} = 0.$$

This is illustrated in the transition diagram of $\{X(t) : t \geq 0\}$ in figure 1.

b) The long-run mean fractions of time spent in each state can be determined by solving the following system of equations:

$$\pi_{I_L} q_{I_L} = \pi_S q_{SI_L} + \pi_{I_H} q_{HI_L}$$

$$\pi_{I_H} q_{I_H} = \pi_S q_{SI_H} + \pi_{I_L} q_{I_L H}$$

$$\pi_S + \pi_{I_H} + \pi_{I_L} = 1,$$

where π_i is the long-run mean fraction of time spent in state i. This results in

$$\pi_S = 0.9233 \quad \pi_{I_H} = 0.0185 \quad \pi_{I_L} = 0.0582$$

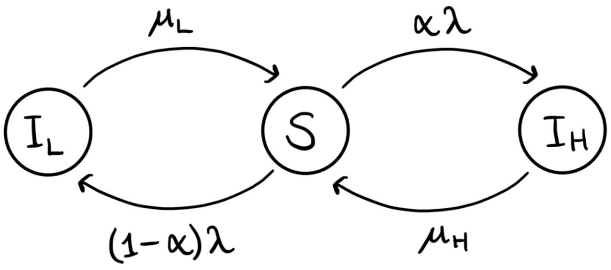

Figure 1: Transition diagram of $\{X(t) : t \geq 0\}$.

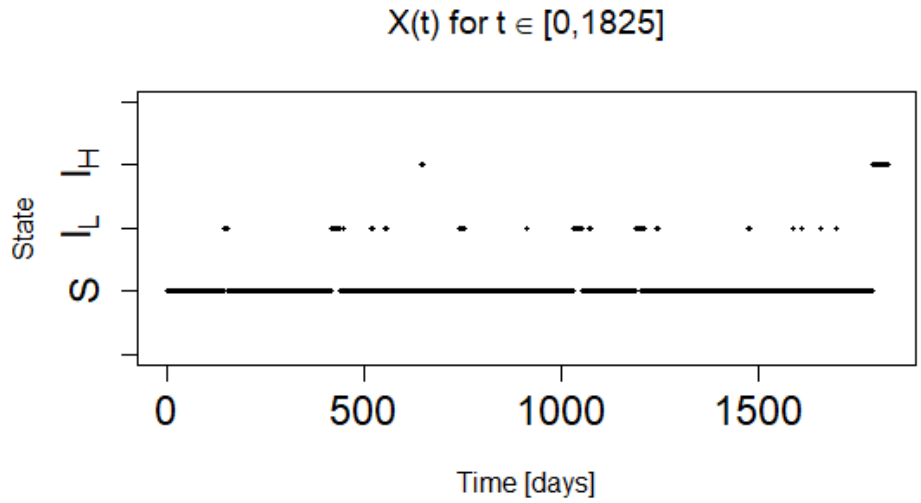

Figure 2: One realization of the continuous-time Markov chain $\{X(t) : t \in [0, 1825]\}$.

In other words an individual is on average infected, either lightly or heavily, 28 days per year, as the mean fraction of time spent in those states sums to 7.67%.

c) In figure 2 one realization of $\{X(t) : t \in [0, 1825]\}$ is shown. It is not possible to tell much about the general tendencies from this single realization, however we can see that the individual experienced two heavy and some light infections over the course of five years. The proportion of time spent in each state seems to correspond well to the results found in **b)**, stating that an individual is 92.33%, lightly infected 5.82% and heavily infected 1.85% of the time.

d) Based on one realization of $\{X(t) : t \in [0, 1000 \cdot 365]\}$, the long-run mean fraction of time that an individual has an infection is estimated to be 7.62%, which corresponds to on average 27.8 days per year. This seems to be a reasonable estimate compared to the result from **b)** where the long-run mean fraction of time spent in either of the infected states was found to be 7.67% or 27.97 days a year on average.

The estimate was found by storing the sums of the sojourn times in each state while simulating the process. After the simulation, the time spent in the infected states was divided by the total simulation time.

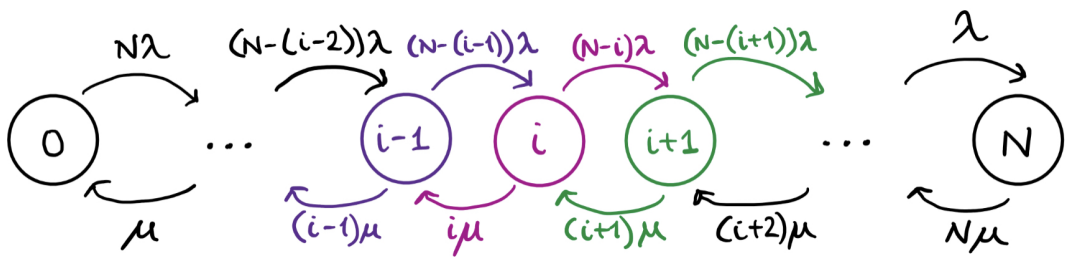

Figure 3: Transition diagram of $\{Y(t) : t \geq 0\}$. N is defined as the number of individual in a population, in this case 5.26 million.

e) To calculate the time between two heavy infections, we consider the time it takes to become heavily infected, giving that the starting state is susceptible. This is equivalent since the jump probability from state I_H to state S is one and the sojourn times are exponentially distributed and therefore memoryless.

We define $T = \min\{t \geq 0 : X(t) = I_H\}$ so that $v_i = E[T|X_0 = i]$ is the expected time until an individual becomes heavily infected, giving starting state i. We then want to solve the following set of equations

$$v_{I_H} = 0$$
$$v_S = E[S_S] + \Pr\{S \to I_L\}v_{I_L} + \Pr\{S \to I_H\}v_{I_H}$$
$$v_{I_L} = E[S_{I_L}] + \Pr\{I_L \to S\}v_S + \Pr\{I_L \to I_H\}v_{I_H}.$$

Here $v_{I_H} = 0$ because the starting state is the same as the desired state. By inserting the jump probabilities

$$v_S = \frac{1}{\lambda} + (1-\alpha)v_{I_L}$$
$$v_{I_L} = \frac{1}{\mu_{I_L}} + \frac{1}{\lambda} + (1-\alpha)v_{I_L},$$

we obtain that the expected time between heavy infections is

$$\mathrm{E}[T|X_0 = S] = v_S = 1063 \text{ days} \approx 2.9 \text{ years}.$$

f) $\{Y(t) : t \geq 0\}$ is a birth and death process because it in a given state i only is possible to transition to either $i+1$ or $i-1$ and because the sojourn times are independent of each other. In each state there are two competing independent processes, which determine the next jump,

$$1)\ T_1 = \text{ "time until next infection" } \sim \text{ Exp}(\lambda_i)$$
$$2)\ T_2 = \text{ "time until next recovery" } \sim \text{ Exp}(\mu_i),$$

so that the sojourn time in state i is $S_i = \min\{T_1, T_2\} \sim \text{ Exp}(\lambda_i + \mu_i)$.

As all individuals in the population become infected and susceptible completely independent of each other, the birth rate, i.e. the infection rate, will decrease as the number of susceptible individuals decreases. Correspondingly, the death rate, i.e. the recovery rate, will increase as the infected population increases.

Let N = 5.26 million be the total population size. Then the birth rate in state i is given by $\lambda_i = (N-i)\lambda$ for $i = 0, .., N$, where λ is the transition rate from susceptible to infected

for one individual. The death rate is similarly given by $\mu_i = i\mu$ for $i = 0, ..., N$, with μ the transition rate from infected to susceptible for one individual. This is illustrated in figure 3.

g) We use Little's law, where L is the average number of individuals with complications from a cold in the hospital, λ_h is the rate of arrival of individuals that requires hospitalization, and W is the average treatment time. We wish to determine W such that L does not exceed the capacity of the hospital. To determine W, we first have to determine λ_h. Since all individuals are independent of each other, we can calculate the long-run mean fraction of time one individual is susceptible and multiply this with the population size to get the average number of individuals susceptible in a given day.

By solving the following set of equations,

$$\pi_S \lambda = \pi_I \mu$$

$$\pi_S + \pi_I = 1,$$

the long-run mean fraction of time an individual is susceptible is determined to be $\pi_S = \frac{\mu}{\lambda+\mu} = \frac{100}{107} = 0.935$. The average number of susceptible individuals per day is then

$$5.26\pi_S = 5.26 \cdot \frac{100}{107} = 4.9159 \text{ million individuals.}$$

Using that the transition rate from susceptible to infected for one individual is $\lambda = 0.01$, the average number of individuals getting infected per day is 49159. As there is a 1% probability that an infection will require hospitalization, $\lambda_h = 491.6$ individuals per day.

Using Little's law, with $W = \lambda_h / L$ with $L = 2000$ and $\lambda_h = 491.59$, the maximal average required treatment time to ensure the average number of individuals in the hospital does not exceed its capacity is then determined to be 4.068 days.

Calibrating Climate Models

a) In figure 4 a prediction of $Y(\theta)$, conditional on the five evaluation points given in the project description, as well as a 90% prediction interval is shown.

b) A plot of the conditional probability $Y(\theta) < 0.30$ as a function of θ given the five evaluation points is shown in figure 5.

c) The prediction of $Y(\theta)$ conditional on the six evaluation points is shown in figure 6. Furthermore, the probability that $Y(\theta) < 0.30$ conditional on these points is shown in figure 7.

In figure 7 it appears that the probabilty of $Y(\theta) < 0.30$ is maximal at $\theta = 0.36$ with $\Pr\{Y(\theta = 0.36) < 0.30\} = 0.18$. Although there also appears to be a relatively high value for $\Pr\{Y(\theta) < 0.30\} = 0.15$ for $\theta = 0.25$, figure 6 shows that there is much higher variance at $\theta = 0.25$ than at $\theta = 0.36$. We would therefore suggest that the scientist use $\theta = 0.36$ to maximize their chances of achieving $Y(\theta) < 0.30$.

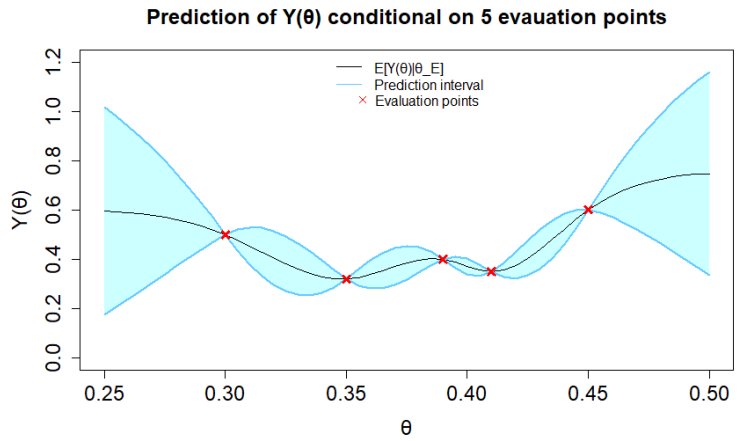

Figure 4: Prediction of $Y(\theta)$ conditional on the five evaluation points $(\theta,\, y(\theta))$: $(0.30, 0.5)$, $(0.35, 0.32)$, $(0.39, 0.40)$, $(0.41, 0.35)$, and $(0.45, 0.60)$. The graph includes a 90% prediction interval.

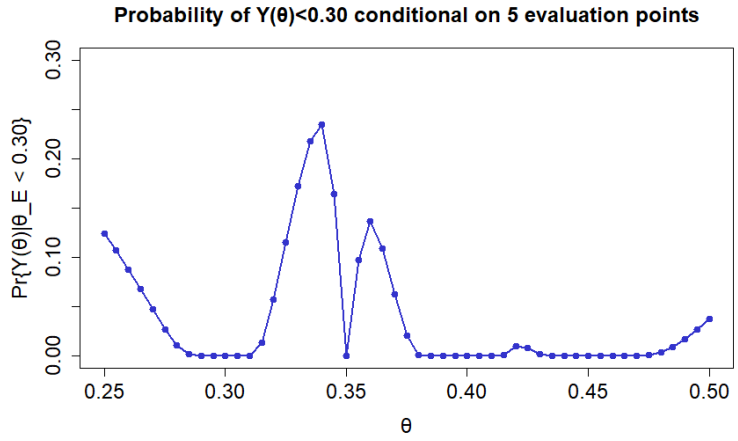

Figure 5: The probability of $Y(\theta) < 0.30$ conditional on the five evaluation points $(\theta,\, y(\theta))$: $(0.30, 0.5)$, $(0.35, 0.32)$, $(0.39, 0.40)$, $(0.41, 0.35)$, and $(0.45, 0.60)$.

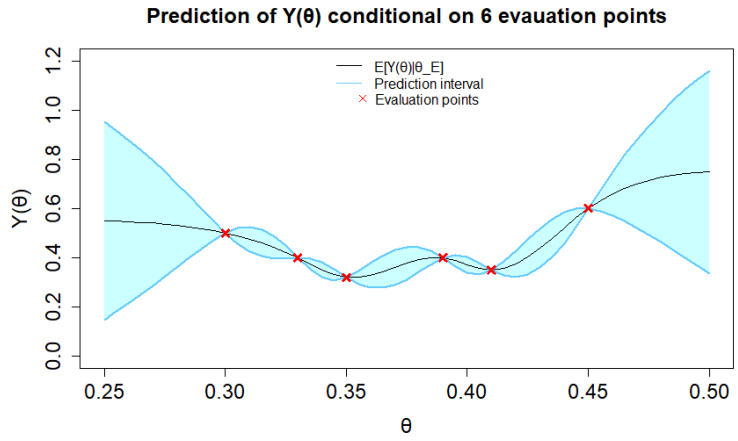

Figure 6: Prediction of $Y(\theta)$ conditional on the six evaluation points $(\theta, y(\theta))$: $(0.30, 0.5)$, $(0.33, 0.40)$, $(0.35, 0.32)$, $(0.39, 0.40)$, $(0.41, 0.35)$, and $(0.45, 0.60)$. The graph includes a 90% prediction interval.

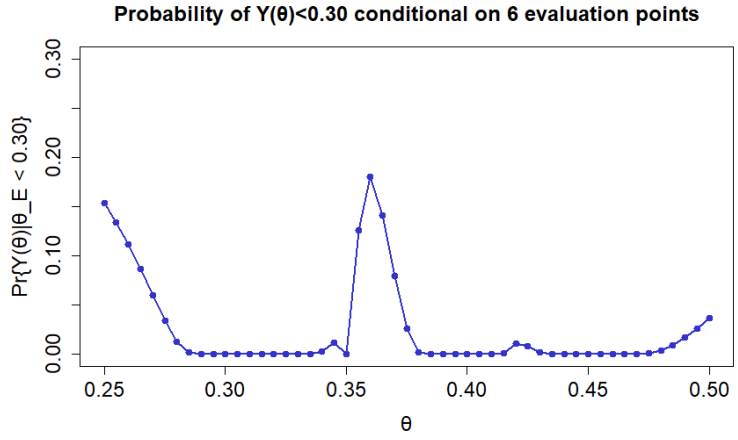

Figure 7: The probability of $Y(\theta) < 0.30$ conditional on the six evaluation points $(\theta, y(\theta))$: $(0.30, 0.5)$, $(0.33, 0.40)$, $(0.35, 0.32)$, $(0.39, 0.40)$, $(0.41, 0.35)$, and $(0.45, 0.60)$.