Project 2

TMA4265

Clara Panchaud
Silje Anfindsen

November 8, 2019

This is the project report for Project 2 in Stochastic Modelling. The computer code for all the problems can be found in the file: "Project2.ipynb".

1 Problem 1: Modelling the common cold

In this problem we will model the common cold, where $X(t)$ denotes the state for an individual at the time t, measured in days.

1.1 a)

A continuous-time Markov chain satisfies the Markov property, has stationary transition probabilities and a discrete state space.

The state space for $X(t)$ is $\{S, I\}$, which is discrete. The waiting times T between the two states are exponentially distributed. This means that T is memoryless, and then $X(t)$ satisfies the Markov property. This also implies that the transition probabilities are stationary. The model makes sense as humans become infected with the common cold independent of the time since their last infection.

The transition rate λ between each infection and μ, between an infection and the susceptible state, are both visualized in the transition diagram in Figure 1.

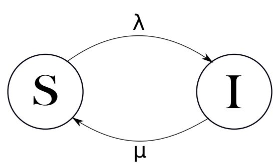

Figure 1: Transition diagram for X(t)

1.2 b)

We will now calculate the long-run mean fraction of time per year that an individual has a cold. Since X(t) is a continuous-time Markov chain without any absorbing states, where the limits satisfy (1), we have a limiting distribution, that also is a stationary distribution which satisfies (2).

$$\pi_S + \pi_I = 1 \tag{1}$$

$$\lambda \pi_S = \mu \pi_I \tag{2}$$

We will use (1) and (2) to find the long-run mean fractions of time in each state.

$$0 = -\lambda\pi_S + \mu(1 - \pi_S)$$
$$0 = \mu - (\lambda + \mu)\pi_S$$
$$\pi_S = \frac{\mu}{\lambda + \mu} \text{ and } \pi_I = \frac{\lambda}{\lambda + \mu}$$

We are interested at the value of π_I which is ~ 0.0654, the fraction of time spent in the infected state. Now, for one year, we multiply this value by 365 days and find that in the long-run an individual has a cold on average 23.87 days per year.

1.3 c1)

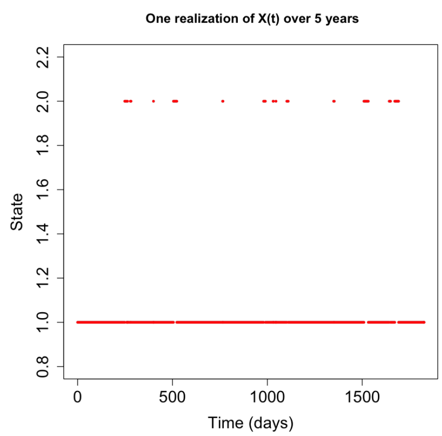

Figure 2: 1 realizations of $X(t)$, $0 \leq t \leq 5 * 365$ days

1.4 c2)

We will now calculate the long-run mean fraction of time per year that an individual has a cold, based on one realization of 1000 years, numerically. The idea is to make a list of all states visited and a list of the arriving times in a states during the period. Then calculate the total time spent in state I during the period in order to find the fraction of time spent in this state per year.

The numerical result is $24, 12$ days which is close to the theoretical value from 1b).

1.5 d)

Let $Y(t)$ denote the number of infected individuals in the population at time t measured in days. The total population, N contains 5.26 million individuals.

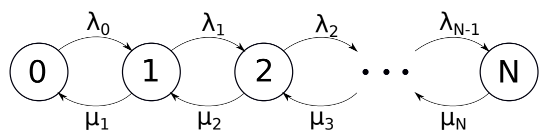

Figure 3: Transition diagram for Y(t)

The process $Y(t)$ is a birth and death process as the number of infected individuals either increase with one infection (birth) or decrease with one recovery (death). All times until next birth or death are independent and exponentially distributed. The birth rates are given by

$$\lambda_i = (N - i)\lambda$$

for $i = 0, 1, 2, .., N - 1$ and the death rates are

$$\mu_i = i\mu$$

for $i = 1, 2, ..., N$

All the assumptions necessary for this birth and death process are already satisfied in the problem introduction. Therefore we do not need to make any more assumptions.

1.6 e)

We assume that $\{Y(t) : t \geq 0\}$ has reached its stationary distribution. For each infection there is a 1% probability that the person will require hospitalization. The hospitals have an average capacity of 2000 individuals. We will use Little's law (3), to calculate the average treatment time required to not exceed the capacity.

L: Average number of people in the system

α : Arrival rate

W: Average time spent in the system for each person

$$L = \alpha W \tag{3}$$

We want to find W when $L = 2000$. This will yield the maximum average time to treat a patient which does not result in exceeding the capacity. Then we find the rate at which people arrive at the hospital, α.

Being at the stationary distribution, we can take the average of the birth rates to be the average number of healthy people multiplied with λ. Recall that the birth rate for the model is $\lambda_i = (N-i)\lambda$, where $N-i$ is the number of susceptible people at state i. Since we have reached the stationary distribution, we can use the result π_I from 1b) to say that the average number of susceptible people is $(1-\pi_I)N$. We know that 1% of these people require hospitalization. So the rate is :

$$\alpha = \frac{1}{100}(1-\pi_I)N\lambda \tag{4}$$

And therefore

$$W = \frac{L}{\alpha} = 4.068 \tag{5}$$

The average treatment time should be maximum 4.068 days.

2 Problem 2: Calibrating climate models

In this problem we will use a Gaussian process $\{Y(\theta) : \theta \in [0,1]\}$ to model the unknown relationship between the parameter value θ and the score $Y(\theta)$.

2.1 a)

In this problem we start by making a grid of 51 parameter values. We want to calculate the conditional means and covariances on the five evaluation points.

In order to do this we first need to find the Gaussian distribution for the two random variables, X. Let X_a be the random variable with the 51 values from the grid, while X_b is the random variable with the five evaluation points. Then X is defined below:

$$\overrightarrow{X} = (\overrightarrow{X_a}, \overrightarrow{X_b}) \sim N_{n_a+n_b}\left(\begin{bmatrix} \overrightarrow{\mu_a} \\ \overrightarrow{\mu_b} \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \right)$$

We can now find the distribution of the conditional variable, X_c.

$$X_a \mid X_b = X_c \sim N_{n_a}(\overrightarrow{\mu_c}, \Sigma_c)$$

Where the conditional mean, $\overrightarrow{\mu_c}$ and the covariance matrix Σ_c are defined below:

$$\overrightarrow{\mu_c} = \overrightarrow{\mu_a} + \Sigma_{ab}\Sigma_{bb}^{-1}(\overrightarrow{X_b} - \overrightarrow{\mu_b})$$

$$\Sigma_c = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

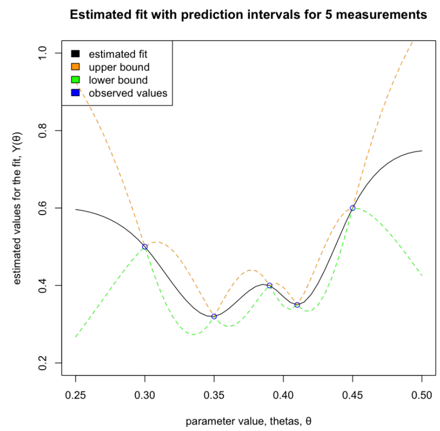

Figure 4: The prediction of $Y(\theta)$ along with the prediction interval conditioned on the 5 evaluation points.

2.2 b)

The scientists would like to achieve a score below 0.30. We used the results from 2.a) to compute the probability that y$(\theta) < 0.30$ given the 5 evaluation points. The results are displayed in Figure 5. We observe that the probability is the highest at θ=0.34. Since we know the values of the 5 evaluation points and they are all above 0.30, their probability is obviously zero.

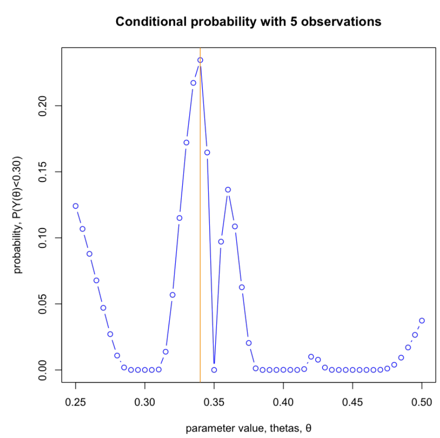

Figure 5: Probability that y$(\theta) < 0.30$ as a function of θ given the 5 evaluation points.

2.3 c)

We have now received a new observation of (θ,y(θ) : (0.33,0.40). We followed the same procedure as in 2.a) and 2.b) to get the prediction, prediction intervals and probability plot. The Figure 6 below shows that at the 6 known points the prediction interval is restricted to the known value, which makes perfect sense. In Figure 7, we can find on the green line that for θ=0.36 we have the maximum probability to achieve y(θ) < 0.30. Therefore we would suggest the scientists to use that θ for their next run of the climate model.

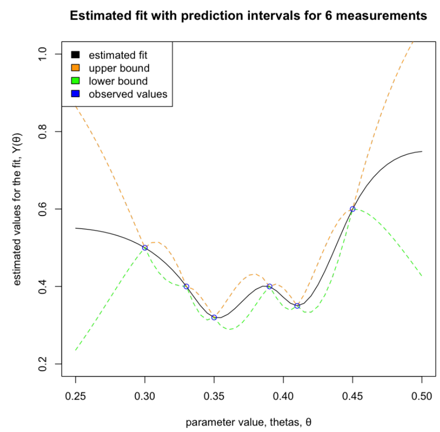

Figure 6: The prediction of $Y(\theta)$ along with the prediction interval conditioned on the 6 evaluation points.

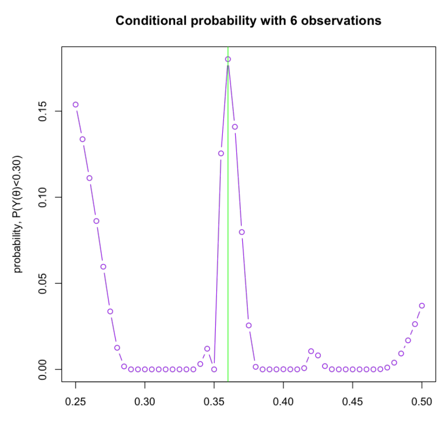

Figure 7: Probability that y(θ) < 0.30 as a function of θ given the 6 evaluation points.