

# Stochastic Modelling - Project 2

Helge Bergo  
Lars Sæle

November 2019

## Problem 1: Modelling the common cold

a)

Based on the fact that we have a state space  $\{S, I\}$ , the individuals in the population are independent and the transition rates follows an exponential distribution, we can state that we have a Markov process. There are no absorbing states, as well as the individual can be sick at any point in time, not at specific discrete time points.

The transition rates are related to the exponential distribution of the time until next infection and time until recovery. As the expected time until next infection is  $\frac{1}{\lambda} = 100$ , the transition rate is  $\lambda = \frac{1}{100}$ . The same goes for time until recovery, and the transition rate is  $\mu = \frac{1}{7}$ .

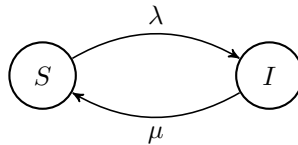


Figure 1: transition diagram

b)

The long-run mean fraction of time per year is calculated by using the fact that the rate in has to be equal to the rate out. Using this, and the fact that  $\pi_S + \pi_I = 1$ , we get the following result:

$$\begin{aligned}\text{"rate in"} &= \text{"rate out"} \\ \pi_S \lambda &= \pi_I \mu \\ \pi_I \mu &= (1 - \pi_I) \lambda \\ \pi_I (\lambda + \mu) &= \lambda \\ \pi_I &= \frac{\lambda}{\lambda + \mu} \\ \pi_I &= \frac{\frac{1}{100}}{\frac{1}{100} + \frac{1}{7}} \\ \pi_I &= \frac{7}{107} \approx 0.0654,\end{aligned}$$

where  $\pi_I$  is the mean fraction of time spent in the infected state. For a year, the mean number of days spent in the infected state is equal to  $\frac{7 \cdot 365}{107} \approx 24$ .

c)

The Markov chain was simulated for 1000 years, and a visualisation is shown of the first 5 years in Figure 2, where state 1 equals susceptible, and 2 infected.

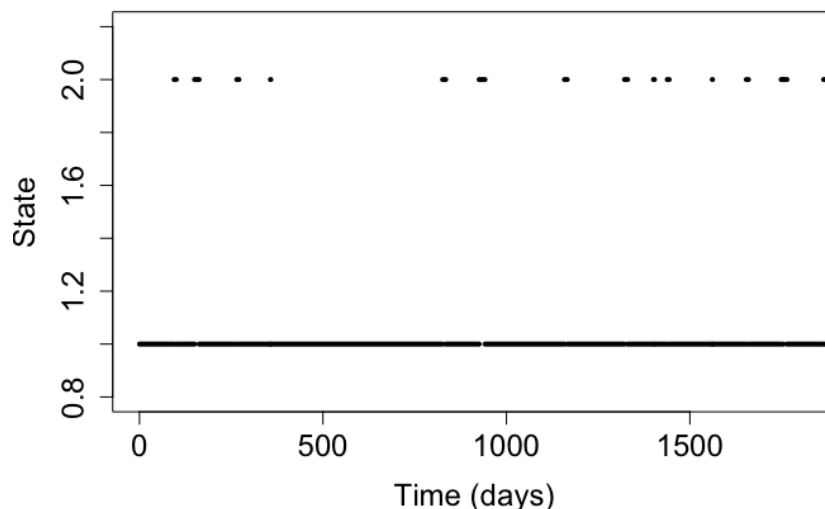


Figure 2: One realisation of  $X(\cdot)$  over 5 years

From the simulation over 1000 years, the long-run mean fraction was calculated to be 0.06534, very close to the theoretical value from **b)**. This was calculated by dividing the time spent as infected by the total running time of the simulation, to get the fraction of infected time.

d)

If  $Y(t)$  is the number of infected individuals at time  $t$ ,  $Y(t)$  can be seen as a birth and death process. This comes from the fact that if in state  $Y(t)$  at time  $t$ , in the interval  $(t, t+h)$ , the number of infected individuals can go up with probability  $\lambda h$  or down with probability  $\mu h$ . So in essence, there is a birth of an infection, or the death of an infection at time  $t$ , where  $t$  is continuous.

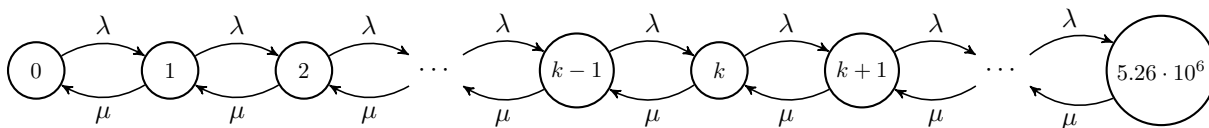


Figure 3: Transition diagram

As seen in Figure 3, we have taken the assumption that  $\lambda_i$  and  $\mu_i$  is constant for each  $i$ .

e)

**Theorem 0.1** *For a stable queuing system,*

$$L = \lambda W, \quad (1)$$

where  $L$  is the average number of individuals in the system,  $\lambda$  is the rate of arrivals and  $W$  is the average time spent by an individual in the system.

From task **b)**, we get the mean fraction of time an individual is sick per year. This result can be used as the value  $L$  from theorem 1 as the average number of infections in the system. The rate of arrivals to the hospital is 1% of these infections, which is  $\lambda$ . This leads to the following result:

$$W = \frac{L}{\lambda} = \frac{\pi_k}{\lambda} = \frac{\frac{7}{107}}{0.1} = 6.54 \quad (2)$$

and we see that the hospital needs to have an average treatment time of 6.5 days to not exceed their capacity.

## Problem 2: Insurance claims

a)

The Gaussian process was simulated by following the procedure in *Note on Gaussian Processes*, and the notation is similar as described there.

The discretized grid of  $\theta$  from 0.25 to 0.50 was constructed, and using a Matern type correlation function with parameter  $\phi_M = 15$ , the block covariance matrices,  $H_A$  and  $H_B$  were created.

$$H_i = |t_i 1'_{ni} - 1_{ni} t'_i| \quad (3)$$

for  $i = A, B$ .

$H_{A,B}$  was created from the differences between the two time step vectors  $t_A$  and  $t_B$

$$H_{A,B} = |t_A 1'_{nB} - 1_{nA} t'_B| \quad (4)$$

After this, the covariance matrices,  $\Sigma$ , were built using the following formulas

$$\Sigma_A = \sigma^2(1 + \phi_M H_A) \exp(-\phi_M H_A) \quad (5)$$

$$\Sigma_B = \sigma^2(1 + \phi_M H_B) \exp(-\phi_M H_B) \quad (6)$$

$$\Sigma_{A,B} = \sigma^2(1 + \phi_M H_{A,B}) \exp(-\phi_M H_{A,B}) \quad (7)$$

finally, the expected value  $y(\theta) = E[x_A|x_B]$  was calculated, together with the variance.

$$E[x_A|x_B] = \mu_A + \Sigma_{A,B} \Sigma_B^{-1} (x_B - \mu_B) \quad (8)$$

$$Var[x_A|x_B] = \Sigma_A - \Sigma_{A,B} \Sigma_B^{-1} \Sigma'_{A,B} \quad (9)$$

From this formula, the diagonal in the matrix is the variance of the system. The 90% prediction intervals were calculated from Equation 10, where  $z_{0.05} = 1.64$ , and the model was plotted with the 5 conditional values, as well as upper and lower prediction intervals, shown in Figure 4. The expected value is a function of  $\theta$  with the given step length of 0.005.

$$E[x_{A,i}|x_B] \pm z_{0.05} \sqrt{\text{Var}(x_{A,i}|x_B)} \quad (10)$$

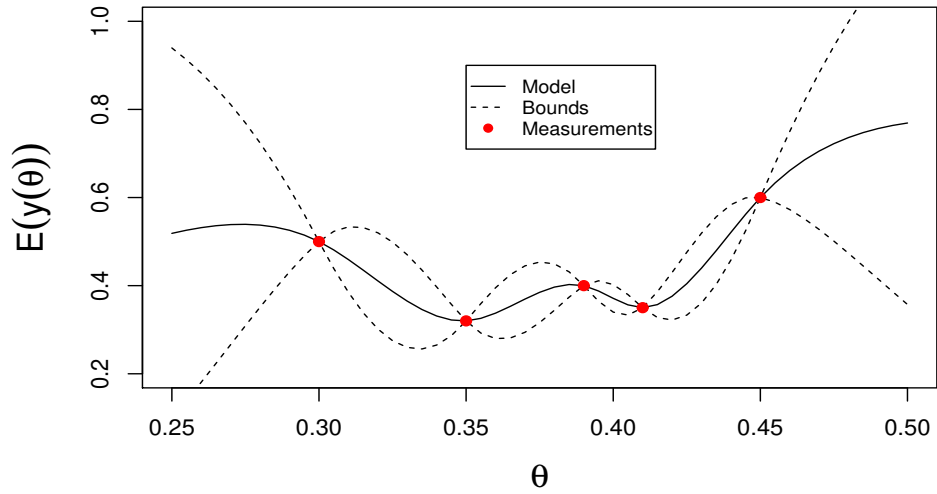


Figure 4: The model of the expected value as a function of  $\theta$ , from the 5 evaluation points.

b)

The conditional probability that  $y(\theta) < 0.3$  was calculated using Equation 11, with a Gaussian distribution function  $\Phi$ . This gives a vector of 51 values, one for each  $\theta$ , and this was plotted as a function of  $\theta$  in Figure 5.

$$P(x_{A,i} < 0.30|x_B) = \Phi\left(\frac{0.30 - E[x_{A,i}|x_B]}{\sqrt{\text{Var}[x_{A,i}|x_B]}}\right) \quad (11)$$

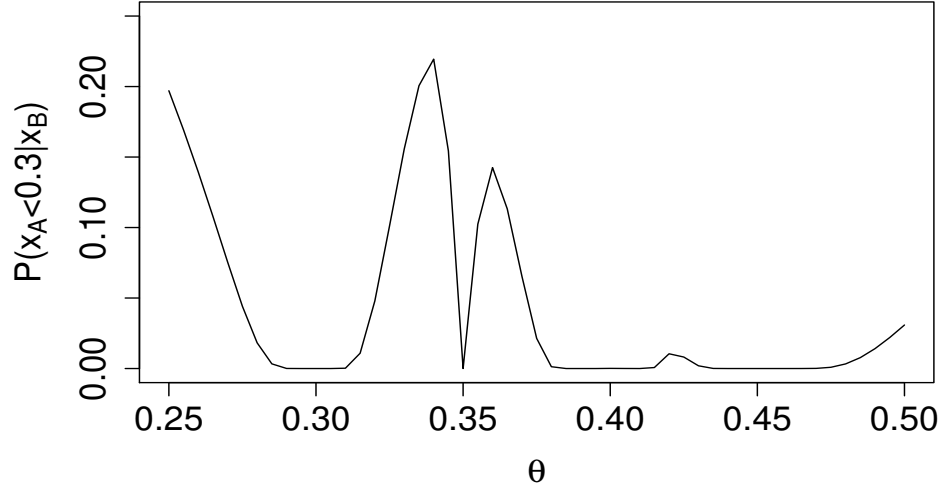


Figure 5: Conditional probability of  $y(\theta) < 0.3$  as a function of  $\theta$  with 5 conditional values.

c)

After the model was run again with the new measurement, 6 conditional values in total, the following plots were produced. The result is similar, but the prediction intervals are much smaller between 0.30 and 0.35. The rest of the model graph is mostly unaffected by the additional conditioning value, since the correlation decreases the longer away from  $\theta = 0.33$  we look.

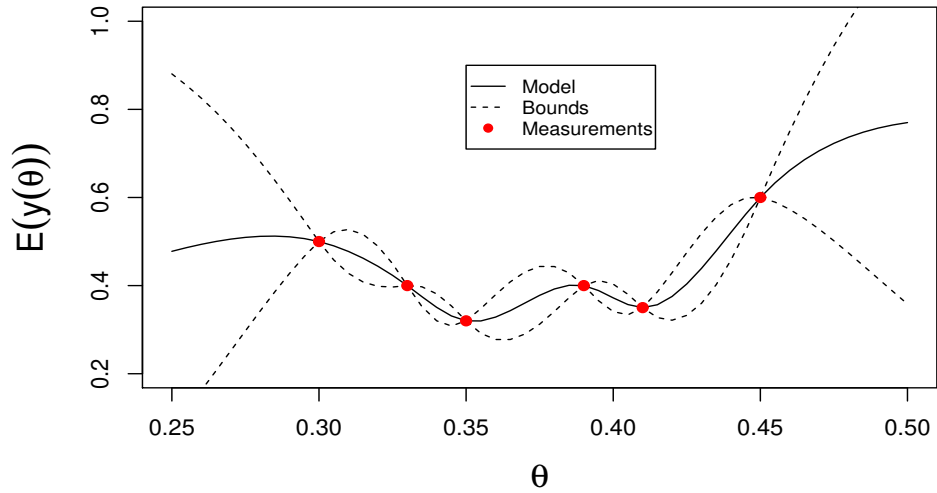


Figure 6: The model of the expected value as a function of  $\theta$ , from the 6 evaluation points.

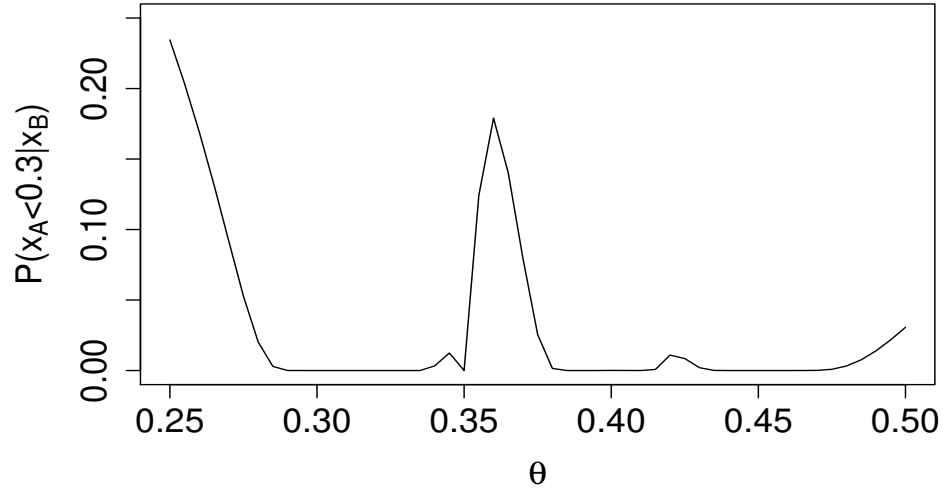


Figure 7: Conditional probability of  $y(\theta) < 0.3$  as a function of  $\theta$  with 6 conditional values.

The best guess for another run of the model is a  $\theta$  value of 0.36, giving a probability of 0.179048 to get  $y(\theta) < 0.30$ . The probability seems higher at the leftmost part of the  $\theta$ -values, but this is on the outer boundary of the model, and therefore not taken into account.