Project: Hidden Markov models and Value of Information

Mina, Henrik

October 2019

Task a)

The marginal probability of $p(x_i = 1)$ for $i = 1, \dots, 50$ will be

$$\begin{aligned} p(x_i = 1) &= \sum_{j=0}^{1} p(x_i = 1, x_{i-1} = j) = \sum_{j=0}^{1} p(x_i = 1 | x_{i-1} = j) p(x_{i-1} = j) \\ &= p(x_i = 1 | x_{i-1} = 0) p(x_{i-1} = 0) + p(x_i = 1 | x_{i-1} = 1) p(x_{i-1} = 1) \\ &= 0.05 p(x_{i-1} = 0) + p(x_{i-1} = 1) \\ &= 0.05(1 - p(x_{i-1} = 1)) + p(x_{i-1} = 1) \\ &= 0.05 + 0.95 p(x_{i-1} = 1), \end{aligned}$$

and this is plotted in Figure 1.

Task b)

The expected prior value will be

$$PV = \max(-100000, -5000 \sum_{i=1}^{50} p(x_i = 1)).$$

We can compute the right hand statement in the max expression to be minus 5000 times the sum of all $p(x_i = 1)$, which is -158 617,6. This is smaller than the value of decision $a = 0$. Therefore, the optimal decision is $a = 0$, to clean the tracks in advance, and the prior value is $PV = -100000$.

Task c)

We can now install sensors at every point x_i which observe information, y_i, according to $p(y_i | x_i) = N(x_i, \tau^2)$, with $\tau = 0.3$. We assume that we observed $\mathbf{y}_D = (0.2, 0.7)$ at points $D = (20, 30)$. The rest of the points in $\mathbf{y}_D$ are simulated

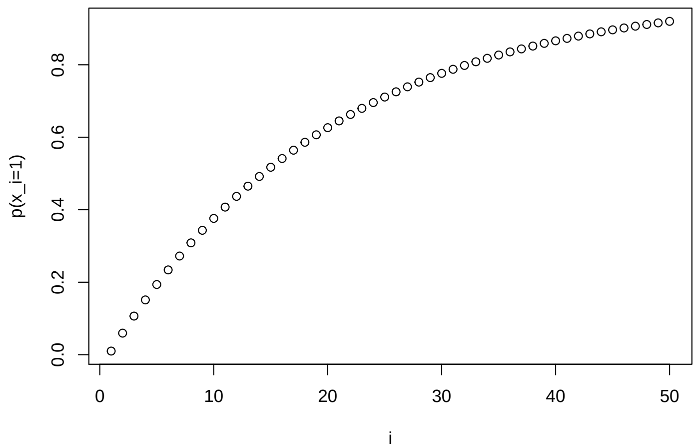

Figure 1: Marginal probabilities of $p(x_i = 1)$.

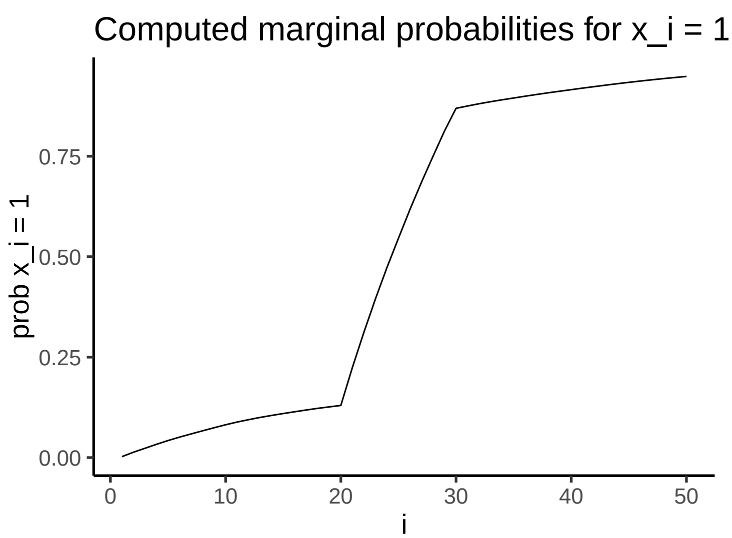

Figure 2: Marginal probabilities of $p(x_i = 1|\mathbf{y}_D)$.

from $N(x_i, 100\tau)$. The resulting marginal probabilities $p(x_i = 1|\mathbf{y}_D)$ are plotted in Figure 2.

Looking at Figure 2, we can see that the probabilities whether a certain location is at high risk or low risk is low in the beginning, and increases as i increases. Also, there is a steep increase between $i = 20$ and $i = 30$, because we here have an increase in y_i from 0.2 to 0.7.

Task d)

Before placing sensor(s), a goal is to find the best design using value of information (VOI) analysis. The expected posterior value (PoV) with information, when a single sensor is placed at location $k = 1, \ldots, n$ is:

$$PoV(k) = \int \max\{-100000, -5000 \sum_{i=1}^{50} p(x_i = 1|y_k)\} p(y_k) dy_k.$$

We then use Monte Carlo sampling of data to approximate the PoV(k) and VOI(k) for all values of $k \in \{1, \ldots, 50\}$, using 20 000 Monte Carlo samples. For each Monte Carlo sample of data, the forward-backward algorithm is run to calculate the posterior marginal probabilities and the integrand required for the PoV.

In Figure 3 the VoI is plotted as a function of single sensor locations $k = 1, \ldots, 50$. The best location of installation is around $k = 32$, although this could vary somewhat if we run the Monte Carlo integration again. The total rental and installation cost of one sensor is 10 000 kroner. This means that the sensor is worth its price, since the maximum VoI is about 11 000 kr, which is greater than the price of the sensor.

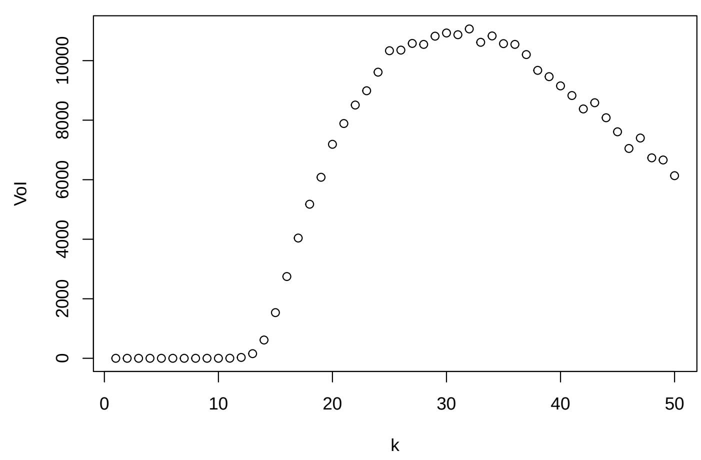

Figure 3: VoI as a function of sensoring location.

Task e)

VOI at the design $D = \{20, 30\}$ is about 12 250 kroner. In Figure 4 the VoI is plotted for placing two sensors at locations $k = 1, \ldots, 50$. On the diagonal, we have plotted the VoI for placing two sensors at the same location. In this case, the information we get is the mean of the two sensor observations. Because both are assumed independent given x_i, with $p(y_i^j|x_i) \sim N(x_i, \tau^2)$, $j = 1, 2$, the mean of the two observations would be distributed as

$$\bar{y}_i = \frac{y_i^1 + y_i^2}{2} \sim p(\bar{y}_i|x_i) = N\left(x_i, \frac{\tau^2}{2}\right).$$

Therefore, placing two sensors in a single location would be the same as placing a single sensor in that location with half the variance of the original two sensors. Looking at the figure, the maximum VoI is obtained by placing two sensors at locations $k = 29$ and $k = 32$, with a VoI of 13 525 kroner. However, this is not completely accurate, and even more than 10 000 Monte Carlo samples should be drawn to achieve an accurate result. Further, placing two sensors at the same location is best for $k = 31$ with a VoI of 11 500 kroner. Still, the VoI by placing two sensors is smaller than 15 000 kroner, so the decision is to not place two sensors, given the quoted cost.

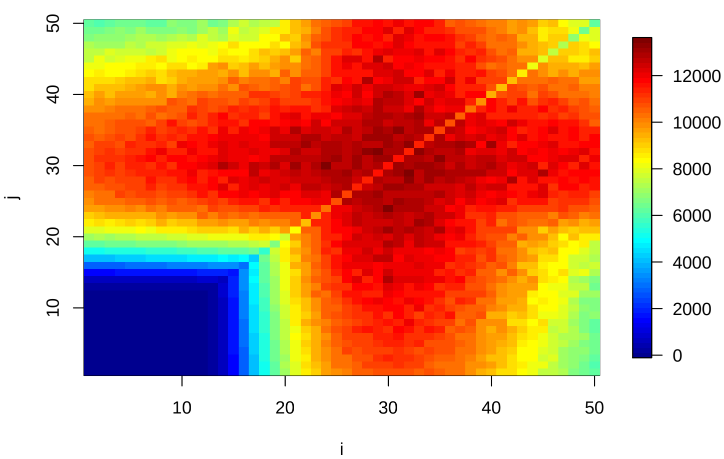

Figure 4: VoI as a function of sensoring location for two sensors. Notice the symmetry in the placement of sensors.

Task f)

We now measure the full road through a satellite. Satellite data is modelled with $p(y_i|x_i) = N(x_i, 1^2)$.

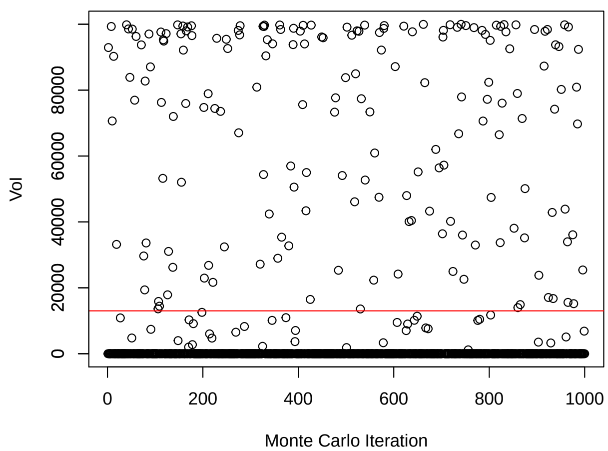

Figure 5: VOI of satellite data, approximated by Monte Carlo sampling. Depicted on Figure the first 1000 samples. Red line represents the mean of the full sample.

Looking at Figure 5 we can see that most of the time the VOI tells us that we are better off cleaning all of the tracks. The rest of the time data is quite spread, which is not strange considering the standard deviation is quite large, namely 1. This is also illustrated with the red line in Figure 5, which represents the mean of the full sample, which is ca 13 000 kr, which is larger than just placing one sensor that has higher accuracy.