Project 2, MA8001

Henrik S. Lie

September 2019

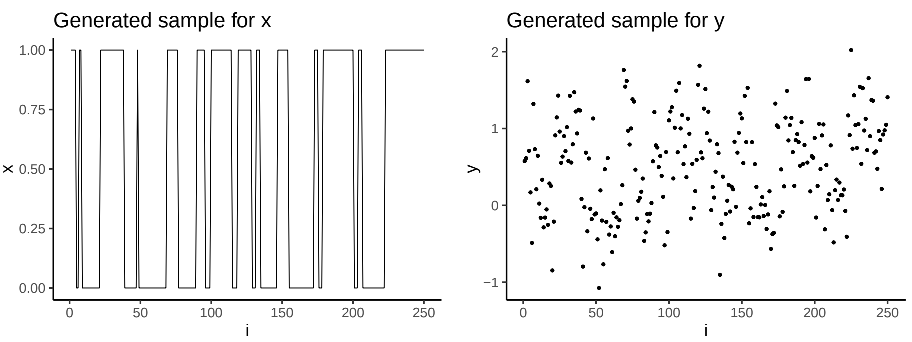
Exercise a)

We start by simulating a markov chain with binary outcome, $x_i \in \{0, 1\}$, $i = 1, \dots, n$ and stationary conditional probabilities $p(x_{i+1} = l | x_i = k) = P(k, l)$.

Data $\mathbf{y} = (y_1, \dots, y_n)$ are conditionally independent, given the variable of interest. We define a Gaussian likelihood, $p(y_i|x_i) = N(x_i, \tau^2), i = 1, \dots, n$.

To generate data, we simulate a markov chain with $n = 250$, $p = P(0, 0) = P(1, 1) = 0.9$, $p(x_1 = 1) = 0.5$. Then, for each x_i, we generate data $y_i = N(x_i, \tau^2)$, $i = 1, \dots, n$, with $\tau = 0.4$.

The generated data are plotted in figure 1. We see that the chain stays in the same value for several iterations before transitioning to another state. The data for $\mathbf{y}$ is also fairly mixed, and it is difficult to determine $\mathbf{x}$ based purely on the plot of $\mathbf{y}$.



(a) Generated sample for $\mathbf{x}$. (b) Generated sample for $\mathbf{y}$.

Figure 1: Generated data for this exercise. $\mathbf{y}$ are now considered known, or observed. $\mathbf{x}$ are "hidden" and to be estimated.

Exercise b)

We now do forward recursion to evaluate the marginal likelihood of the data, $p(\mathbf{y})$.

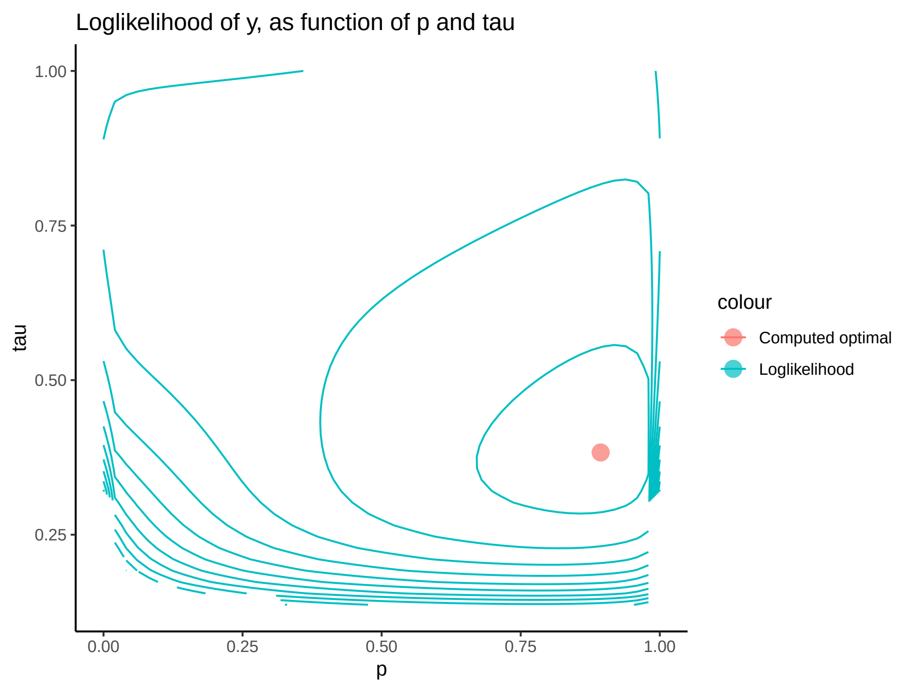


Figure 2: Contour plot of the loglikelihood $p(\mathbf{y}$, with $\mathbf{y}$ from a). The indicated point is the optimal value for p and τ.

We first initialize the first iteration by setting

$$C_1 = \left(\sum_{i=0}^{1} p(y_1|x_1 = i) \times p(x_1 = i) \right)^{-1}$$

$$p(x_1 = i|y_1) = C_1 \times p(y_1|x_1 = i) \times p(x_1 = i), i \in \{0, 1\}.$$

Then, we do forward recursion for $t = 2, \ldots, n$ by the formula

$$C_t = \left(\sum_{i=0}^{1} \sum_{j=0}^{1} p(y_t|x_t = j) \times P(i,j) \times p(x_{t-1} = i|\mathbf{y}_{1:(t-1)}) \right)^{-1},$$

$$p(x_{t-1} = i, x_t = j|\mathbf{y}_{1:t}) = C_t \times p(y_t|x_t = j) \times P(i,j) \times p(x_{t-1} = i|\mathbf{y}_{1:(t-1)}),$$

$$p(x_t = i|\mathbf{y}_{1:t}) = \sum_{i=0}^{1} p(x_{t-1} = i, x_t = j|\mathbf{y}_{1:t}).$$

Then, finally, we arrive at the expression $p(\mathbf{y}) = \prod_{t=1}^{n} 1/C_t$.

We then compute the marginal likelihood model for a grid of values for p and τ, given the data generated in a). Then, we plot the log of the marginal likelihood in figure 2. The loglikelihood is then optimized using the R function `optim`. The resulting values are $p = 0.9001172$, $\tau = 0.4188034$, which are almost equal to the "true" values set in a).

Exercise c)

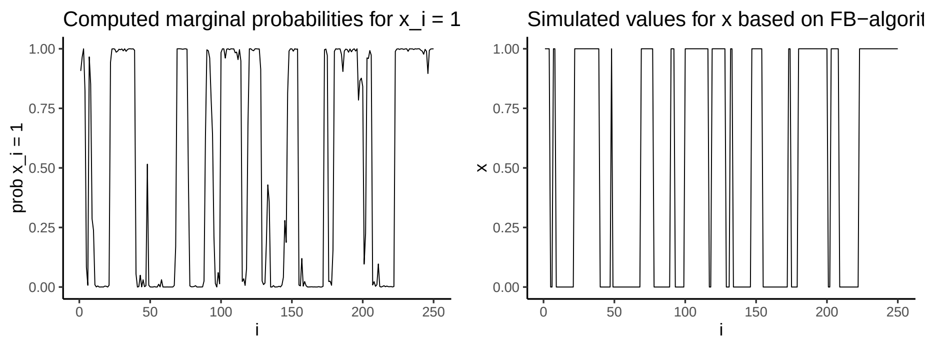
We then use the backward recursion to find the marginal probabilities $p(x_i = 1|y_1, \ldots, y_n)$, for all $i = 1, \ldots, n$. We first note that $p(x_n = i|\mathbf{y})$ is known from the forward recursion, and this is used as initialization for the backward recursion. Then, we can compute the backward probabilities by the following equations for $t = n, \ldots, 2$

$$\begin{aligned}
p(x_{t-1} = i, x_t = j|\mathbf{y}) &= \frac{p(x_{t-1} = i, x_t = j|\mathbf{y}_{1:t})}{p(x_t = i|\mathbf{y}_{1:t})} \times p(x_t = i|\mathbf{y}), \\
p(x_{t-1} = i|\mathbf{y}) &= \sum_{j=0}^{1} p(x_{t-1} = i, x_t = j|\mathbf{y}), \\
p(x_t = j|x_{t-1} = i, \mathbf{y}) &= \frac{p(x_{t-1} = i, x_t = j|\mathbf{y})}{p(x_{t-1} = i|\mathbf{y})}.
\end{aligned}$$

Then, we note that the full posterior model can be written by the backward probabilities as

$$p(\mathbf{X} = \mathbf{x}|\mathbf{y}) = p(x_1|\mathbf{y}) \times \prod_{t=2}^{n} p(x_t|x_{t-1}, \mathbf{y}).$$

The probabilities for $p(x_i = 1|\mathbf{y})$ are plotted in figure 3a.



(a) Marginal probabilities for $p(x_i = 1|\mathbf{y})$. (b) Generated sample of $\mathbf{x}$ based on $p(\mathbf{x}|\mathbf{y})$.

Figure 3: Marginal probability and one generated sample based on the backward recursion algorithm.

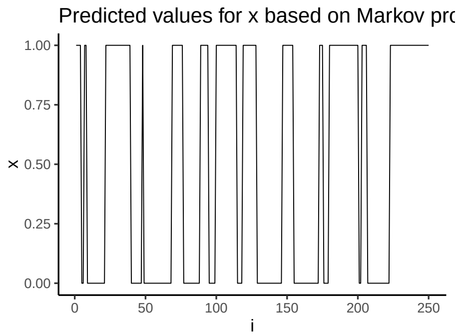
Having computed the probabilities by the backward recursion algorithm, we are also able to simulate values for $\mathbf{x}$. First, we sample x_1^s from $p(x_1|\mathbf{y})$, and then we can sample $x_i^s \sim p(x_i|x_{i-1}^s, \mathbf{y})$ for $i = 2, \ldots, n$. The resulting sample $\mathbf{x}^s$ is plotted in figure 3b.

Exercise d)

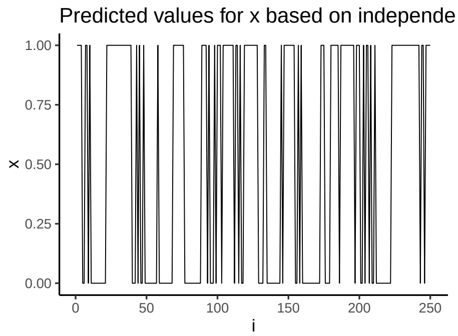
We now want to compare results between predicted values for $\mathbf{x}$ with and without the markov assumption. If we assume markov dependence, $x_i = 1$

if $p(x_i = 1|\mathbf{y}) > 0.5$, and $x_i = 0$ if $p(x_i = 1|\mathbf{y}) < 0.5 \Rightarrow p(x_i = 0|\mathbf{y}) > 0.5)$. The resulting prediction is plotted in figure 4a.

If we assume no markov dependence, then x_i will be predicted based solely on the individual observations y_i. This means that x_i will receive the value that is most likely based on given y_i. Because y_i has a Gaussian likelihood, it is symmetric. Hence, assuming no markov dependence, we predict $x_i = 0$ if $y_i < 0.5$ and $x_i = 1$ if $y_i > 0.5$. The resulting prediction for $\mathbf{x}$ is plotted in figure 4b.



(a) Predicted values of $\mathbf{x}$ based on Markov assumption.



(b) Predicted values of $\mathbf{x}$ based on no dependence in Markov chain.

Figure 4: Comparison of predicted values for $\mathbf{x}$ using the Markov chain property and simply using the likelihood model at each individual location.

Comparing figure 4 to figure 1a, we see that the markov assumption provides a better prediction than independence assumption. The independence assumption has too frequent shifts in the values of $\mathbf{x}$, and evidently does not capture that the underlying $\mathbf{x}$ follows a markov chain. A comparison between the probabilities of $p(x_i = 1|\mathbf{y})$ for markov and independence assumption is also included in figure 5 for reference.

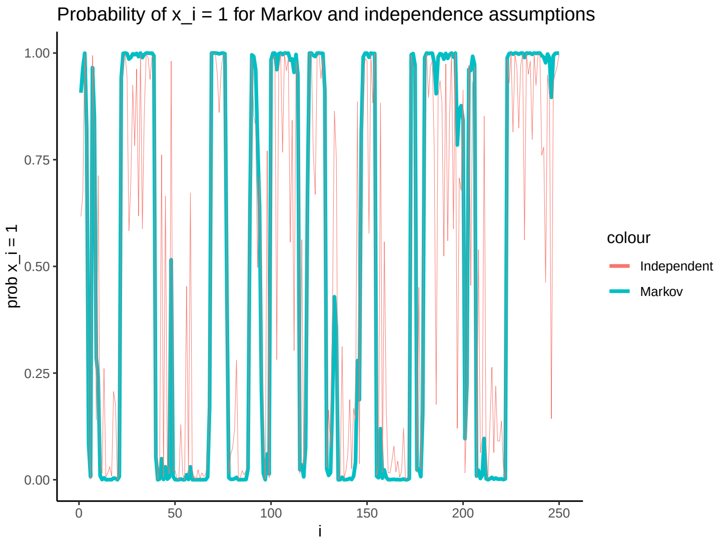


Figure 5: Comparison of probability for $x_i = 1$ for Markov chain model and no Markov dependence model.