Obligatory assignment 2 MVE550, autumn 2024

Petter Mostad

November 22, 2024

1. A biologist is investigating the frequency with which sea-living animals of a certain species develop a certain disease, and how this frequency depends on the concentration of a certain pollutant and the temperature. Based on experience from similar contexts, she uses a model where an animal exposed to the pollutant concentration x and the temperature y has a probability $p = f(x, y, \theta_1, \theta_2, \theta_3)$ of developing the disease, where

$$f(x, y, \theta_1, \theta_2, \theta_3) = \frac{\exp\left(e^{\theta_1} x + e^{\theta_2}(y - \theta_3)^2\right) - 1}{\exp\left(e^{\theta_1} x + e^{\theta_2}(y - \theta_3)^2\right) + 1}.$$

Here, $\theta = (\theta_1, \theta_2, \theta_3)$ are the parameters of the model. Each of them can take on any real value. A flat prior is assumed for all of them.

The data is given in the file "dataAssignment2.txt". It can be read into `R` with the command `data = read.table("dataAssignment2.txt", header=TRUE)` and consists of a matrix where each row i contains observed values (x_i, y_i, z_i) for an animal i: x_i is the pollutant concentration the animal was exposed to, y_i the temperature it was exposed to, while $z_i = 1$ indicates that the animal had the disease and $z_i = 0$ indicates it did not.

(a) Visualize the data, for example with
`plot(data$x, data$y, col=ifelse(data$z, "red", "blue"))`.

(b) Using the model above and the function f, write down the likelihood of the data (i.e., a formula for the probability of the data given the parameters of the model). Also, write down a function that is proportional to the posterior density for the parameters.

(c) Write an R function that takes as input values for two parameters $\theta = (\theta_1, \theta_2, \theta_3)$ and $\theta^* = (\theta_1^*, \theta_2^*, \theta_3^*)$ and computes a function that is equal to

$$\log\left(\frac{\pi(\theta^* \mid \text{data})}{\pi(\theta \mid \text{data})}\right),$$

i.e., the logarithm of the quotient of the posterior densities for θ^* and θ.

(d) Implement an MCMC algorithm that generates a Markov chain of length 10000 with limiting distribution equal to the posterior for θ.

Use a proposal distribution which adds to each parameter a normally distributed variable with expectation zero and standard deviation 0.4. Find a starting value for the chain by studying what values for θ might be reasonable for the given data. Produce trace plots (plots mapping simulated values for θ_i against its index i) for the parameters θ_1, θ_2, and θ_3.

(e) Compute numerically the predicted probability that an animal at pollutant concentration $x = 0.7$ and temperature $y = 19$ will have the disease. Also, compute the predicted probability that among 10 animals exposed to this temperature and this pollutant concentration, 9 will have the disease.

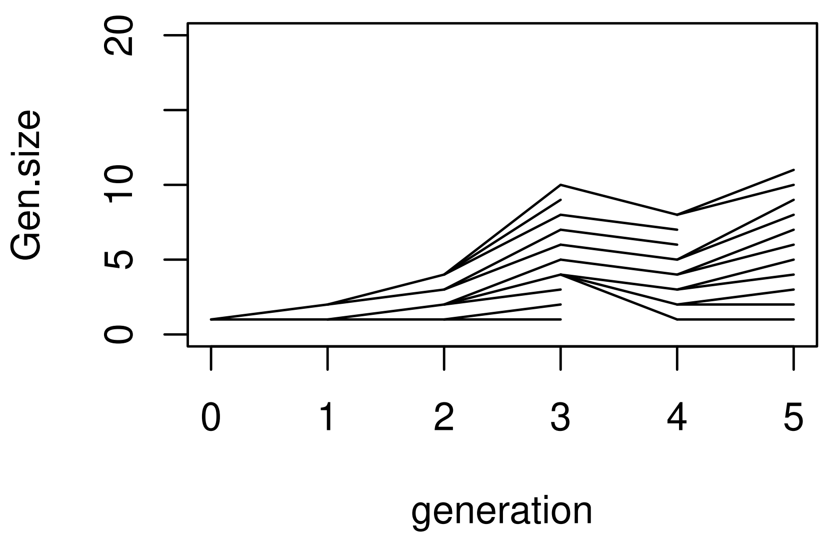

Figure 1: The 5 first steps of the Branching process for question 1

2. (a) Consider a branching process where the offspring distribution is a Poisson distribution with parameter $\lambda = 1.4$. Use R to compute[1] the probability that such a process will go extinct.

   (b) Consider the process whose first 5 steps are pictured in Figure 1.

[1]For example, you may consider the R function `uniroot`.

Assume its offspring distribution is Poisson(1.4). What is the probability that the continuation of this process will go extinct?

(c) Now, consider instead a branching process with a Poisson(λ) offspring distribution where λ is unknown. We assume a prior $\pi(\lambda) \propto_\lambda 1/\lambda$. Assume now that Figure 1 depicts a realization of this process. What is the resulting likelihood for λ? What is the posterior distribution for λ?

(d) We want to compute the probability of extinction of a branching process of the type of question (c), taking into account the uncertainty in λ. The extinction probability can then be written as an integral of a product of two functions of λ. Write down this integral and compute its value with R using numerical integration.

(e) Consider the process in in Figure 1, depicting a realization of a branching process with a Poisson(λ) offspring distribution with unknown λ, as in question (c). Compute the probability that the continuation of this process will go extinct.