Assignment 2, MVE550, autumn 2022

Lisa Samuelsson - *lisasam@student.chalmers.se*
Isak Palenius - *palenius@student.chalmers.se*
Oskar Giljegård - *oskargi@student.chalmers.se*

November 25, 2022

1. Consider a discrete-time Markov chain with possible states 1,2,3. The transition matrix P is unknown, but we have observed the first 10 steps of a realization of the chain: 1,2,3,2,3,1,2,1,3,2. Let P_1, P_2, and P_3, be the rows of P and we assume the priors $P_i \sim$ Dirichlet(1,1,1) for i = 1,2,3.

All code for this question is included in the submitted file `A2Q1.R`.

(a) Given the data, describe the posterior distributions for P_1, P_2, and P_3, and thus for P. Compute the expectation of the posterior for P.

The given prior means

$$\alpha = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Based on our transitions we know

$$c = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$$

The posterior is going to be $Dirichlet(\alpha_i + c_i)$ meaning that

$P_1|data \sim Dirichlet(1, 3, 2)$
$P_2|data \sim Dirichlet(2, 1, 3)$
$P_3|data \sim Dirichlet(2, 3, 1)$

(b) Write an R function simulate1 that simulates a continuation of the chain above into a chain of length 400, as follows: Sequentially, for each of the added steps, one should

- **simulate a new state for the chain, using the expected value of P based on all observed and simulated values so far**

- **update a count matrix with the new simulated value. Let the function plot the result of the simulation.**

See provided R code.

(c) Adapt simulate1 to simulate2 so that instead of plotting, it returns $E(P_{12}$ | actual and simulated data). Make a histogram of 1000 such values.

Since the posterior of P used in (b) will always be the expected value of P given all the data, we can just index P at position $1, 2$. Computing this value 1000 times gives the histogram in Figure 1.
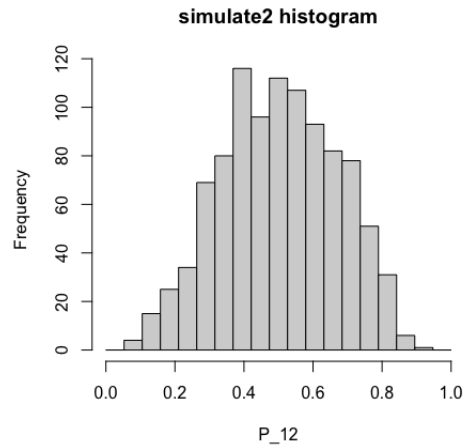
Figure 1: The histogram generated by the R function simulate2.

(d) Adapt simulate2 to simulate3 so that when you simulate a new state for the chain, you use the expected value of P based only on the actual 10 data values. Make a new histogram as in (c). Describe and explain any differences between the histograms in (c) and (d).

We remove the code that continuously updates P so that only the actual data is used. The posterior for P is then calculated after all values have been simulated to find P_{12}. This gives the histogram in Figure 2.
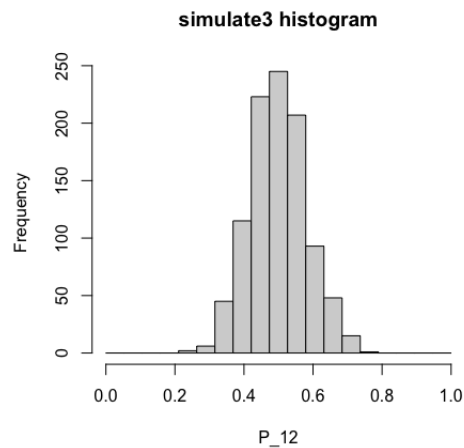

Figure 2: The histogram generated by the R function simulate3.

Here in Figure 2 we see that the histogram becomes much more narrow. We would expect the values to be much closer to what the observed data is, which is $\frac{3}{6} = \frac{1}{2}$ since our P_{12} has that value. Essentially, we are sampling values where going from 1 to 2 has a $\frac{1}{2}$ probability of being selected. We then check the frequency of going from 1 to 2 in all our samples. Due to the law of large numbers, we expect this to be close to $\frac{1}{2}$.

The histogram in Figure 1 is much wider. This is because previously generated values will reinforce the probability of getting those values. For example, if we go from 1 to 2 many times near the start that will dramatically increase the value P_{12} and decrease P_{11} and P_{13}. This makes us more likely to go from 1 to 2 in the future, creating a positive feedback loop. Therefore, the probability of the final P_{12} being further away from the original $\frac{1}{2}$ is greater than in Figure 2.

(e) Adapt simulate3 to simulate4 by starting with sampling a P from the posterior found in (a). Then you use this sampled value everytime you simulate new states in the chain. Make a histogram as in (c) and (d). Describe similarities and differences between the three histograms. Can you describe theoretically why these results are obtained? Can you explain what would happen with your results if the value 400 were increased or decreased?
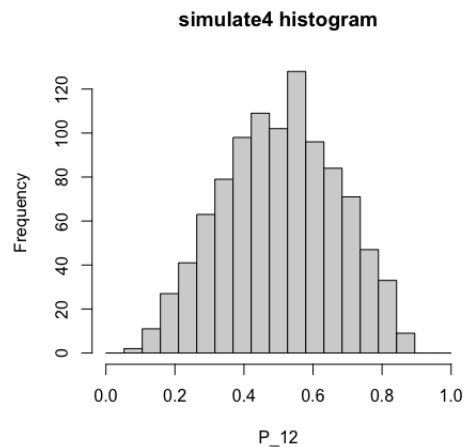

Figure 3: The histogram generated by the R function simulate4.

The histogram from Figure 3 is very similar to the one in Figure 1; they are both much wider than the one in Figure 2. This holds true even when increasing the number of steps simulated. Since we sample from Dirichlet distributions in Figure 3 it is possible to get a starting probability for the transitions from state 1 that are different from the expected transitions used in Figure 2. This allows the frequencies calculated to differ more just like in Figure 1.
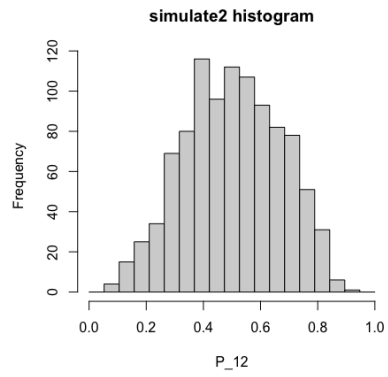
(a) Histogram generated by simulate2.
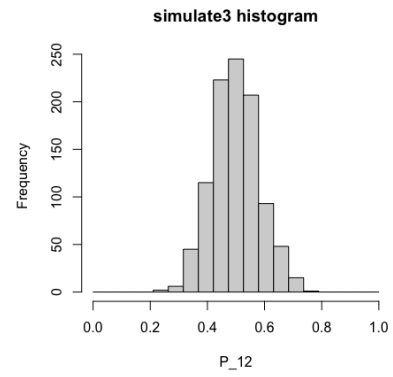
(b) Histogram generated by simulate3.
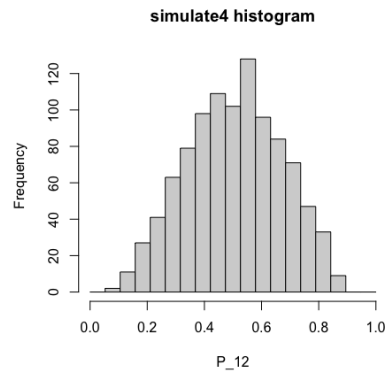
(c) Histogram generated by simulate4.

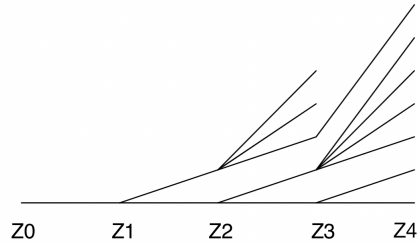Figure 4: The three histograms generated.

Figure 5: The branching process described in Question 2.

2. Assume a branching process has been observed for $Z_0, ..., Z_4$, and it looks like Figure 5. Assume the offspring distribution is Poisson with expectation λ. Assume that we use the improper prior $\pi(\lambda) \propto_\lambda 1/\lambda$ for λ. Compute the probability of extinction as follows:

All code for this question is included in the submitted file `A2Q2.R`.

(a) Using the information about the offspring distribution appearing in Figure 5, compute the posterior distribution for λ.

State	Number of incoming branches
$Z_0 = 1$	none
$Z_1 = 1$	$X_1 = 1$
$Z_2 = 2$	$X_1 = 2$
$Z_3 = 5$	$X_1 = 3, X_2 = 2$
$Z_4 = 7$	$X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 4, X_5 = 2$

All seen data: 0, 0, 1, 1, 2, 2, 2, 3, 4

9 values seen that sum to 15.

$$\pi(\lambda|data) \propto_\lambda \pi(data|\lambda)\pi(\lambda)$$
$$= \pi(0|\lambda) \overset{data}{\cdots} \pi(4|\lambda)\pi(\lambda)$$
$$\propto_\lambda e^{-9\lambda}\lambda^{15}\frac{1}{\lambda}$$
$$= e^{-9\lambda}\lambda^{14}$$
$$\propto_\lambda Gamma(\lambda; 15, 9)$$

(b) Consider the continuation of the branching process pictured in Figure 5 into generations Z_5, Z_6, etc. Implement in R a function that takes as input a value for λ and outputs the probability that this branching process will become extinct.

This was the function implemented in R:

```
1  get_prob_of_extinction = function(lambda) {
2    if (lambda <= 1){
3      return (1)
4    }
5    n <- 10
6    G <- Vectorize(function(s) {sum(s^(0:n)*dpois(0:n, lambda))})
7    f <- function(s) (G(s)-s)^2
8    return (optimize(f, lower=0, upper=0.999)[[1]])
9  }
```

According to Theorem 4.2 in Dobrow, the extinction probability is equal to 1 if $\mu \leq 1$, where μ is the expected value, which in this case represents the expected number of offspring generated after a branching has occured. Since the expected value of a Poisson distribution is λ, i.e. $\mu = \lambda$, we can in the code check if $\lambda \leq 1$ and return probability 1 if that is true.

If $\lambda > 1$, more calculation is needed. Theorem 4.2 in Dobrow also states that the probability of eventual extinction is the smallest positive root of the equation $s = G(s)$, where $G(s)$ is the probability generating function of the offspring distribution.

We used the R function `optimize`, and chose to optimize over $(G(s) - s)^2$ instead of over $G(s) - s$ since `optimize` tries to find the minimum value, which could be negative, but we are only interested in the intersection of $G(s)$ and s, i.e. where $G(s) - s = 0$. Using `optimize` in the interval $s \in [0, 0.99]$ ensures that we avoid the root at $s = 1$.

(c) Now compute the probability of extinction for the branching process taking the uncertainty in λ into account: Write down an integral representing this probability in terms of the results from (a) and (b). Then compute the integral using numerical integration.

Let $e(\lambda)$ be the extinction probability for a given λ. We can then find the probability of extinction taking the uncertainty into account by computing

$$E(e(\lambda)) = \int_0^\infty e(\lambda)\pi(\lambda)d\lambda$$

Using the posterior for λ computed in (a) and the function for extinction probability created in (b) we get that the probability of extinction is 0.4002142.

(d) Use simulation to check your result in (c): Simulate from the posterior found in (a), and combine with the code found in (b).

We simulated the probability of extinction using 100 generations, and ran the simulation 1000 times. In order to make the code more efficient we stop the simulation when the population is 0 since it will never recover from that. We

also decided to stop the simulation if the current population is greater than 100, since it is very likely that it will survive at that point. From this we obtained a probability of extinction = 0.383.

(e) What is the maximum likelihood estimate for λ? What is the probability of extinction for the branching process if you use this estimate for λ in our computation?

We obtain the maximum likelihood estimate for λ by maximizing the Gamma function with parameters $\alpha = 15$, $\beta = 9$. By doing this, we found that the maximum likelihood estimate for λ is $\lambda = 1.555553$. When using this λ in the R function written for Question (2.b) we found that the probability of extinction for the branching process is 0.3829624.