BDA HM4

Piero Birello

January 2025

8.1

```
1   "y","s"
2   "1","Enzo"
3   "0","Enzo"
4   "1","Enzo"
5   "1","Enzo"
6   "1","Enzo"
7   "1","Enzo"
8   "1","Enzo"
9   "1","Enzo"
10  "1","Enzo"
11  "0","Salvo"
12  "0","Salvo"
13  "0","Salvo"
14  "1","Salvo"
15  "0","Salvo"
16  "1","Salvo"
17  "0","Salvo"
18  "0","Salvo"
19  "1","Salvo"
20  "0","Salvo"
21  "0","Salvo"
22  "1","Salvo"
23  "0","Salvo"
24  "1","Ugo"
25  "0","Ugo"
26  "1","Ugo"
```

All diagnostic statistics show that the MCMC chains are consistent and representative of the posterior distributions. Namely, for all parameters $\theta_1, \theta_2, \theta_3$: trace plots for different chains are well overlapped; autocorrelation is almost null for all chains; the shrink factor is close to 1; posterior estimates show very little variation (see Fig. 1). Looking at the posterior estimates, θ_1 and θ_2 have HDI that are not centered in 0.5. In particular, θ_1's HDI does not contain 0.5 at all, so that we may defend the hypothesis that the coin is biased. Enzo actually obtained one zero only out of nine trials. On the other hand, the posterior for θ_2 has high density at values lower than 0.5. The posterior for θ_3 does not deviate much from the prior distribution, since Ugo only had three trials.

8.2

The function `smryMCMC` saves summary statistics for the posterior distributions of $\theta_1, \theta_2, \theta_3, \theta_1-\theta_2, \theta_1-\theta_3, \theta_2-\theta_3$. In particular, default statistics include: mean, median, mode, effective sample size, mass contained in the HDI
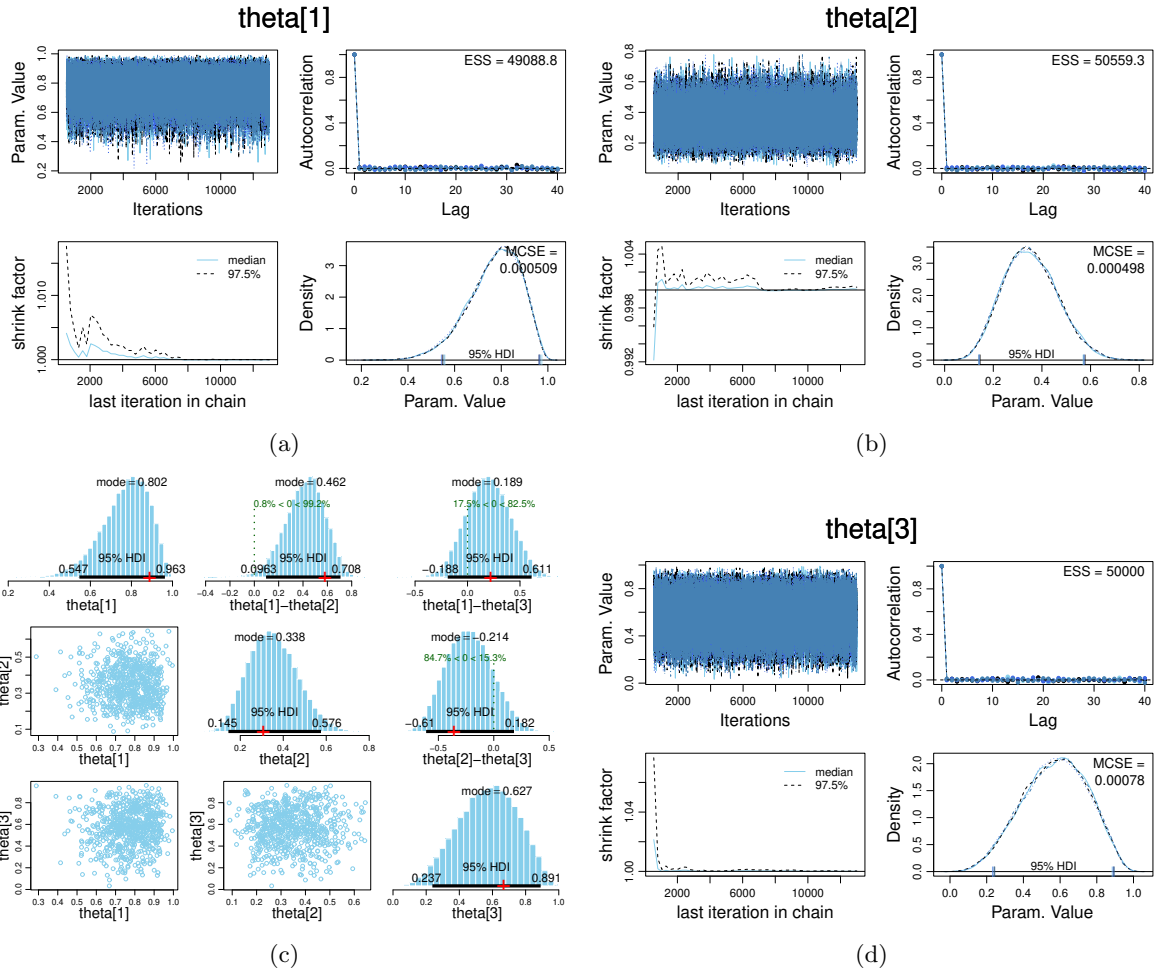
theta[1]

(a)

theta[2]

(b)

(c)

theta[3]

(d)

Figure 1: Exercise 8.1. Diagnostic for θ_1 (a), θ_2 (b), θ_3 (d). Posterior distributions for the three parameters and their pairwise differences (c).
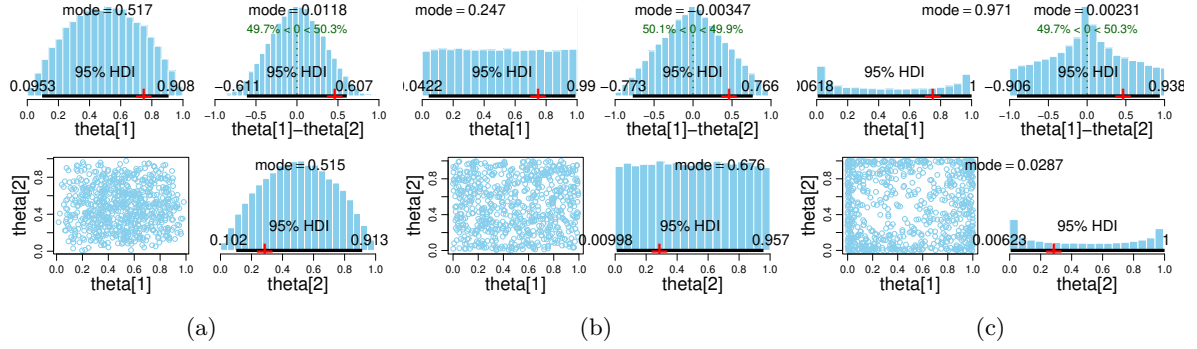
Figure 2: Exercise 8.4. Sampling of two prior identical theta distributions and their difference with parameters $(2, 2)$ (a), $(1, 1)$ (b), $(0.5, 0.5)$ (c).

to display (95%), lower and upper bound of the HDI. One can set to non null values the parameters `rope`, `ropeDiff` and `compValDiff` to save further summary statistics. `rope` applies to the posterior distributions for single parameters. It considers a rope -indeed- spanning a given range (e.g., 0.45 to 0.55, an interval around 0.5) and computes the mass below its lower bound, inside it and after its upper bound. The other two parameters apply to the posterior distributions of parameter differences. `ropeDiff` plays the same role as `rope`. `compValDiff` sets a value we're interested in knowing the mass below and after. In our case, putting `compValDiff` to 0 allows us to obtain the probability than one coin has higher or lower θ than the other one.

8.3

The script saves two interim data files. The first one is saved by the function `genMCMC`. It contains the actual MCMC sampled chain. Saving this file is particularly useful when the sampling process takes long time and is computationally costly. We run the sampling algorithm once and save its results, so that we'll be able to process the synthetic data later on. The second file is saved by `smryMCMC`. This file, as mentioned in the previous section, contains only the summary statistics of the chains saved by `genMCMC`. This is fast to read and gives us an overview of the relevant estimates we are interested in.

8.4

Sampling from the prior distributions can be obtained by passing the MCMC sampler empty data, i.e., commenting out the line setting the y vector in the script. We can exploit this to sample prior distributions with different parameters. In Fig. 2 we set parameters $(2, 2)$ (a), $(1, 1)$ (b), $(0.5, 0.5)$ (c). While the first one shows a bell shape, the second one is a flat distribution and the third one has high values at its extremes. Note that for (a) and (c) the sampled mode is not meaningful, since there exist more than one (or infinite) maxima. In all cases, the distribution for the difference is peaked at 0, since the two priors are identical. However, the shape of such prior on the difference changes with the distributions' parameters.