Bayesian inference 2017
Exercise session 4 (27.–30.11.2017)

1. Normal distribution with a known mean but an unknown variance. Although this may seem a bit strange model, it turns out to be an useful building block later when we consider posterior inference for multiparameter models.

Assume observations $Y_i, \ldots, Y_n$ from the normal distribution with the known mean $\theta_0 \in \mathbb{R}$ but the unknown variance σ^2:

$$\begin{aligned} Y_i &\sim N(\theta_0, \sigma^2) \quad \text{for all } i = 1, \ldots, n \\ \sigma^2 &\sim \chi^{-2}(\nu_0, \sigma_0^2) \end{aligned}$$

The conjugate prior used for the variance is a so-called (scaled) inverted chi-squared distribution with a density function

$$p(\sigma^2) = \frac{(\nu_0/2)^{-\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma_0^2)^{\nu_0/2} (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0 \sigma_0^2}{2\sigma^2}\right).$$

This is an inverse gamma distribution with a slightly more convenient parameterization.

Show that the posterior distribution for the variance is

$$\sigma^2 | \mathbf{Y} \sim \chi^{-2}(\nu_n, \sigma_n^2),$$

where

$$\begin{aligned} \nu_n &= \nu_0 + n, \\ \sigma_n^2 &= \frac{\nu_0 \sigma_0^2 + nv}{\nu_0 + n} \end{aligned}$$

and

$$v = \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta_0)^2.$$

Remember that the mean θ_0 is assumed to be known, so it is handled here as a constant, not as a parameter.

2. Continued from Exercise 2 of the previous week. You have marked down your waiting times for one week with the following results:

```
y <- c(1.36, 7.47, 7.31, 7.48, 10.33)
```

and decide to model them with the following model:

$$\begin{aligned} Y_i &\sim \mathrm{U}(0, \theta) \quad \text{for all } i = 1, \ldots, n \\ \theta &\sim \mathrm{Pareto}(b, K) \end{aligned}$$

with prior parameters $b = 1$ and $K = 1$.

Assume that next monday you sleep through an alarm (again!). You run to the tram stop and realize that you may end up being late from your Bayesian inference lecture. This would of course be a disaster! But luckily the lecturer isn't always so precise, so also he may end being few minutes late.

After some super-fast calculations your realize that if the tram arrives within the next 5 minutes, you will not miss the beginning of the lecture, but if it takes more than 11 minutes, you will surely miss it.

So luckily you still have some minutes left to calculate what is the probability of you making it to the lecture on time!

(a) You are in a hurry, so maybe it is better to start with a little bit of an approximation. A simplest way is to get a predictive distribution for a new observation $\tilde{Y}$ (your waiting time this morning; this assumed to be an observation from the same uniform distribution $U(0, \theta)$, and independent of the previous observations given the parameter) plug a posterior mode into the sampling distribution of it. What is the posterior mode (the maximum value of the posterior distribution):

$$\hat{\theta}_{\text{MAP}}(\mathbf{y}) := \underset{\theta}{\text{argmax}}\; p(\theta|\mathbf{y})?$$

Using $p(\tilde{y}|\hat{\theta}_{\text{MAP}}(\mathbf{y}))$ as your predictive distribution, compute the probabilities that your waiting time is smaller than 5 minutes and that it is greater than 11 minutes.

(b) A slightly more advanced plug-in estimate for the predictive distribution is obtained by using the posterior mean (the expected value of the posterior distribution):

$$\hat{\theta}_{\text{Bayes}}(\mathbf{y}) := E[\theta \mid \mathbf{Y} = \mathbf{y}]$$

instead of the posterior mode. Compute the aforementioned probabilities using $p(\tilde{y}|\hat{\theta}_{\text{Bayes}}(\mathbf{y}))$ as your predictive distribution.

(c) A fully Bayesian way to define the predictive distribution is to average over all the possible parameter values:

$$p(\tilde{y}|\mathbf{y}) = \int p(\tilde{y}|\theta)p(\theta|\mathbf{y})\,\mathrm{d}\theta.$$

Derive (Now you can actually solve the integral by really integrating. Jau!) this posterior predictive distribution $p(\tilde{y}|\mathbf{y})$, and compute the aforementioned probabilities using it.

(d) Draw all your predictive distributions into the same plot, and compare the curves and the computed probabilities. Well, after all this you will probably end up being late from the lecture anyway, but in case you did all these computations super-fast, which one these do you think would be a best way to estimate your probabilities, and why so?

3. A simple demonstration of the strong law of law large numbers.

(a) Simulate[1] a sample of 10^5 points from the posterior distribution[2] Pareto$(c, K + n)$ of the Exercise 2 of the previous week. Plot (an example plot is found on Figure 1) a trace of the cumulative mean[3] for the first $10^2, 10^3, 10^4$ and 10^5 points[4]. What is the mean

$$\frac{1}{S}\sum_{s=1}^{S}\theta_s$$

of your whole sample $\theta_1, \ldots, \theta_S$?

Repeat the simulation 10 times, and compare the sample means to the expected value of the posterior distribution

$$E[\theta \mid \mathbf{Y} = \mathbf{y}],$$

which they should approximate.

[1]You can use `VGAM::rpareto` in this exercise set, no need to implement any sampler.

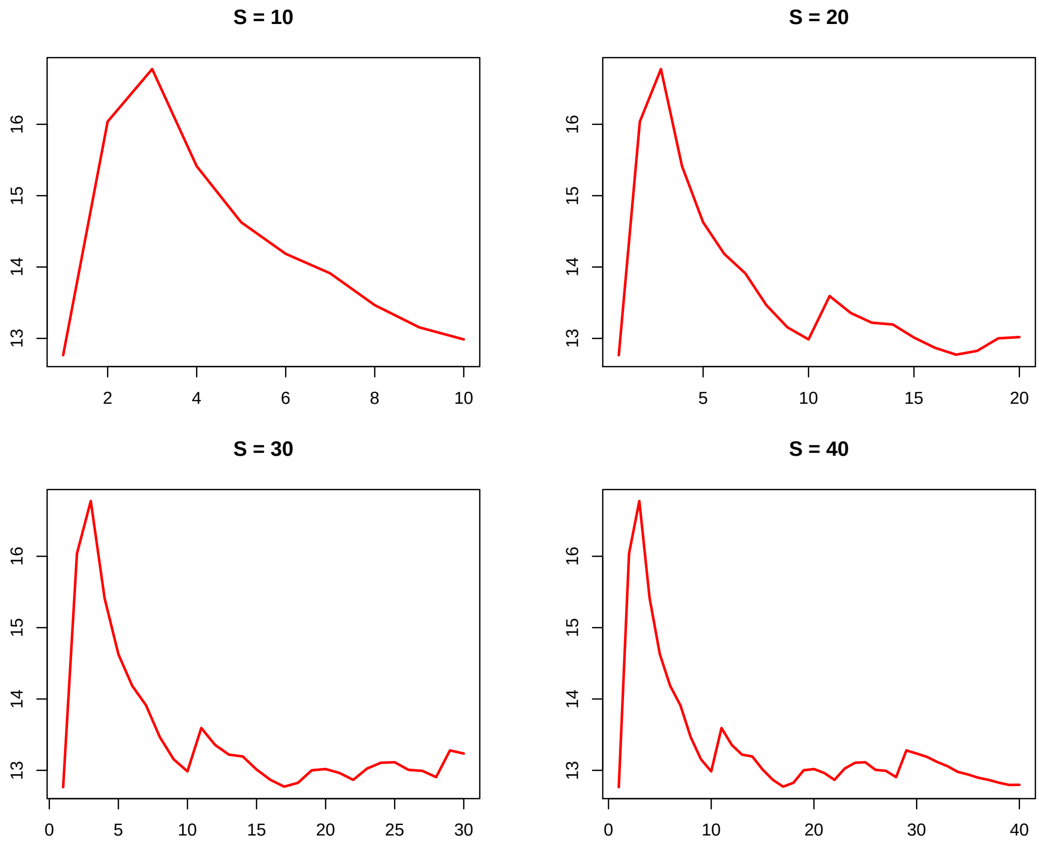
[2]With a smaller data set with a sample size $n = 5$.

[3]So that first you plot the first value of your sample, then the mean of the two first values, then the mean of three first values and so on.

[4]By setting the graphical parameters with the command `par(mfrow = c(2,2), mar = c(2,2,4,4))` you can conveniently draw 4 plots into a 2x2 grid.

(b) Simulate a sample of 10^5 points from the priori distribution Pareto$(1, 1)$, and draw the aforementioned plots from your sample. Repeat this 10 times, and compare the sample means. How would you explain this!?

Kuva 1: An example picture for the the exercise 3 with a smaller sample size.



4. A warm-up exercise for using Stan.

(a) Write the model of Exercise 2 of the previous week using Stan. Fit the model, and draw a histogram of your sample from the posterior distribution. Draw the true posterior density, which you solved last week, to verify that the simulated and the analytically solved posterior match each other. If you are not happy with the accuracy of your sample, try increasing the sample size of the simulation, and see what happens.

Approximate also the probabilities $P(\theta > 11 \,|\, \mathbf{Y} = \mathbf{y})$ and $P(11 < \theta < 13 \,|\, \mathbf{Y} = \mathbf{y})$ based on your simulated sample, and compare them to the true probabilities you solved last week.

(b) Sample also from the posterior predictive distribution for the new observation $\tilde{Y}$ using your sample from the posterior distribution. Plot a histogram from your sample and draw the posterior predictive density $p(\tilde{y}|\mathbf{y})$, which you solved analytically on Exercise 2c) of this week, on top of it.

Approximate also the probabilities $P(\tilde{Y} < 5 \,|\, \mathbf{Y} = \mathbf{y})$ and $P(\tilde{Y} > 11 \,|\, \mathbf{Y} = \mathbf{y})$ based on your simulated sample, and compare them to the analytical results you computed in Exercise 2c) of this week.