Bayes-päättely, 2. harjoitukset (8.–10.11.2017)

1. Continued from the 3:rd exercise of the previous week. Assume that the sampling distribution of the random variable $\mathbf{Y} = (Y_1, \dots, Y_d)$ is a multinomial distribution:

$$p(\mathbf{y}|\boldsymbol{\theta}) = \binom{n}{y_1 \dots y_d} \prod_{i=1}^{d} \theta_i^{y_i},$$

and the prior distribution for the parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ is a Dirichlet distibution with the parameter $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$:

$$p(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_d)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_d)} \prod_{i=1}^{d} \theta_i^{\alpha_i - 1}.$$

(a) Prove that the marginal likelihood $p(\mathbf{y})$ is a Dirichlet-multinomial distribution with parameters n and $\boldsymbol{\alpha}$:

$$p(\mathbf{y}) = \frac{n!\,\Gamma(\alpha_0)}{\Gamma(\alpha_0 + n)} \prod_{i=1}^{d} \frac{\Gamma(\alpha_i + y_i)}{y_i!\,\Gamma(\alpha_i)},$$

where

$$\alpha_0 = \sum_{i=1}^{d} \alpha_i.$$

(b) Now that you have solved the normalizing constant $p(\mathbf{y})$, derive the posterior distribution from the exercise 3 of last week using the full Bayes formula:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}.$$

(c) Assume a new observation $\tilde{\mathbf{Y}}$ from the multinomial distribution with the same parameter θ, but a different sample size m:

$$\tilde{\mathbf{Y}} \sim \text{Multinom}(m, \boldsymbol{\theta}).$$

Assume also that the observations are independent given the parameter:

$$\tilde{\mathbf{Y}} \perp\!\!\!\perp \mathbf{Y} \,|\, \boldsymbol{\theta}.$$

Derive a posterior predictive distribution $p(\tilde{\mathbf{y}}|\mathbf{y})$ for this observation.

2. Continued from the thumbtack tossing example of the lectures. So our model is:

$$\begin{aligned} Y &\sim \text{Binom}(n, \theta) \\ \theta &\sim \text{Beta}(\alpha, \beta), \end{aligned}$$

where Y is a number of times the thumbtack lands point up.

Before we have any observations, it may be sensible to set a symmetric prior distribution which does not give extra probability to neither failures nor successes.

We will try three different prior distributions, which have the same expected value $E\theta = \frac{1}{2}$:

- $\text{Beta}(0.01, 0.01)$,
- $\text{Beta}(1, 1)$,

- Beta$(5, 5)$.

(a) Plot (with `R` or the programming language of your choice) the density functions of these prior distributions onto the same plot with different colors.

(b) Now it is a time to generate some data. Find a thumbtack (or some other object that can land on only two possible ways and is unlikely to break or break anything when it lands).

Throw it 3 times, and count how many times it lands point up. Now you have the observation y from the binomial distribution $\text{Bin}(3, \theta)$.

Plot your posterior distributions for the success probability θ with all the different priors into the new plot.

Compute (you can use `R`, no need to compute this with pen and paper) also the probabilities

$$P(\theta < 0.5 \,|\, Y = y)$$

using these different posterior distributions.

How do your posterior inferences differ with the different priors? How would you explain the differences?

(c) Throw a thumbtack 22 times more, and count the number of successes. Now you have an observation from the binomial distribution $\text{Bin}(25, \theta)$. Plot your new posteriors and compute the probabilities $P(\theta < 0.5 \,|\, Y = y)$ with the new posteriors. How do your results differ from the results with only 3 throws, and how would you explain the differences?

3. (Continued from the previous exercise.) Assume we that we throw the same thumbtack 10 more times. We can model this with a random variable $\tilde{Y} \sim \text{Bin}(10, \theta)$, which is independent of the first 25 throws given the parameter:

$$\tilde{Y} \perp\!\!\!\perp Y \,|\, \theta.$$

(a) Plot[1] marginal likelihoods $p(\tilde{y})$ for $\tilde{Y}$ with the priors of the previous exercise into the same plot[2]. How different priors weight different values of $\tilde{Y}$?

(b) Plot the posterior predictive distributions $p(\tilde{y}|y)$ with the priors of the previous exercises, and with the data of 3 throws from the last exercise into a new plot. Draw also a probability mass function of the binomial distribution $\text{Bin}(10, \hat{\theta})$. Which of the posteriors is closest to this binomial distribution? Why do you think it is so?

(c) Repeat part (b) with a data set of all 25 throws from the last exercise (Now the maximum likelihood estimate is of course $\hat{\theta} = y/25$). How increasing the sample size affects the behaviour of the posterior distributions.

(d) Actually throw the thumbtack 10 more times, and count the number of times it lands point up. Which probability did the posterior predictive distribution computed on the basis of your first 25 throws give to this number of successes?

[1] Beta-binomial distribution is not found on base R, but you can write it yourself or use `dbetabinom` from the package `VGAM` (just check that it has the same parametrization that you expect).

[2] When plotting probability mass functions (which are discrete), the plotting option `type = 'b'`, which draws points and connects them by lines is convenient (bar plot or line plot (plot of vertical lines) look normally nicer, but if you want several distributions into the same picture, they overlay each other).

4.

(a) Compute 95% (equal-tailed) credible intervals for the parameter θ using the posterior distributions for 3 and 25 observations from the exercise 2. Use prior parameters $\alpha = 1$ and $\beta = 1$. Which one of the credible intervals is wider, and why?

(b) Plot these posterior densities (into their own plots), and color the area under the curve of the densities on the credible intervals (or otherwise mark the credible intervals into the plot).