Bayesian inference 2017
Exercise session 1 (6.–9.11.2017)

1. Assume that we burn out n light bulbs, mark down how long they last (in hours), and denote their lifespans as $y_1, \ldots, y_n$. Assume that the lifespans of the light bulbs are independent, and follow the exponential distribution with parameter $\theta > 0$:

$$Y_1, \ldots, Y_n \,|\, \theta \sim Exp(\theta) \perp\!\!\!\perp .$$

This means that

$$f_{Y_i|\Theta}(y_i|\theta) = \theta e^{-\theta y_i}$$

for all $i \in 1, \ldots, n$.

(a) Derive the maximum likelihood estimate for the parameter θ.

(b) Assume that the prior distribution is gamma distribution $\mathrm{Gam}(\alpha, \beta)$:

$$f_\Theta(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}.$$

Derive the posterior distribution $f_{\Theta|\mathbf{Y}}$ for parameter θ with this prior distribution.

(c) Let's choose the hyperparameters $\alpha = \beta = 0.01$. For three light bulbs we observe lifespans y_1, y_2, y_3 with a mean $\bar{y} = \frac{1}{3}\sum_{i=1}^{3} y_i = 1000$. Plot a posterior density[1] with `R` (or programming language of your choice). Further, we observe two other data sets with the same mean, but in the second data set the sample size is $n = 5$, and in the third data set the sample size is 25. Plot posterior density functions of these data sets into the same plot as the first one with different colors. Plot also the density of the prior distribution (`R` function `dgamma` gives density function of the gamma distribution) into the same plot. Based on this plot, what happens to your uncertainty about the true parameter value when the sample size grows?

2. Continued from the previous exercise.

(a) Derive the distribution for the lifespan $\tilde{Y}$ of the next light bulb given the first n observations, when we assume that the new observation is from the same distribution as the old ones: $\tilde{Y}, Y_1, \ldots, Y_n \,|\, \theta \sim Exp(\theta) \perp\!\!\!\perp$. In other words, derive the posterior predictive distribution $f_{\tilde{Y}|\mathbf{Y}}(\tilde{y}|\mathbf{y})$. Prove that its functional form is (as function of $\tilde{y}$):

$$f(x) = \frac{a\lambda^a}{(x+\lambda)^{a+1}},$$

where

$$a := \alpha + n,$$
$$\lambda := \beta + \sum_{i=1}^{n} y_i.$$

[1] Most probably you want limit your x-axis to an very short interval, such as (0, 0.005).

(b) Plot posterior predictive densities (there is not a function for the density function of this distribution in R, but you can write your own!) with three different data sets from the last exercise (for example on the interval (0, 5000)).

A simple (non-Bayesian) way to predict new values $\tilde{Y}$ is to plug a point estimate as a parameter value of the original density function of $\tilde{Y}$: plot also into the same plot the density function of exponential distribution $\text{Exp}(\hat{\theta})$, where $\hat{\theta}$ is a maximum likelihood estimate for the parameter θ computed in the last exercise.

Which of these distributions has lightest tail? Which has heaviest tail? How would you explain this?

3. Generalization of the binomial distribution into the case with more than two different outcomes of the experiment. Assume that random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ (for which holds that $\sum_{i=1}^{d} Y_i = n$) follows a multinomial distribution with parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $\sum_{i=1}^{d} \theta_i = 1$, written concisely as $\mathbf{Y} \,|\, \boldsymbol{\theta} \sim \text{Multin}(n, \boldsymbol{\theta})$. Probability mass function of $\boldsymbol{Y}$ is

$$f_{\mathbf{Y}|\boldsymbol{\Theta}}(\mathbf{y}|\boldsymbol{\theta}) = \binom{n}{y_1, \ldots, y_d} \prod_{i=1}^{d} \theta_i^{y_i}.$$

Derive the posterior distribution $f_{\boldsymbol{\Theta}|\mathbf{Y}}$ for the parameter $\boldsymbol{\theta}$, when the prior distribution for parameter vector $\boldsymbol{\theta}$ is Dirichlet distribution[2] $\text{Dir}(\boldsymbol{\alpha})$, which means that

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_d)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_d)} \prod_{i=1}^{d} \theta_i^{\alpha_i - 1},$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$.

4. Continued from the previous exercise.

(a) Consider a Dirichlet distribution with 3 possible outcomes ($d = 3$). Generate[3] random samples of 5000 points from the prior distribution $\text{Dir}(\boldsymbol{\alpha})$ with parameters $\boldsymbol{\alpha} = (1/3, 1/3, 1/3)$, $\boldsymbol{\alpha} = (1, 1, 1)$ ja $\boldsymbol{\alpha} = (10, 10, 10)$, and draw 3-dimensional scatterplots of these samples[4]. Where the probability mass of the prior distribution is concentrated with different values of parameter $\boldsymbol{\alpha}$?

(b) Generate a random sample of 5000 points from the posterior distribution $f_{\Theta|\mathbf{Y}}(\boldsymbol{\theta}|\mathbf{y})$, when prior distribution is set to $\text{Dir}(1, 1, 1)$ and the observed data is $\mathbf{y} = (30, 15, 2)$. Plot this sample as a 3-dimensional scatterplot. How the probability mass is concentrated compared to the prior distribution simulated in (a)?

[2] Dirichlet distribution is a generalization of the beta distribution into more than two dimensions: two-dimensional Dirichlet distribution $\text{Dir}(\alpha_1, \alpha_2)$ is a same thing as the beta distribution $\text{Beta}(\alpha_1, \alpha_2)$.

[3] R package `gtools` contains a function `rdirichlet`.

[4] R contains a function `scatterplot3d`, but you get a fancier 3-d scatterplot (which you can also rotate) with `scatter3d` function of the package `car` (requires also package `rgl` to work).