11th November 2016

Assignment 1

(Deadline for report: 18.11.2016)

Saving figures in R

For saving plots in R:

```
x<- c(1:10)
pdf("filename.pdf") #start editing the file filename.pdf
plot(x, x^2)        # plot something
dev.off()           # clode the file filename.pdf
```

The file will be saved to your current working directory, check your which is your current working directory using `getwd()`. Other then `pdf` format one can also use `jpeg` or `png`, but don't forget `dev.off()` once your done.

Bayesian statistics, prior distribution part 1

1 Assume that you want to investigate the proportion (θ) of a common mutation in a population. In a random sample of 50 individuals seven of them was mutated.

(a) Find the likelihood function $P(x_1, \dots, x_n|\theta)$.

(b) Assume a uniform prior for θ, and compute the posterior of θ.

(c) Derive $E(\theta)$ and $Var(\theta)$ of the posterior distribution $f(\theta|x_1, \dots, x_n)$. Give the 95% posterior interval of θ.

(d) If we instead can assume that the mutation is rare, use a suitable conjugate prior for $\theta \sim Beta(\alpha, \beta)$ and compute the posterior of θ. Give the 95% posterior interval of θ. Hint, use Figure 6.1 in the text book for choosing the prior.

1. Let $(X_1, X_2, ..., X_n)$ be at independent random sample from $N(\mu, \sigma^2)$. Assume σ^2 is known, and put a Normal prior on μ, i.e. $\mu \sim N(\mu_0, \tau_0^2)$. Recall that the posterior distribution $f(\mu|x_1, ..., x_2) \propto f(x_1, \dots, x_2|\mu, \sigma^2) \cdot f(\mu)$.

(a) Find the likelihood function $f(x_1, \dots, x_n|\mu, \sigma^2)$.

(b) The posterior distribution of θ is a normal distribution $N(\mu_n, \tau_n^2)$, where $\mu_n = w\bar{x} + (1-w)\mu_0$, $\frac{1}{\tau_0^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2}$ and $w = \frac{n/\sigma^2}{n/\sigma^2 + 1/\tau_0^2}$. Use the posterior distribution to calculate a 95% posterior interval of μ after observing a random sample of 15 observations where $\bar{x} = 13.2$, $\sigma^2 = 0.4$, $\mu_0 = 10$ and $\tau_0^2 = 8$.

Bayesian statistics, prior distribution part 2

1. In this exercise, we study the effect of the maternal condition known as placenta previa. In a German study of 980 births where the mother had placenta previa, 437 of the babies were female. This imposes a likelihood of the data

$$n_f \sim Bin(n, \theta),$$

here n_f is number of girls born, n is total number of babies and θ the probability of a baby being a girl.

- Assuming a uniform prior plot the posterior distribution of θ.
- Compute the posterior probability that θ is less then the population average 0.485.
- Find a Beta prior such that the posterior probability that $\theta > 0.5$ is larger then 0.5. Plot the prior.

2. A classical data (first studied in 1898) is number of deaths in the Prussian army due to kicks from horses:

1875	1876	1877	1878	1879	1880	1881	1882	1883	1884
3	5	7	9	10	18	6	14	11	9
1885	1886	1887	1888	1889	1890	1891	1892	1893	1894
5	11	15	6	11	17	12	15	8	4

A common model for this type of count data is to assume that the number of deaths, Y, are Poisson distributed:

$$P(Y = k; \theta) = \frac{\theta^k e^{-\theta}}{k!}.$$

(a) Assuming a uniform (improper) prior on θ, plot the posterior distribution of θ.

A conjugate prior for θ is $\Gamma(\alpha, \beta)$, the density of a gamma distribution is

$$f(\theta; \alpha, \beta) = \Gamma(\theta; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta},$$

where Γ is the gamma function. In R the gamma density is given by `dgamma`, use `help(dgamma)` to examine the input arguments.

Suppose that your data consists of n observations $\{y_1, y_2, \ldots, y_n\}$. If assume a $\Gamma(\alpha_0, \beta_0)$ prior on θ then the posterior distribution is

$$f(\theta | \{y_1, y_2, \ldots, y_n\}, \alpha_0, \beta_0) = \Gamma(\theta; \alpha = \alpha_0 + \sum_{i=1}^{n} Y_i, \beta = \beta_0 + n).$$

(a) Plot the posterior distribution of θ, i.e. the expected number of yearly deaths in the Prussian army, given a $\Gamma(1, 0.1)$ prior. Give the 95% posterior interval of θ. Hint, use `qgamma`.

(b) For a fixed α_0 does a higher value of β_0 indicate more or less information prior information about θ?

3. In this exercise, we will study the content of hemoglobin (Hb) in blood. For an adult person we can assume that the expected value of the hemoglobin content (mmol/L) has a normal distribution. Due to differences between the genders we use different distributions for the hemoglobin content for the two genders. For an adult male the expected value of the Hb content is a r.v.

$$\mu_{male} \sim N(\mu_0 = 9.65, \sigma_0 = 2.41).$$

For an adult female the expected value of the Hb content is a r.v.

$$\mu_{female} \sim N(\mu_0 = 8.65, \sigma_0 = 2.16).$$

A random sample of Hb content from a person, of unknown gender, gave the following result:

$$x = (\,7.66,\ 10.14,\ 8.59,\ 9.47,\ 9.49).$$

We assume that the observations are distributed as follows

$$x_1, \ldots, x_n | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2),$$

where $\sigma = 1.5$. Recall that the normal distribution $N(\mu, \sigma)$ has density function

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(\,x - \mu)^2\right],$$

for $-\infty < x < \infty$ and that $\mu > 0$, and $\sigma^2 > 0$.

(a) Find the likelihood function $f(x_1, \ldots, x_n|\mu, \sigma^2)$.

(b) Find the posterior distribution $f(\mu|x_1, \ldots, x_n)$ under both the male and female prior μ.

(c) Find $E(\mu)$ and $Var(\mu)$ of the posterior distribution $f(\mu|x_1, \ldots, x_n)$, for both male and female.

(d) Give the 95% posterior interval of μ_{male} and μ_{female}. What do you think, is the person a male or a female?