16th November 2016

Monte Carlo and Stan

Hand in a **short report** with the answer to the three questions.

Computer exercises

1. Often for a real application a measured distribution is a mixture of several distribution. In this exercise you will simulate from a mixture distribution and compute some basic statistics. A normal mixture model with two component can be written as

$$Y = wX_1 + (1-w)X_2,$$

where $w \sim Bin(1, \theta)$, $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$.

   (a) Assume $\theta = 0.3, \mu_1 = 0.1, \sigma_1 = 1, \mu_2 = 0.5$ and $\sigma_2 = 3$, Plot a histogram of 10000 simulation of Y.

   (b) Compute the expectation and variance of Y using Monte Carlo method.

2. In this exercise, we look at the same data as in lab 1. Recall that

$$\mu_{male} \sim N(\mu_0 = 9.65, \sigma_0 = 2.41),$$
$$\mu_{female} \sim N(\mu_0 = 8.65, \sigma_0 = 2.16).$$

We had a sample:

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

where

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

and $\sigma = 1.5$.

All the exercises below should be done using Rstan.
HINT: Look at the computer code from lecture 6, and do separate codes for μ_{male} and μ_{female}.

   (a) Plot histogram of the posterior samples for μ under both the male and female prior μ.

   (b) Estimate the posterior mean and variance of μ for both male prior and female prior.

   (c) Give the estimated 95% posterior interval of both μ_{male} and μ_{female}.

   (d) Now Assume that also σ is unknown and put a $\Gamma(3, 2)$ prior on it instead. Give the posterior mean and 95% confidence interval of σ. What is now the posterior 95% posterior interval for μ_{male} and μ_{female}?

3. In Lab 2, we examined the number of deathsin 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

Now we want to examine if time had an effect on the expected number of deaths. For this we assume the following model:

$$P(Y = k; \theta_i) = \frac{\theta_i^k e^{-\theta_i}}{k!},$$
$$\theta_i = \exp\left(\beta_0 + (year_i - 1875)\beta_1\right),$$
$$\beta_0 \sim N(0, \sigma_0^2),$$
$$\beta_1 \sim N(0, \sigma_0^2)$$

where $\sigma_0 = 10^6$ (a non-informative prior).

(a) Write the **stan** code that generates samples from the posterior distribution. **HINT:** Combine (not illiterately) the codes about regression and the code about Poisson distribution, from the lecture. To get θ from β_0 and β_1 use the following stan code:

```
transformed parameters{
vector[n] theta;
theta = exp(time * beta1 + beta0);
}
```

(b) Plot the histogram of β_0 and β_1. What is the posterior mean of β_0 and β_1.

(c) Is there a 95% significant effect of time? (Does the 95% confidence interval of β_1 contains the zero-value.)