6 Week 6

6.1 Monte Carlo error

R-template `ex_monte_carlo_error_template.Rmd`.

The purpose of this exercise is to examine the properties of Monte Carlo approximation. We will continue from the model answers of the Markov chain exercise. Your tasks are the following:

1. Sample 100 independent realizations of length 2000 chains from the Markov chain defined in exercise 3.1 (that is; $\theta^{(1)}, \dots, \theta^{(2000)}$) using each of the combinations of ϕ and σ^2 in the rows of the below table

Var[$\theta^{(i)}$]	ϕ	σ^2	Corr[$\theta^{(i)}, \theta^{(i+1)}$]
1	0	1	0
1	0.5	0.75	0.5
1	0.89	0.2	0.89
1	0.1	0.99	0.1

2. With each of the chains approximate $E[\theta^{(i)}]$, Pr($\theta^{(i)} > 0.5$) and Pr($\theta^{(i)} > 2$) using Monte Carlo with the $n = 10$, $n = 100$ and $n = 1000$ last samples. Hence, you will construct 100 independent Monte Carlo approximations for the mean and two probabilities of θ corresponding to Markov chain sample sizes 10, 100 and 1000.

3. Examine the calculated Monte Carlo approximations, compare them to the exact answers for $\theta \sim N(0, 1)$ (which is the distribution of $\theta^{(i)}$ in the limit of $i \to \infty$) and answer to the following questions:
   - How does the Monte Carlo estimate of $E[\theta^{(i)}]$ behave with respect to the number of samples and with respect to the autocorrelation of the Markov chain?
   - How does the Monte Carlo estimate of Pr($\theta^{(i)} > 0.5$) behave with respect to the number of samples and with respect to the autocorrelation of the Markov chain?
   - How does the Monte Carlo estimate of Pr($\theta^{(i)} > 2$) behave with respect to the number of samples and with respect to the autocorrelation of the Markov chain?
   - What kind of general conclusions can you make based on these results?

Grading The results do not need to be presented exactly similarly as in the model results as long as they are presented so that they support the claims made. This gives 2 points. After that 2 points from each question if correct qualitative results have been described.

6.2 Bio-assay experiment

R-template `ex_bio-assay.R`
Data file `ex_bio-assay.dat`

An experiment on a drug's toxicity has been performed on a group of animals according to Table 1. We assume that the number of deaths is Binomially distributed given a lethality parameter θ_i so that

$$y_i|\theta_i \sim \text{Bin}(n_i, \theta_i)$$

where θ_i depends on the dose level x_i through logistic regression model

$$\text{logit}(\theta_i) = \alpha + \beta x_i.$$

Here $\text{logit}(\theta_i) = \log(\theta_i/(1-\theta_i))$. We will assume a vague (practically uniform) prior for the parameters

$$\alpha \sim N(0, 10^6)$$
$$\beta \sim N(0, 10^6).$$

The posterior is then

$$p(\alpha, \beta|y, n, x) \propto p(\alpha)p(\beta) \prod_{i=1}^{n} p(y_i|\alpha, \beta, n_i, x_i)$$

Your tasks are to

1. Implement the model in Stan and sample from the posterior distribution of the parameters α and β. Check for convergence and report the convergence test results. Visualize the posterior of α and β

2. visualize the posterior of $\theta(x)$ as a function of the dose levels $x \in (-0.8, 0.8)$ (see help from linear regression exercise)

3. calculate the posterior covariance between α and β. How does this differ from the prior covariance between α and β, why?

4. LD50 is an index which tells the dose level with which half of the test animals are expected to die

$$\text{LD50} : \text{E}\left[\frac{y_i}{n_i}\right] = \frac{1}{2}$$

   - write the LD50 dose level as a function of α and β.
   - visualize the posterior distribution of LD50 and calculate its posterior mean and variance and *upper* 95% credible interval.

5. visualize the posterior distribution of θ with dose level (log g/ml) 0.

Table 1: Bio-assay data.

Dose, x_i (log g/ml)	Number of animals, n_i	Number of deaths, y_i
-0.86	5	0
-0.30	5	1
-0.05	5	3
0.73	5	5

6. sample from the posterior predictive distribution of an outcome for a new experiment with dose level $x = 0$ and 5 test animals. Calculate the posterior probability that more than 3 animals out of 5 would die with this dose level $p(\tilde{y} > 3|\tilde{n} = 5, \tilde{x} = 0, x, y)$.

Grading: This exercise gives in total 20 points so that steps 1 and 2 give 4 points each and steps 3-6 give 3 points each.