Laboratory Session 06 : May 20, 2025

Exercises due : June 8, 2024

# Exercise 1 : Quadratic Bayesian regression using `Stan`

- The following $n = 9$ data sample

```
##    X      Y       sigma_i
##    2   105.5203   15.40459
##    3   227.4457   36.05480
##    4   405.6937   51.84040
##    5   661.2858   94.66744
##    6   886.1422  123.79828
##    7  1189.6514  216.99935
##    8  1631.8262  173.83507
##    9  1951.8381  318.46511
##   10  2528.5246  427.27788
```

  is believed to follow

$$Y = \beta_0 + \beta_1 \cdot X + \beta_2 \cdot X^2$$

- Let the prior distributions for $\beta$ parameters be,
  - $\beta_0$ normal distribution centered at 5 with a standard deviation of 1.
  - $\beta_1$ uniform distribution between -1 and 1.
  - $\beta_2$ normal distribution centered at 30 with a standard deviation of 15.

  Build a `Stan` model and run a MCMC to obtain the posterior distribution of the $\beta$ parameters using also the $\sigma_i$ uncertainty for each measurement.

- Compute the 95% credibility interval for each $\beta$ parameter.

- Draw the fitted quadratic funtion (using the mean values of the $\beta$ parameters from MCMC) together with data points and their uncertainties.

# Exercise 2

- A study on water quality of streams, a high level of bacter X was defined as a level greater than 100 per 100 ml of stream water. $n = 115$ samples were taken from streams having a high environmental impact on pandas. Out of these, $y = 10$ had a high bacter X level.

- Letting $p$ be the probability that a sample of water taken from the stream has a high bacter X level,

(a) Find the frequentist estimator for $p$.

(b) Using a $\texttt{Beta}(1, 11)$ prior for $p$, calculate and posterior distribution $P(p \mid y)$.

(c) Find the bayesian estimator for $p$, the posterior mean and variance, and a 95% credible interval.

(d) Test the hypothesis

$$H_\circ : p = 0.1 \text{ versus } H_1 : p \neq 0.1$$

at 5% level of significance with both the frequentist and bayesian approach.

- A new measurement, performed one month later on $n = 165$ water samples, gives $y = 9$ high bacter X level

(e) Find the frequentist estimator for $p$.

(f) Find a bayesian estimator for $p$, assuming both a $\texttt{Beta}(1, 10)$ prior for $p$, and assuming the posterior probability of the older measurement as the prior for the new one.

(g) Find the bayesian estimator for $p$, the posterior mean and variance, and a 95% credible interval.

(h) Test the hypothesis

$$H_\circ : p = 0.1 \text{ versus } H_1 : p \neq 0.1$$

at 5% level of significance with both the frequentist and bayesian approach.

## Exercise 3

- analyze the data of Exercise 2 and solve points (b) and (c) building and running a MCMC using `Stan`.