

ST509 - Final Project

Submission due on June 14th (Thu) 7:00pm

Task

- Simple linear regression model:

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \dots, n \quad (1)$$

with ϵ_i being random error.

- **Write a short simulation paper** (≤ 3 pages) to compare the performance of various interval estimates of β .

The paper contain:

- Methods to be compared.
 - Simulation set-up
 - Performance measure
 - Summary of results (Table or figures)
 - Interpretations and discussions
- You should submit the following via email (to `sjshin@korea.ac.kr`):
 - Report (`ST509_final_report_yourStdID.pdf`)
 - Ready-to-run r-code (`ST509_final_code_yourStudentID.R` or `.zip`) .
 - The subject line of the email must be `ST509_final_yourStudentID`.

- Mean Regression

- (Frequentist) The conventional least square regression solves

$$\min_{\alpha, \beta} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

which essentially estimate the conditional mean of y given x when $E(\varepsilon) = 0$.

- (Bayesian) Bayesian approach can be applied by assuming the following:

$$\text{Likelihood: } y_i | \alpha, \beta \stackrel{\text{ind}}{\sim} N(\alpha + \beta x_i, \sigma^2)$$

$$\text{Prior: } \alpha \sim N(0, \sigma_\alpha^2); \quad \beta \sim N(0, \sigma_\beta^2) \quad \text{and} \quad \frac{1}{\sigma^2} \sim \text{Gamma}(a, b)$$

Bayesian estimates of α and β are simply the corresponding posterior means, which can be easily obtained by the **Gibbs sampling**.

- Median Regression

- (Frequentist) Under presence of outliers, one can solve

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha, \beta}{\operatorname{argmin}} \sum_{i=1}^n |y_i - \alpha - \beta x_i|$$

which essentially estimates the conditional median of y given x when the median of ε is 0.

- (Bayesian) Bayesian median regression can be done by assuming:

$$\text{Likelihood: } y_i | \alpha, \beta \stackrel{\text{ind}}{\sim} DE(\alpha + \beta x_i, \sigma) \tag{2}$$

where $DE(\mu, \sigma)$ denotes the double exponential distribution with location and scale parameter μ and σ , respectively. The PDF is given by

$$f(y; \mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|y - \alpha - \beta x|}{\sigma}\right)$$

Posterior means of α and β under (2) can be thought as a Bayesian median regression estimate. Posterior sampling can be done via **Metropolis-Hasting algorithm within Gibbs sampling**.