Week 4 Lab MAST90125: Bayesian Statistical learning

Question One

We have seen residual plots in lecture. One example is a case where observations y_i are simulated using

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad i = 1, \cdots, n, \tag{1}$$

but the model fitted was

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \cdots, n. \tag{2}$$

Simulate 2000 datapoints according to (1) and fit the model (2) to this data and answer the following.

a) For your simulated data, generate replicate data from the posterior predictive distribution, and construct two test statistics such that one will suggest poor model fit and the other will suggest good model fit.

   Hint: In week 2 lab, we found a Bayesian interpretation of a t-test in the context of estimating a mean. Generalise this to the regression case.

b) Perform a marginal check, that is calculate

$$p_i = \Pr(y_i^{\text{rep}} \leq y_i | y_1, \cdots, y_n)$$

   Comment on the distribution of p_i, including a discussion of whether this marginal check was appropriate for checking model plausibility in this example. Graphical summaries may prove useful.

Question Two

In week 2 lab, we looked at the posterior distribution for the parameters of a normal distribution assuming Jeffreys' priors. For this example, determine

$$\text{i) } \text{Var}(\mu|\mathbf{y}), \quad \text{ii) } E(\text{Var}(\mu|\mathbf{y})), \quad \text{iii) } \text{Var}(E(\mu|\mathbf{y})), \quad \text{with } \mathbf{y} = (y_1, \cdots, y_n)^\top$$

and using the law of total variance, deduce what this implies about $\text{Var}(\mu)$.

Hint: If z is drawn from a student-t distribution with ν degrees of freedom, then $E(z) = 0$ and $\text{Var}(z) = \frac{\nu}{\nu-2}$.