Q6. Bayes Estimation of Two-Phase Linear Regression Model

Linear regression is a commonly used staitstical strategy in many fields, and one basic assumption is that the regression coefficients are constant and thus can be estimated using the data. However, in practice, there can be chaning points in the coeffeicients, and we can treat the location of changing point, coefficients before and after seperately as parameters. This extension of standard linear regression is called two-phased linear regression model.

In the simplest case with only on predictor and one changing point, the parameter of interests are changing point m, intercept and slope before changing α_1, β_1, slope after chaning β_2 and measurement error σ^2. When we have prior knowledge on the parameters, we can characterize them as prior distribution, and calculate the best estimate of parameters from the posterior distribution given data observed, so that we utilize both previous knowledge and data observation to the best extent.

Depending on situation, the procedure of geting estimation of parameters from the posterior distribution can vary. And some of the estimation is only valid when the prior distributions are correctly or not too far from the true one. In the two-phase linear regression model, we can have posterior mean $\alpha_1, \beta_1, \beta_2$ as reasonable Bayesian estimation. However, the posterior mean of m is very sensitive to prior, while the posterior mode can act as a robust estimation.

A nice application of bayes estimation of two-phase linear regression is the regression of height on age. From biological research, we know that human's height increases the fastest during childhood and puberty, and the growing speed will drop to nearly zero after adulthood. Accordingly, we can set the priors to be $m \sim N(20, 2)$, $\beta_2/const. \sim Beta(1, 100)$, $\beta_1/const. \sim Beta(5, 1)$, $f(\sigma^2) \propto \frac{1}{\sigma^2}$. The Bayesian estimation with these science-based priors and data combined will give us robust parameter estiamtions, and are more conveincing than depending all on data itself.