Problem from STAT5020

1. Suppose we have a collection of binary responses y_i, $i = 1, \cdots, n$, and associated k-dimensional predictor variables $\mathbf{x}_i$. Define the latent variable y_i^* as

$$y_i^* = \mathbf{x}_i^T \beta + \epsilon_i, \quad i = 1, \cdots, n,$$

where the ϵ_i are independent mean-zero errors having cumulative distribution function F, and β is a k-dimensional regression parameter. Consider the model

$$Y_i = \begin{cases} 0, & \text{if } Y_i^* \geq 0 \\ 1, & \text{if } Y_i^* < 0 \end{cases}$$

(a) Show that if F is the standard normal distribution, this model is equivalent to the usual probit model for $p_i = P(Y_i = 1)$.

(b) Under a $N(\mu, \Sigma)$ prior for β, find the full conditional distributions for β and the Y_i^*, $i = 1, \cdots, n$.

(c) How would you modify your computational approach if, instead of the probit model, we wished to fit the logit (logistic regression) model?

$$\log\left(\frac{p_i}{1-p_i}\right) = X_i^T \beta + \epsilon_i.$$

2. In Bayesian model comparison:

(a) Use a concrete example to illustrate how to implement the path sampling procedure for computing Bayes factor in the context of multilevel structural equation models (SEMs).

(b) Discuss other Bayesian model comparison statistics in the comparison of multilevel SEMs.

$$u_{gi} = v_g + \lambda_{1g} w_{1gi} + \epsilon_{1gi}$$

$$v_g = \mu + \lambda_2 w_{2g} + \epsilon_{2g}.$$

$$\eta_{1gi} = \alpha_{1g} + \beta_{1gi1} + \beta_{1gi2} + \alpha_{1gi}$$