Computational Statistics
Analysis of Labor Force Participation Data using the Probit Model

We consider the data `partitipation` in R package `Ecdat`. It is cross-section of 872 observations about participation of women to the labor force in Switzerland to be explained by several covariates (log of nonlabour income, years of formal education, number of children, etc.). We will analyze these data using the Probit model. We recall that this model can be written as follows:

$$y_i^* = x_i^T \beta + u_i, \quad u_i \sim N(0,1)$$

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq 0 \\ 1 & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, n$, where x denotes the vector of covariates, y the response, and y^* a latent variable. We thus have

$$P(y_i = 1 | x_i) = P(y_i^* \leq 0 | x_i) = \Phi(x_i^T \beta).$$

1. Apply the Probit model to this data set (use function `glm` with argument `family=binomial("probit")`.

2. Compute 95% confidence intervals on the coefficients using three methods:

   (a) The profiling method, using function `confint`;

   (b) The normal approximation method. For this, write a function that computes the log-likelihood, optimize it using function `optim` (check that you get the same estimates as in Question 1), and compute confidence intervals using the observed information matrix.

   (c) The bootstrap percentile method.

3. We now perform a Bayesian analysis. We assume a normal prior $\beta \sim N(\beta_0, \mathbf{B}_0)$. We want to compute credible intervals on the coefficients.

   (a) Try using the rejection sampling. Does it work? Why?

(b) Using a Gibbs sampler; in this case we generate in turn the vector $\mathbf{y}^* = (y_1^*, \ldots, y_n^*)$ of latent variables using its distribution given β and $\mathbf{y} = (y_1, \ldots, y_n)$, and β from its distribution given $\mathbf{y}^*$ and $\mathbf{y}$. (Remark that, assuming the latent variable to be observed, the estimation for β becomes a linear regression problem). Use function `rtruncnorm` from package `truncnorm` to draw from the truncated normal distribution).