Bayesian Statistics, Assignment 2
Collegio Carlo Alberto
Due date: November 25^{th}, 2022

Question 1 (15 points): Probit regression. (Hoff 6.3)

A panel study followed $n = 25$ married couples over a period of five years. One item of interest is the relationship between divorce rates and the various characteristics of the couples. For example, the researchers would like to model the probability of divorce as a function of age differential, recorded as the man's age minus the woman's age. The data can be found in the file `divorce.RData`. We will model these data with probit regression, in which a binary variable Y_i is described in terms of an explanatory variable x_i via the following latent variable model:

$$
\begin{aligned}
Z_i &= \beta x_i + \epsilon_i \\
Y_i &= \mathbb{1}_{(c,+\infty)}(Z_i),
\end{aligned}
$$

where β and c are unknown coefficients, $\epsilon_1, \ldots, \epsilon_n \overset{\text{iid}}{\sim} N(0,1)$ and $\mathbb{1}_{(c,+\infty)}(z) = 1$ if $z > c$ and equals zero otherwise. In the following, since the covariates x_i are known, they will be treated as constants and so not explicitly written in the conditioning part.

a) Assuming $\beta \sim N(0, \sigma_\beta^2)$, obtain the full conditional distribution $p(\beta \mid y_{1:n}, z_{1:n}, c)$.

b) Assuming $c \sim N(0, \sigma_c^2)$, show that $p(c \mid y_{1:n}, z_{1:n}, \beta)$ is a constrained normal density, i.e. proportional to a normal density but constrained to lie in an interval. Similarly, show that $p(z_i \mid y_{1:n}, z_{-i}, \beta, c)$ is proportional to a normal density but constrained to be either above c or below c, depending on y_i.

   Hint: A constrained, or truncated, normal random variable V is obtained by restricting a normally distributed random variable $N(\mu, \tau)$ to lie in an interval (a, b), with possibly $a = -\infty$ or $b = \infty$. We use the notation $V \sim TN_{(a,b)}(\mu, \tau^2)$. It holds:

   - $p(v; \mu, \tau^2, a, b) = \dfrac{1}{C} \dfrac{1}{\sqrt{2\pi\tau^2}} \exp\left\{ -\dfrac{1}{2\tau^2} (v - \mu)^2 \right\} \mathbb{1}_{(a,b)}(v)$, where $C = \Phi\left(\dfrac{b-\mu}{\tau}\right) - \Phi\left(\dfrac{a-\mu}{\tau}\right)$, being $\Phi(\cdot)$ the cdf of the standard normal distribution. By definition, it holds $\Phi\left(\dfrac{b-\mu}{\tau}\right) = 1$ if $b = \infty$ and $\Phi\left(\dfrac{a-\mu}{\tau}\right) = 0$ if $a = -\infty$.
   - Sampling can be performed thanks to the function `rtruncnorm(n, a, b, mean, sd)` from the package `rtruncnorm` [https://cran.r-project.org/web/packages/truncnorm/truncnorm.pdf]. This function receives in input the number of desired samples (n) and the four parameters specifying the distribution of V: a, b, μ, τ. Pay attention that it takes as last inputs the mean μ and the standard deviation τ (not the variance τ^2!) of the un-truncated normal density.

c) Letting $\sigma_\beta^2 = \sigma_c^2 = 16$, implement a Gibbs sampling scheme that approximates the joint posterior distribution of $Z_{1:n}$, β, and c. After a burnin of $1,000$, run the Gibbs sampler long enough so that the effective sample sizes of all unknown parameters are greater than $1,000$ (including the Z_i's). Compute the autocorrelation function of the parameters and discuss the mixing of the Markov chain.

d) Obtain a 95% posterior credible interval for β, as well as $\mathbb{P}(\beta > 0 \mid y_{1:n})$.

Question 2 (15 points): Hierarchical modeling. (adapted from Hoff 8.3)

The file `schools.RData` gives weekly hours spent on homework for students sampled from eight different schools. Obtain posterior distributions for the true means for the eight different schools using a hierarchical normal model with the following prior parameters:

$$\mu_0 = 7,\ \gamma_0^2 = 5, \quad \eta_0 = 2,\ \tau_0^2 = 10, \quad \nu_0 = 2,\ \sigma_0^2 = 15.$$

That is,

$$\begin{aligned} y_{1,j}, \ldots, y_{n_j,j} \mid \theta_j, \sigma^2 &\overset{\text{iid}}{\sim} N\left(\theta_j, \sigma^2\right), \quad j = 1, \ldots, 8, \\ \theta_1, \ldots, \theta_8 \mid \mu, \tau^2 &\overset{\text{iid}}{\sim} N\left(\mu, \tau^2\right), \\ \mu \sim N(\mu_0, \gamma_0^2), \quad 1/\tau^2 &\sim \text{Gamma}(\eta_0/2, \eta_0\tau_0^2/2), \quad 1/\sigma^2 \sim \text{Gamma}(\nu_0/2, \nu_0\sigma_0^2/2). \end{aligned}$$

a) Run a Gibbs sampling algorithm to approximate the posterior distribution of $\{\theta_1, \ldots, \theta_8, \mu, \sigma^2, \tau^2\}$. Assess the convergence of the Markov chain, and find the effective sample size for $\{\theta_1, \ldots, \theta_8, \mu, \sigma^2, \tau^2\}$. Run the chain long enough so that the effective sample sizes are all above $1,000$, after a burnin of $1,000$.

b) Compute posterior means and 95% confidence regions for $\{\mu, \sigma^2, \tau^2\}$. Also, compare the posterior densities to the prior densities, and discuss what was learned from the data. (For the density of the inverse-Gamma distribution you can use the function `dinvgamma(x, shape, rate)` from the library `invgamma`).

c) Plot the posterior density of $R = \dfrac{\tau^2}{\sigma^2 + \tau^2}$ and compare it to a plot of the prior density of R (obtained via MC). Describe the evidence for between-school variation.

d) Compute the posterior probability that, if we were to observe a new school with school-specific parameter θ_9, $\theta_9 > \theta_7$, as well as the posterior predictive probability that a new observation from this school would be greater than a new observation from school 7.

e) Plot the sample averages $\bar{y}_1, \ldots, \bar{y}_8$ against the posterior expectations of $\theta_1, \ldots, \theta_8$, and describe the relationship. Also compute the sample mean of all observations and compare it to the posterior mean of μ.

Question 1 (15 points): Probit regression. (Hoff 6.3)

A panel study followed $n = 25$ married couples over a period of five years. One item of interest is the relationship between divorce rates and the various characteristics of the couples. For example, the researchers would like to model the probability of divorce as a function of age differential, recorded as the man's age minus the woman's age. The data can be found in the file `divorce.RData`. We will model these data with probit regression, in which a binary variable Y_i is described in terms of an explanatory variable x_i via the following latent variable model:

$$Z_i = \beta x_i + \epsilon_i$$
$$Y_i = \mathbb{1}_{(c,+\infty)}(Z_i),$$

where β and c are unknown coefficients, $\epsilon_1, \ldots, \epsilon_n \overset{\text{iid}}{\sim} N(0, 1)$ and $\mathbb{1}_{(c,+\infty)}(z) = 1$ if $z > c$ and equals zero otherwise. In the following, since the covariates x_i are known, they will be treated as constants and so not explicitly written in the conditioning part.

a) Assuming $\beta \sim N(0, \sigma_\beta^2)$, obtain the full conditional distribution $p(\beta \mid y_{1:n}, z_{1:n}, c)$.

a. We already know:

- $p(z_i \mid \beta)$ $\forall i=1\ldots m$: $Z_i(\omega) \mid \beta = \beta x_i + \varepsilon_i(\omega) \sim \beta x_i + N(0,1) \sim N(\beta x_i, 1) \Rightarrow Z_i \mid \beta \sim N(\beta x_i, 1) \Rightarrow p(z_i \mid \beta) = (2\pi)^{-1/2} \exp\{-\frac{1}{2}(z_i - x_i\beta)^2\}$
- $p(y_i \mid c, z_i)$ $\forall i=1\ldots m$: $y_i(\omega) \mid c, \underbrace{z_i}_{\text{it means } Z_i = z_i} = \mathbb{1}_{(c,+\infty)}(z_i) = \begin{cases} 1 & \text{if } z_i > c \\ 0 & \text{otherwise} \end{cases}$, constant r.v. $\Rightarrow y_i \mid c, z_i \sim p(y_i \mid c, z_i) =$

$= \mathbb{P}(Y_i = y_i \mid c, z_i) = \mathbb{P}(\mathbb{1}_{(c,+\infty)}(z_i) = y_i) = \begin{cases} \mathbb{P}(\mathbb{1}_{(c,+\infty)}(z_i) = 1) & \text{if } y_i = 1 \\ \mathbb{P}(\mathbb{1}_{(c,+\infty)}(z_i) = 0) & \text{if } y_i = 0 \\ 0 & \text{else} \end{cases} = \begin{cases} \mathbb{P}(z_i > c) & \text{if } y_i = 1 \\ \mathbb{P}(\{z_i > c\}^c) & \text{if } y_i = 0 \\ 0 & \text{else} \end{cases} = \mathbb{1}_{\{1\}}(y_i) \cdot \mathbb{P}(z_i > c) + \mathbb{1}_{\{0\}}(y_i) \cdot \mathbb{P}(\{z_i > c\}^c)$

$p(\beta \mid y_{1:m}, z_{1:m}, c) = p(\beta, y_{1:m}, z_{1:m}, c)_1 / \cancel{p(y_{1:m}, z_{1:m}, c)} \cdot \frac{p(\beta, c, z_{1:m})_2}{p(\beta, c, z_{1:m})_1} \cdot \frac{p(\beta, c)_3}{p(\beta, c)_2} \cdot \frac{\cancel{p(c)}}{p(c)_3} \propto_\beta p(y_{1:m} \mid \cancel{\beta}, c, z_{1:m}) \cdot p(z_{1:m} \mid \beta, \cancel{c}) \cdot p(\beta \mid \cancel{c}) \propto_\beta p(z_{1:m} \mid \beta) \cdot p(\beta)$

$\Rightarrow p(\beta \mid y_{1:m}, z_{1:m}, c) \propto \prod_{i=1}^m p(z_i \mid \beta) \cdot p(\beta \mid c) \propto$

$\propto \exp\{-\frac{1}{2}\sum_{i=1}^m (z_i - x_i\beta)^2\} \cdot \exp\{-\frac{1}{2\sigma_\beta^2} \cdot \beta^2\} =$

$= \exp\{-\frac{1}{2}(\sum_{i=1}^m x_i^2 \cdot \beta + \cancel{\sum_{i=1}^m z_i^2} - 2\sum_{i=1}^m x_i z_i \cdot \beta + \frac{\beta^2}{\sigma_\beta^2})\} =$

$= \exp\{-\underbrace{(\sum_{i=1}^m x_i^2 + \frac{1}{\sigma_\beta^2})}_{=:(\sigma_{\beta,m}^2)^{-1}} \cdot \frac{\beta^2}{2} + \underbrace{\sum_{i=1}^m x_i z_i}_{=:\mu_{\beta,m}/\sigma_{\beta,m}^2} \beta\}$

$\Rightarrow \beta \mid y_{1:m}, z_{1:m}, c \sim N(\mu_{\beta,m}, \sigma_{\beta,m}^2)$, $\begin{cases} \sigma_{\beta,m}^2 = (\frac{1}{\sigma_\beta^2} + \sum_{i=1}^m x_i^2)^{-1} \\ \mu_{\beta,m} = \sigma_{\beta,m}^2 \cdot \sum_{i=1}^m x_i z_i \end{cases}$ $\square$

b) Assuming $c \sim N(0, \sigma_c^2)$, show that $p(c \mid y_{1:n}, z_{1:n}, \beta)$ is a constrained normal density, i.e. proportional to a normal density but constrained to lie in an interval. Similarly, show that $p(z_i \mid y_{1:n}, z_{-i}, \beta, c)$ is proportional to a normal density but constrained to be either above c or below c, depending on y_i.

b.1 $p(c \mid y_{1:m}, z_{1:m}, \beta) = p(c, y_{1:m}, z_{1:m}, \beta)_1 / \cancel{p(y_{1:m}, z_{1:m}, \beta)} \cdot \frac{p(\beta, c, z_{1:m})_2}{p(\beta, c, z_{1:m})_1} \cdot \frac{p(c, \beta)_3}{p(c, \beta)_2} \cdot \frac{\cancel{p(\beta)}}{p(\beta)_3} \propto_c$

$\propto_c p(y_{1:m} \mid \cancel{\beta}, c, z_{1:m}) \cdot p(z_{1:m} \mid \beta, \cancel{c}) \cdot p(c \mid \cancel{\beta}) \propto p(y_{1:m} \mid c, z_{1:m}) \cdot p(c)$. $\quad c_i := \begin{cases} 1 & \text{if } y_i = 0 \\ c & \text{if } y_i = 1 \end{cases}$

$\Rightarrow p(c \mid y_{1:m}, z_{1:m}, \beta) \propto \prod_{i=1}^m p(y_i \mid c, z_i) \cdot p(c) =$

$= (2\pi)^{-1/2} \cdot \frac{1}{\sigma_c} \cdot \exp\{-\frac{1}{2} \cdot \frac{1}{\sigma_c^2} \cdot c^2\} \cdot \prod_{i=1}^m (\mathbb{1}_{\{1\}}(y_i) \cdot \underbrace{\mathbb{P}(z_i > c)}_{= \mathbb{1}_{(-\infty, z_i)}(c)} + \mathbb{1}_{\{0\}}(y_i) \cdot \underbrace{\mathbb{P}(\{z_i > c\}^c)}_{= \mathbb{1}_{[z_i, +\infty)}(c)}) = \ldots \cdot \prod_{i=1}^m \mathbb{1}_{(-\infty, z_i)^{c_i}}(c) =$

$= \ldots \cdot \mathbb{1}_{\bigcap_{i=1}^m (-\infty, z_i)^{c_i}}(c)$

$= \frac{1}{\sqrt{2\pi\sigma_c^2}} \cdot \exp\{-\frac{1}{2\sigma_c^2} \cdot c^2\} \cdot \mathbb{1}_{\bigcap_{i=1}^m (-\infty, z_i)^{c_i}}(c)$

$\Rightarrow c \mid y_{1:m}, z_{1:m}, \beta \sim TN_{a_m, b_m}(0, \sigma_c^2)$, $a_m := \min\{z_i \mid i \in \{1, \ldots, m\}, y_i = 0\}$, $b_m = \max\{z_i \mid i \in \{1, \ldots, m\}, y_i = 1\}$.

b.2. Similarly $p(z_i \mid y_{1:m}, z_{-i}, \beta, c) = p(z_i, y_{1:m}, z_{-i}, \beta, c)_1 \Big/ p(\cancel{y_{1:m}, z_{-i}}, \beta, c) \cdot \dfrac{p(z_i, z_{-i}, \beta, c)_2}{p(z_i, z_{-i}, \beta, c)_1} \cdot \dfrac{p(z_i, \beta, c)_3}{p(z_i, \beta, c)_2} \cdot \dfrac{\cancel{p(\beta, c)}}{p(\beta, c)_3} \propto_{z_i}$

$$\propto_{z_i} p(y_{1:m} \mid z_{1:m}, \cancel{\beta}, c) \cdot p(z_{-i} \mid \cancel{z_i}, \beta, \cancel{c}) \cdot p(z_i \mid \beta, \cancel{c}) =$$

$$\propto_{z_i} p(y_{1:m} \mid z_{1:m}, c) \cdot \cancel{p(z_{-i} \mid \beta)} \cdot p(z_i \mid \beta) \propto_{z_i}$$

$$\propto_{z_i} p(y_{1:m} \mid z_{1:m}, c) \cdot p(z_i \mid \beta) \propto$$

$$\propto \prod_{j=1}^{m} p(y_j \mid z_j, c) \cdot p(z_i \mid \beta) \propto_{z_i}$$

$$\propto_{z_i} p(y_i \mid z_i, c) \cdot p(z_i \mid \beta) \propto$$

$$\propto \Big(y_i \cdot \underbrace{\mathbb{1}_{(-\infty, z_i)}(c)}_{\substack{= \begin{cases} 1 & \text{if } c < z_i \\ 0 & \text{otherwise} \end{cases} \\ = \mathbb{1}_{(c, +\infty)}(z_i)}} + (1 - y_i) \cdot \underbrace{\mathbb{1}_{(-\infty, z_i)^c}(c)}_{\substack{= \begin{cases} 1 & \text{if } c \geq z_i \\ 0 & \text{otherwise} \end{cases} \\ = \mathbb{1}_{(-\infty, c]}(z_i)}} \Big) \cdot \mathbb{1}_{\{0,1\}}(y_i) \cdot \exp\left(-\frac{1}{2} (z_i - x_i \beta)^2 \right) =$$

$$= \begin{cases} \mathbb{1}_{(c, +\infty)}(z_i) \cdot \exp\left(-1/2 \, (z_i - x_i \beta)^2 \right) & \text{if} \quad y_i = 1 \\ \mathbb{1}_{(-\infty, c]}(z_i) \cdot \exp\left(-1/2 \, (z_i - x_i \beta)^2 \right) & \text{if} \quad y_i = 0 \end{cases}$$

As before, this conditional density is proportional to the kernel of a gaussian (evaluated on z_i) multiplied by an indicator function (also evaluated on z_i) which constrains the domain to be $(c, +\infty)$ or $(-\infty, c]$ (equivalently $(-\infty, c)$, with the same motivation given above) depending on y_i.

In particular, completing to a density what we found

$$p(z_i \mid y_{1:m}, z_{-i}, \beta, c) = \begin{cases} \dfrac{1}{\underbrace{\phi\left(\frac{+\infty - x_i \beta}{1}\right)}_{= 1} - \phi\left(\frac{c - x_i \beta}{1}\right)} \cdot \dfrac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} (z_i - x_i \beta)^2 \right\} \cdot \mathbb{1}_{(c, +\infty)}(z_i) & \text{if} \quad y_i = 1 \\[2ex] \dfrac{1}{\phi\left(\frac{c - x_i \beta}{1}\right) - \underbrace{\phi\left(\frac{-\infty - x_i \beta}{1}\right)}_{= 0}} \cdot \dfrac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} (z_i - x_i \beta)^2 \right\} \cdot \mathbb{1}_{(-\infty, c)}(z_i) & \text{if} \quad y_i = 0 \end{cases}$$

$$\Rightarrow \quad z_i \mid y_{1:m}, z_{-i}, \beta, c \sim \begin{cases} TN_{(c, +\infty)}(x_i \beta, 1) & \text{if} \quad y_i = 1 \\ TN_{(-\infty, c)}(x_i \beta, 1) & \text{if} \quad y_i = 0 \end{cases}$$