Hints for assignment 2: MAST900125

John Holmes

THE UNIVERSITY OF MELBOURNE

Question Two:

- In question 2b), you are told to use the output from part a) to define priors to use in part b). The motivation for this is that we can revisit a comment made in lecture 3.
  - But the posterior can also be written as,
    $$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}_2|\boldsymbol{\theta})p(\mathbf{y}_1|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y}_2,\mathbf{y}_1)} = \frac{p(\mathbf{y}_2|\boldsymbol{\theta})p(\mathbf{y}_1,\boldsymbol{\theta})}{p(\mathbf{y}_2,\mathbf{y}_1)} = \frac{p(\mathbf{y}_2|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_1)p(\mathbf{y}_1)}{p(\mathbf{y}_2,\mathbf{y}_1)}$$
    $$= \frac{p(\mathbf{y}_2|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y}_1)}{p(\mathbf{y}_2|\mathbf{y}_1)},$$
  - such that we can calculate the posterior from the first sample only, $p(\boldsymbol{\theta}|\mathbf{y}_1)$, and then use $p(\boldsymbol{\theta}|\mathbf{y}_1)$ as a prior when analysing the second sample.
  - Hence you can think of the informative prior as representing the information gained from all past experiments of the same phenomena.

Viewing question 2a as a Bayesian:

- In question 2 a), you were asked to fit a model using `lm`. From a Bayesian viewpoint (see lecture 12), this is equivalent to assuming $p(\beta) \propto 1, p(\tau) \propto \tau^{-1}$.
  - With these priors, the joint distribution $p(\mathbf{y}, \mathbf{X}, \beta, \tau)$ is,

$$\begin{aligned}
\frac{\tau^{n/2}}{(2\pi)^{n/2}} e^{-\frac{\tau(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2}} \times \tau^{-1} &= \frac{\tau^{n/2-1}}{(2\pi)^{n/2}} e^{-\frac{\tau(\mathbf{y}-\mathbf{X}\beta+\mathbf{X}\hat{\beta}-\mathbf{X}\hat{\beta})'(\mathbf{y}-\mathbf{X}\beta+\mathbf{X}\hat{\beta}-\mathbf{X}\hat{\beta})}{2}} \\
&= \frac{\tau^{n/2-1}}{(2\pi)^{n/2}} e^{-\frac{\tau(\mathbf{y}-\mathbf{X}\hat{\beta})'(\mathbf{y}-\mathbf{X}\hat{\beta})}{2}} e^{-\frac{\tau(\mathbf{X}\beta-\mathbf{X}\hat{\beta})'(\mathbf{X}\beta-\mathbf{X}\hat{\beta})}{2}} e^{\tau(\mathbf{y}-\mathbf{X}\hat{\beta})'(\mathbf{X}\beta-\mathbf{X}\hat{\beta})} \\
&= \frac{\tau^{n/2-1}}{(2\pi)^{n/2}} e^{-\frac{\tau(n-p)s^2}{2}} e^{-\frac{\tau(\beta-\hat{\beta})'(\mathbf{X}'\mathbf{X})(\beta-\hat{\beta})}{2}}
\end{aligned}$$

as $\mathbf{X}'(\mathbf{y}-\mathbf{X}\hat{\beta}) = \mathbf{0}$ by definition and $s^2 = \dfrac{(\mathbf{y}-\mathbf{X}\hat{\beta})'(\mathbf{y}-\mathbf{X}\hat{\beta})}{n-p}$.

Viewing question 2a as a Bayesian:

- In question 2 a), you were asked to fit a model using `lm`. From a Bayesian viewpoint (see lecture 12), this is equivalent to assuming $p(\beta) \propto 1, p(\tau) \propto \tau^{-1}$.
  - If we extract the kernel of β from the joint distribution,

$$e^{-\frac{\tau(\beta-\hat{\beta})'(\mathbf{X}'\mathbf{X})(\beta-\hat{\beta})}{2}}$$

    we can deduce that $p(\beta|\tau,\hat{\beta})$ is normal with mean $\hat{\beta}$ and variance $\frac{(\mathbf{X}'\mathbf{X})^{-1}}{\tau}$.
  - Now marginalise β out of the joint distribution,

$$p(\tau,\mathbf{y},\mathbf{X}) = \frac{\tau^{\frac{n}{2}-1}e^{-\frac{\tau(n-p)s^2}{2}}}{(2\pi)^{n/2}}\int e^{-\frac{\tau(\beta-\hat{\beta})'(\mathbf{X}'\mathbf{X})(\beta-\hat{\beta})}{2}}d\beta = \frac{\tau^{\frac{n}{2}-1}e^{-\frac{\tau(n-p)s^2}{2}}}{(2\pi)^{n/2}}\left(\frac{2\pi}{\tau}\right)^{\frac{p}{2}}\det(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}$$

$$\propto \quad \tau^{(n-p)/2-1}e^{-\frac{\tau(n-p)s^2}{2}}$$

    so we can deduce that $p(\tau|s^2) = \text{Ga}(\frac{n-p}{2}, \frac{(n-p)s^2}{2})$.

Using question 2a in 2b:

- The idea in question two b) is to use the posteriors implied in a) and priors in b). Remember the parameters to be estimated in b) are $\boldsymbol{\beta}, \tau$. Hence we need a joint prior $p(\boldsymbol{\beta}, \tau)$.
- By the laws of probability, we know $p(\boldsymbol{\beta}, \tau) = p(\boldsymbol{\beta}|\tau)p(\tau)$. More importantly, in lecture 12 we determined posteriors of the form $\boldsymbol{\beta}|\tau, \hat{\boldsymbol{\beta}}$ and $\tau|s^2$. These are your new priors:
  - $p(\boldsymbol{\beta}|\tau) = \mathcal{N}(\hat{\boldsymbol{\beta}}_1, (\mathbf{X}_1'\mathbf{X}_1)^{-1}/\tau)$
  - $p(\tau|s^2) = \text{Ga}(\frac{n_1 - p}{2}, \frac{(n_1 - p)s^2}{2})$,

  where the subscript 1 indicates the results are from group 1 (analysed in 2a) alone.
- Note these priors look just like the $p(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{K}/\tau_\beta)$ considered in lecture 13, except τ_e, τ_β have been merged.

Modifying Lecture 13 results for question 2b

- In lecture 13, you were told the joint distribution implied for a regression where you assume $p(\beta) = \mathcal{N}(\beta_0, \mathbf{K}/\tau_\beta)$ and gamma priors for the precisions was,

$$\left(\frac{\tau_e}{2\pi}\right)^{n/2} e^{-\frac{\tau_e(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2}} \times \left(\frac{\tau_\beta}{2\pi}\right)^{p/2} \det(\mathbf{K})^{-1/2} e^{-\frac{\tau_\beta(\beta-\beta_0)'\mathbf{K}^{-1}(\beta-\beta_0)}{2}} \times \frac{\gamma_\beta^{\alpha_\beta}}{\Gamma(\alpha_\beta)} \tau_\beta^{\alpha_\beta-1} e^{-\gamma_\beta\tau_\beta} \times \frac{\gamma_e^{\alpha_e}}{\Gamma(\alpha_e)} \tau_e^{\alpha_e-1} e^{-\gamma_e\tau_e}.$$

- For question 2 b), drop the prior for τ_β, and in all other places in the joint distribution, replace τ_e, τ_β with τ. Hence the joint distribution becomes,

$$\left(\frac{\tau}{2\pi}\right)^{n/2} e^{-\frac{\tau(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)}{2}} \times \left(\frac{\tau}{2\pi}\right)^{p/2} \det(\mathbf{K})^{-1/2} e^{-\frac{\tau(\beta-\beta_0)'\mathbf{K}^{-1}(\beta-\beta_0)}{2}} \times \frac{\gamma^{\alpha}}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\gamma\tau},$$

where $\mathbf{K} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}$, $\beta_0 = \hat{\beta}_1$, $\alpha = (n_1 - p)/2$ and $\gamma = (n_1 - p)s^2/2$.

Modifying Lecture 13 results for Question 2b

- In lecture 13, we determined that the conditional posteriors for τ_e, τ_β were:
  - The component of the joint distribution that is a function of τ_e is,
    $$\left(\frac{\tau_e}{2\pi}\right)^{n/2} e^{-\frac{\tau_e(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}} \frac{\gamma_e^{\alpha_e}}{\Gamma(\alpha_e)} \tau_e{}^{\alpha_e-1} e^{-\gamma_e\tau_e} \propto \tau_e^{\alpha_e+n/2-1} e^{-\tau_e(\gamma_e+(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2)},$$
    a gamma kernel, meaning $p(\tau_e|\mathbf{y}, \boldsymbol{\beta}, \mathbf{X}) = \mathrm{Ga}(\alpha_e + \frac{n}{2}, \gamma_e + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2})$.
  - The component of the joint distribution that is a function of τ_β is,
    $$\left(\frac{\tau_\beta}{2\pi}\right)^{p/2} e^{-\frac{\tau_\beta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2}} \times \frac{\gamma_\beta^{\alpha_\beta}}{\Gamma(\alpha_\beta)} \tau_\beta^{\alpha_\beta-1} e^{-\gamma_\beta\tau_\beta} \propto \tau_\beta^{\alpha_\beta+p/2-1} e^{-\tau_\beta(\gamma_\beta+(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)/2)},$$
    a gamma kernel, meaning $p(\tau_\beta|\mathbf{y}, \boldsymbol{\beta}, \mathbf{K}) = \mathrm{Ga}(\alpha_\beta + \frac{p}{2}, \gamma_\beta + \frac{(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2})$.
- Since the difference between Lecture 13 and Question 2 is the merging of τ_e, τ_β, the conditional posterior for τ in 2b) must be,
  $$p(\tau|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\beta}_0, \mathbf{K}) = \mathrm{Ga}\left(\alpha + \frac{n+p}{2}, \gamma + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2}\right).$$

Modifying Lecture 13 results for question 2b

- In lecture 13, we determined that the conditional posteriors for β was:
  - The component of the joint distribution that is a function of β is,

$$
\begin{aligned}
e^{-\frac{\tau_e(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2}} e^{-\frac{\tau_\beta(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{K}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)}{2}} &\propto e^{-\frac{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})\boldsymbol{\beta}}{2}} e^{\frac{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{y}+\tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)}{2}} e^{\frac{(\tau_e\mathbf{X}'\mathbf{y}+\tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)'\boldsymbol{\beta}}{2}} \\
&= e^{-\frac{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})\boldsymbol{\beta}}{2}} e^{\boldsymbol{\beta}'(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})(\tau_e\mathbf{X}'\mathbf{X}+\tau_\beta\mathbf{K}^{-1})^{-1}(\tau_e\mathbf{X}'\mathbf{y}+\tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)}
\end{aligned}
$$

  which corresponds to a normal kernel, such that $p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_0, \mathbf{K}, \tau_e, \tau_\beta)$ is multivariate-normal with mean $=(\tau_e\mathbf{X}'\mathbf{X} + \tau_\beta\mathbf{K}^{-1})^{-1}(\tau_e\mathbf{X}'\mathbf{y} + \tau_\beta\mathbf{K}^{-1}\boldsymbol{\beta}_0)$ and variance-covariance matrix $(\tau_e\mathbf{X}'\mathbf{X} + \tau_\beta\mathbf{K}^{-1})^{-1}$.

- Since the difference between lecture 13 and question 2 is the merging of τ_e, τ_β, the conditional posterior for $\boldsymbol{\beta}$ in question 2b) must be,

$$
\begin{aligned}
p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_0, \mathbf{K}, \tau) &= \mathcal{N}((\tau\mathbf{X}'\mathbf{X} + \tau\mathbf{K}^{-1})^{-1}(\tau\mathbf{X}'\mathbf{y} + \tau\mathbf{K}^{-1}\boldsymbol{\beta}_0), (\tau\mathbf{X}'\mathbf{X} + \tau\mathbf{K}^{-1})^{-1}) \\
&= \mathcal{N}((\mathbf{X}'\mathbf{X} + \mathbf{K}^{-1})^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{K}^{-1}\boldsymbol{\beta}_0), (\mathbf{X}'\mathbf{X} + \mathbf{K}^{-1})^{-1}/\tau).
\end{aligned}
$$

Modifying Lecture 13 results for question 2b

- We have now defined the conditional posteriors required to construct a Gibbs sampler for question 2b).
- For the coding, look at the mixed model code given in lab 7 and modify. You need to make the following changes to the function:
  - Add an argument $\boldsymbol{\beta}_0$.
  - Add an argument $\mathbf{K}$.
  - Remove the arguments related to τ_β.
  - As all β co-efficients have a normal prior, remove the argument related to fixed co-efficients.
  - Modify the code updating $\boldsymbol{\beta}$ to correspond to the new posterior.

Question 3c

- In lecture 14, we discussed generalised linear models. In particular, we looked at probit regression, because we could implement a Gibbs sampler for this problem.
  - Lets determine a Gibbs sampler for probit regression. We know $\mathbf{z} = \mathbf{X}\beta + \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $y_i = 1$ if $z_i \geq 0$ and 0 otherwise and assume $p(\beta) \propto 1$.
  - The joint distribution $p(\mathbf{y}, \mathbf{z}, \beta)$ is
  $$\prod_{i=1}^{n} \mathbb{1}_{\mathrm{sign}(z_i)=\mathrm{sign}(y_i-1/2)} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_i-\mathbf{x}_i'\beta)^2}{2}}.$$
  - The kernel of β is
  $$\prod_{i=1}^{n} e^{-\frac{(z_i-\mathbf{x}_i'\beta)^2}{2}} = e^{-\frac{\sum_{i=1}^{n}(z_i-\mathbf{x}_i'\beta)^2}{2}} = e^{-\frac{(\mathbf{z}-\mathbf{X}\beta)'(\mathbf{z}-\mathbf{X}\beta)}{2}} \propto e^{-\frac{\beta'(\mathbf{X}'\mathbf{X})\beta - 2\beta'(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}}{2}},$$
  which implies that the conditional posterior $p(\beta|\mathbf{z})$ is $\mathcal{N}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}, (\mathbf{X}'\mathbf{X})^{-1})$.
  - Then we just cycle between sampling from $p(\beta|\mathbf{z})$ and $p(z_i|y_i)$; $i = 1, \ldots, n$.

Question 3c

- In Question 3, what is the difference?
  - We are now assuming $\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2\mathbf{I})$ and $\tau_u = 1/\sigma_u^2$.
- If you look at the conditional posterior for $\boldsymbol{\beta}$ determined for probit regression, it looks just like normal linear regression, but with $\tau = 1$.
- Question 3 looks like a mixed model, so finding the conditional posteriors of $\boldsymbol{\beta}, \mathbf{u}, \tau_u$ conditional on $\mathbf{z}$ should match results in lecture 13, except $\tau_e = 1$. This means assume a gamma prior for τ_u.

Question 3c

- In particular, we want to look at the part of lecture 13, where $\boldsymbol{\beta}$ was split into $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.
  - Now consider the special case of the linear mixed model where $\boldsymbol{\beta}$ can be split into $\begin{pmatrix}\boldsymbol{\beta}_1\end{pmatrix}$ such that *a priori* $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are independent, that is $p(\boldsymbol{\beta}_1) \propto 1$, $p(\boldsymbol{\beta}_2) = \mathcal{N}(\mathbf{0}_{p_2}, \sigma^2_\beta \mathbf{K}_2)$.
  - If we look at the kernel of the prior of $\boldsymbol{\beta}_2$, $e^{-\tau_\beta \boldsymbol{\beta}_2' \mathbf{K}_2^{-1} \boldsymbol{\beta}_2/2}$. This can be extended to incorporate the prior for $\boldsymbol{\beta}$ by assuming

$$p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \propto e^{-\frac{\tau_\beta (\boldsymbol{\beta}_1' \quad \boldsymbol{\beta}_2') \begin{pmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{K}_2^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}}{2}} = e^{-\tau_\beta \boldsymbol{\beta}' \mathbf{K}^{-1} \boldsymbol{\beta}/2},$$

where $\mathbf{K}^{-1} = \begin{pmatrix} \mathbf{0}_{p_1 \times p_1} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{K}_2^{-1} \end{pmatrix}$.

Question 3c

- In question 3, the $\boldsymbol{\beta}$ is $\boldsymbol{\beta}_1$, the $\mathbf{u}$ is $\boldsymbol{\beta}_2$, $\mathbf{K}_2$ is $\mathbf{I}$, τ_β is τ_u and as previously noted $\tau_e = 1$. Hence the conditional posteriors
  - As a result, the conditional posteriors are modified as follows when fitting a linear mixed model,

$$p(\tau_\beta|\cdot) \;=\; \mathrm{Ga}(\alpha_\beta + p_2/2, \gamma_\beta + \boldsymbol{\beta}_2'\mathbf{K}_2^{-1}\boldsymbol{\beta}_2/2)$$

$$p\left(\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \middle| \cdot \right) = \mathcal{N}\left(\tau_e \begin{pmatrix} \tau_e\mathbf{X}_1'\mathbf{X}_1 & \tau_e\mathbf{X}_1'\mathbf{X}_2 \\ \tau_e\mathbf{X}_2'\mathbf{X}_1 & \tau_e\mathbf{X}_2'\mathbf{X}_2 + \tau_\beta\mathbf{K}_2^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix}, \begin{pmatrix} \tau_e\mathbf{X}_1'\mathbf{X}_1 & \tau_e\mathbf{X}_1'\mathbf{X}_2 \\ \tau_e\mathbf{X}_2'\mathbf{X}_1 & \tau_e\mathbf{X}_2'\mathbf{X}_2 + \tau_\beta\mathbf{K}_2^{-1} \end{pmatrix}^{-1} \right)$$

in question three become:

$$p(\tau_u|\cdot) \;=\; \mathrm{Ga}(\alpha_u + q/2, \gamma_u + \mathbf{u}'\mathbf{u}/2)$$

$$p\left(\begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{u} \end{pmatrix} \middle| \cdot \right) = \mathcal{N}\left(\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{X} + \tau_u\mathbf{I}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\mathbf{z} \\ \mathbf{Z}'\mathbf{z} \end{pmatrix}, \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{X} + \tau_u\mathbf{I}^{-1} \end{pmatrix}^{-1} \right)$$

Question 3c coding hint

- Note, the model described in question 3 has already been coded up for the case of continuous data in Lab 7 (see `normalmm.Gibbs`). You need to modify this.
  - Remove references to τ_e so this is now fixed at 1.
  - Add lines into the iteration for loop to sample $\mathbf{z}_i; 1, \ldots n$ from the truncated normal posterior. Question 2b is essentially practice to work out what these lines should be.
- In lecture 14 we showed the conditional posterior for z_i in probit regression was,
  - But how can we learn z_i? By the rules of probability, the posterior for z_i is,

$$p(z_i|y_i, \mathbf{X}, \beta) = \frac{p(y_i, z_i|\mathbf{X}, \beta)}{\Pr(y_i|\mathbf{X}, \beta)} = \begin{cases} \frac{1}{\Phi(\mathbf{x}_i'\beta)\sqrt{2\pi}} e^{-\frac{(z_i - \mathbf{x}_i'\beta)^2}{2}} & \text{If } y_i = 1 \text{ and } z_i \geq 0. \\ \frac{1}{(1-\Phi(\mathbf{x}_i'\beta))\sqrt{2\pi}} e^{-\frac{(z_i - \mathbf{x}_i'\beta)^2}{2}} & \text{If } y_i = 0 \text{ and } z_i \leq 0. \\ 0 & \text{otherwise} \end{cases}$$

  - Which corresponds to a truncated normal distribution defined on $(0, \infty)$ if $y_i = 1$ and $(-\infty, 0)$ if $y_i = 0$.

Question 3c coding hint

- In question 3, note $\mathbf{z} = \mathbf{X}\beta + \epsilon$ has become $\mathbf{z} = \mathbf{X}\beta + \mathbf{Zu} + \epsilon$
- This changes the conditional posterior for z_i to,

$$p(z_i|y_i, \mathbf{X}, \boldsymbol{\beta}, \mathbf{u}) = \frac{p(y_i, z_i|\mathbf{X}, \boldsymbol{\beta}, \mathbf{u})}{\Pr(y_i|\mathbf{X}, \boldsymbol{\beta}, \mathbf{u})} = \begin{cases} \frac{1}{\Phi(\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{Z}_i\mathbf{u})\sqrt{2\pi}} e^{-\frac{(z_i-\mathbf{x}_i'\boldsymbol{\beta}-\mathbf{Z}_i\mathbf{u})^2}{2}} & \text{If } y_i = 1 \text{ and } z_i \geq 0. \\ \frac{1}{(1-\Phi(\mathbf{x}_i'\boldsymbol{\beta}+\mathbf{Z}_i\mathbf{u}))\sqrt{2\pi}} e^{-\frac{(z_i-\mathbf{x}_i'\boldsymbol{\beta}-\mathbf{Z}_i\mathbf{u})^2}{2}} & \text{If } y_i = 0 \text{ and } z_i \leq 0. \\ 0 & \text{otherwise} \end{cases}$$

- Which corresponds to a truncated normal distribution defined on $(0, \infty)$ with parameters $\mu = \mathbf{x}_i'\beta + \mathbf{Z}_i\mathbf{u}$ and $\sigma^2 = 1$ if $y_i = 1$ and $(-\infty, 0)$ with parameters $\mu = \mathbf{x}_i'\beta + \mathbf{Z}_i\mathbf{u}$ and $\sigma^2 = 1$ if $y_i = 0$.