Winter 22 – STAT206B Homework 4 Solution

1. (Wasserman, 2003) A random variable Z has an *inverse Gaussian distribution* if it has density

$$f(z \mid \theta_1, \theta_2) \propto z^{-3/2} \exp\left\{-\theta_1 z - \frac{\theta_2}{z} + 2\sqrt{\theta_1\theta_2} + \log(\sqrt{2\theta_2})\right\}, \quad z > 0,$$

where $\theta_1 > 0$ and $\theta_2 > 0$ are parameters. It can be shown that $E(Z) = \sqrt{\theta_2/\theta_1}$ and $E(1/Z) = \sqrt{\theta_1/\theta_2} + 1/(2\theta_2)$.

(a) Let $\theta_1 = 1.5$ and $\theta_2 = 2$. Draw a sample of size 1,000 using the independence-Metropolis-Hastings method with a Gamma distribution as the proposal density (note that in an independence-Metropolis-Hastings $q(\theta^*|\theta) = q(\theta^*)$). To assess the accuracy of the method, compare the mean of Z and $1/Z$ from the sample to the theoretical means. Try different Gamma distributions to see if you can get an accurate sample.

Solution: The theoretical means are $\mathrm{E}(Z) = 1.1547$ and $\mathrm{E}(1/Z) = 1.405$. I tried three different gamma distributions as a proposal distribution to simulate Z using the independence-Metropolis-Hastings method, (a) $\mathrm{Gamma}(1, 1)$, (b) $\mathrm{Gamma}(0.1, 0.1)$, (c) $\mathrm{Gamma}(5, 1)$. The results are shown in the table below with acceptance ratios.

(n, x)	$\mathrm{E}(Z)$	$\mathrm{E}(1/Z)$	acceptance ratio
$\mathrm{Gamma}(1, 1)$	1.136	1.154	0.553
$\mathrm{Gamma}(0.1, 0.1)$	1.138	1.099	0.11
$\mathrm{Gamma}(5, 1)$	1.667	0.709	0.076
RW with $\mathrm{N}(0, 0.5)$	1.200	1.137	0.628

Proposal distributions (a) and (b) yield similar estimates. Proposal distribution (c) has a very low acceptance ratio, meaning that it does not produce proposals that represent the target distribution. The estimates are off from their theoretical means. I then compare the proposal densities to the target density in Figure 1. The figure shows that $\mathrm{Gamma}(5, 1)$ has very large discrepancy from the target density. It explains a reason that $\mathrm{Gamma}(5, 1)$ produces poor estimates.

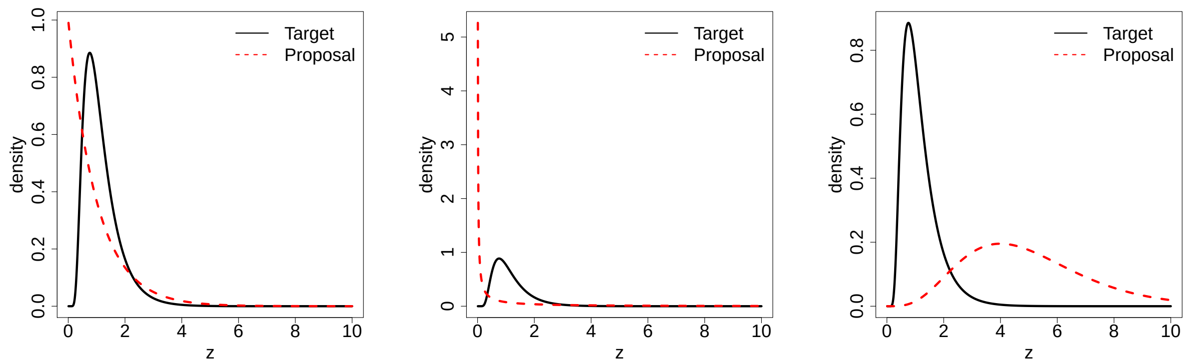
(b) Draw a sample of size 1,000 using the random-walk Metropolis method. Since $z > 0$ we cannot just use a Normal density. Let $W = \log(Z)$. Find the density of W. Use the random-walk Metropolis method to get a sample $W_1, \ldots, W_M$ and let $Z_i = e^{W_i}$. Assess the accuracy of the simulation as in the previous part.

Solution: Using that the density of W is given by

$$f_W(w) = f_Z(z(w))|dw/dz| \propto \exp\left\{\frac{3}{2}w - \theta_1 \exp(w) - \frac{\theta_2}{\exp(w)}\right\}\exp(w).$$

Estimates of $\mathrm{E}(Z)$ and $\mathrm{E}(1/Z)$ are given in the table above. They seem reasonable.

2. Consider i.i.d. data $x_1, \ldots, x_n$ such that $x_i|\nu, \theta \sim \mathrm{Gamma}(\nu, \theta)$ where $E(x_i) = \nu/\theta$, and assign priors $\nu \sim \mathrm{Gamma}(3, 1)$ and $\theta \sim \mathrm{Gamma}(2, 2)$.



(a) $\text{Gamma}(1,1)$ (b) $\text{Gamma}(0.1,0.1)$ (c) $\text{Gamma}(5,1)$

Figure 1: Comparison of the proposal densities to the target density. $\text{Gamma}(1,1)$, $\text{Gamma}(0.1,0.1)$ and $\text{Gamma}(5,1)$ are used as a proposal density. The target density is the inverse-Gaussian with $\theta_1 = 1.5$ and $\theta_2 = 2.0$.

(a) Develop a Metropolis-within-Gibbs algorithm to sample from $p(\nu, \theta | x_1, \ldots, x_n)$ using the full conditional distributions $p(\theta|\nu, x_1, \ldots, x_n)$ and $p(\nu|\theta, x_1, \ldots, x_n)$. For the second full conditional, use a random walk proposal on $\log(\nu)$.

Solution: Let $\nu \sim \text{Gamma}(a, b)$ and $\theta \sim \text{Gamma}(\alpha, \beta)$. The joint posterior distribution of θ and ν satisfies

$$\begin{aligned}
p(\theta, \nu \mid \boldsymbol{x}) &\propto p(\boldsymbol{x} \mid \nu, \theta) p(\nu, \theta) \\
&\propto \frac{\theta^{n\nu}}{\{\Gamma(\nu)\}^n} \exp\left\{ (\nu - 1) \sum_{i=1}^n \log(x_i) - \theta \sum_{i=1}^n x_i \right\} \\
&\qquad \times \nu^{a-1} \exp(-b\nu) \theta^{\alpha-1} \exp(-\beta\theta) \\
&\propto \frac{\nu^{(a-1)} \theta^{(\alpha+n\nu-1)}}{\{\Gamma(\nu)\}^n} \\
&\qquad \times \exp\left\{ \nu \left(\sum_{i=1}^n \log(x_i) - b \right) - \theta \left(\sum_{i=1}^n x_i + \beta \right) \right\}.
\end{aligned}$$

Thus, the full conditional of θ is

$$p(\theta \mid \nu, \boldsymbol{x}) \propto \theta^{n\nu+\alpha-1} \exp\left\{ -\theta \left(\sum_{i=1}^n x_i + \beta \right) \right\}, \quad \theta \in \mathbb{R}^+.$$

This, $\theta \mid \nu, \boldsymbol{x} \sim \text{Gamma}(n\nu + \alpha, \sum_{i=1}^n x_i + \beta)$. The full conditional of ν is

$$p(\nu \mid \theta, \boldsymbol{x}) \propto \frac{\nu^{a-1} \theta^{n\nu}}{\{\Gamma(\nu)\}^n} \exp\left\{ \nu \left(\sum_{i=1}^n \log(x_i) - b \right) \right\}, \quad \nu \in \mathbb{R}^+,$$

which is not a member of a known family of distributions.
We use the following Metropolis-within-Gibbs algorithm to sample (ν, θ) from $p(\nu, \theta \mid \boldsymbol{x})$.

Step 0 Initialize $\nu^{(0)}$ and $\theta^{(0)}$.

Step 1 Iterate the following steps for $t = 1, \ldots, T$ for large enough T.

Step 1-1 Draw $\theta^{(t)}$ given $\boldsymbol{x}$ and $\nu^{(t-1)}$ from

$$\theta \mid \nu^{(t-1)}, \boldsymbol{x} \sim \text{Gamma}\left(n\nu^{(t-1)} + \alpha, \sum_{i=1}^{n} x_i + \beta\right).$$

Step 1-2 Use a MH algorithm with random walk proposal to draw $\nu^{(t)}$ given $\boldsymbol{x}$ and $\theta^{(t)}$ from

$$p(\nu \mid \theta, \boldsymbol{x}) \quad \propto \quad \frac{\nu^{a-1}\theta^{n\nu}}{\{\Gamma(\nu)\}^n} \exp\left\{\nu\left(\sum_{i=1}^{n}\log(x_i) - b\right)\right\}.$$

Specifically, let $\xi = \exp(\log(\nu) + \epsilon)$ with $\epsilon \sim \mathrm{N}(0, v^2)$, where v^2 is pre-specified. We let $\nu^{(t)} = \xi$ with probability $\min(\alpha, 1)$, where

$$\alpha = \frac{\frac{\xi^a \times (\theta^{(t)})^{n\xi}}{\{\Gamma(\xi)\}^n} \exp\left\{\xi\left(\sum_{i=1}^{n}\log(x_i) - b\right)\right\}}{\frac{(\nu^{(t-1)})^a \times (\theta^{(t)})^{n\nu^{(t-1)}}}{\{\Gamma(\nu^{(t-1)})\}^n} \exp\left\{\nu^{(t-1)}\left(\sum_{i=1}^{n}\log(x_i) - b\right)\right\}}.$$

Note that the update of ξ is done on the logarithm scale. Do not forget the Jacobian! Otherwise, we let $\nu^{(t)} = \nu^{(t-1)}$.

(b) Develop a Metropolis-Hastings algorithm that jointly proposes $\log(\nu)$ and $\log(\theta)$ using a Gaussian random walk centered on the current value of the parameters. Tune the variance-covariance matrix of the proposal using a test run that proposes the parameters independently (but evaluates acceptance jointly).

Let $w = \log(\theta)$ and $v = \log(\nu)$, and rewrite the posterior as:

$$p(w, v \mid \boldsymbol{x}) \propto \exp\left\{av + (\alpha + ne^v)w - n\log\left(\Gamma(e^v)\right) + e^v\left(\sum_{i=1}^{n}\log(x_i) - b\right) - e^w\left(\beta + \sum_{i=1}^{n}x_i\right)\right\}$$

We use the following MH algorithm to sample (ν, θ) from $p(\nu, \theta \mid \boldsymbol{x})$.

Step 0 Initialize $w^{(0)}$ and $v^{(0)}$.

Step 1 Iterate the following step for $t = 1, \ldots, T$ for large enough T.

Step 1-1 Let $\xi_1 = w^{(t-1)} + \epsilon_1$ and $\xi_2 = v^{(t-1)} + \epsilon_2$ where $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2] \sim \mathrm{N}_2(0, \Omega)$. Ω is calibrated based on a test run (see the provided code and check how I calibrated Ω) and pre-specified. We let $w^{(t)} = \xi_1$ and $v^{(t)} = \xi_2$ with probability $\min(\alpha, 1)$, where

$$\begin{aligned}\alpha \quad = \quad & \frac{\exp\left\{a\xi_2 + (\alpha + ne^{\xi_2})\xi_1 - n\log\left(\Gamma(e^{\xi_2})\right)\right\}}{\exp\left\{av^{(t-1)} + (\alpha + ne^{v^{(t-1)}})w^{(t-1)} - n\log\left(\Gamma(e^{v^{(t-1)}})\right)\right\}} \\ & \times \frac{\exp\left\{e^{\xi_2}\left(\sum_{i=1}^{n}\log(x_i) - b\right) - e^{\xi_1}\left(\beta + \sum_{i=1}^{n}x_i\right)\right\}}{\exp\left\{e^{v^{(t-1)}}\left(\sum_{i=1}^{n}\log(x_i) - b\right) - e^{w^{(t-1)}}\left(\beta + \sum_{i=1}^{n}x_i\right)\right\}}.\end{aligned}$$

Otherwise, let $w^{(t)} = w^{(t-1)}$ and $v^{(t)} = v^{(t-1)}$.

Step 1-2 Let $\theta^{(t)} = \exp(w^{(t)})$ and $\nu^{(t)} = \exp(v^{(t)})$

(c) Develop a Metropolis algorithm that jointly proposes $\log(\nu)$ and $\log(\theta)$ using independent proposals based on the Laplace approximation of the posterior distribution of $\log(\nu)$ and $\log(\theta)$.

Solution: To do a Laplace approximation of $p(w, v \mid \boldsymbol{x})$, we let

$$h(w, v) = av + (\alpha + ne^v)w - n\log\left(\Gamma(e^v)\right) + e^v\left(\sum_{i=1}^n \log(x_i) - b\right) - e^w\left(\beta + \sum_{i=1}^n x_i\right).$$

Find $\hat{w}$ and $\hat{v}$ that maximize $h(w, v)$. We then approximate $p(w, v \mid \boldsymbol{x})$ with $\mathrm{N}_2([\hat{w}, \hat{v}]', A^{-1})$, where A is the Hassian matrix of $h(w, v)$ evaluated at $(\hat{w}, \hat{v})$ and use it as a proposal distribution for a MH algorithm.

We use the following MH algorithm to sample (ν, θ) from $p(\nu, \theta \mid \boldsymbol{x})$.

Step 0 Initialize $w^{(0)}$ and $v^{(0)}$.

Step 1 Iterate the following steps for $t = 1, \ldots, T$ for large enough T.

Step 1-1 Draw $\boldsymbol{\xi} = (\xi_1, \xi_2)$ from $\mathrm{N}_2([\hat{w}, \hat{v}]', A^{-1})$ and let $w^{(t)} = \xi_1$ and $v^{(t)} = \xi_2$ with probability $\min(\alpha, 1)$, where

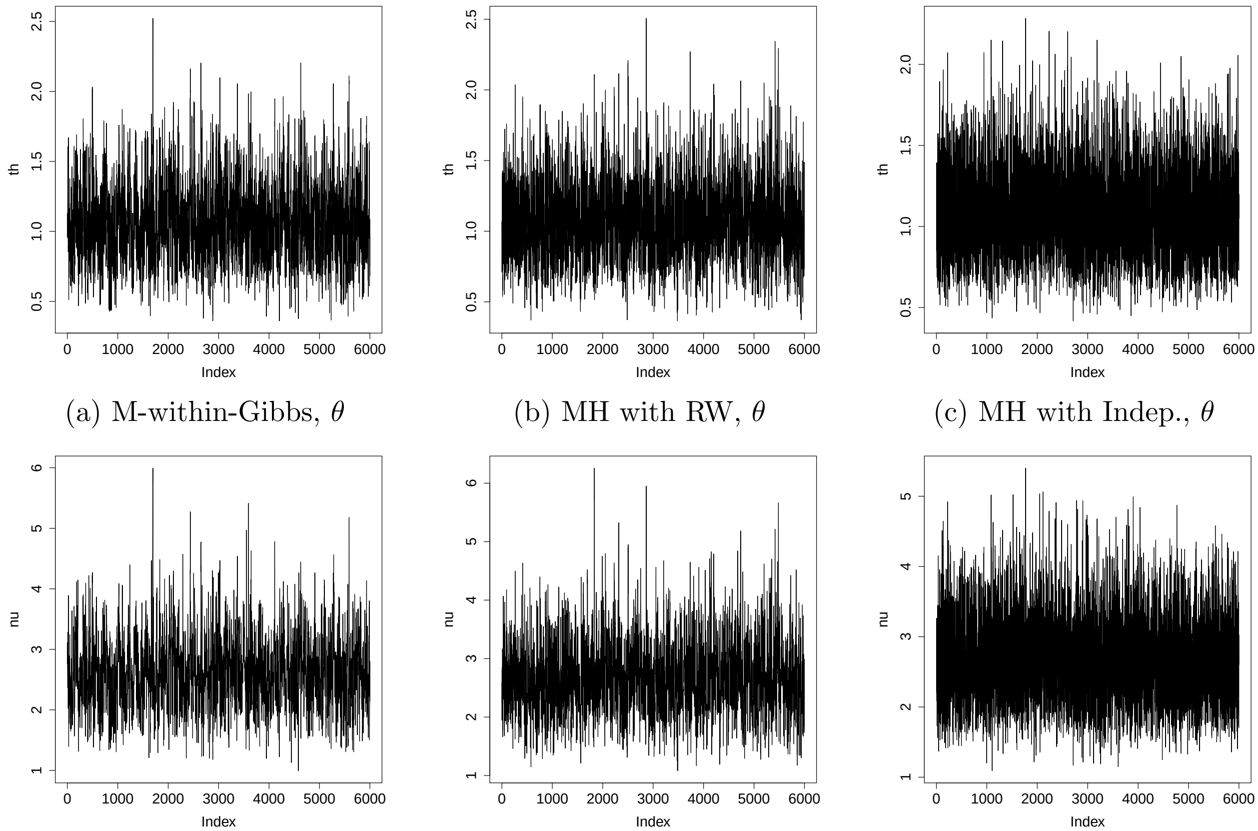
$$\begin{aligned}\alpha &= \frac{\exp\left\{a\xi_2 + (\alpha + ne^{\xi_2})\xi_1 - n\log\left(\Gamma(e^{\xi_2})\right)\right\}}{\exp\left\{av^{(t-1)} + (\alpha + ne^{v^{(t-1)}})w^{(t-1)} - n\log\left(\Gamma(e^{v^{(t-1)}})\right)\right\}} \\ &\times \frac{\exp\left\{e^{\xi_2}\left(\sum_{i=1}^n \log(x_i) - b\right) - e^{\xi_1}\left(\beta + \sum_{i=1}^n x_i\right)\right\}}{\exp\left\{e^{v^{(t-1)}}\left(\sum_{i=1}^n \log(x_i) - b\right) - e^{w^{(t-1)}}\left(\beta + \sum_{i=1}^n x_i\right)\right\}} \\ &\times \frac{\exp\left(-\frac{1}{2}([w^{(t-1)}, v^{(t-1)}] - [\hat{w}, \hat{v}])' A([w^{(t-1)}, v^{(t-1)}] - [\hat{w}, \hat{v}])\right)}{\exp\left(-\frac{1}{2}\boldsymbol{\xi}' A \boldsymbol{\xi}\right)}.\end{aligned}$$

Otherwise, let $w^{(t)} = w^{(t-1)}$ and $v^{(t)} = v^{(t-1)}$.

Step 1-2 Let $\theta^{(t)} = \exp(w^{(t)})$ and $\nu^{(t)} = \exp(v^{(t)})$

(d) Run each of the algorithms for the dataset in `my-data.txt` and compute the effective sample sizes associated with each parameter under each of the samplers. Also, construct trace and autocorrelation plots. Report posterior means for each of the parameters of interest, along with 95% symmetric credible intervals. Discuss.

Solution: Effective sample sizes and posterior inference on θ and ν are reported in the table below. Trace plots and autucorrelation plots are in Figures 2 and 3, respectively. For each algorithm, first 15000 iterations were discarded for burn-in and every 5th draws were taken for Monte Carlo sample of size 6000. The MH algorithm with an independence proposal based on the Laplace approximation has the largest effective sample size. Autocorrelation and trace plots do not show an evidence of bad mixing or poor convergence for all three algorithms. The posterior inference on θ and ν are similar under the three algorithms.



(a) M-within-Gibbs, θ (b) MH with RW, θ (c) MH with Indep., θ

(d) M-within-Gibbs, ν (e) MH with RW, ν (f) MH with Indep., ν

Figure 2: [Q2] Trace plots

Algorithm	Effective SS		Posterior Inference	
	θ	ν	θ	ν
(a) Metropolis-within-Gibbs	603.90	515.79	1.07 (0.60, 1.63)	2.64 (1.59, 3.87)
(b) MH with RW	874.00	609.85	1.07 (0.59, 1.67)	2.66 (1.55, 3.97)
(c) MH with Indep.	3000	3000	1.08 (0.61, 1.67)	2.66 (1.61 3.98)

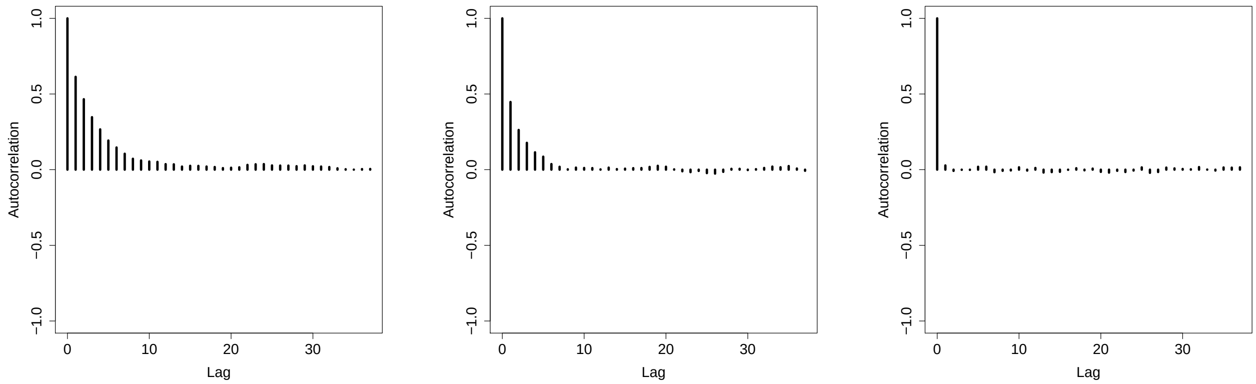
Note that proposals in part (c) that are generated from a distribution on the Laplace approximation were almost always accepted. The proposal distribution based on the Laplace approximation worked well for this example, but it may not be a case all the time.

3. (Robert and Casella) Consider a random effects model,

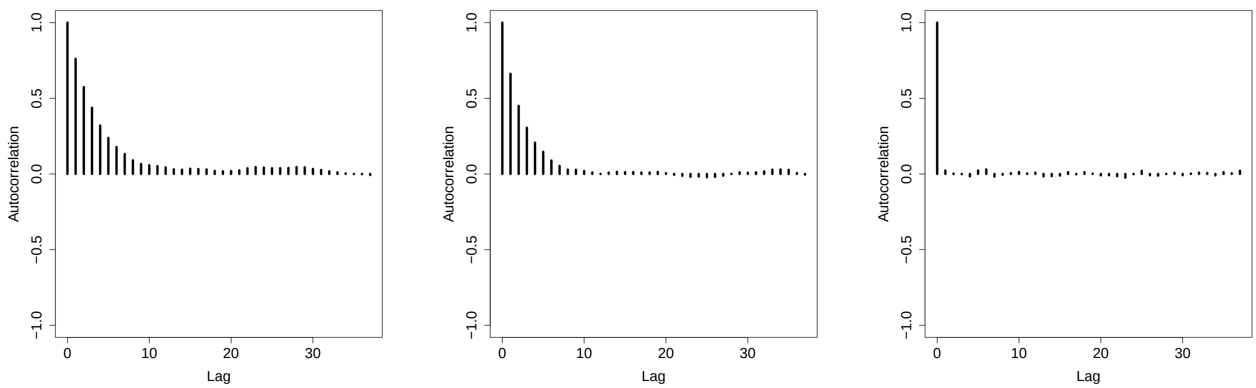
$$y_{i,j} = \beta + u_i + \epsilon_{i,j}, \;\; i = 1 : I, j = 1 : J,$$

where $u_i \sim N(0, \sigma^2)$ and $\epsilon_{i,j} \sim N(0, \tau^2)$. Assume a prior of the form

$$\pi(\beta, \sigma^2, \tau^2) \propto \frac{1}{\sigma^2 \tau^2}.$$



(a) M-within-Gibbs, θ (b) MH with RW, θ (c) MH with Indep., θ



(d) M-within-Gibbs, ν (e) MH with RW, ν (f) MH with Indep., ν

Figure 3: [Q2] Autocorrelation plots

(a) Find the full conditional distributions:

Solution: The joint posterior distribution of all parameters $\beta, \boldsymbol{u}, \sigma^2, \tau^2$ is

$$p(\beta, \boldsymbol{u}, \sigma^2, \tau^2 \mid \boldsymbol{y}) \propto (\tau^2)^{-IJ/2-1}(\sigma^2)^{-I/2-1} \exp\left\{-\sum_{i,j}\frac{(y_{i,j}-\beta-u_i)^2}{2\tau^2} - \sum_i \frac{u_i^2}{2\sigma^2}\right\}.$$

i. $\pi(u_i \mid \boldsymbol{y}, \beta, \tau^2, \sigma^2)$;

$$p(u_i \mid \beta, \sigma^2, \boldsymbol{u}_{-i}, \tau^2, \boldsymbol{y}) \propto \exp\left[-\frac{1}{2}\left\{\left(\frac{J}{\tau^2}+\frac{1}{\sigma^2}\right)u_i^2 - \frac{\sum_j(y_{i,j}-\beta)}{\tau^2}u_i\right\}\right],$$

where $\boldsymbol{u}_{-i}$ is the vector of $\boldsymbol{u}$ but excludes u_i. The full conditional is

$$u_i \mid \boldsymbol{y}, \beta, \tau^2, \sigma^2 \sim \mathrm{N}\left(\left(\frac{J}{\tau^2}+\frac{1}{\sigma^2}\right)^{-1}\frac{\sum_j(y_{i,j}-\beta)}{\tau^2}, \left(\frac{J}{\tau^2}+\frac{1}{\sigma^2}\right)^{-1}\right).$$

ii. $\pi(\beta \mid \boldsymbol{y}, \boldsymbol{u}, \tau^2, \sigma^2)$;

$$p(\beta \mid \sigma^2, \tau^2, \boldsymbol{y}) \propto \exp\left[-\frac{1}{2\tau^2}\left\{IJ\beta^2 - 2\sum_{i,j}(y_{i,j} - u_i)\beta\right\}\right],$$

$$\beta \mid \tau^2, \boldsymbol{y}, \boldsymbol{u} \sim \mathrm{N}\left(\frac{\sum_{i,j}(y_{i,j} - u_i)}{IJ}, \frac{\tau^2}{IJ}\right).$$

iii. $\pi(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{u}, \beta, \tau^2)$;

$$p(\sigma^2 \mid \beta, \boldsymbol{u}, \tau^2, \boldsymbol{y}) \propto (\sigma^2)^{-I/2-1} \exp\left\{-\sum_i \frac{u_i^2}{2\sigma^2}\right\}.$$

$$\sigma^2 \mid \boldsymbol{u} \sim \mathrm{IG}\left(\frac{I}{2}, \sum_i \frac{u_i^2}{2}\right).$$

iv. $\pi(\tau^2 \mid \boldsymbol{y}, \boldsymbol{u}, \beta, \sigma^2)$.

$$p(\tau^2 \mid \beta, \sigma^2, \boldsymbol{u}, \boldsymbol{y}) \propto (\tau^2)^{-IJ/2-1} \exp\left\{-\sum_{i,j} \frac{(y_{i,j} - \beta - u_i)^2}{2\tau^2}\right\}.$$

$$\tau^2 \mid \beta, \boldsymbol{u}, \boldsymbol{y} \sim \mathrm{IG}\left(\frac{IJ}{2}, \sum_{i,j} \frac{(y_{i,j} - \beta - u_i)^2}{2}\right).$$

(b) Find $\pi(\beta, \tau^2, \sigma^2 \mid y)$ up to a proportionality constant.

Solution:

$$\begin{aligned}
\pi(\beta, \tau^2, \sigma^2 \mid \boldsymbol{y}) &= \int p(\beta, \boldsymbol{u}, \sigma^2, \tau^2 \mid \boldsymbol{y}) d\boldsymbol{u} \\
&\propto \int (\tau^2)^{-IJ/2-1}(\sigma^2)^{-I/2-1} \exp\left\{-\sum_{i,j}\frac{(y_{i,j} - \beta - u_i)^2}{2\tau^2} - \sum_i \frac{u_i^2}{2\sigma^2}\right\} d\boldsymbol{u} \\
&= \left(\tau^2\right)^{-IJ/2-1} \left(\sigma^2\right)^{-I/2-1} \left(\frac{J}{\tau^2} + \frac{1}{\sigma^2}\right)^{-I/2} \\
&\quad \times \exp\left\{-\frac{\sum_{ij}(y_{i,j} - \beta)^2}{2\tau^2} + \frac{1}{2}\left(\frac{J}{\tau^2} + \frac{1}{\sigma^2}\right)^{-1} \sum_i \left(\frac{\sum_j (y_{i,j} - \beta)}{\tau^2}\right)^2\right\}
\end{aligned}$$

(c) Find $\pi(\sigma^2, \tau^2 | y)$ up to a proportionality constant and show that this posterior is not integrable since, for $\tau \neq 0$, it behaves like σ^{-2} in a neighborhood of 0.

Solution:

$$\pi(\tau^2,\sigma^2 \mid \boldsymbol{y}) \propto (\tau^2)^{-\frac{I(J-1)}{2}+1} (\sigma^2)^{-1} (J\sigma^2+\tau^2)^{\frac{I+1}{2}} \exp\left\{-\frac{1}{2\tau^2}\sum_{i,j} y_{ij}^2\right\}$$
$$\exp\left\{\frac{\sigma^2}{2\tau^2(J\sigma^2+\tau^2)}\sum_i\left(\sum_j y_{i,j}\right)^2 + \frac{1}{2IJ(J\sigma^2+\tau^2)}\left(\sum_{i,j} y_{i,j}\right)^2\right\}.$$

notice that, for a fixed τ, as $\sigma^2 \to 0$ $\pi(\tau^2,\sigma^2 \mid \boldsymbol{y}) \propto (\sigma^2)^{-1}$, which is not integrable w.r.t. σ^2.

Note: This problem shows that even though the full conditional posteriors exist and the Gibbs sampling could be easily implemented, the joint posterior distribution does not exist. Users should be aware of the risks of using the Gibbs sampler in situations like this!

4. (Carlin, Gelfand and Smith, 1992) Let $y_1,\dots,y_n$ be a sample from a Poisson distribution for which there is a suspicion of a change point m along the observation process where the means change, $m=1,\dots,n$. Given m, $y_i \sim \text{Poi}(\theta)$, for $i=1,\dots,m$ and $y_i \sim \text{Poi}(\phi)$, for $i = m+1,\dots,n$. The model is completed with independent prior distributions $\theta \sim \text{Gamma}(\alpha,\beta)$, $\phi \sim \text{Gamma}(\gamma,\delta)$ and m uniformly distributed over $\{1,\dots,n\}$ where α,β,γ and δ are known constants. Implement a Gibbs sampling algorithm to obtain samples from the joint posterior distribution. Run the Gibbs sampler to apply this model to the data `mining-data.r` which consists of counts of coal mining disasters in Great Britain by year from 1851 to 1962.

Solution: The joint posterior distribution of θ,ϕ,m is

$$p(\theta,\phi,m \mid \boldsymbol{y}) \propto \theta^{\alpha+\sum_{i=1}^m y_i-1}\exp\{-(\beta+m)\theta\}\phi^{\gamma+\sum_{i=m+1} y_i-1}\exp\{-(\delta+n-m)\phi\}.$$

The full conditionals are

- $\theta \mid m,\boldsymbol{y} \sim \text{Gamma}(\alpha+\sum_{i=1}^m y_i, \beta+m)$.
- $\phi \mid m,\boldsymbol{y} \sim \text{Gamma}(\gamma+\sum_{i=m+1}^n y_i, \delta+n-m)$.
- $p(m \mid \theta,\phi,\boldsymbol{y}) \propto \theta^{\alpha+\sum_{i=1}^m y_i-1}\phi^{\gamma+\sum_{i=m+1}^n y_i-1}\exp\{-(\beta+m)\theta-(\delta+n-m)\phi\}, \quad m = 1,\dots,n.$

  Or we can integrate out θ and ϕ and sample m from $p(m \mid \boldsymbol{y})$ (called collapsed Gibbs) instead of $p(m \mid \theta,\phi,y)$,

$$\begin{aligned} p(m \mid \boldsymbol{y}) &= \int p(m,\theta,\phi \mid \boldsymbol{y})d\phi d\theta \\ &\propto \underbrace{\frac{\Gamma(\alpha+\sum_{i=1}^m y_i)\Gamma(\gamma+\sum_{i=m+1}^n y_i)}{(\beta+m)^{\alpha+\sum_{i=1}^m y_i}(\delta+n-m)^{\gamma+\sum_{i=m+1}^n y_i}}}_{=z_m}. \end{aligned}$$

Draw $m \in \{1,\dots,n\}$ from a multinomial distribution with sample size 1 and probabilities $z_m/\sum_{\ell=1}^n z_\ell$.

Let $\alpha = \beta = 0.1$ for the prior of θ and $\gamma = \delta = 0.1$ for the prior of ϕ. The prior means and variances of both θ and ϕ are 1 and 10, respectively, i.e., fairly noninformative priors. The first 3000 draws were discarded for burn-in and every 3rd draw was taken for the posterior sample of size 5000. The data is illustrated in Fig 4(a), where the red vertical line represents an MAP estimate of m. Panels (b)-(d) have the posterior distributions of m, θ and ϕ, respectively, where the vertical lines are MAP estimate of m in (b) and posterior mean and 95% credible estimates in (c) and (d). The MAP estimate of m is 41. The posterior means of θ and ϕ are 3.12 and 0.92, with 95% posterior credible interval estimates (2.59, 3.71) and (0.71, 1.16), respectively. The 95% credible intervals do not overlap, possibly indicating that there is a significant decrease in the mean counts of coal mining disasters.

5. Souza (1999) considers a number of hierarchical models to describe the nutritional pattern of pregnant women. One of the models adopted was a hierarchical regression model where

$$\begin{aligned} y_{i,j} &\sim \mathrm{N}(\alpha_i + \beta_i t_{i,j}, \sigma^2), \\ (\alpha_i, \beta_i)' \mid \alpha, \beta &\sim \mathrm{N}_2((\alpha, \beta)', diag(\tau_\alpha^2, \tau_\beta^2)), \\ (\alpha, \beta)' &\sim \mathrm{N}_2((0, 0)', diag(P_\alpha^2, P_\beta^2). \end{aligned}$$

Here $y_{i,j}$ and $t_{i,j}$ are the jth weight measurement and visit time of the ith woman with $j = 1 : n_i$ and $i = 1 : I$ for $I = 68$ pregnant women. Here $n = \sum_{i=1}^I n_i = 415$. For unknown scale parameters, we assume a priori independence and place a Gamma prior,

$$\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma), \quad \tau_\alpha^2 \sim \mathrm{IG}(a_\alpha, b_\alpha), \text{ and } \tau_\beta^2 \sim \mathrm{IG}(a_\beta, b_\beta).$$

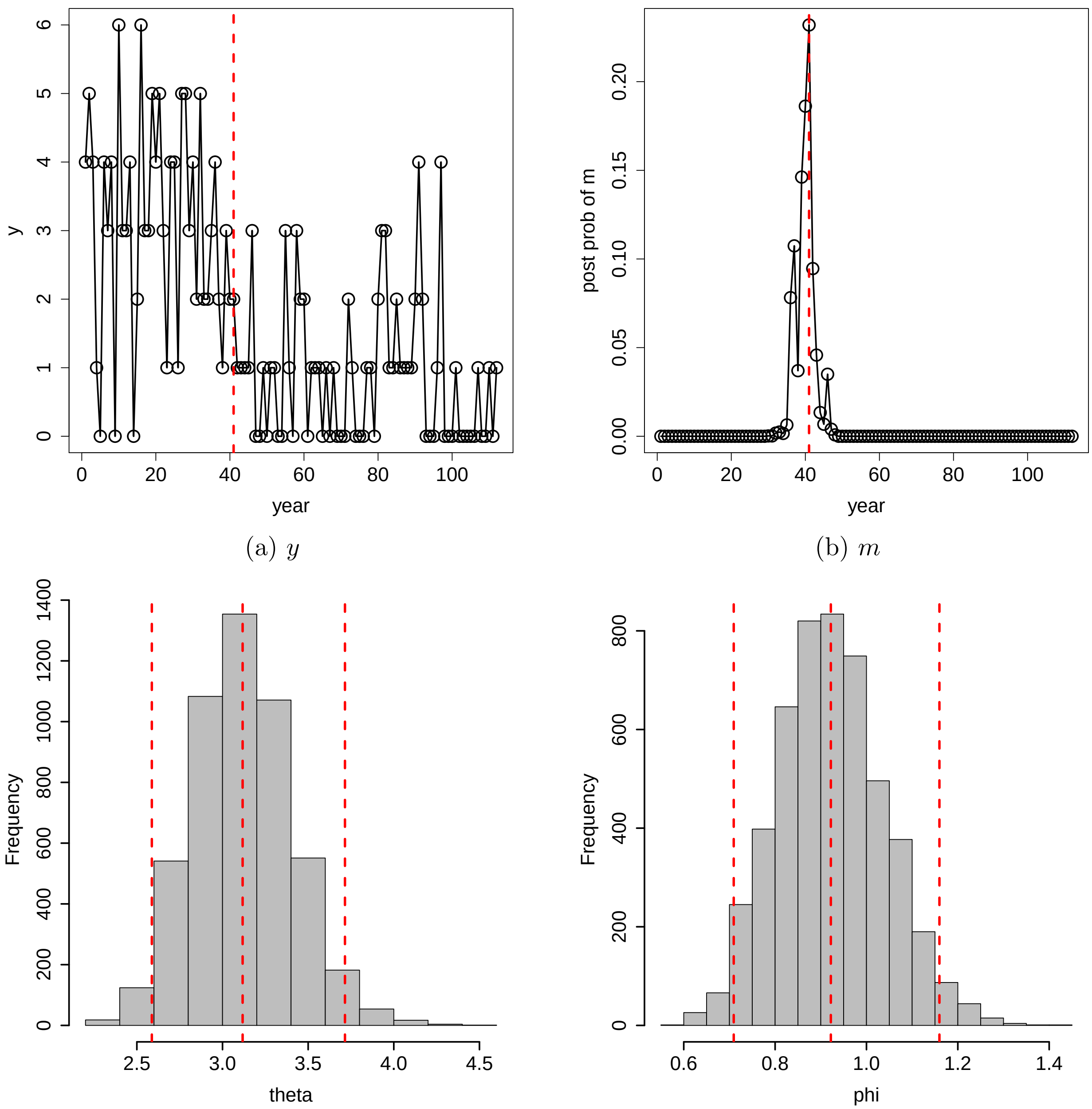
Hyperparameters, $a_\sigma, b_\sigma, a_\alpha, b_\alpha, a_\beta, b_\beta$, P_α^2, P_β^2 are fixed.

(a) Find the joint posterior distribution of all random parameters (up to proportionality)
Solution: Let θ denote a vector of parameters $\theta = ((\alpha_i, \beta_i)_{i=1}^I, \alpha, \beta, \sigma^2, \tau_\alpha, \tau_\beta)$. The joint distribution of θ is

$$\begin{aligned} p(\theta \mid \boldsymbol{y}) \quad \propto \quad & (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{i,j} - (\alpha_i + \beta_i t_{i,j}))^2 \right\} \\ & \times (\tau_\alpha)^{-I/2} (\tau_\beta)^{-I/2} \exp\left[-\frac{1}{2} \left\{ \frac{1}{\tau_\alpha} \sum_{i=1}^I (\alpha_i - \alpha)^2 + \frac{1}{\tau_\beta} \sum_{i=1}^I (\beta_i - \beta)^2 \right\} \right] \\ & \times \exp\left\{ -\frac{1}{2} (\frac{1}{P_\alpha} \alpha^2 + \frac{1}{P_\beta} \beta^2) \right\} (\sigma^{-2})^{a_\sigma - 1} (\tau_\alpha^{-1})^{a_\alpha - 1} (\tau_\beta^{-1})^{a_\beta - 1} \\ & \times \exp\left\{ -(b_\sigma \sigma^{-2} + b_\alpha \tau_\alpha^{-1} + b_\beta \tau_\beta^{-1}) \right\}. \end{aligned}$$

(b) Find the full conditional distributions of $\alpha, \beta, \tau_\alpha, \tau_\beta$, σ^2, α_i, β_i, and (α_i, β_i).
Solution: The full conditionals are

- $\alpha \mid \boldsymbol{y}, \theta_{-\alpha} \sim \mathrm{N}((I/\tau_\alpha + 1/P_\alpha)^{-1} (\sum_{i=1}^I \alpha_i)/\tau_\alpha, (I/\tau_\alpha + 1/P_\alpha)^{-1})$



(a) y (b) m

(c) θ (d) ϕ

Figure 4: [Mining Data]

- $\beta \mid \boldsymbol{y}, \theta_{-\beta} \sim \mathrm{N}\left((I/\tau_\beta + 1/P_\beta)^{-1}(\sum_{i=1}^{I}\beta_i/\tau_\beta), (I/\tau_\beta + 1/P_\beta)^{-1}\right)$
- $\tau_\alpha^{-1} \mid \boldsymbol{y}, \theta_{-\tau_\alpha} \sim \mathrm{Gamma}\left(a + \frac{I}{2},\ b + \frac{1}{2}\sum_{i=1}^{I}(\alpha_i - \alpha)^2\right)$
- $\tau_\beta^{-1} \mid \boldsymbol{y}, \theta_{-\tau_\beta} \sim \mathrm{Gamma}\left(a + \frac{I}{2},\ b + \frac{1}{2}\sum_{i=1}^{I}(\beta_i - \beta)^2\right)$
- $\sigma^{-2} \mid \boldsymbol{y}, \theta_{-\sigma^{-2}} \sim \mathrm{Gamma}\left(a + \frac{n}{2},\ b + \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{i,j} - (\alpha_i + \beta_i t_{i,j}))^2\right)$

- $\alpha_i \mid \boldsymbol{y}, \theta_{-\alpha_i} \overset{indep}{\sim} \mathrm{N}((1/\tau_\alpha + n_i/\sigma^2)^{-1}\{\alpha/\tau_\alpha + \frac{1}{\sigma^2}\sum_{j=1}^{n_i}(y_{i,j} - \beta_i t_{i,j})\},\ (1/\tau_\alpha + n_i/\sigma^2)^{-1})$
- $\beta_i \mid \boldsymbol{y}, \theta_{-\beta_i} \overset{indep}{\sim} \mathrm{N}((1/\tau_\beta + \frac{1}{\sigma^2}\sum_{j=1}^{n_i} t_{i,j}^2)^{-1}\{\beta/\tau_\beta + \frac{1}{\sigma^2}\sum_{j=1}^{n_i} t_{i,j}(y_{i,j} - \alpha_i)\},\ (1/\tau_\beta + \frac{1}{\sigma^2}\sum_{j=1}^{n_i} t_{i,j}^2)^{-1})$
- Alternatively, we may draw (α_i, β_i) as follows; let $n_i \times 2$ design matrix $\boldsymbol{\xi}_i = (\mathbf{1}_{n_i}, \boldsymbol{t}_i)$ and $T = \mathrm{diag}(\tau_\alpha, \tau_\beta)$. We find $(\alpha_i, \beta_i) \overset{indep}{\sim} \mathrm{N}_2(\boldsymbol{\mu}_i, \Sigma_i)$, where

$$\boldsymbol{\mu}_i = \Sigma_i\{\boldsymbol{\xi}_i' \boldsymbol{y}_i/\sigma^2 + T^{-1}(\alpha, \beta)'\},\ \text{and}\ \Sigma_i = (\boldsymbol{\xi}_i'\xi_i/\sigma^2 + T^{-1})^{-1}.$$

(c) Describe a Gibbs sampling algorithm to obtain samples from the joint posterior distribution.

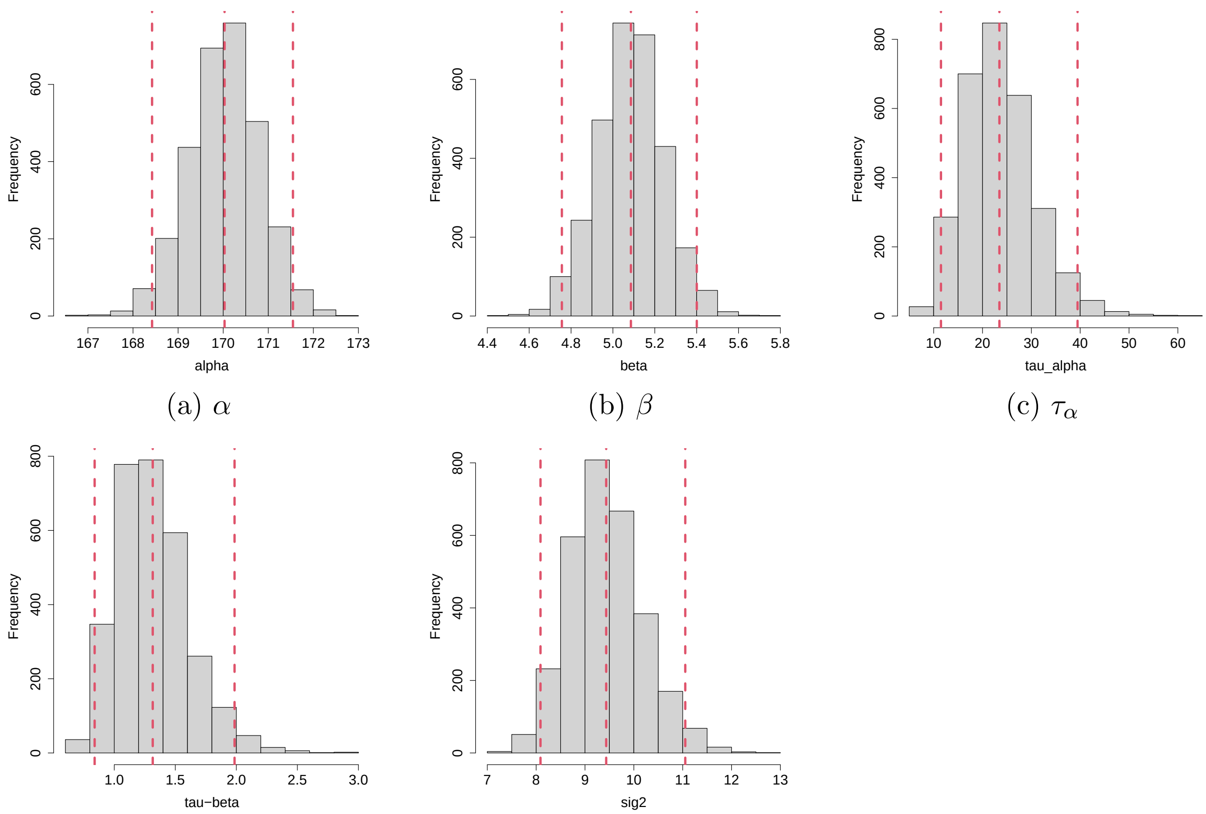
We specify initial values of the random parameters, $\alpha^{(0)}$, $\beta^{(0)}$, $\tau_\alpha^{(0)}$, $\tau_\beta^{(0)}$, $\sigma^{2,(0)}$, $\alpha_i^{(0)}$ and $\beta_i^{(0)}$. OLS estimates of a linear regression for each patient are used for $\alpha_i^{(0)}$ and $\beta_i^{(0)}$, and the means of $\alpha_i^{(0)}$ and $\beta_i^{(0)}$ are to specify $\alpha^{(0)}$ and $\beta^{(0)}$.

(d) Specify values of the fixed hyperparameters, and justify your choices.

We let $P_\alpha = P_\beta = 1000$ and $a_\sigma = b_\sigma = a_\alpha = b_\alpha = a_\beta = b_\beta = 0.001$ to reflect vague prior information in the prior.

(e) Implement your sampling algorithm in part (c) and summarize the posterior distribution using point estimates and interval estimates of the parameters. Provide interpretations of the inference in layman's word.

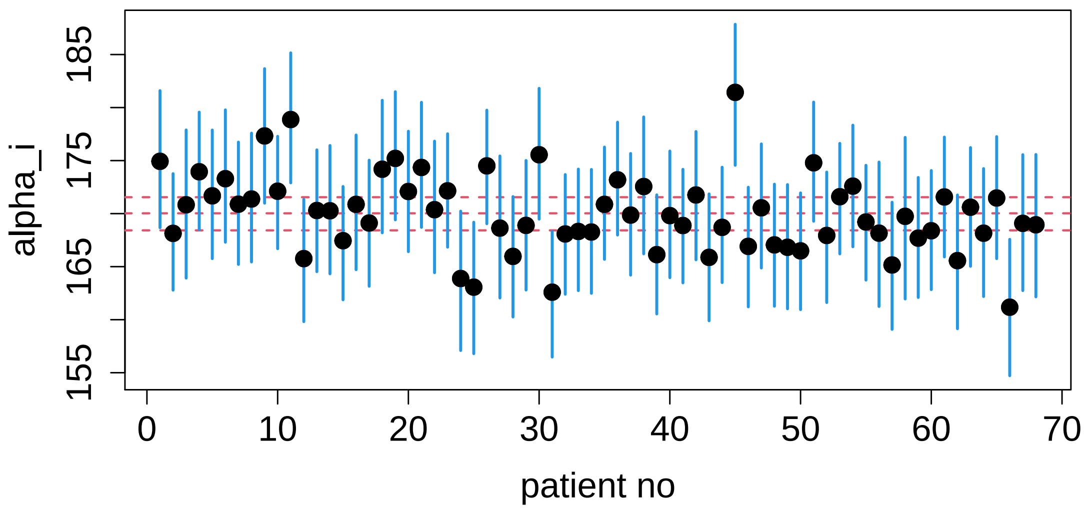
Figure 5 illustrates posterior distributions of random parameters α, β, τ_α, τ_β and σ^2. Their posterior mean estimates and 95% credible interval estimates are represented with red vertical lines. Especially, posterior mean estimates of α and β are $(\hat{\alpha}, \hat{\beta}) = (170.03, 5.09)$ with their interval estimates (168.42, 171.55) and (4.76, 5.40). It implies that a woman's weight is estimated to be 170.03 pounds at time $t = 0$ and increase 5.09 pounds a month on average. The model includes a regression model with subject specific intercept α_i and slope β_i. Their posterior estimates are shown in Figure 6.



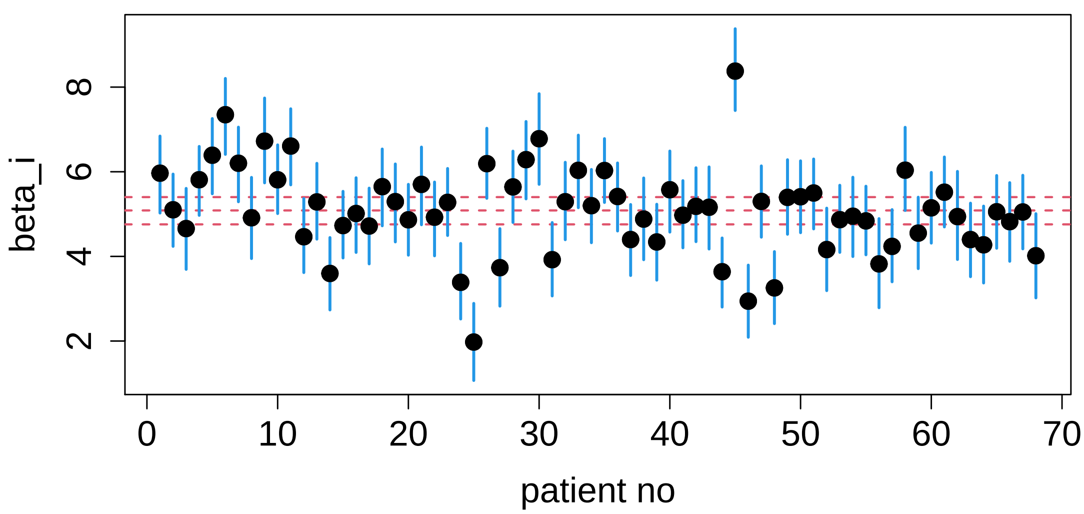
(a) α (b) β (c) τ_α

(d) τ_β (e) σ^2

Figure 5: [Souza] Posterior distributions of random parameters α, β, τ_α, τ_β and σ^2. In each plot, red vertical lines represent posterior mean estimate and 95% credible interval estimate.



(a) α_i



(b) β_i

Figure 6: [Souza] Posterior estimates of α_i and β_i in panels (a) and (b), respectively. Black dots represent posterior mean estimates, and blue vertical lines 95% credible interval estimates. Red horizontal lines are posterior point estimates of α and β with their 95% posterior credible interval estimates.