Bayesian Statistics III/IV (MATH3341/4031) Epiphany term, 2022

Draft script 1: Gibbs sampler for Univariate Normal mixture models

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Aim: To see how the concepts of: semi-conjugate priors, augmentation/imputation, Gibbs sampler MCMC methods, work together

References:

- Handouts from Michaelmas term:
  - Handout 6: Conjugate and semi-conjugate priors
  - Handout 16: Hierarchical Bayes modeling

  From: `https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/tree/main/Lecture_handouts#details-about-lecture-material`
- Lecture notes from Epiphany term (Gibbs sampler material)

Source code

- Gibbs_example_1.R
  - `https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/master/Lecture_handouts/Rscripts/Gibbs_on_Bayesian_univariate_Normal_mixture_model/Gibbs_example_1.R`
- Gibbs_example_2.R
  - `https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/master/Lecture_handouts/Rscripts/Gibbs_on_Bayesian_univariate_Normal_mixture_model/Gibbs_example_2.R`

The tasks

Consider a sequence of n observables $y = (y_1, ..., n_n)$ independently drawn from

$$
\begin{aligned}
y_i|k, \theta_{1:k} \overset{\text{ind}}{\sim} f(y_i|k, \varpi_{1:k}, \theta_{1:k}) &= \sum_{j=1}^{k} \varpi_j f_j(y_i|\theta_j)\, \mathrm{d}y_i && (1)\\
&= \sum_{j=1}^{k} \varpi_j \mathrm{N}(y_i|\mu_j, \sigma_j^2)\, \mathrm{d}y_i, \text{ for } i = 1, ..., n && (2)
\end{aligned}
$$

I need to learn the unknown quantities $(\varpi_{1:k}, \theta_{1:k})$, while the rest of the quantities are considered as known/fixed.

1. Specify the Bayesian model by choosing a prior model for $(\varpi_{1:k}, \theta_{1:k})$ that consists of semi-conjugate priors; that is

$$
\begin{cases}
\varpi_{1:k}|k & \sim \mathrm{Di}(\delta, ..., \delta) \\
\mu_j|\sigma_j^2 & \sim \mathrm{N}(\xi, \sigma_j^2/\kappa) \qquad \text{for } j = 1, ..., k \\
\sigma_j^2|\beta & \sim \mathrm{IG}(\alpha, \beta) \qquad \text{for } j = 1, ..., k
\end{cases}
$$

2. Compute the Blocks of the Gibbs sampler targeting the posterior distribution that results from Part 1.
3. Assume that a hyper-prior β is unknown. Specify a semi-conjugate hyper-prior distribution for β, compute the full conditional posterior, and re-write part 2.
4. Code in R a (systematic) Gibbs sampler based on Part 2; run the code to generate the Markov chain; and draw the trace plots of the sequence of random numbers for each unknown parameter.
5. Has the chain converged? If not why? Suggest remedies. Apply these remedies by repeating Parts 1-4.

Solution to tasks 1 & 2

The likelihood is

$$f\left(y|\varpi_{1:k},\mu_{1:k},\sigma^2_{1:k}\right)=\prod_{i=1}^{n}\sum_{j=1}^{k}\varpi_j\mathrm{N}(y_i|\mu_j,\sigma_j^2)$$

and there is no way I can factorize and specify/design any conjugate priors or even semi-conjugate priors. I cannot design a Gibbs sampler here.

I will try to extend the sampling space by employing augmentation / imputation of the statistical model such that the augmented likelihood (the likelihood of the augmented model) can be properly factorized in a manner that I can specify / derive conjugate or at least semi-conjugate priors.

Recall Handout 16:

- It is natural to regard the group label z_i for the ith observation as a latent allocation variable: then z_i is supposed to be distributed as $z_i\sim f(z_i)=\varpi_{z_i}$ for $z_i\in\{1,...,k\}$, and y_i is supposed to be distributed as $y_i|z_i,\theta_{z_i}\sim f_{z_i}(y_i|z_i,\theta_{z_i}):=\mathrm{N}(y_i|\mu_{z_i},\sigma^2_{z_i})$, for $i=1,...,n$; i.e.

$$\begin{cases}y_i|z_i,\mu_{z_i},\sigma^2_{z_i} & \sim f_j(y_i|\mu_{z_i},\sigma^2_{z_i})\\ z_i & \sim f(z_i)\end{cases}\implies\begin{cases}y_i|z_i,\mu_{z_i},\sigma^2_{z_i} & \sim \mathrm{N}(y_i|\mu_{z_i},\sigma^2_{z_i})\\ z_i & \sim f(z_i):=\varpi_{z_i}\end{cases}\tag{3}$$

as

$$\sum_{\forall z_i\in\{1,...,k\}}f\left(y_i,z_i|\mu_{1:k},\sigma^2_{1:k}\right)=\sum_{\forall z_i\in\{1,...,k\}}f\left(z_i\right)f\left(y_i|z_i,\mu_{z_i},\sigma^2_{z_i}\right)=\sum_{j=1}^{k}\varpi_j\mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)$$

So I consider augmented statistical model

$$\begin{cases}y_i|z_i,\mu_{z_i},\sigma^2_{z_i} & \sim \mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)\\ z_i & \sim f(z_i):=\varpi_{z_i}\end{cases}$$

with augmented likelihood

$$\begin{aligned}f\left(y_{1:n},z_{1:n}|\varpi_{1:k},\mu_{1:k},\sigma^2_{1:k}\right)&=\prod_{i=1}^{n}\varpi_{z_i}\mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)\\&=\prod_{j=1}^{k}\left(\varpi_j\right)^{n_j}\prod_{i=1}^{n}\mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)\end{aligned}$$

where $n_j=\sum_{i=1}^{n}1\left(z_i=j\right)$ is the number of observables $\{y_i\}$ in sub-group j.

The augmented likelihood can possibly factorized in a manner that I can compute full conjugate priors (recall the examples in Handout 6).

Full conditionals and Gibbs update for $z_i|...$

It is

$$\begin{aligned}f\left(z_i|y_{1:n},z_{-i},\varpi_{1:k},\mu_{1:k},\sigma^2_{1:k}\right)&\propto\prod_{i=1}^{n}\varpi_{z_i}\mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)\propto\varpi_{z_i}\mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)\\&\propto\frac{\varpi_{z_i}\mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)}{\sum_{\forall z_i}\varpi_{z_i}\mathrm{N}\left(y_i|\mu_{z_i},\sigma^2_{z_i}\right)}1\left(z_i\in\{1,...,k\}\right)\end{aligned}$$

So the full conditional posterior distribution for z_j is

$$\mathsf{P}\left(z_i = j | y_{1:n}, z_{-j} \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right) = \frac{\varpi_{z_i} \mathrm{N}\left(y_i | \mu_{z_i}, \sigma^2_{z_i}\right)}{\sum_{j=1}^{k} \varpi_j \mathrm{N}\left(y_i | \mu_j, \sigma^2_j\right)} 1\left(z_i \in \{1, ..., k\}\right)$$

Hence the Gibbs update for z_i is

$$z_i | ... \sim \mathsf{P}\left(z_i = j | z_{-j}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right)$$

Semi-conjugate prior, full conditionals, and Gibbs update for $\varpi_{1:k}$

To find the semi-conjugate prior for $\varpi_{1:k}$, I consider all the parameters but $\varpi_{1:k}$ as fixed and hence the likelihood kernel becomes

$$f\left(y_{1:n}, z_{1:n} | \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right) = \prod_{i=1}^{n} \varpi_{z_i} \mathrm{N}\left(y_i | \mu_{z_i}, \sigma^2_{z_i}\right) \quad \propto \prod_{j=1}^{k} \left(\varpi_j\right)^{n_j} \; = \prod_{j=1}^{k-1} \left(\varpi_j\right)^{n_j} \left(\varpi_k\right)^{n_k}$$
$$\propto \prod_{j=1}^{k-1} \left(\varpi_j\right)^{n_j} \left(1 - \sum_{j=1}^{k-1} \varpi_j\right)^{n - \sum_{j=1}^{k-1} n_j}$$

leading to a semi-conjugate prior

$$\pi\left(\varpi_{1:k}\right) \propto \prod_{j=1}^{k-1} \left(\varpi_j\right)^{\tau_j} \left(1 - \sum_{j=1}^{k-1} \varpi_j\right)^{\overbrace{\tau_0 - \sum_{j=1}^{k-1} \tau_j}^{\tau_k =}} \propto \prod_{j=1}^{k} \left(\varpi_j\right)^{\tau_j + 1 - 1}$$

hence

$$\pi\left(\varpi_{1:k}\right) = \mathrm{Di}\left(\varpi_{1:k} | \delta, ..., \delta\right)$$

So by reparametrizing and simplifying as $\tau_j \leftarrow \delta - 1$ I get a semi-conjugate prior

$$\varpi_{1:k} \sim \mathrm{Di}\left(\delta, ..., \delta\right)$$

The full conditional distribution of $\varpi_{1:k}$ given the data and the rest parameters $z_{1:n}, \mu_{1:k}, \sigma^2_{1:k}$ is, according to the Bayesian theorem

$$\begin{aligned}
\pi\left(\varpi_{1:k} | y_{1:n}, \mu_{1:k}, \sigma^2_{1:k}\right) \propto & f\left(y_{1:n}, z_{1:n} | \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right) \mathrm{Di}\left(\varpi_{1:k} | \delta, ..., \delta\right) \\
\propto & \prod_{j=1}^{k} \left(\varpi_j\right)^{n_j} \prod_{j=1}^{k} \left(\varpi_j\right)^{\delta - 1} \\
\propto & \prod_{j=1}^{k} \left(\varpi_j\right)^{n_j + \delta - 1} \\
\propto & \mathrm{Di}\left(\varpi_{1:k} | \delta + n_1, ..., \delta + n_k\right)
\end{aligned}$$

Hence the full conditional distribution of $\varpi_{1:k}$ given the data and the rest parameters is

$$\varpi_{1:k} | y_{1:n}, z_{1:n}, \mu_{1:k}, \sigma^2_{1:k} \sim \mathrm{Di}\left(\delta + n_1, ..., \delta + n_k\right)$$

Semi-conjugate prior, full conditionals, and Gibbs update for $\left(\mu_j, \sigma_j^2\right)$

To find the semi-conjugate prior for $\left(\mu_j, \sigma_j^2\right)$, I consider all the parameters but $\left(\mu_j, \sigma_j^2\right)$ as fixed and hence the likelihood kernel becomes

$$\begin{aligned}
f\left(y_{1:n}, z_{1:n}|\varpi_{1:k}, \mu_{1:k}, \sigma_{1:k}^2\right) = \prod_{i=1}^{n} \varpi_{z_i} \mathrm{N}\left(y_i|\mu_{z_i}, \sigma_{z_i}^2\right) &\propto \prod_{\forall i: z_i=j} \mathrm{N}\left(y_i|\mu_j, \sigma_j^2\right) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n_j}{2}} \exp\left(-\frac{1}{2} \sum_{\forall i: z_i=j} \frac{\left(y_i-\mu_j\right)^2}{\sigma_j^2}\right) \\
&\propto \left(\frac{1}{\sigma_j^2}\right)^{\frac{n_j}{2}} \exp\left(-\frac{1}{2}\frac{n_j}{\sigma_j^2}\mu_j^2\right) \exp\left(-\frac{1}{2}\frac{n_j}{\sigma_j^2}y_i^2 + \frac{n_j}{\sigma_j^2}\mu_j y_i\right)
\end{aligned}$$

leading to a semi-conjugate prior

$$\begin{aligned}
\pi\left(\mu_j, \sigma_j^2\right) &\propto \left(\sqrt{\frac{1}{\sigma^2}} \exp\left(-\frac{1}{2}\frac{1}{\sigma^2}\mu^2\right)\right)^{\tau_0} \exp\left(\mu\frac{1}{\sigma^2}\tau_1 - \frac{1}{2}\frac{1}{\sigma^2}\tau_2\right) \\
&\propto \underbrace{\left(\frac{1}{\sigma^2/\tau_0}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{1}{\sigma^2/\tau_0}(\mu - \frac{\tau_1}{\tau_0})^2\right)}_{\propto \mathrm{N}(\mu|\frac{\tau_1}{\tau_0}, \frac{\sigma^2}{\tau_0})} \underbrace{(\frac{1}{\sigma^2})^{\frac{(\tau_0-3)}{2}+1} \exp\left(-\frac{1}{\sigma^2}\frac{1}{2}(\tau_2 - \frac{\tau_1^2}{\tau_0})\right)}_{\propto \mathrm{IG}(\sigma^2|\frac{\tau_0-3}{2}, \frac{1}{2}(\tau_2-\frac{\tau_1^2}{\tau_0}))}
\end{aligned}$$

So by reparametrizing the fixed parameters in a more convenient manner I get

$$\pi\left(\mu_j, \sigma_j^2\right) = \mathrm{N}\left(\mu_j|\xi, \frac{\sigma^2}{\kappa}\right) \mathrm{IG}\left(\sigma_j^2|\alpha, \beta\right)$$

That is

$$\begin{cases}
\mu_j|\sigma_j^2 & \sim \mathrm{N}\left(\xi, \frac{\sigma^2}{\kappa}\right) \qquad \text{for } j = 1, ..., k \\
\sigma_j^2 & \sim \mathrm{IG}\left(\sigma_j^2|\alpha, \beta\right) \qquad \text{for } j = 1, ..., k
\end{cases}$$

The full conditional distribution of $\left(\mu_j, \sigma_j^2\right)$ given the data and the rest parameters $z_{1:n}, \varpi_{1:k}, \mu_{-j}, \sigma_{-j}^2$ is, according to the Bayesian theorem

$$\begin{aligned}
\pi\left(\mu_j, \sigma_j^2|y_{1:n}, z_{1:n}\varpi_{1:k}, \mu_{-j}, \sigma_{-j}^2\right) \propto & f\left(y_{1:n}, z_{1:n}|\varpi_{1:k}, \mu_{1:k}, \sigma_{1:k}^2\right) \pi\left(\mu_j, \sigma_j^2\right) \\
\propto & \prod_{\forall i: z_i=j} \mathrm{N}\left(y_i|\mu_j, \sigma_j^2\right) \mathrm{N}\left(\mu_j|\xi, \frac{\sigma^2}{\kappa}\right) \mathrm{IG}\left(\sigma_j^2|\alpha, \beta\right) \\
\propto & \left(\frac{1}{\sigma_j^2}\right)^{\frac{n_j}{2}} \exp\left(-\frac{1}{2} \sum_{\forall i: z_i=j} \frac{\left(\mu_j - y_i\right)^2}{\sigma_j^2}\right) \\
& \times \left(\frac{1}{\sigma_j^2/\kappa}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{1}{\sigma_j^2/\kappa}\left(\mu_j - \xi\right)^2\right) \left(\frac{1}{\sigma_j^2}\right)^{\alpha+1} \exp\left(-\beta\frac{1}{\sigma_j^2}\right)
\end{aligned}$$

where

$$
\begin{aligned}
\pi\left(\mu_j|y_{1:n},\sigma_j^2,...\right) &\propto \exp\left(-\frac{1}{2}\sum_{\forall i:z_i=j}\frac{(\mu_j-y_i)^2}{\sigma_j^2}\right)\exp\left(-\frac{1}{2}\frac{1}{\sigma_j^2/\kappa}(\mu_j-\xi)^2\right)\\
&\propto \exp\left(-\frac{1}{2}\frac{\left(\mu_j-\frac{\sum_{i:z_i=j}y_i-\xi\kappa}{n_j+\kappa}\right)^2}{\frac{\sigma_j^2}{n_j+\kappa}}\right)\\
&\propto \mathrm{N}\left(\mu_j|\frac{\sum_{i:z_i=j}y_i-\xi\kappa}{n_j+\kappa},\frac{\sigma_j^2}{n_j+\kappa}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\pi\left(\sigma_j^2|y_{1:n},...\right) \propto& \int \pi\left(\mu_j,\sigma_j^2|y_{1:n},z_{1:n},\varpi_{1:k},\mu_{-j},\sigma_{-j}^2\right)\mathrm{d}\mu_j\\
\propto& \int \exp\left(-\frac{1}{2}\sum_{\forall i:z_i=j}\frac{(\mu_j-y_i)^2}{\sigma_j^2}-\frac{1}{2}\frac{1}{\sigma_j^2/\kappa}(\mu_j-\xi)^2\right)\mathrm{d}\mu_j\\
&\times\left(\frac{1}{\sigma_j^2}\right)^{\frac{n_j}{2}}\left(\frac{1}{\sigma_j^2/\kappa}\right)^{\frac{1}{2}}\left(\frac{1}{\sigma_j^2}\right)^{\alpha+1}\exp\left(-\beta\frac{1}{\sigma_j^2}\right)\\
\propto& \exp\left(-\frac{1}{\sigma_j^2}\frac{1}{2}\sum_{i:z_i=j}(y_i-\mu_j)^2\right)\\
&\times\left(\frac{1}{\sigma_j^2}\right)^{\frac{n_j}{2}+\frac{1}{2}+\alpha+1}\exp\left(-\beta\frac{1}{\sigma_j^2}\right)\\
\propto& \left(\frac{1}{\sigma_j^2}\right)^{\frac{n_j}{2}+\frac{1}{2}+\alpha+1}\exp\left(-\frac{1}{\sigma_j^2}\left[\beta+\frac{1}{2}\sum_{i:z_i=j}(y_i-\mu_j)^2\right]\right)\\
\propto& \mathrm{IG}\left(\sigma_j^2|\alpha+\frac{n_j}{2},\beta+\frac{1}{2}\sum_{i:z_i=j}(y_i-\mu_j)^2\right)
\end{aligned}
$$

Hence the full conditional distribution of $\left(\mu_j,\sigma_j^2\right)$ given the data and the rest parameters is

$$
\begin{cases}
\mu_j|\sigma_j^2 & \sim \mathrm{N}\left(\frac{\sum_{i:z_i=j}y_i-\xi\kappa}{n_j+\kappa},\frac{\sigma_j^2}{n_j+\kappa}\right)\\
\sigma_j^2 & \sim \mathrm{IG}\left(\sigma_j^2|\alpha+\frac{n_j}{2},\beta+\frac{1}{2}\sum_{i:z_i=j}(y_i-\mu_j)^2\right)
\end{cases}
$$

To sum-up, considering the above semi-conjugate priors, the Bayesian model is

$$
\begin{cases}
y_i|z_i,\varpi_{1:k},\mu_{1:k},\sigma_{1:k}^2 & \sim f_{z_i}(y_i|\varpi_{z_i},\mu_{z_i},\sigma_{z_i}^2) \qquad \text{for } i=1,...,n\\
z_i & \sim f(z_i):=\varpi_{z_i} \qquad \text{for } i=1,...,n\\
\varpi_{1:k} & \sim \mathrm{Di}(\delta,...,\delta)\\
\mu_j|\sigma_j^2 & \sim \mathrm{N}(\xi,\sigma_j^2/\kappa) \qquad \text{for } j=1,...,k\\
\sigma_j^2|\beta & \sim \mathrm{IG}(\alpha,\beta) \qquad \text{for } j=1,...,k
\end{cases}
$$

And the full conditional posterior distributions for the Gibbs sampler are

$$\begin{aligned}
\varpi_{1:k}|y_{1:n}, z_{1:n}, \mu_{1:k}, \sigma^2_{1:k}, \beta &\sim \text{Di}\left(\delta + n_1, ..., \delta + n_k\right) \\
\mu_j|y_{1:n}, z_{1:n}, \varpi_{1:k}, \sigma^2_{1:k}, \beta &\sim \text{N}\left(\frac{\sum_{i:z_i=j} y_i - \xi\kappa}{n_j + \kappa}, \frac{\sigma_j^2}{n_j+\kappa}\right), \text{ for } j = 1, ..., k \\
\sigma_j^2|y_{1:n}, \varpi_{1:k}, \sigma^2_{-j}, \beta &\sim \text{IG}\left(a + \frac{n_j}{2}, \beta + \frac{1}{2}\sum_{i:z_i=j}(y_i - \mu_j)^2\right), \text{ for } j = 1, ..., k \\
z_i|y_{1:n}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}, \beta &\sim \pi(z_i = j|y...) = \frac{\frac{w_j}{\sigma_j}\exp\left(-\frac{1}{2}\frac{(y_i-\mu_j)^2}{\sigma_j^2}\right)}{\sum_{j'=1}^k \frac{w_{j'}}{\sigma_{j'}}\exp\left(-\frac{1}{2}\frac{(y_i-\mu_{j'})^2}{\sigma^2_{j'}}\right)}; \text{ for } i = 1, ..., n
\end{aligned}$$

Solution to task 3

Well, here the prior $\sigma_j^2|\beta \sim \text{IG}(\alpha, \beta) \quad$ for $j = 1, ..., k$ acts as a 'sampling distribution' for us to derive the 'likelihood' and hence find the semi-conjugate prior for β. So

$$\pi\left(\sigma^2_{1:k}|\alpha, \beta\right) \propto \prod_{j=1}^{k} \beta^\alpha \exp\left(-\frac{\beta}{\sigma_j^2}\right) \propto \beta^{k\alpha}\exp\left(-\beta\frac{1}{\prod_{j=1}^k \sigma_j^2}\right)$$

leading to a hyper-prior

$$\pi\left(\beta\right) \propto \beta^{\tau_0+1-1}\exp\left(-\tau_1\beta\right) \propto \text{G}\left(\beta|\tau_0+1, \tau_1\right)$$

so by reparametrizing as $g \leftarrow \tau_0 + 1$ and $h \leftarrow \tau_1$ it is

$$\beta \sim \text{G}\left(g, h\right)$$

According to the Bayes theorem the full conditional posterior of β given the data $y_{1:n}$ and all the rest parameters is

$$\pi\left(\beta|y_{1:n}, ...\right) \propto \prod_{j=1}^{k}\beta^\alpha\exp\left(-\frac{\beta}{\sigma_j^2}\right)\beta^g\exp\left(-\beta h\right)$$

Hence

$$\beta|y_{1:n}, ... \sim \text{Ga}\left(g + k\alpha, h + \sum_{j=1}^{k}\frac{1}{\sigma_j^2}\right)$$

To sum-up, considering the above semi-conjugate priors, the Bayesian model is

$$\begin{cases}
y_i|z_i, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k} & \sim f_{z_i}(y_i|\varpi_{z_i}, \mu_{z_i}, \sigma^2_{z_i}) \qquad \text{for } i = 1, ..., n \\
z_i & \sim f(z_i) := \varpi_{z_i} \qquad \text{for } i = 1, ..., n \\
\varpi_{1:k} & \sim \text{Di}(\delta, ..., \delta) \\
\mu_j|\sigma_j^2 & \sim \text{N}(\xi, \sigma_j^2/\kappa) \qquad \text{for } j = 1, ..., k \\
\sigma_j^2|\beta & \sim \text{IG}(\alpha, \beta) \qquad \text{for } j = 1, ..., k \\
\beta & \sim \text{G}\left(g, h\right)
\end{cases}$$

And the full conditional posterior distributions for the Gibbs sampler are

$$\varpi_{1:k}|y_{1:n}, z_{1:n}, \mu_{1:k}, \sigma^2_{1:k}, \beta \sim \text{Di}\left(\delta + n_1, ..., \delta + n_k\right)$$

$$\mu_j|y_{1:n}, z_{1:n}, \varpi_{1:k}, \sigma^2_{1:k}, \beta \sim \text{N}\left(\frac{\sum_{i:z_i=j} y_i - \xi\kappa}{n_j + \kappa}, \frac{\sigma_j^2}{n_j + \kappa}\right), \text{ for } j = 1, ..., k$$

$$\sigma_j^2|y_{1:n}, \varpi_{1:k}, \sigma^2_{-j}, \beta \sim \text{IG}\left(a + \frac{n_j}{2}, \beta + \frac{1}{2}\sum_{i:z_i=j}(y_i - \mu_j)^2\right), \text{ for } j = 1, ..., k$$

$$z_i|y_{1:n}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}, \beta \sim \pi(z_i = j|y...) = \frac{\frac{w_j}{\sigma_j}\exp\left(-\frac{1}{2}\frac{(y_i-\mu_j)^2}{\sigma_j^2}\right)}{\sum_{j'=1}^{k}\frac{w_{j'}}{\sigma_{j'}}\exp\left(-\frac{1}{2}\frac{(y_i-\mu_{j'})^2}{\sigma_{j'}^2}\right)}; \text{ for } i = 1, ..., n$$

$$\beta|y_{1:n}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k} \sim \text{Ga}\left(g + k\alpha, h + \sum_{j=1}^{k}\frac{1}{\sigma_j^2}\right)$$

Solution to task 4

See the R script Gibbs_example_1.R

- `https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/master/Lecture_handouts/Rscripts/Gibbs_on_Bayesian_univariate_Normal_mixture_model/Gibbs_example_1.R`

Solution to task 5

No because it has not explored all the sampling space with positive mass. Due to the non-differentiability of the mixture model under consideration, I would expect the marginal posteriors of each unknown parameter to have k modes.

For instance in the produced plots of μ_1, I see that the chain has explored only one mode.

Remedies:

- Include more blocks, such as Metropolis random walk updates with Normal proposal distributions of large variance with purpose to be able to pass the zero-mass barrier. This is difficult to work due to the large sampling space.
- Include a Metropolis-Hastings proposal targeting the joint posterior $\pi\left(z_{1:n}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}|y_{1:n}\right)$ with a proposal distribution $q\left(\mathfrak{p}\right) = \frac{1}{k!}$ that proposes a random permutation of the labels of the mixture components. Then:

  At state $\left(z_{1:n}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right)$

  1. draw $\mathfrak{p}(1:k) \sim q\left(\cdot\right)$
  2. Accept $\left(\mathfrak{p}(z_{1:n}), \varpi_{\mathfrak{p}(1:k)}, \mu_{\mathfrak{p}(1:k)}, \sigma^2_{\mathfrak{p}(1:k)}\right)$ as the next state with probability $a = \min\left(1, r\right)$ otherwise reject

However it is

$$
\begin{aligned}
r &= \frac{f\left(y_{1:n}, \mathfrak{p}(z_{1:n})|\varpi_{\mathfrak{p}(1:k)}, \mu_{\mathfrak{p}(1:k)}, \sigma^2_{\mathfrak{p}(1:k)}\right)\pi\left(\varpi_{\mathfrak{p}(1:k)}\right)\pi\left(\mu_{\mathfrak{p}(1:k)}, \sigma^2_{\mathfrak{p}(1:k)}\right)\pi\left(\beta\right)}{f\left(y_{1:n}, z_{1:n}|\varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right)\pi\left(\varpi_{1:k}\right)\pi\left(\mu_{1:k}, \sigma^2_{1:k}\right)\pi\left(\beta\right)} \frac{q\left(1:k\right)}{q\left(\mathfrak{p}(1:k)\right)} \\
&= \frac{f\left(y_{1:n}, \mathfrak{p}(z_{1:n})|\varpi_{\mathfrak{p}(1:k)}, \mu_{\mathfrak{p}(1:k)}, \sigma^2_{\mathfrak{p}(1:k)}\right)\pi\left(\varpi_{\mathfrak{p}(1:k)}\right)\pi\left(\mu_{\mathfrak{p}(1:k)}, \sigma^2_{\mathfrak{p}(1:k)}\right)\pi\left(\beta\right)}{f\left(y_{1:n}, z_{1:n}|\varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right)\pi\left(\varpi_{1:k}\right)\pi\left(\mu_{1:k}, \sigma^2_{1:k}\right)\pi\left(\beta\right)} = 1
\end{aligned}
$$

because the mixture is invariant to component permutations. So this is equivalent to just randomly permuting the components at each Gibbs iteration with probability 1.

So Gibbs:

$$
\begin{aligned}
\varpi_{1:k}|y_{1:n}, z_{1:n}, \mu_{1:k}, \sigma^2_{1:k}, \beta &\sim \text{Di}\left(\delta + n_1, ..., \delta + n_k\right) \\
\mu_j|y_{1:n}, z_{1:n}, \varpi_{1:k}, \sigma^2_{1:k}, \beta &\sim \text{N}\left(\frac{\sum_{i:z_i=j} y_i - \xi\kappa}{n_j + \kappa}, \frac{\sigma_j^2}{n_j + \kappa}\right), \text{ for } j = 1, ..., k \\
\sigma_j^2|y_{1:n}, \varpi_{1:k}, \sigma^2_{-j}, \beta &\sim \text{IG}\left(a + \frac{n_j}{2}, \beta + \frac{1}{2}\sum_{i:z_i=j}(y_i - \mu_j)^2\right), \text{ for } j = 1, ..., k \\
z_i|y_{1:n}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}, \beta &\sim \pi(z_i = j|y...) = \frac{\frac{w_j}{\sigma_j}\exp\left(-\frac{1}{2}\frac{(y_i - \mu_j)^2}{\sigma_j^2}\right)}{\sum_{j'=1}^k \frac{w_{j'}}{\sigma_{j'}}\exp\left(-\frac{1}{2}\frac{(y_i - \mu_{j'})^2}{\sigma_{j'}^2}\right)}; \text{ for } i = 1, ..., n \\
\beta|y_{1:n}, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k} &\sim \text{Ga}\left(g + k\alpha, h + \sum_{j=1}^k \frac{1}{\sigma_j^2}\right) \\
1:k &\leftarrow \mathfrak{p}(1:k), \text{ where } \mathfrak{p}(.) \text{ is a random permutation of the components}
\end{aligned}
$$

See the R script Gibbs_example_2.R

- `https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/master/Lecture_handouts/Rscripts/Gibbs_on_Bayesian_univariate_Normal_mixture_model/Gibbs_example_2.R`

Formulas

$$-\frac{1}{2}\sum_{i=1}^{n}\frac{(x-\mu_i)^2}{\sigma_i^2} = -\frac{1}{2}\frac{(x-\hat{\mu})^2}{\hat{\sigma}^2} + C(\hat{\mu},\hat{\sigma}^2)$$

$$\hat{\sigma}^2 = (\sum_{i=1}^{n}\frac{1}{\sigma_i^2})^{-1}; \qquad \hat{\mu} = \hat{\sigma}^2(\sum_{i=1}^{n}\frac{\mu_i}{\sigma_i^2}); \qquad C(\hat{\mu},\hat{\sigma}^2) = \underbrace{\frac{1}{2}\frac{(\sum_{i=1}^{n}\frac{\mu_i}{\sigma_i^2})^2}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}} - \frac{1}{2}\sum_{i=1}^{n}\frac{\mu_i^2}{\sigma_i^2}}_{=\text{independent of } x}$$

$\mathrm{Di}_k(a)$ denotes the Dirichlet distribution with PDF

$$\mathrm{Di}_k(\theta|a) = \begin{cases} \frac{\Gamma(\sum_{j=1}^{k} a_j)}{\prod_{j=1}^{k}\Gamma(a_j)}\prod_{j=1}^{k}\theta_j^{a_j-1} & \text{, if } \theta \in \Theta \\ 0 & \text{, otherwise} \end{cases}$$

$\mathrm{Ga}(\alpha,\beta)$ is the Gamma distribution with shape and rate parameters α and β, and PDF

$$f_{\mathrm{Ga}(\alpha,\beta)}(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}1(x>0)$$

The inverse Gamma distr.: $x \sim \mathrm{IG}(a,b)$ has pdf

$$f(x) = \frac{b^a}{\Gamma(a)}x^{-a-1}\exp(-\frac{b}{x})1_{(0,+\infty)}(x)$$