Bayesian Statistics III/IV (MATH3341/4031) Michaelmas term, 2021

Handout 16: Hierarchical Bayes modeling

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

Aim

To be able to specify and analyze a Hierarchical Bayesian, as well as to extend previously introduces concepts in the Hierarchical Bayes framework.

Basic reading list:

- Berger, J. O. (2013; Section 4.6). Statistical decision theory and Bayesian analysis. Springer.
- Robert, C. (2007, Sections 10.1-10.3). The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media.
- Robert, C. P., & Reber, A. (1998). Bayesian modelling of a pharmaceutical experiment with heterogeneous responses. Sankhyā: The Indian Journal of Statistics, Series B, 145-160.

R scripts:

Bayesian Normal Mixture model (Appendix A): `http://htmlpreview.github.io/?https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/main/Computer_practical/Normal_Mixture_model/Bayesian_Normal_Mixture_Model.nb.html`

Bayesian Variable Selection in Logistic regression model (Appendix B): `http://htmlpreview.github.io/?https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/main/Computer_practical/Bernoulli_regression_model_variable_selection/Bernoulli_Regression_Model_VS_full.nb.html`

Random effect model (Appendix C): `https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/main/Lecture_handouts/Rscripts/Hierarchical_bayes/HierarchicalBayesPharmaceutical.R`

1 Hierarchical Bayesian Model

A hierarchical Bayesian model involves several levels of conditional distributions. It can be hierarchical due to the sampling distribution modeling or due to the decomposition of the prior information.

Definition 1. A hierarchical Bayes model is a Bayesian statistical model with sampling distribution $y \sim f(y|\theta)$ and prior $\theta \sim \pi(\theta)$, where the prior distribution $\pi(\theta)$ is decomposed in conditional distributions. The Bayesian model is

$$\begin{cases} y|\theta \sim f(y|\theta), \text{ sampling distribution} & \begin{cases} y|z,\theta \sim f(y|z,\theta) \\ z|\theta \sim f(z|\theta) \end{cases} \\ \theta \sim \pi(\theta) \text{ marginal prior specified} \overset{\text{extend space}}{\Longrightarrow} & \begin{cases} \theta|\phi_1 \sim \pi_1(\theta|\phi_1) & \text{1st level prior} \\ \phi_1|\phi_2 \sim \pi_2(\phi_1|\phi_2) & \text{2nd level hyper-prior} \\ \vdots & \\ \phi_j|\phi_{j+1} \sim \pi_{j+1}(\phi_j|\phi_{j+1}) & j\text{th level hyper-prior} \\ \vdots & \\ \phi_{m-1}|\phi_m \sim \pi_m(\phi_{m-1}|\phi_m) & m\text{th level hyper-prior} \end{cases} \end{cases} \quad (1)$$

Remark 2. The joint distribution $p(y, z, \theta, \phi_1, ..., \phi_j, ...\phi_{m-1})$ has pdf

$$p(y, z, \theta, \phi_1, ..., \phi_j, ...\phi_{m-1}) = f(y|z, \theta)f(z|\theta)\pi_1(\theta|\phi_1)\pi_2(\phi_1|\phi_2)\pi_3(\phi_2|\phi_3)...\pi(\phi_{m-1}|\phi_m)$$

Remark 3. Hierarchical Bayesian model is simply a special type of Bayesian model, where

$$\begin{cases} y|\theta & \sim f(y|\theta) \\ \theta|\phi & \sim \pi(\theta|\phi) \\ \phi|\phi_m & \sim \pi(\phi|\phi_m) \end{cases} \tag{2}$$

for $\phi = (\phi_1, ..., \phi_{m-1})$, and ϕ_m fixed hyper-parameter.

Remark 4. The Bayesian model with sampling distribution $y \sim f(y|\theta)$ and prior $\theta \sim \pi(\theta)$, can be recovered from (2) by marginalizing the prior as

$$\begin{aligned} f(y|z, \theta) &= \int_\Phi f(y|z, \theta)f(z|\theta)\mathrm{d}z \\ \pi(\theta|\phi_m) &= \int_\Phi \pi(\theta|\phi)\pi(\phi|\phi_m)\mathrm{d}\phi = \int_{\Phi_1 \times \Phi_{m-1}} \pi(\theta|\phi_1)\pi(\phi_1|\phi_2)\mathrm{d}\phi_1...\pi(\phi_{m-1}|\phi_m)\mathrm{d}\phi_{m-1}, \end{aligned} \tag{3}$$

where ϕ_m is just a fixed hyper-parameter. Hence, hierarchical models are indeed included in the Bayesian paradigm.

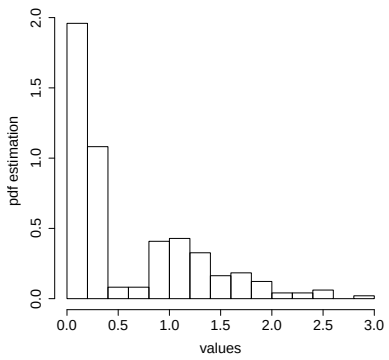
Note 5. A hierarchical Bayesian model can be used as a mean to specify more diverse priors. This is achieved by setting ϕ to be a random hyper-parameter with $\phi|\phi_m \sim \pi_2(\phi|\phi_m)$ instead of setting ϕ to have a fixed value.

Note 6. A hierarchical Bayesian model can be used when the sampling distribution or the prior distributions justify a certain structure.

Example 7. [Normal mixture model] Consider the following application where our concern is the distribution of enzymatic activity in the blood, for an enzyme involved in the metabolism of carcinogenic substances, among a group of $n = 245$ unrelated individuals; aka cluster analysis.

Our observables are $y = (y_1, ..., y_n)$ with $n = 245$. In the Boxplot on the right, we can clearly see that the distribution is multimodal, suggesting the existence of subpopulations/groups.

Interest lies on identifying subgroups of slow or fast metabolizers as a marker of genetic polymorphism in the general population. As we are interested in learning/identifying the sub-populations as the identity/label of the group from which each observation is drawn is unknown.



Histogram of Enzyme dataset which is available from: `https://people.maths.bris.ac.uk/~mapjg/mixdata`

Questions of interest:

- How many groups exist?
- To which group each observartion belongs?

As for the sampling model, we can assume that the i-th observation y_i is randomly drawn from the j-th group which has proportion ϖ_j in the population and which is distributed according to the sampling distribution $y_i|\theta_j \sim f_j(y_i|\theta_j)$.

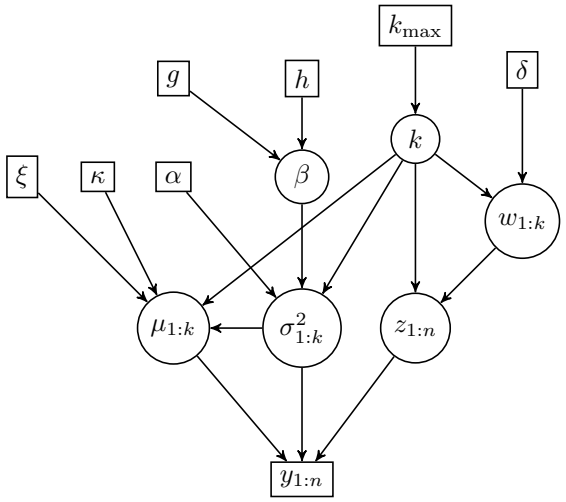
For simplicity, let's assume that all groups are Normally distributed but with different parameter values $\{\theta_j\}$; hence j-th group is $y_i|\mu_j, \sigma_j^2 \sim \mathrm{N}(y_i|\mu_j, \sigma_j^2)$ with $\theta_j = (\mu_j, \sigma_j^2)$.

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.

$$\begin{cases} y_i|z_i, \mu_{z_i}, \sigma^2_{z_i} & \sim f_j(y_i|\mu_{z_i}, \sigma^2_{z_i}) \\ z_i & \sim f(z_i) \end{cases} \implies \begin{cases} y_i|z_i, \mu_{z_i}, \sigma^2_{z_i} & \sim \mathrm{N}(y_i|\mu_{z_i}, \sigma^2_{z_i}) \\ z_i & \sim f(z_i) := \varpi_{z_i} \end{cases} \tag{4}$$

To complete the Bayesian model, we specify priors on the unknown quantities: Given there are k groups, $\varpi_{1:k} \sim \mathrm{Di}(\delta)$ for the group proportions, $\mu_j \sim \mathrm{N}(\xi_j, \sigma^2_j/\kappa)$ for the mean, and $\sigma^2_j \sim \mathrm{Ga}(\alpha, \beta)$ for the variances. Assume we wish a more spread prior for σ^2_j (for some reason), and hence we specify a hyper-prior on β as $\beta \sim \mathrm{Ga}(g, h)$. As the number of the groups is unknown, we assign prior $k \sim \pi(k) \in \mathrm{U}_{\text{discr}}(1, k_{\max})$.

$$\begin{cases} y_i|z_i, k, \varpi, \mu, \sigma^2 & \sim f_{z_i}(y_i|\varpi_{z_i}, \mu_{z_i}, \sigma^2_{z_i}) \\ & \quad \text{for } i = 1, ..., n \\ z_i|k & \sim f(z_i) := \varpi_{z_i} \\ & \quad \text{for } i = 1, ..., n \\ \varpi|k & \sim \mathrm{Di}(\delta, ..., \delta) \\ \mu_j|\sigma^2_j, k & \sim \mathrm{N}(\xi, \sigma^2_j/\kappa) \\ & \quad \text{for } j = 1, ..., k \\ \sigma^2_j|\beta, k & \sim \mathrm{IG}(\alpha, \beta) \\ & \quad \text{for } j = 1, ..., k \\ \beta & \sim \mathrm{Ga}(g, h) \\ k & \sim \mathrm{U}_{\text{discr}}(1, k_{\max}) \end{cases} \tag{5}$$



A Directed Acyclic Graph representation of the hierarchical model (5). Circles correspond to random quantities. Squares correspond to fixed quantities.

The joint distribution has pdf

$$p(y_{1:n}, z_{1:n}, k, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}) = \overbrace{\prod_{i=1}^{n} \mathrm{N}(y_i|\mu_{z_i}, \sigma^2_{z_i})}^{f(y_{1:n}|z_{1:n}, \mu_{1:k}, \sigma^2_{1:k})} \overbrace{\prod_{i=1}^{n} \varpi_{z_i}}^{f(z_{1:n}|k)} \overbrace{\mathrm{Di}(\varpi_{1:k}|\delta)}^{\pi(\varpi_{1:k}|k)} \overbrace{\prod_{j=1}^{k} \mathrm{N}(\mu_k|\xi_k, \sigma^2_k/\kappa)}^{\pi(\mu_{1:k}|\sigma^2_{1:k}, k)} \overbrace{\prod_{j=1}^{k} \mathrm{Ga}(\sigma^2_k|\alpha, \beta)}^{\pi(\sigma^2_{1:k}|\beta, k)} \overbrace{\mathrm{Ga}(\beta|g, h)}^{\pi(\beta)} \overbrace{\frac{1}{|k_{\max}|}}^{\pi(k)}$$

Appendix A

The joint posterior $\pi(k, \varpi, \mu, \sigma^2, z|y)$ can be computed with the Bayesian theorem, and factorized as

$$\begin{aligned} \pi(k, \varpi, \mu, \sigma^2, \beta, z|y) &= \frac{p(y, z, k, \varpi, \mu, \sigma^2, \beta)}{\int p(y, z, k, \varpi, \mu, \sigma^2, \beta) \mathrm{d}(z, k, \varpi, \mu, \sigma^2, \beta)} && (6) \\ &= \pi(z|y, k, \varpi, \mu, \sigma^2) \pi(\mu_{1:k}, \sigma^2_{1:k}|y, k) \pi(\varpi_{1:k}|y, k) \pi(k|y) && (7) \end{aligned}$$

where to infer the number of groups from $\pi(k|y)$, the proportions in each group from $\pi(\varpi_{1:k}|y, k)$, the moments of each group from $\pi(\mu_{1:k}, \sigma^2_{1:k}|y, k)$, and the allocation of each observation to each group with $\pi(z|y, k, \varpi, \mu, \sigma^2)$.

As the required integrals are intractable, we can resolve to numerical methods, etc... Monte Carlo e.g. via JAGS....

Note 8. A particularly appealing aspect of hierarchical models is that they allow for conditioning on all levels, and this easy decomposition of the posterior. Consider the Bayesian hierarchical model (2) a parametric model $f(y|\theta)$ with a

hierarchical prior $\theta \sim \pi_1(\theta|\phi)$, and $\phi \sim \pi(\phi)$. The posterior distribution of θ is

$$\pi(\theta|y) = \int_{\Phi} \pi(\theta|y,\phi)\pi(\phi|y)\mathrm{d}\phi \tag{8}$$

where

$$\pi(\theta|y,\phi) = \frac{f(y|\theta)\pi_1(\theta|\phi)}{f_1(y|\phi)}; \qquad \pi(\phi|y) = \frac{f_1(y|\phi)\pi_2(\phi)}{f(y)};$$
$$f_1(y|\phi) = \int_{\Theta} f(y|\theta)\pi_1(\theta|\phi)\mathrm{d}\theta; \qquad f(y) = \int_{\Theta} f_1(y|\phi)\pi_2(\phi)\mathrm{d}\phi$$

Remark 9. Note 8 has important consequences in terms of the computation of Bayes estimators, since it shows that $\pi(\theta|y)$ can be simulated by generating, first, ϕ from $\pi(\phi|y)$ and then θ from $\pi(\theta|y,\phi)$, if these two conditional distributions are easier to work with. (Snapshot from Term 2).

Note 10. Hierarchical decomposition (2) may facilitate the computation of intractable posterior moments. Let h be a function $h : \Theta \to \mathbb{R}$, then

$$\mathrm{E}_{\pi}(h(\theta)|y) = \mathrm{E}_{\pi}\left(\mathrm{E}_{\pi}\left(h(\theta)|y,\phi\right)\,|y\right).$$

If $\mathrm{E}_{\pi}(h(\theta)|y) = \int h(\theta)\pi(\theta|y)\mathrm{d}\theta$ is intractable and θ has high dimensionality, one could possibly try to specify the prior decomposition $\pi(\theta) = \int_{\Phi} \pi_1(\theta|\phi)\pi_2(\phi|\phi_m)\mathrm{d}\phi$ in (3) such that $\mathrm{E}_{\pi}\left(h(\theta)|y,\phi\right)$ can be computed analytically, and ϕ has low dimensionality. In that case one would have to compute the equivalent but lower dimensional (and hence easier) integral $\mathrm{E}_{\pi}\left(\mathrm{E}_{\pi}\left(h(\theta)|y,\phi\right)\,|y\right) = \int \mathrm{E}_{\pi}\left(h(\theta)|y,\phi\right)\pi(\phi|y)\mathrm{d}\phi$.

Example 11. (Cont. Example 7) From another point of view, recall that the compound distribution $f(y_i|k,\varpi_{1:k}\theta_{1:k})$ of (4) is mixture model of distribution

$$y_i|k,\theta_{1:k} \sim f(y_i|k,\varpi_{1:k},\theta_{1:k}) \;= \sum_{j=1}^{k} \varpi_j f_j(y_i|\theta_j) \;= \sum_{j=1}^{k} \varpi_j \mathrm{N}(y_i|\mu_j,\sigma_j^2),\ \text{for}\, i = 1,...,n \tag{9}$$

is a suitable sampling distribution for modeling heterogeneous populations. Then by marginalizing, we can get the equivalent model

$$\begin{cases} y_i|k,\varpi,\mu,\sigma^2 & \sim f(y_i|k,\varpi,\mu,\sigma^2) := \sum_{j=1}^{k} \varpi_j \mathrm{N}(y_i|\mu_j,\sigma_j^2),\ \text{for}\, i = 1,...,n \\ \varpi|k & \sim \mathrm{Di}(\delta) \\ \mu_j|\sigma_j^2,k & \sim \mathrm{N}(\mu_j|\xi,\sigma_j^2),\ \text{for}\, j = 1,...,k \\ \sigma_j^2|k & \sim \mathrm{Ga}(a,\beta),\ \text{for}\, j = 1,...,k \\ \beta & \sim \mathrm{Ga}(g,h) \\ k & \sim \mathrm{U}_{\mathrm{discr}}(1,k_{\max}) \end{cases} \tag{10}$$

The joint distribution that admits pdf

$$p(y,k,\varpi,\mu,\sigma^2,\beta) = f(y|k,\varpi,\mu,\sigma^2)\pi(\varpi|k)\pi(\mu|\sigma^2,k)\pi(\sigma^2|k,\beta)\pi(\beta)\pi(k).$$

The posterior $\pi(k,\varpi,\mu,\sigma^2|y)$ can be computed with the Bayesian theorem, and factorized as

$$\pi(k,\varpi,\mu,\sigma^2,\beta|y) \quad = \frac{p(y,\varpi,\mu,\sigma^2,\beta)}{\int p(y,k,\varpi,\mu,\sigma^2,\beta)\mathrm{d}(k,\varpi,\mu,\sigma^2,\beta)} \tag{11}$$
$$= \pi(\mu_{1:k},\sigma^2_{1:k}|y,k)\pi(\varpi_{1:k}|y,k)\pi(k|y) \tag{12}$$

Models (5) and (10) are the same model in the sense that posterior (11) is the marginal of the posterior (6).

Example 12. [Variable selection in logistic regression] Consider the 'Challenger O-ring' example from the Computer practicals. Let y_i denote the presence of a defective O-ring in the ith flight (0 for absence, and 1 for presence). Appendix B

Assume that y_i can be modeled as observations generated independently from a Bernoulli distribution with parameter p_i. Here, p_i denotes the relative frequency of defective O-rings at flight i. We study if 'presence of a defective O-ring' (y) depends on the 'temperature' (t), or the 'pressure' (s).

Let t_i denote the temperature (in F) in the platform, and let s_i denote the Leak check pressure (in PSI) before the ith flight. Here are some possible models of interest:

$$\mathscr{M}^I: \ p(t;\beta_{\mathscr{M}^I},\mathscr{M}^I)=\frac{\exp(\beta_0)}{1+\exp(\beta_0)} \qquad ;\mathscr{M}^{IV}: \ p(t;\beta_{\mathscr{M}^{IV}},\mathscr{M}^{IV})=\frac{\exp(\beta_0+\beta_1 t+\beta_2 s)}{1+\exp(\beta_0+\beta_1 t+\beta_2 s)}$$

$$\mathscr{M}^{II}: \ p(t;\beta_{\mathscr{M}^{II}},\mathscr{M}^{II})=\frac{\exp(\beta_0+\beta_1 t)}{1+\exp(\beta_0+\beta_1 t)} \qquad ;\mathscr{M}^{V}: \ p(t;\beta_{\mathscr{M}^{V}},\mathscr{M}^{V})=\frac{\exp(\beta_0+\beta_1 t+\beta_2 s+\beta_3 ts)}{1+\exp(\beta_0+\beta_1 t+\beta_2 s+\beta_3 ts)}$$

$$\mathscr{M}^{III}: \ p(t;\beta_{\mathscr{M}^{III}},\mathscr{M}^{III})=\frac{\exp(\beta_0+\beta_2 s)}{1+\exp(\beta_0+\beta_2 s)} \qquad \text{etc...}$$

The Bayesian hierarchical model under consideration is:

$$\begin{cases} y_i|\theta\sim f(y_i|\theta) :: & \left\{ y_i|\mathscr{M},\beta_{\mathscr{M}}\sim \text{Br}\left(y_i|\ \frac{\exp(x_{i,\mathscr{M}}^\top\beta_{\mathscr{M}})}{1+\exp(x_{i,\mathscr{M}}^\top\beta_{\mathscr{M}})}\right),\ x_{i,\mathscr{M}}^\top\beta_{\mathscr{M}}=\sum_{j\in\mathscr{M}}x_{i,j}^\top\beta_j \quad \text{for } \ i=1,...,n \right. \\ \theta|\phi_1\sim\pi(\theta|\phi_1) :: & \begin{cases} \beta_j|\mathscr{M}\sim \text{N}(\beta_j|\mu_0,\sigma_0^2) \ \text{ for all } j\in\mathscr{M} \\ \begin{cases} \mathscr{M} & =\{j\in\{1,...,d\},\ \text{s.t.}\ \gamma_j=1\} \\ \gamma_j|\varpi & \sim \text{Br}(\varpi),\ \ j=1,...,d \end{cases} \end{cases} \\ \phi_1|\phi_2\sim\pi(\phi_1|\phi_2) :: & \left\{ \varpi \ \sim \text{Be}(a_0,b_0) \right. \end{cases} \tag{13}$$

where $\theta=(\mathscr{M},\beta_{\mathscr{M}})$, $\phi_1=\varpi$, and $\phi_2=(a_0,b_0)$. Above, in the prior we considered an extra level of uncertainty by considering $\varpi\sim\text{Be}(a_0,b_0)$.

Here we added an additional level of uncertainty, and set $\varpi\sim\text{Be}(a_0,b_0)$ which creates a more diverse prior model, compared to the computer practical handout example where we had set $\varpi=0.5$.

Now the joint probability distribution has pdf

$$p(y,\beta_{\mathscr{M}},\mathscr{M},\varpi)=\underbrace{\prod_{i=1}^n \text{Br}\left(y_i|\ \frac{\exp(x_{i,\mathscr{M}}^\top\beta_{\mathscr{M}})}{1+\exp(x_{i,\mathscr{M}}^\top\beta_{\mathscr{M}})}\right)}_{f(y|\theta)}\underbrace{\prod_{j\in\mathscr{M}}\text{N}\left(\beta_j|\mu_0,\sigma_0^2\right)\prod_{j=1}^d \text{Br}(\gamma_j|\varpi)}_{\pi(\theta|\phi_1)}\underbrace{\text{Be}(\varpi|a_0,b_0)}_{\pi(\phi_1|\phi_2)}$$

Example 13. Robert and Reber (1998) considers an experiment under which rats are intoxicated by a substance, then treated by either a placebo or a drug. (See: `https://www.jstor.org/stable/pdf/25053027.pdf`) Appendix C

Statistical model ($f(y|\theta)$): The model associated with this experiment is a linear additive model effect: given x_{ij} , y_{ij} and z_{ij}, jth responses of the ith rat at the control, intoxication and treatment stages, respectively. The statistical

model was specified such as that $(1 \leq i \leq I)$

$$
\begin{aligned}
x_{i,j} &\sim \mathrm{N}(\theta_i, \sigma_c^2) && , 1 \leq j \leq J_i^c \\
y_{i,j} &\sim \mathrm{N}(\theta_i + \delta_i, \sigma_a^2) && , 1 \leq j \leq J_i^a, \\
z_{i,j} &\sim \mathrm{N}(\theta_i + \delta_i + \xi_i, \sigma_t^2) && , 1 \leq j \leq J_i^t,
\end{aligned}
$$

where θ_i is the average control measurement, δ_i the average intoxication effect and ξ_i the average treatment effect for the ith rat, the variances of these measurements being constant for the control, the intoxication and the treatment effects. An additional (observed) variable is w_i, which is equal to 1 if the rat is treated with the drug, and 0 otherwise.

Prior model $\pi(\theta|\phi)$: The different individual averages are related through a common (conjugate) prior distribution,

$$
\begin{aligned}
&\theta_i \sim \mathrm{N}(\mu_\theta, \sigma_\theta^2), && \delta_i \sim \mathrm{N}(\mu_\delta, \sigma_\delta^2), && \xi_i | w_i \sim \begin{cases} \mathrm{N}(\mu_P, \sigma_P^2) & , w_i = 0 \\ \mathrm{N}(\mu_D, \sigma_D^2) & , w_i = 1 \end{cases} \\
&\sigma_c \sim \pi(\sigma_c) \propto \frac{1}{\sigma_c}, && \sigma_a \sim \pi(\sigma_a) \propto \frac{1}{\sigma_a}, && \sigma_t \sim \pi(\sigma_t) \propto \frac{1}{\sigma_t}, \qquad (14)
\end{aligned}
$$

This modeling seems to describe the natural phenomenon realistically enough, in the sense of the responses x_{ij} , y_{ij} and z_{ij}.

Hyper-priors $\pi(\phi|\phi_m)$: For the higher levels of prior ($\pi(\phi|\phi_m)$ in Eq 2), they considered improper (Jeffrey's) hyper-priors.

$$
(\mu_\theta, \sigma_\theta) \sim \pi(\mu_\theta, \sigma_\theta) \propto \frac{1}{\sigma_\theta}, \qquad (\mu_\delta, \sigma_\delta) \sim \pi(\mu_\delta, \sigma_\delta) \propto \frac{1}{\sigma_\delta}, \qquad (\mu_P, \sigma_P) \sim \pi(\mu_P, \sigma_P) \propto \frac{1}{\sigma_P}, \qquad (15)
$$

$$
(\mu_D, \sigma_D) \sim \pi(\mu_D, \sigma_D) \propto \frac{1}{\sigma_D}. \qquad (16)
$$

The priors in lines (14), (15) and (16) are improper non-informative priors. One could have have specify proper priors, like Normal-Inverse Gamma which are conjugate, however in that case he/she should have to specify the values for the fixed hyper-parameters.

2 Non-identifiability issue

Definition 14. A parametric model for which an element of the parametrisation is redundant is said to be non-identified.

Remark 15. Let Bayesian model $(f(y|\theta), \pi(\theta))$, where $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$, and assume that the parametric model does not depend on θ_1; i.e. $f(y|\theta_1, \theta_2) = f(y|\theta_2)$. Because the likelihood does not depend on θ_1 suggests that y does not provide information about θ_1 directly.

Remark 16. Bayesian analysis of a non-identified model is always possible if a suitable prior $\Pi(\theta_1, \theta_2)$ on all the parameters is specified. For instance, if one specifies a priori that learning the value of θ_2 may change his belief about θ_1, via $\pi(\theta_1|\theta_2) \neq \pi(\theta_1)$.

Factorize the prior distribution as $\pi(\theta_1, \theta_2) = \pi(\theta_1|\theta_2)\pi(\theta_2)$. Then, we have the following PDF/PMF

$$\begin{aligned}
\pi(\theta_1, \theta_2|y) &\propto f(y|\theta_1, \theta_2)\pi(\theta_1, \theta_2) = f(y|\theta_2)\pi(\theta_1|\theta_2)\pi(\theta_2) \implies \\
\pi(\theta_1, \theta_2|y) &= \pi(\theta_2|y)\pi(\theta_1|\theta_2) \implies \\
\pi(\theta_2|y) &= \frac{f(y|\theta_2)\pi(\theta_2)}{\int_{\Theta_2} f(y|\theta_2)\pi(\theta_2)\mathrm{d}\theta_2} \quad . \\
\pi(\theta_1|y, \theta_2) &= \pi(\theta_1|\theta_2) \qquad (17) \\
\pi(\theta_1|y) &= \int_{\Theta_2} \pi(\theta_1|\theta_2)\pi(\theta_2|y)\mathrm{d}\theta_2 \qquad (18)
\end{aligned}$$

θ_1 is non-identifiable parameter from the data y, because y provides no direct information about θ_1. Inference about θ_1 based on marginal posterior $\pi(\theta_1|y)$ depends on y but the information provided about θ_1 comes indirectly through the marginal posterior of θ_2, see (18). Equivalently, (18) implies that y provides no information about θ_1 given θ_2.

If we a priori specify that learning the value of θ_2 does not change our belief about θ_1 $\pi(\theta_1|\theta_2) = \pi(\theta_1)$, then (18) becomes $\pi(\theta_1|y) = \pi(\theta_1)$ and hence data y provide no information about θ_1 at all.

Example 17. (Cont Example 11) It is not difficult to understand that the Bayesian model as defined in Example 7 is non-identifiable. For simplicity we focus on the Bayesian mixture of $k = 2$ components with Appendix A

$$\begin{aligned}
y|\varpi, \mu, \sigma^2 &\sim f(y|\varpi, \mu, \sigma^2) := \varpi_1 \mathrm{N}(y|\mu_1, \sigma_1^2) + \varpi_2 \mathrm{N}(y|\mu_2, \sigma_2^2) \\
\pi(\varpi, \mu, \sigma^2) &= \underbrace{\mathrm{N}(\mu_1|\xi, \sigma_1^2)\mathrm{N}(\mu_2|\xi, \sigma_2^2)}_{\pi(\mu|\sigma^2)}\underbrace{\mathrm{IG}(\sigma_1^2|\alpha, \beta)\mathrm{IG}(\sigma_2^2|\alpha, \beta)}_{\pi(\sigma^2)}\mathrm{Di}(\varpi|\delta) \qquad (19)
\end{aligned}$$

which leads to a posterior such as

$$\pi(\varpi, \mu, \sigma^2|y) \propto \left[\varpi_1 \mathrm{N}(y|\mu_1, \sigma_1^2) + \varpi_2 \mathrm{N}(y|\mu_2, \sigma_2^2)\right] \mathrm{N}(\mu_1|\xi, \sigma_1^2)\mathrm{N}(\mu_2|\xi, \sigma_2^2)\mathrm{IG}(\sigma_1^2|\alpha, \beta)\mathrm{IG}(\sigma_2^2|\alpha, \beta)\mathrm{Di}(\varpi|\delta)$$

Here the parametrization is non-identifiable because the symmetry in the sampling distribution

$$\varpi_1 \mathrm{N}(y|\mu_1, \sigma_1^2) + \varpi_2 \mathrm{N}(y|\mu_2, \sigma_2^2) = \varpi_2 \mathrm{N}(y|\mu_2, \sigma_2^2) + \varpi_1 \mathrm{N}(y|\mu_1, \sigma_1^2)$$

and the naive prior in (19) produce a posterior such that

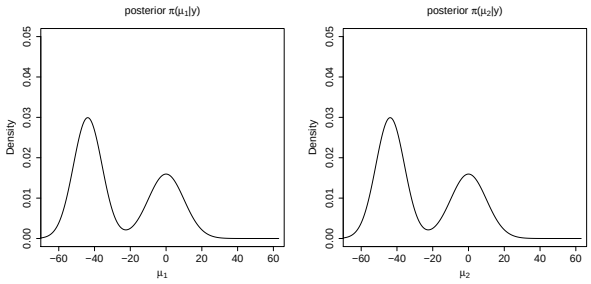
$$\pi\left(\varpi = \begin{pmatrix}\varpi_1 \\ \varpi_2\end{pmatrix}, \mu = \begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix}, \sigma^2 = \begin{pmatrix}\sigma_1^2 \\ \sigma_2^2\end{pmatrix}|y\right) = \pi\left(\varpi = \begin{pmatrix}\varpi_2 \\ \varpi_1\end{pmatrix}, \mu = \begin{pmatrix}\mu_2 \\ \mu_1\end{pmatrix}, \sigma^2 = \begin{pmatrix}\sigma_2^2 \\ \sigma_1^2\end{pmatrix}|y\right)$$

This parametrization is not meaningful for parametric inference. In Bayesian stats this can be resolved for instance by changing the prior and imposing an identifiability constrain as

$$\pi^*(\mu|\sigma^2) = \frac{\mathrm{N}(\mu_1|\xi, \sigma_1^2)\mathrm{N}(\mu_2|\xi, \sigma_2^2)1\left(\mu_1 \leq \mu_2\right)}{\int \mathrm{N}(\mu_1|\xi, \sigma_1^2)\mathrm{N}(\mu_2|\xi, \sigma_2^2)1\left(\mu_1 \leq \mu_2\right)\mathrm{d}(\mu_1, \mu_2)} \propto \pi(\mu|\sigma^2)1\left(\mu_1 \leq \mu_2\right)$$

The non-identifiable model produces marginal posterior in Figures 1a and 1b.

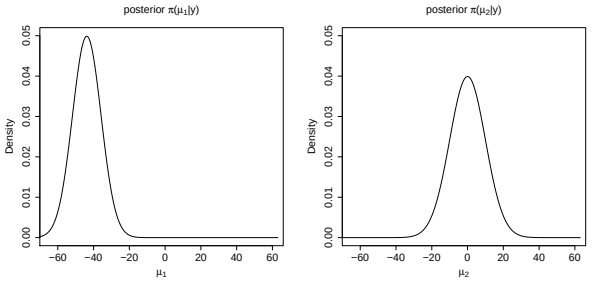
$$\begin{cases} y_i|\varpi,\mu,\sigma^2 & \sim \varpi_1 \mathrm{N}(y|\mu_1,\sigma_1^2)+\varpi_2 \mathrm{N}(y|\mu_2,\sigma_2^2) \\ \varpi & \sim \mathrm{Di}(\delta) \\ \mu|\sigma^2 & \sim \mathrm{N}(\mu_1|\xi,\sigma_1^2)\mathrm{N}(\mu_2|\xi,\sigma_2^2) \\ \sigma^2 & \sim \mathrm{Ga}(\sigma_1^2|\alpha,\beta)\mathrm{Ga}(\sigma_1^2|\alpha,\beta) \end{cases} \tag{20}$$



(a) Marginal posterior $\pi(\mu_1|y)$ (b) Marginal posterior $\pi(\mu_2|y)$

After non-identifiability is resolved, the identifiable model produces marginal posterior in Figures 1c and 1d.

$$\begin{cases} y_i|\varpi,\mu,\sigma^2 & \sim \varpi_1 \mathrm{N}(y|\mu_1,\sigma_1^2)+\varpi_2 \mathrm{N}(y|\mu_2,\sigma_2^2) \\ \varpi & \sim \mathrm{Di}(\delta,...,\delta) \\ \mu|\sigma^2 & \sim \mathrm{N}(\mu_1|\xi,\sigma_1^2)\mathrm{N}(\mu_2|\xi,\sigma_2^2)1\left(\mu_1 \leq \mu_2\right) \\ \sigma^2 & \sim \mathrm{Ga}(\sigma_1^2|\alpha,\beta)\mathrm{Ga}(\sigma_1^2|\alpha,\beta) \end{cases} \tag{21}$$



(c) Marginal posterior $\pi(\mu_1|y)$ (d) Marginal posterior $\pi(\mu_2|y)$

3 Augmentation

Definition 18. Augmentation technique on a Bayesian model $\left(y \sim F\left(y|\theta\right),\ \theta \sim \pi\left(\theta\right)\right)$ is when the vector (y,θ) is imputed by auxiliary random quantities $x \sim Q(\cdot)$ where the distribution $Q(\cdot)$ is such that the joint distribution of (y,θ,x) admits the joint distribution of (y,θ) as its marginal, e.g.

$$p\left(y,\theta\right)=\int p\left(y,\theta,x\right)\mathrm{d}x$$

Remark 19. According to Definition 18, the posterior of θ given the observables y remains the same; e.g.

$$\pi\left(\theta|y\right)=\int \pi\left(\theta,x|y\right)\mathrm{d}x=\frac{\int p\left(y,\theta,x\right)\mathrm{d}x}{\int p\left(y,\theta,x\right)\mathrm{d}\theta\mathrm{d}x}$$

Example 20. (Cont. Examples 7 and 11) Imputation of the hierarchical model (10) by considering the latent variables $z_i \sim f(z_i)=\varpi_{z_i}$ for $z_i \in \{1,...,k\}$ leads to the augmented hierarchical model (5).

Remark 21. Software JAGS, aiming at facilitating Bayesian computations, requires as inputs only Bayesian hierarchical models where the dimension of the unknown quantities is fixed (not random). Consequently many hierarchical models for model comparison such as (5), (10), and 13 cannot be used directly. This limitation can be addressed via augmentation, i.e. the imputation of the hierarchical model with suitable random variables such that the resulting joint posterior distribution admits the posterior distribution as its marginal.

Example 22. (Cont Example 12) Notice that the length of the join unknown parameter vector

$$(\mathscr{M},\beta_{\mathscr{M}},\gamma_{1:d},\varpi)$$

is not fixed but random because $\mathscr{M} = \{j \in \{1, ..., d\}, \text{ s.t. } \gamma_j = 1\}$ is random/unknown. Current versions of JAGS work only when the join unknown parameter vector of the Bayesian hierarchical model is fixed. Augent the hierarchical model properly so that it can be used with RJAGS.

Solution. We augment model (13) with auxiliary random quantities such that $\beta_j|\mathscr{M} \sim \text{N}(\beta_j|\mu_0, \sigma_0^2)$ for all $j \notin \mathscr{M}$, i.e.

$$\begin{cases} \left\{ y_i|\mathscr{M}, \beta_{\mathscr{M}} \sim \text{Br}\left(y_i| \frac{\exp(x_{i,\mathscr{M}}^\top \beta_{\mathscr{M}})}{1+\exp(x_{i,\mathscr{M}}^\top \beta_{\mathscr{M}})} \right), \; x_{i,\mathscr{M}}^\top \beta_{\mathscr{M}} = \sum_{j \in \mathscr{M}} x_{i,j}^\top \beta_j \quad \text{for } \; i = 1, ..., n \right. \\ \\ \left\{ \beta_j|\mathscr{M} \sim \text{N}(\beta_j|\mu_0, \sigma_0^2) \;\; \text{for all } j \notin \mathscr{M} \right. \\ \\ \begin{cases} \beta_j|\mathscr{M} \sim \text{N}(\beta_j|\mu_0, \sigma_0^2) \;\; \text{for all } j \in \mathscr{M} \\ \\ \begin{cases} \mathscr{M} & = \{j \in \{1, ..., d\}, \text{ s.t. } \gamma_j = 1\} \\ \gamma_j|\varpi & \sim \text{Br}(\varpi), \;\; j = 1, ..., d \end{cases} \end{cases} \\ \\ \left\{ \varpi \quad \sim \text{Be}(a_0, b_0) \right. \end{cases}$$

then

$$p(y, \beta_{\mathscr{M}} \beta_{-\mathscr{M}}, \mathscr{M}, \varpi) = p(y, \beta_{\mathscr{M}}, \mathscr{M}, \varpi) \prod_{j \notin \mathscr{M}} \text{N}\left(\beta_j|\mu_0, \sigma_0^2\right)$$

and

$$\int p(y, \beta_{\mathscr{M}} \beta_{-\mathscr{M}}, \mathscr{M}, \varpi) \text{d}\beta_{-\mathscr{M}} = p(y, \beta_{\mathscr{M}}, \mathscr{M}, \varpi) \int \prod_{j \notin \mathscr{M}} \text{N}\left(\beta_j|\mu_0, \sigma_0^2\right) \text{d}\beta_{-\mathscr{M}} = p(y, \beta_{\mathscr{M}}, \mathscr{M}, \varpi)$$

A Appendix: Bayesian Normal Mixture model

A.1 Case where k is a fixed known parameter

The complete R / RJAGS code is available from

- `http://htmlpreview.github.io/?https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/main/Computer_practical/Normal_Mixture_model/Bayesian_Normal_Mixture_Model.nb.html`
- `https://raw.githubusercontent.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/main/Computer_practical/Normal_Mixture_model/enz.dat`

For simplification, consider the hierarchical model where the number of groups k is a fixed/known quantity.

$$\begin{cases} y_i|z_i,k,\varpi,\mu,\sigma^2 & \sim f_{z_i}(y_i|\varpi_{z_i},\mu_{z_i},\sigma^2_{z_i}) \\ & \quad \text{for } i=1,...,n \\ z_i|k & \sim f(z_i):=\varpi_{z_i} \\ & \quad \text{for } i=1,...,n \\ \varpi|k & \sim \text{Di}(\delta,...,\delta) \\ \mu_j|\sigma^2_j,k & \sim \text{N}(\xi,\sigma^2_j/\kappa) \\ & \quad \text{for } j=1,...,k \\ \sigma^2_j|\beta,k & \sim \text{IG}(\alpha,\beta) \\ & \quad \text{for } j=1,...,k \\ \beta & \sim \text{Ga}(g,h) \end{cases} \tag{22}$$

Note 23. The RJAGS code for the analysis of the data set in the Example 7 is provided in Algorithm 1, however here the number of groups k is fixed/known quantity and not random.

Algorithm 1 [Bayesian Normal Mixture model] RJAGS script for the analysis of the data set in the Example 7 .

```
rm(list=ls())
# Load rjags
library("rjags")
# define the Bayesian hierarchical model in JAGS syntax
jags_model <- "
  model{
    #sampling distribution
    for (i in 1:n) {
      y[i] ~ dnorm(mu[zeta[i]], tau[zeta[i]])
      zeta[i] ~ dcat(varpi[])
    }
    #prior distribution
    varpi ~ ddirich(delta)
    for (i in 1:k) {
      mu[i] ~ dnorm(xi,kappa)
      tau[i] ~ dgamma(alpha,beta)
      sigma[i] <- 1/sqrt(tau[i])
    }
    beta ~ dgamma(g,h)
}
"
```

Example 24. (Cont. Example 7) As we will see in term 2, although (5) and 10 are equivalent in the sense that they produce the same inference, the expended hierarchical model (6) is computational convenient compared to hierarchical model (11) in the sense that its priors are conditional conjugate.

For simplicity assume that the number of groups is known and fixed to k. The full conditional posteriors in model (6) are:

$$w|y... \sim \text{Di}\left(\delta + n_1, ..., \delta + n_k\right); \text{ where } n_j = \sum_{i=1}^{n} 1\left(z_i = j\right)$$

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

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

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

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

which can be used in Monte Carlo integration.

Model (11) does not produce full conditional posteriors of standard form, due to the summation in the likelihood.

A.2 Case where k is a random/unknown parameter

The complete R / RJAGS code is available from

- `http://htmlpreview.github.io/?https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/main/Computer_practical/Normal_Mixture_Model_unknown_number_of_components/Bayesian_Normal_Mixture_Model_unknown_number_of_components.nb.html`
- `https://raw.githubusercontent.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/main/Computer_practical/Normal_Mixture_Model_unknown_number_of_components/enz.dat`

Note 25. Consider the hierarchical model 5.

Note 26. Notice that the length of the join unknown parameter vector

$$\left(z_{1:n}, k, \varpi_{1:k}, \mu_{1:k}, \sigma_{1:k}^2\right)$$

is not fixed but random because k is random/unknown. Current versions of JAGS work only when the join unknown parameter vector of the Bayesian hierarchical model is fixed.

Note 27. It is given that if $c_j \sim \text{Ga}(\delta_j, 1)$ for $j = 1, ..., k$ then

$$\varpi_{1:k} = \left(\frac{c_1}{\sum_{j=1}^{k} c_j}, ..., \frac{c_k}{\sum_{j=1}^{k} c_j}\right) \sim \text{Di}(\delta_1, ..., \delta_k).$$

Note 28. To make JAGS work on our problem we augment $\left(z_{1:n}, k, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}\right)$ with additional random variables

$$
\begin{aligned}
c_j|k &\sim \text{Ga}(\delta, 1) && \text{for } j = k+1, ..., k_{\max} \\
\mu_j|\sigma^2_j, k &\sim \text{N}(\xi, \sigma^2_j/\kappa) && \text{for } j = k+1, ..., k_{\max} \\
\sigma^2_j|\beta, k &\sim \text{IG}(\alpha, \beta) && \text{for } j = k+1, ..., k_{\max}
\end{aligned}
$$

Precisely, we consider the augmented Bayesian hierarchical model

Note 29. Consider augmentation where (5) is augmented by $\mu_j, \sigma^2_j|k \sim \text{N}(\xi, \sigma^2_j/\kappa)\text{IG}(\alpha, \beta)$ for $j = k+1, ..., k_{\max}$, and $c_j \sim \text{Ga}(\delta, 1)$ for $j = k+1, ..., k_{\max}$

$$
\left\{
\begin{array}{lll}
y_i|z_i, k, \varpi, \mu, \sigma^2 & \sim f_{z_i}(y_i|\varpi_{z_i}, \mu_{z_i}, \sigma^2_{z_i}) & \text{for } i = 1, ..., n \\
z_i|k & \sim f(z_i) := \varpi_{z_i} & \text{for } i = 1, ..., n \\
\varpi_{1:k} & = \left(\frac{c_1}{\sum_{j=1}^k c_j}, ..., \frac{c_k}{\sum_{j=1}^k c_j}\right) & \\
 & & \\
c_j|k & \sim \text{Ga}(\delta, 1) & \text{for } j = 1, ..., k_{\max} \\
\mu_j|\sigma^2_j, k & \sim \text{N}(\xi, \sigma^2_j/\kappa) & \text{for } j = k+1, ..., k_{\max} \\
\sigma^2_j|\beta, k & \sim \text{IG}(\alpha, \beta) & \text{for } j = k+1, ..., k_{\max} \\
 & & \\
\mu_j|\sigma^2_j, k & \sim \text{N}(\xi, \sigma^2_j/\kappa) & \text{for } j = 1, ..., k \\
\sigma^2_j|\beta, k & \sim \text{IG}(\alpha, \beta) & \text{for } j = 1, ..., k \\
\beta & \sim \text{Ga}(g, h) & \\
k & \sim \text{U}_{\text{discr}}(1, k_{\max}) &
\end{array}
\right. \tag{23}
$$

then

$$
p(y_{1:n}, z_{1:n}, k, \varpi_{1:k}, \mu_{1:k_{\max}}, \sigma^2_{1:k_{\max}}, c_{k+1:1:k_{\max}}) = p(y_{1:n}, z_{1:n}, k, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k}) \prod_{j=k+1}^{k_{\max}} \text{N}\left(\mu_j|\xi, \sigma^2_j/\kappa\right) \text{Ga}(\sigma^2_k|\alpha, \beta)\text{Ga}(c_j|\delta, 1)
$$

and

$$
\int p(y_{1:n}, z_{1:n}, k, \varpi_{1:k}, \mu_{1:k_{\max}}, \sigma^2_{1:k_{\max}}, c_{k+1:1:k_{\max}})\text{d}\left(\mu_{k+1:k_{\max}}, \sigma^2_{k+1:k_{\max}}, c_{k+1:1:k_{\max}}\right) = p(y_{1:n}, z_{1:n}, k, \varpi_{1:k}, \mu_{1:k}, \sigma^2_{1:k})
$$

The RJAGS code for the analysis of the data set in the Example 12 is provided in Algorithm 2.

Algorithm 2 [Bayesian Normal Mixture model] RJAGS script for the analysis of the data set in the Example 7 .

```
rm(list=ls())
# Load rjags
library("rjags")
# define the Bayesian hierarchical model in JAGS syntax
jags_model <- "
  model{
  # sampling distribution
  for (i in 1:n) {
    y[i] ~ dnorm(mu[zeta[i]], tau[zeta[i]])
    zeta[i] ~ dcat( varpi ) }
    # within model parameter priors
    for (i in 1:length(pk) ) {
      x[i] ~ dgamma(delta[i],1)
      ind[i] <- (i<=k)
    }
    varpi <- (x*ind)/sum(x*ind)
    for (i in 1:length(pk) ) {
      mu[i] ~ dnorm(xi,kappa)
      tau[i] ~ dgamma(alpha,beta)
      sigma[i] <- 1/sqrt(tau[i])
    }
    beta ~ dgamma(g,h)
    k~dcat(pk)
  }
  "
```

B Appendix: Bayesian Variable Selection in Logistic regression model

The complete R / RJAGS code is available from

- `http://htmlpreview.github.io/?https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/main/Computer_practical/Bernoulli_regression_model_variable_selection/Bernoulli_Regression_Model_VS_full.nb.html`

Note 30. Notice that the length of the join unknown parameter vector

$$(\mathscr{M}, \beta_{\mathscr{M}}, \gamma_{1:d}, \varpi)$$

is not fixed but random because $\mathscr{M} = \{j \in \{1, ..., d\}, \text{ s.t. } \gamma_j = 1\}$ is random/unknown. Current versions of JAGS work only when the join unknown parameter vector of the Bayesian hierarchical model is fixed.

Note 31. To make JAGS work on our problem we 'cheat' by augmenting $(\mathscr{M}, \beta_{\mathscr{M}}, \gamma_{1:d}, \varpi)$ with additional random variables $\beta_j | j \notin \mathscr{M} \sim \text{N}(\beta_j | \mu_0, \sigma_0^2)$. Precisely, we consider the augmented Bayesian hierarchical model

$$\begin{cases} y_i | \mathscr{M}, \beta_{\mathscr{M}} & \sim \text{Bernoulli}(p(x_i; \mathscr{M}, \beta_{\mathscr{M}})), \quad \text{for, } i = 1, ..., n \\ p(x_i; \mathscr{M}, \beta_{\mathscr{M}}) & = \frac{\exp(x_{i,\mathscr{M}}^\top \beta_{\mathscr{M}})}{1+\exp(x_{i,\mathscr{M}}^\top \beta_{\mathscr{M}})}; \text{ where } x_{i,\mathscr{M}}^\top \beta_{\mathscr{M}} = \sum_{j \in \mathscr{M}} x_{i,j} \beta_j \\ \beta_j | j \notin \mathscr{M} & \sim \text{N}(\beta_j | \mu_0, \sigma_0^2), \ \ j \notin \mathscr{M} \\ \beta_j | j \in \mathscr{M} & \sim \text{N}(\beta_j | \mu_0, \sigma_0^2), \ \ j \in \mathscr{M} \\ \mathscr{M} & = \{j \in \{1, ..., d\}, \text{ s.t. } \gamma_j = 1\} \\ \gamma_j & \sim \text{Bernoulli}(\varpi), \ \ j = 1, ..., d \\ \varpi & \sim \text{Be}(a_0, b_0) \end{cases}$$

Now we work on the augmented vector $(\mathscr{M}, \beta_{\mathscr{M}}, \beta_{\mathscr{M}^\complement}, \gamma_{1:d}, \varpi)$. The augmented hierarchical model admits the hierarchical model as its marginal because

$$\pi(\mathscr{M}, \beta_{\mathscr{M}}, \gamma_{1:d}, \varpi | y) = \int \pi(\mathscr{M}, \beta_{\mathscr{M}}, \beta_{\mathscr{M}^\complement}, \gamma_{1:d}, \varpi | y) \, \mathrm{d}\beta_{\mathscr{M}^\complement}$$

Note 32. The RJAGS code for the analysis of the data set in the Example 12 is provided in Algorithm 3.

Algorithm 3 [Bayesian Mixed Effect Model] RJAGS script for the analysis of the data set in the Example 7 .

```
rm(list=ls())
# Load rjags
library("rjags")
# define the Bayesian hierarchical model in JAGS syntax
hierarhicalmodel<-"
model {
  # sampling distribution
  for (i in 1:n) {
    eta[i] <- inprod(X[i,],beta*ind)
    mean[i] <- exp( eta[i] ) / ( 1 + exp( eta[i] ) )
    y[i] ~ dbern(mean[i])
  }
  # within model prior + augmentation
  for ( j in 1:dmax ) {
    beta[j] ~ dnorm( 0 , 0.1 )
  }
  # marginal model prior
  ind[1] <- 1
  for (j in 2:dmax) {
    ind[j] ~ dbern( pp )
  }
  # hyper-prior
  pp ~ dbeta(1.0,1.0)
}
"
```

C Appendix: Random effect model

The complete R / RJAGS code is available from

- `https://github.com/georgios-stats/Bayesian_Statistics_Michaelmas_2021/blob/main/Lecture_handouts/Rscripts/Hierarchical_bayes/HierarchicalBayesPharmaceutical.R`

Note 33. The RJAGS code for the analysis of the data set in the Example 13 is provided in Algorithm 4.

Algorithm 4 [Random effect model] RJAGS script for the analysis of the data set in the Example 13.

```
rm(list=ls())
# Load rjags
library("rjags")
# define the Bayesian hierarchical model in JAGS syntax
hierarhicalmodel <- "
  model {
     for ( i in 1 :  I ) {
      for ( j in 1 :  J ) {
        x[i,j] ~ dnorm( theta[i] , tau_c )
        y[i,j] ~ dnorm( theta[i] + delta[i] , tau_a )
        z[i,j] ~ dnorm( theta[i] + delta[i] + xi[i] , tau_t )
      }
      theta[i] ~ dnorm( mu_theta , tau_theta )
      delta[i] ~ dnorm( mu_delta , tau_delta )
      w_ind[i] <- ifelse(w[i] == 0, 0, 1)
      xi[i] <- w_ind[i]*x_d +(1-w_ind[i])*x_p
    }
    sig2_c <- 1/tau_c tau_c ~ dgamma(0.01 , 0.01)
    sig2_a = 1/tau_a tau_a ~ dgamma(0.01 , 0.01)
    sig2_t <- 1/tau_t tau_t ~ dgamma(0.01 , 0.01)
    mu_theta ~ dnorm( 0 , 0.001 )
    sig2_theta = 1/tau_theta
    tau_theta ~ dgamma(0.01 , 0.01)
    mu_delta ~ dnorm( 0 , 0.001 )
    sig2_delta <- 1/tau_delta
    tau_delta ~ dgamma(0.01 , 0.01)
    x_p ~ dnorm( mu_p , tau_p )
    x_d ~ dnorm( mu_d , tau_d )
    mu_p ~ dnorm( 0 , 0.001 )
    sig2_p <- 1/tau_p
    tau_p ~ dgamma(0.01 , 0.01)
    mu_d ~ dnorm( 0 , 0.001 )
    sig2_d <- 1/tau_d
    tau_d ~ dgamma(0.01 , 0.01)
  }
"
```

Example 34. (Cont...) You may use

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

The joint posterior pdf of $\vartheta = (\theta_{1:I}, \delta_{1:I}, \xi_{1:I}, \sigma_c^2, \sigma_a^2, \sigma_t^2, \sigma_\theta^2, \sigma_\delta^2, \sigma_P^2, \sigma_D^2, \mu_\theta, \mu_\delta, \mu_P, \mu_D)$ given obs. x, y, z is

$$\pi(\vartheta|x,y,z) \propto \prod_{i=1}^{I} \left[\exp\left(-\frac{(\theta_i - \mu_\theta)^2}{2\sigma_\theta^2} - \frac{(\delta_i - \mu_\delta)^2}{2\sigma_\delta^2} \right) \prod_{j=1}^{J_i^c} \exp\left(-\frac{(x_{i,j} - \theta_i)^2}{2\sigma_c^2} \right) \times \prod_{j=1}^{J_i^a} \exp\left(-\frac{(y_{i,j} - \theta_i - \delta_i)^2}{2\sigma_a^2} \right) \right.$$
$$\left. \times \prod_{j=1}^{J_i^t} \exp\left(-\frac{(z_{i,j} - \theta_i - \delta_i - \xi_i)^2}{2\sigma_t^2} \right) \times \prod_{w_i=0} \exp\left(-\frac{(\xi_i - \mu_P)^2}{2\sigma_P^2} \right) \prod_{w_i=0} \exp\left(-\frac{(\xi_i - \mu_D)^2}{2\sigma_D^2} \right) \right]$$
$$\times \sigma_c^{-\sum_i J_i^c - 1} \sigma_a^{-\sum_i J_i^a - 1} \sigma_t^{-\sum_i J_i^t - 1} \sigma_\theta^{I-1} \sigma_\delta^{I-1} \sigma_D^{I_D - 1} \sigma_P^{I_P - 1}.$$

The joint posterior distributions is not of standard form, and its pdf is intractable. However the full conditionals are of standard form. For instance, the full conditional posterior distribution density

$$\pi(\delta_{1:I}|x_{\text{all}}, y_{\text{all}}, z_{\text{all}}, \theta_{1:I}, \xi_{1:I}, \sigma_c^2, \sigma_a^2, \sigma_t^2, \sigma_\theta^2, \sigma_\delta^2, \sigma_P^2, \sigma_D^2, \mu_\theta, \mu_\delta, \mu_P, \mu_D)$$
$$\propto \prod_{i=1}^{I} \left[\exp\left(-\frac{(\delta_i - \mu_\delta)^2}{2\sigma_\delta^2} \right) \times \prod_{j=1}^{J_i^a} \exp\left(-\frac{(y_{i,j} - \theta_i - \delta_i)^2}{2\sigma_a^2} \right) \times \prod_{j=1}^{J_i^t} \exp\left(-\frac{(z_{i,j} - \theta_i - \delta_i - \xi_i)^2}{2\sigma_t^2} \right) \right]$$
$$\propto \prod_{i=1}^{I} \left[\exp\left(-\frac{(\delta_i - \mu_\delta)^2}{2\sigma_\delta^2} - \sum_{j=1}^{J_i^a} \frac{(\delta_i - (y_{i,j} - \theta_i))^2}{2\sigma_a^2} - \sum_{j=1}^{J_i^t} \frac{(\delta_i - (z_{i,j} - \theta_i - \xi_i)}{2\sigma_t^2} \right) \right]$$
$$\propto \prod_{i=1}^{I} \left[\exp\left(-\frac{(\delta_i - \mu_{\delta,i}^*)^2}{2\left(\sigma_{\delta,i}^*\right)^2} + \text{const...} \right) \right] \propto \prod_{i=1}^{I} \left[\exp\left(-\frac{(\delta_i - \mu_{\delta,i}^*)^2}{2\left(\sigma_{\delta,i}^*\right)^2} + \text{const...} \right) \right]$$
$$\propto \prod_{i=1}^{I} \mathrm{N}\left(\delta_i | \mu_{\delta,i}^*, \left(\sigma_{\delta,i}^*\right)^2 \right)$$

with

$$\delta_i | \text{rest}, ... \overset{\text{ind}}{\sim} \mathrm{N}\left(\mu_{\delta,i}^*, \left(\sigma_{\delta,i}^*\right)^2 \right), \forall i = 1, ..., n$$

where

$$\left(\sigma_{\delta,i}^*\right)^2 = \left(\frac{1}{\sigma_\delta^2} + \frac{1}{\sigma_a^2} J_i^a + \frac{1}{\sigma_t^2} J_i^t \right)^{-1}; \quad \mu_{\delta,i}^* = \left(\sigma_{\delta,i}^*\right)^2 \left(\frac{\mu_\delta}{\sigma_\delta^2} + \frac{\sum_{j=1}^{J_i^a} y_{i,j} - J_i^a \theta_i}{\sigma_a^2} + \frac{\sum_{j=1}^{J_i^a} y_{i,j} - J_i^t \theta_i - J_i^t \xi_i}{\sigma_t^2} \right)$$

Notice that δ_i are a postriori independent given all the resp unknown parameters $\left(\theta_{1:I}, \xi_{1:I}, \sigma_c^2, \sigma_a^2, \sigma_t^2, \sigma_\theta^2, \sigma_\delta^2, \sigma_P^2, \sigma_D^2, \mu_\theta, \mu_\delta, \mu_P, \mu_D\right)$. Notice that the prior $\delta_i \sim \mathrm{N}(\mu_\delta, \sigma_\delta^2)$ in Example 13 is conditional conjugate prior of δ_i.

Try to compute the rest

$$\pi(\theta_{1:I}|\text{rest}, ...) \sim ?; \quad \pi(\sigma_t^2|\text{rest}, ...) \sim ?; \quad \pi(\sigma_c^2|\text{rest}, ...) \sim ?; \quad \pi(\sigma_a^2|\text{rest}, ...) \sim ?;$$
$$\pi(\xi_{1:I}|\text{rest}, ...) \sim ?; \quad \pi(\sigma_\theta^2|\text{rest}, ...) \sim ?; \quad \pi(\sigma_\delta^2|\text{rest}, ...) \sim ?; \quad \pi(\sigma_P^2|\text{rest}, ...) \sim ?, \text{ etc...}$$

See the solutions in: Robert, C. P., & Reber, A. (1998) from the link (`https://www.jstor.org/stable/pdf/25053027.pdf`).