1. BDA3 Problem 5.3. Instead of printing the table, you can plot the pair-wise probabilities of one being better than the other as an 8×8 matrix.

3. Hierarchical models and multiple comparisons:

   (a) Reproduce the computations in Section 5.5 for the educational testing example. Use the posterior simulations to estimate (i) for each school j, the probability that its coaching program is the best of the eight; and (ii) for each pair of schools, j and k, the probability that the coaching program in school j is better than that in school k.

   (b) Repeat (a), but for the simpler model with τ set to ∞ (that is, separate estimation for the eight schools). In this case, the probabilities (ii) can be computed analytically.

   (c) Discuss how the answers in (a) and (b) differ.

   (d) In the model with τ set to 0, the probabilities (i) and (ii) have degenerate values; what are they?

School	treatment effect, y_j	of effect estimate, σ_j
A	28	15
B	8	10
C	−3	16
D	7	11
E	−1	9
F	1	11
G	18	10
H	12	18

(a) $y_j \mid \theta_j \sim N(\theta_j, \sigma_j^2)$

$\theta_j \sim N(\mu, \tau^2)$

$\pi(\mu \mid \tau^2) \propto 1 \quad \pi(\tau^2) \propto 1$

$$f(\vec{\theta}, \mu, \tau \mid \vec{y}) \propto \prod_j p(y_j \mid \theta_j) \prod_j (\theta_j) \cdot \pi(\mu \mid \tau^2) \cdot \pi(\tau^2)$$

$$\propto \exp\left(\sum_{i=1}^{J} -\frac{(y_i - \theta_i)^2}{2\sigma_i^2}\right) \cdot (\tau^2)^{-\frac{J}{2}} \exp\left(-\frac{\sum_{i=1}^{J}(\theta_i - \mu)^2}{2\tau^2}\right)$$

$$\pi(\theta_j \mid \mu, \tau^2, y) \propto \exp\left(-\frac{(y_j - \theta_j)^2}{2\sigma_j^2}\right) \cdot \exp\left(-\frac{(\theta_j - \mu)^2}{2\tau^2}\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}\right)\theta_j^2 - 2\left(\frac{1}{\sigma_j^2} y_j + \frac{\mu}{\tau^2}\right)\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}\right)\left(\theta_j - \frac{1/\sigma_j^2\, y_j + 1/\tau^2 \cdot \mu}{1/\sigma_j^2 + 1/\tau^2}\right)^2\right)\right)$$

$$\sim N\left(\frac{1/\sigma_j^2\, y_j + 1/\tau^2 \cdot \mu}{1/\sigma_j^2 + 1/\tau^2}, \frac{1}{1/\sigma_j^2 + 1/\tau^2}\right)$$

$$\pi(\mu \mid \theta_j, \mu, \tau^2, y) \propto \exp\left(-\frac{J\mu^2 - 2\sum_{i=1}^{J}\theta_i \mu}{2\tau^2}\right) \propto \exp\left(-\frac{(\mu - \bar{\theta})^2}{2\tau^2/J}\right)$$

$$\sim N(\bar{\theta}, \tau^2/J)$$

$$\pi(\tau^2 \mid y, \vec{\theta}, \mu) \propto (\tau^2)^{-\frac{J}{2}} \exp\left(-\frac{\sum_{i=1}^{J}(\mu-\theta_i)^2}{2\tau^2}\right)$$

$$\sim IG\left(\frac{J}{2}-1, \frac{\sum_{i=1}^{n}(\theta_i-\mu)^2}{2}\right)$$

Therefore, we got a full conditional distribution, then we can implement a Gibbs Sampler.

(i)

Table 1: Probability of Best School - Prior on Tau

A	B	C	D	E	F	G	H
0.257	0.11	0.093	0.096	0.04	0.066	0.197	0.142

(ii)

Table 2: Pairwise Comparison - Prior on Tau (Row Index larger than Column Index)

	A	B	C	D	E	F	G	H
A	0.000	0.626	0.719	0.660	0.733	0.719	0.548	0.610
B	0.374	0.000	0.587	0.515	0.640	0.608	0.389	0.474
C	0.281	0.413	0.000	0.416	0.508	0.495	0.313	0.408
D	0.340	0.485	0.584	0.000	0.609	0.566	0.367	0.473
E	0.267	0.360	0.492	0.391	0.000	0.473	0.253	0.373
F	0.281	0.392	0.505	0.434	0.527	0.000	0.303	0.407
G	0.452	0.611	0.687	0.633	0.747	0.697	0.000	0.581
H	0.390	0.526	0.592	0.527	0.627	0.593	0.419	0.000

(b) In this case, instead of giving a prior distribution of τ. I will fix τ as a very big number, like $1e^6$

Table 3: Probability of Best School - Tau Very Big

A	B	C	D	E	F	G	H
0.539	0.041	0.033	0.048	0.002	0.006	0.17	0.161

Table 4: Pairwise Comparison - Tau Very Big (Row Index larger than Column Index)

	A	B	C	D	E	F	G	H
A	0.000	0.885	0.927	0.875	0.954	0.941	0.730	0.764
B	0.115	0.000	0.742	0.553	0.774	0.703	0.266	0.435
C	0.073	0.258	0.000	0.313	0.473	0.421	0.143	0.269
D	0.125	0.447	0.687	0.000	0.729	0.652	0.229	0.413
E	0.046	0.226	0.527	0.271	0.000	0.436	0.086	0.252
F	0.059	0.297	0.579	0.348	0.564	0.000	0.139	0.299
G	0.270	0.734	0.857	0.771	0.914	0.861	0.000	0.625
H	0.236	0.565	0.731	0.587	0.748	0.701	0.375	0.000

(c) If τ^2 is very large, then actually we are estimating each of them independently, so we don't consider the inner relationship between schools anymore, it's like:

$x_i \overset{ind}{\sim} N(\mu_i, \sigma_i^2)$ and calculate $P(x_i > x_j)$

Analytically: $\mu_i = y_j$, $\sigma_i^2 = (\sigma_i)^2$

$$x_i \sim N(y_i, \sigma_i^2), \quad x_j \sim N(y_j, \sigma_j^2)$$

$$x_i - x_j \sim N(y_i - y_j, \sigma_i^2 + \sigma_j^2)$$

$$P(x_i - x_j > 0) = P\Bigg(\underbrace{\frac{(x_i - x_j) - (y_i - y_j)}{\sqrt{\sigma_i^2 + \sigma_j^2}}}_{= z} > -\frac{y_i - y_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \Bigg)$$

$$z \sim N(0, 1) \Rightarrow P(x_i - x_j > 0) = 1 - \Phi\left(-\frac{y_i - y_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right)$$

(d) If we set $\tau = 0$, that means they are from the "same school". So, there are no best school because all the school are from the same "common school". And the probability that one is greater than the other should all be around 0.5. I set $\tau = 1e\text{-}20$ to get the result:

Table 5: Probability of Best School - Tau Very Small

A	B	C	D	E	F	G	H
0.12	0.123	0.13	0.144	0.123	0.122	0.117	0.122

Table 6: Pairwise Comparison - Tau Very Small (Row Index larger than Column Index)

	A	B	C	D	E	F	G	H
A	0.000	0.491	0.500	0.508	0.497	0.494	0.500	0.496
B	0.509	0.000	0.524	0.506	0.526	0.506	0.497	0.508
C	0.500	0.476	0.000	0.493	0.514	0.497	0.475	0.490
D	0.492	0.494	0.507	0.000	0.504	0.517	0.495	0.505
E	0.503	0.474	0.486	0.496	0.000	0.509	0.481	0.493
F	0.506	0.494	0.503	0.483	0.491	0.000	0.500	0.478
G	0.500	0.503	0.525	0.505	0.519	0.500	0.000	0.518
H	0.504	0.492	0.510	0.495	0.507	0.522	0.482	0.000

So the probability of each one being best is almost the same, and each one is better compared with another school is also same, 0.5.

2. BDA3 Problem 5.8

8. Discrete mixture models: if $p_m(\theta)$, for $m = 1, \ldots, M$, are conjugate prior densities for the sampling model $y|\theta$, show that the class of finite mixture prior densities given by

$$p(\theta) = \sum_{m=1}^{M} \lambda_m p_m(\theta)$$

is also a conjugate class, where the λ_m's are nonnegative weights that sum to 1. This can provide a useful extension of the natural conjugate prior family to more flexible distributional forms. As an example, use the mixture form to create a bimodal prior density for a normal mean, that is thought to be near 1, with a standard deviation of 0.5, but has a small probability of being near -1, with the same standard deviation. If the variance of each observation $y_1, \ldots, y_{10}$ is known to be 1, and their observed mean is $\overline{y} = -0.25$, derive your posterior distribution for the mean, making a sketch of both prior and posterior densities. Be careful: the prior and posterior mixture proportions are different.

$P_1(\theta) \sim N(1, 0.5^2)$ $\quad P_2(\theta) \sim N(-1, 0.5^2)$. let $\lambda_1 = 0.8, \lambda_2 = 0.2$

$$f(y|\theta) = \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi \cdot 1^2}} \exp\left(-\frac{(y_i-\theta)^2}{2}\right) = (2\pi)^{5} \cdot \exp\left(-\frac{\sum y_i^2 - 2\theta\sum_{i=1}^n y_i + n\theta^2}{2}\right)$$

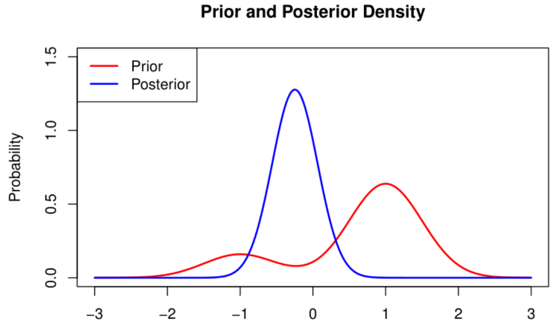
$$\pi(\theta|y) \propto \exp\left(-\frac{10\theta^2 - 2\theta\sum_{i=1}^n y_i}{2 \cdot 1^2}\right) \cdot \left(0.8\exp\left(-\frac{(\theta-1)^2}{2\times 0.5^2}\right) - \frac{(\theta+1)^2}{2\times 0.5^2} 0.2\right)$$

$$\propto 0.8\exp\left(-\frac{10\theta^2 - 20\theta\bar{y} + \frac{1}{4}\theta^2 - \frac{1}{2}\theta}{2}\right) + 0.2\exp\left(-\frac{10\theta^2 - 20\theta\bar{y} + \frac{1}{4}\theta^2 + \frac{1}{2}\theta}{2}\right)$$

$$= 0.8\exp\left(-\frac{\frac{41}{4}\theta^2 - \frac{41}{2}\bar{y}\theta}{2}\right) + 0.2\exp\left(-\frac{\frac{41}{4}\theta^2 - \frac{39}{2}\theta\bar{y}}{2}\right)$$

$$= 0.8\exp\left(-\frac{(\theta-\bar{y})^2 - \bar{y}^2}{2\times 4/41}\right) + 0.2\exp\left(-\frac{(\theta - \frac{39}{41}\bar{y})^2 - (\frac{39}{41}\bar{y})^2}{2\times 4/41}\right)$$

$$= 0.8\exp\left(\frac{41\bar{y}^2}{4}\right) N\left(\bar{y}, \frac{4}{41}\right) + 0.2\exp\left(\frac{41}{4}\cdot\left(\frac{39}{41}\right)^2\bar{y}^2\right) N\left(\frac{39}{41}\bar{y}, \frac{4}{41}\right)$$



Here is the density plot of the prior and posterior density.

3. BDA3 Problem 5.13

13. Hierarchical binomial model: Exercise 3.8 described a survey of bicycle traffic in Berkeley, California, with data displayed in Table 3.3. For this problem, restrict your attention to the first two rows of the table: residential streets labeled as 'bike routes,' which we will use to illustrate this computational exercise.

  (a) Set up a model for the data in Table 3.3 so that, for $j = 1, \dots, 10$, the observed number of bicycles at location j is binomial with unknown probability θ_j and sample size equal to the total number of vehicles (bicycles included) in that block. The parameter θ_j can be interpreted as the underlying or 'true' proportion of traffic at location j that is bicycles. (See Exercise 3.8.) Assign a beta population distribution for the parameters θ_j and a noninformative hyperprior distribution as in the rat tumor example of Section 5.3. Write down the joint posterior distribution.

  (b) Compute the marginal posterior density of the hyperparameters and draw simulations from the joint posterior distribution of the parameters and hyperparameters, as in Section 5.3.

  (c) Compare the posterior distributions of the parameters θ_j to the raw proportions, (number of bicycles / total number of vehicles) in location j. How do the inferences from the posterior distribution differ from the raw proportions?

  (d) Give a 95% posterior interval for the average underlying proportion of traffic that is bicycles.

  (e) A new city block is sampled at random and is a residential street with a bike route. In an hour of observation, 100 vehicles of all kinds go by. Give a 95% posterior interval for the number of those vehicles that are bicycles. Discuss how much you trust this interval in application.

  (f) Was the beta distribution for the θ_j's reasonable?

(a) $f(y_j \mid \theta_j) \sim \text{Binomial}(n_j, \theta_j)$

$\theta_j \mid \alpha, \beta \sim \text{Beta}(\alpha, \beta)$

$\pi(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}$

$$\pi(\theta_j, \alpha, \beta \mid y_j) \propto \theta_j^{y_j}(1-\theta_j)^{n_j-y_j} \cdot \left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]^{-1} \cdot \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}(\alpha+\beta)^{-\frac{5}{2}}$$

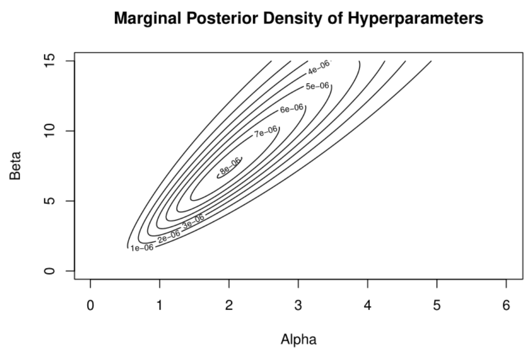
(b) $$\pi(\alpha, \beta \mid y_j) \propto \prod_j \left[\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)^{-1} \int_0^1 \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}\, d\theta_j\right]$$

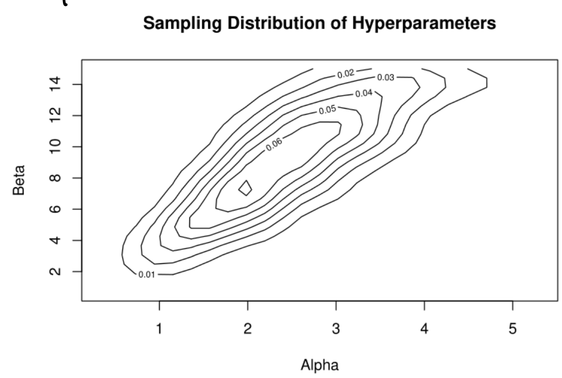
$$\cdot (\alpha+\beta)^{-\frac{5}{2}}$$

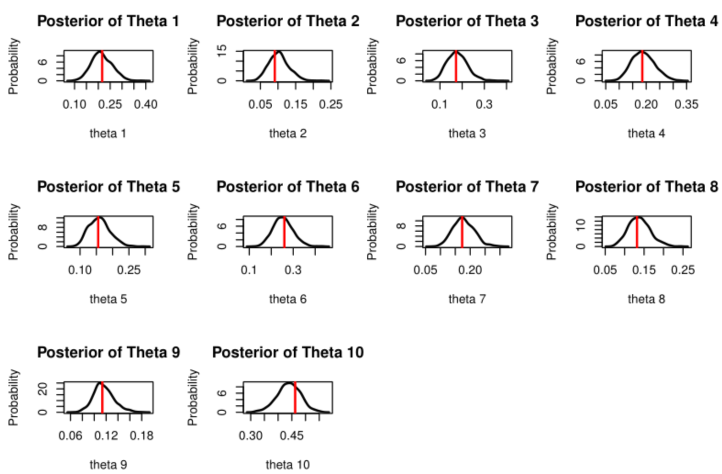
$$= (\alpha+\beta)^{-\frac{5}{2}} \left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)^{-N} \prod_{i=1}^{N}\left(\frac{\Gamma(\alpha+\beta+n_j)}{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}\right)^{-1}$$

$$\pi(\theta_j|\alpha,\beta,y) \propto \theta_j^{y_j+\alpha-1}(1-\theta_j)^{n_j-y_j+\beta-1}$$

Method: Draw (α,β) from $\pi(\alpha,\beta|y)$, then draw $\theta|\alpha,\beta,y$ based on the sample (α,β)







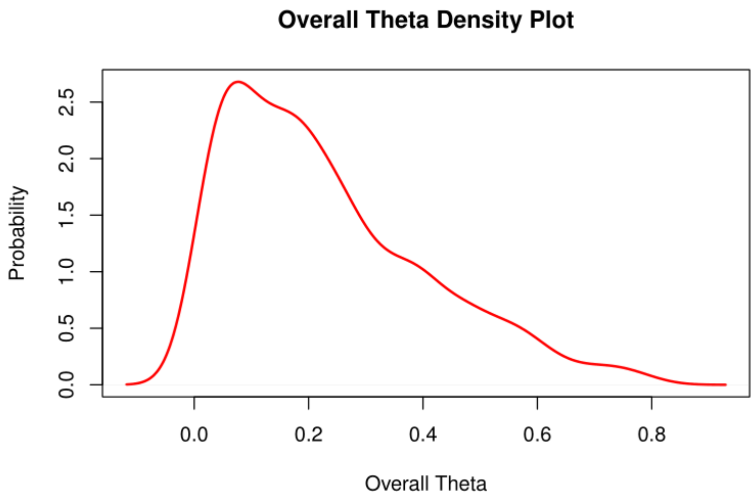
(c) The red line indicates the true θ_j, and black is my posterior distribution.

It seems most of the red line is around the mean or mode of this distribution. But for θ_{10}, when the true fraction is very large, the posterior doesn't look good but also okay because of the affect of prior.

(d). We have got a posterior sample from $\pi(\alpha, \beta | y)$,

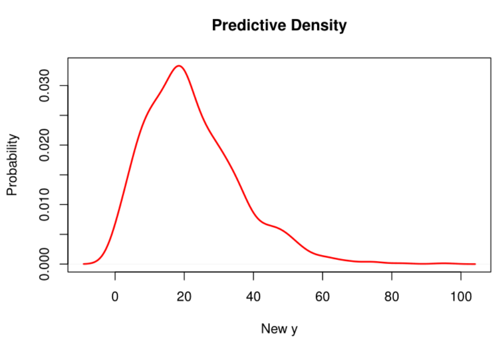
so $\theta_j | \alpha, \beta, y \sim Beta(\alpha_{pos}, \beta_{pos})$

We can use α, β to generate overall somple of θ_j.




```
##   2.5% 97.5%
## 0.015 0.653
```

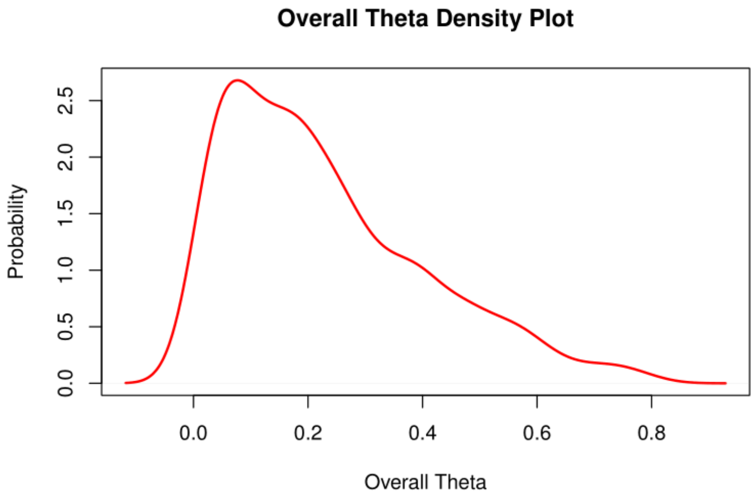
(e) Generate another (α, β) pair then generate another θ_j then generage $y_j \sim Bin(n_j, \theta_j)$




```
##  2.5% 97.5%
##     2    53
```

95% credible interve) should be from 2 to 53. It's so wide, so maybe I will not trust it when I am doing a relatively accurate estimation

(f)




```
##  2.5% 97.5%
## 0.015 0.653
```

I think it's reasonable since $\theta_j \in (0,1)$ and we can adjust the shape by adjusting α and β. However, I believe a mixture of beta distribution will be better since we can see that some θ_j is very big compared with others.

4. Consider the time between consecutive eruptions of a volcano, $t_1, ..., t_n$. Assume they follow a exponential distribution with parameters λ_i, i.e.,

$$p(t_i|\lambda_i) = \lambda_i \exp(-\lambda_i t_i)\mathbf{1}_{t_i>0}$$

Assume the following priors

$$\lambda_i \sim_{ind} Gamma(\alpha, \beta), \quad i = 1, ..., n$$

and that $\alpha \sim Gamma(a_\alpha, b_\alpha)$ and $\beta \sim Gamma(a_\beta, b_\beta)$. *For this question, you do not need to implement the sampler.*

(a) Obtain the full conditional of $\lambda_i, i = 1, ..., n$.

(b) Obtain $p(\alpha, \beta|t_1, ..., t_n)$.

(c) How would you draw samples from the joint posterior distributions of all the parameters in this model?

(d) How would you infer the time to the next eruption?

(a). $f(t_i|\lambda_i) = \lambda_i \exp(-\lambda_i t_i)$ $\quad f(\vec{t}|\vec{\lambda}) = \prod_{i=1}^n \lambda_i \exp(-\sum_{i=1}^n \lambda_i t_i)$

$\pi(\lambda_i|\alpha.\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} \exp(-\beta\lambda_i)$

$\pi(\alpha) \propto \alpha^{a_\alpha-1} \exp(-b_\alpha \alpha)$

$\pi(\beta) \propto \beta^{a_\beta-1} \exp(-b_\beta \beta)$

$\pi(\alpha.\beta, \vec{\lambda} | \vec{t}) \propto \prod_{i=1}^n \lambda_i \exp(-\sum_{i=1}^n \lambda_i t_i) \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \left(\prod_{i=1}^n \lambda_i\right)^{\alpha-1} \exp(-\beta\sum_{i=1}^n \lambda_i)\cdot$

$\alpha^{a_\alpha-1} \exp(-b_\alpha\alpha)\, \beta^{a_\beta-1} \exp(-b_\beta\beta)$

$= \left(\prod_{i=1}^n \lambda_i\right)^\alpha \exp\left(-\sum_{i=1}^n \lambda_i(t_i+\beta)\right) \cdot \left(\frac{\beta^\alpha}{\Gamma\alpha}\right)^n \alpha^{a_\alpha-1} \beta^{a_\beta-1} \exp(-b_\alpha\alpha - b_\beta\beta)$

$\pi(\lambda_i|\text{others}) \propto \lambda_i^\alpha \exp(-\lambda_i(t_i+\beta))$

$\sim \text{Gamma}(\alpha+1, \beta+t_i)$

(b) $p(\alpha,\beta \mid t) \propto \int_0^\infty \cdots \int_0^\infty \pi(\vec{\lambda}, \alpha, \beta \mid \vec{t})\, d\lambda_1 d\lambda_2 \cdots d\lambda_n$

$$\propto \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \alpha^{a_\alpha - 1} \beta^{a_\beta - 1} \exp(-b_\alpha \alpha - b_\beta \beta) \prod_{i=1}^n \int_0^\infty \lambda_i^\alpha \exp(-\lambda_i(\beta + t_i))\, d\lambda_i$$

$$= \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \alpha^{a_\alpha - 1} \beta^{a_\beta - 1} \exp(-b_\alpha \alpha - b_\beta \beta) \prod_{i=1}^n \frac{\Gamma(\alpha+1)}{(\beta + t_i)^{\alpha+1}}$$

$$= \frac{\beta^{\alpha n}}{\prod_{i=1}^n (\beta + t_i)^{\alpha+1}} \cdot \alpha^{a_\alpha + n - 1} \beta^{a_\beta - 1} \exp(-b_\alpha \alpha - b_\beta \beta)$$

(c) first, since I have the marginal distribution of $p(\alpha,\beta \mid t)$, I can use grid approximation method or Metropolis-Hasting since it's not in a closed form. Then, I get $(\alpha,\beta)^{(1)}$, $(\alpha,\beta)^{(2)}$ .... $(\alpha,\beta)^N$, N is the sample size. Then, based on each $(\alpha,\beta)^{(i)}$, I can sample $\lambda_i^{(i)}$ by full conditional distribution, $\text{Gamma}(\alpha^{(i)}+1, \beta^{(i)}+t_i)$.

Then we got $(\alpha, \beta, \vec{\lambda})^{(i)}$, it's a joint posterior.

(d). We want to make a prediction, so we get M posterior samples $(\alpha,\beta)^{(1,2\ldots M)}$, then generate a new $\lambda^{(i)} \mid (\alpha,\beta)^{(i)}$ by $\lambda \sim \text{Gamma}(\alpha^{(i)}, \beta^{(i)})$ then Generate $t^{(i)} \mid \lambda^{(i)} \sim \text{expo}(\lambda^{(i)})$. then we get a predictive sample, based on which, we can make inference like 95% credible interval and mean or variance.

5. In this question, we will analyze simulated movie rating data using a Normal-Inverse-Wishart model. Consider the dataset being $X \in \mathbb{R}^{n \times p}$ for n raters and p movies. Assume every rater assigns a score to each of the p movies and the score has been pre-processed so that they can be modeled as Normally distributed random variables. We consider the model

$$X_{ij} \sim_{ind} N(u_i^T v_j, \sigma^2), \quad i = 1, ..., n, j = 1, ..., p$$

where $u_i \in \mathbb{R}^K$ is a K dimensional latent user feature vector, and $v_j \in \mathbb{R}^K$ is a K dimensional latent movie feature vector. We put independent Gaussian priors on u_i and v_j so that

$$u_i \sim_{ind} N(\mu_u, \Sigma_u), \quad i = 1, ..., n \qquad \text{iid?}$$

?

$$v_j \sim_{ind} N(\mu_v, \Sigma_v), \quad j = 1, ..., p$$

For the hyperpriors, we let

? $\to \Sigma_\mu$?

$$\mu|\Sigma_\mu ? \quad \mu_u \sim N(\mu_0, \Sigma_u/\kappa_0), \quad \Lambda_u \sim InvWishart(\nu_0, \Lambda_0)$$

$$v|\Sigma_v ? \quad \mu_v \sim N(\mu_0, \Sigma_v/\kappa_0), \quad \Lambda_v \sim InvWishart(\nu_0, \Lambda_0)$$

with fixed parameters $\mu_0 = 0, \nu_0 = p+1, \Lambda_0 = I, \sigma = 1$. For this particular problem, consider the simulated data in *SimMovieRating.csv* on Canvas with $n = 100$ raters and $p = 30$ movies.

(a) Write out the steps of a Gibbs sampler to analyze the data.

(b) Implement the sampler for $K = 2, 4, 6, 8, 10$ and plot the posterior means of the $p \times K$ matrix $V = (v_1, v_2, ..., v_p)$ in each case.

(a) $u_i : k \times 1 \quad v_i : k \times 1$

$$\pi(u_i | \mu_u, \Sigma_u) \propto |\Sigma_u|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(u_i - \mu_u)^T \Sigma_\mu^{-1}(u_i - \mu_u)\right)$$

$$\pi(v_i | \mu_v, \Sigma_v) \propto |\Sigma_v|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v_i - \mu_v)^T \Sigma_v^{-1}(v_i - \mu_v)\right)$$

$$\pi(\mu_u) \propto \left|\frac{\Sigma_u}{k_0}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu_u - \mu_0)^T k_0 \Sigma_u^{-1}(\mu_u - \mu_0)\right)$$

$$\pi(\mu_v) \propto \left|\frac{\Sigma_v}{k_0}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu_v - \mu_0)^T k_0 \Sigma_v^{-1}(\mu_v - \mu_0)\right)$$

$$\pi(\Sigma_u) \propto |\Sigma_u|^{-\frac{\nu_0 + K + 1}{2}} \exp\left(-\frac{1}{2} tr(\Lambda_0 \Sigma_\mu^{-1})\right)$$

$$\pi(\Sigma_v) \propto |\Sigma_v|^{-\frac{\nu_0 + k + 1}{2}} \exp\left(-\frac{1}{2} tr(\Lambda_0 \Sigma_v^{-1})\right)$$

$$P(x_{ij} \mid u_i, v_j) \propto \exp\left(-\frac{(x_{ij} - u_i^T v_j)^2}{2}\right), \text{ given } \sigma^2 = 1$$

When $\Lambda_0 = I$, $\mu_0 = 0$, $\nu_0 = p+1$ we have:

$$\pi(\mu_u) \propto |\Sigma_u|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mu_u^T k_0 \Sigma_u^{-1} \mu_u\right)$$

$$\pi(\mu_v) \propto |\Sigma_v|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mu_v^T k_0 \Sigma_v^{-1} \mu_v\right)$$

$$\pi(\Sigma_u) \propto |\Sigma_u|^{-\frac{p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr(\Sigma_u^{-1})\right)$$

$$\pi(\Sigma_v) \propto |\Sigma_v|^{-\frac{p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr(\Sigma_v^{-1})\right)$$

$$\pi(u, v, \mu_u, \mu_v, \Sigma_u, \Sigma_v \mid x) \propto \exp\left(-\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - u_i^T v_j)}{2}\right)$$

$$|\Sigma_u|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu_u)^T k_0 \Sigma_u^{-1}(\mu_u)\right) |\Sigma_v|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu_v)^T k_0 \Sigma_v^{-1}(\mu_v)\right)$$

$$\prod_{i=1}^{n}\left(|\Sigma_u|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(u_i - \mu_u)^T \Sigma_\mu^{-1}(u_i - \mu_u)\right)\right)$$

$$\prod_{j=1}^{p}\left(|\Sigma_v|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(v_j - \mu_v)^T \Sigma_v^{-1}(v_j - \mu_v)\right)\right)$$

$$|\Sigma_u|^{-\frac{p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr(\Sigma_\mu^{-1})\right) |\Sigma_v|^{-\frac{p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr(\Sigma_v^{-1})\right)$$

Full Conditional Distribution:

$$\pi(u_i \mid \text{others}) \propto \exp\left(-\frac{\sum_{j=1}^{p}(x_{ij} - u_i^T v_j)^2}{2}\right) \exp\left(-\frac{(u_i - \mu_u)^T \Sigma_u^{-1}(u_i - \mu_u)}{2}\right)$$

$$= \exp\Bigg(\underbrace{-\frac{(x_{i\cdot} - u_i^T V)(x_{i\cdot} - u_i^T V)^T}{2} - \frac{(u_i - \mu_u)^T \Sigma_u^{-1}(u_i - \mu_u)}{2}}_{\star}\Bigg)$$

$$\star = -\frac{1}{2} tr\left((x_{i:} - u_i^T V)(x_{i:} - u_i^T V)^T + \Sigma_u^{-1} (u_i - \mu_u)^T (u_i - \mu_u) \right)$$

$$= -\frac{1}{2} tr\left(-x_{i:} V^T u_i + x_{i:} x_{i:}^T - u_i^T V x_{i:}^T + u_i^T V V^T u_i + \Sigma_u^{-1} u_i^T u_i - \Sigma_u^{-1} u_i^T \mu_u \right.$$

constant

$$\left. - \Sigma_u^{-1} \mu_u^T u_i + \Sigma_u^{-1} \mu_u^T \mu_u \right)$$

Constant

$$= -\frac{1}{2} tr\left(u_i^T V V^T u_i + \Sigma_u^{-1} u_i^T u_i - u_i^T (V x_{i:}^T + \mu_u \Sigma_u^{-1}) \right.$$

$$\left. - (x_{i:} V^T + \mu_u^T \Sigma_u^{-1}) u_i + \text{constant} \right)$$

$$= -\frac{1}{2} tr\left(u_i^T (V V^T + \Sigma_u^{-1}) u_i - u_i^T (V x_{i:}^T + \mu_u \Sigma_u^{-1}) - (x_{i:} V^T + \mu_u^T \Sigma_u^{-1}) u_i + c \right)$$

let $A^T A = V V^T + \Sigma_u^{-1}$:

$$\Rightarrow -\frac{1}{2} tr\left((A u_i)^T (A u_i) - (A u_i)^T (A^{-1})^T (V x_{i:}^T + \mu_u \Sigma_u^{-1}) - (x_{i:} V^T + \mu_u \Sigma_u) A^{-1} A u_i + c \right)$$

$$= -\frac{1}{2} tr\left(\left[A u_i - (A^{-1})^T (V x_{i:}^T + \mu_u \Sigma_u^{-1})\right]^T \left[A u_i - (A^{-1})^T (V x_{i:}^T + \mu_u \Sigma_u^{-1})\right] \right)$$

$$= -\frac{1}{2} tr\left(\left\{ A\left[u_i - (V V^T + \Sigma_u^{-1})^{-1} (V x_{i:}^T + \mu_u \Sigma_u^{-1})\right] \right\}^T \right.$$

$$\left. \left\{ A\left[u_i - (V V^T + \Sigma_u^{-1})^{-1} (V x_{i:}^T + \mu_u \Sigma_u^{-1})\right] \right\} \right)$$

$$= -\frac{1}{2} tr\left(\left[u_i - (V V^T + \Sigma_u^{-1})^{-1} (V x_{i:}^T + \mu_u \Sigma_u^{-1})\right]^T (V V^T + \Sigma_u^{-1})^{-1} \right.$$

$$\left. \left[u_i - (V V^T + \Sigma_u^{-1})^{-1} (V x_{i:}^T + \mu_u \Sigma_u^{-1})\right] \right)$$

$$\sim MVN\left((V V^T + \Sigma_u^{-1})^{-1} (V x_{i:}^T + \mu_u \Sigma_u^{-1}),\ (V V^T + \Sigma_u^{-1})^{-1} \right)$$

$$\pi(V_j \mid \text{others}) \propto \exp\left(-\frac{\sum_{j=1}^{p}(X_{ij} - u_i^T V_j)}{2}\right)\exp\left(-\frac{1}{2}(V_j - \mu_v)^T \Sigma_V^{-1}(V_j - \mu_v)\right)$$

$$\propto \exp\Big(-\frac{1}{2}\underbrace{\Big[(X_{\cdot j} - U^T V_j)^T(X_{\cdot j} - U^T V_j) + (V_j - \mu_v)^T \Sigma_v^{-1}(V_j - \mu_v)\Big]}_{\star\star}\Big)$$

$\star\star = tr(\star\star)$ since it's a scalar:

$$\star\star = tr\Big(-V_j^T U X_{\cdot j} + V_j^T U U^T V_j - X_{\cdot j}^T U^T V_j + V_j^T \Sigma_V^{-1} V_j$$

$$- \mu_V^T \Sigma_V^{-1} V_j - V_j \Sigma_V^{-1} \mu_V + \mu_V^T \Sigma_V^{-1} \mu_V + X_{\cdot j}^T X_{\cdot j}\Big)$$

constant

$$= tr\Big(V_j^T(UU^T + \Sigma_V^{-1})V_j - V_j^T(UX_{\cdot j} + \Sigma_V^{-1}\mu_u)$$

$$- (X_{\cdot j}^T U^T + \mu_u^T \Sigma_V^{-1})V_j + C \Big)$$

let $B^T B = UU^T + \Sigma_V^{-1}$

$$\star\star = tr\Big((BV_j)^T(BV_j) - (BV_j)^T(B^T)^{-1}(UX_{\cdot j} + \Sigma_V^{-1}\mu_u)$$

$$- (X_{\cdot j}^T U^T + \mu_V^T \Sigma_V^{-1}) B^{-1}(BV_j) + C \Big)$$

$$= tr\Big(\big[BV_j - (B^T)^{-1}(UX_{\cdot j} + \Sigma_V^{-1}\mu_V)\big]^T\big[BV_j - (B^T)^{-1}(UX_{\cdot j} + \Sigma_V^{-1}\mu_V)\big]\Big)$$

$$= tr\Big(\big[V_j - (UU^T + \Sigma_V^{-1})^{-1}(UX_{\cdot j} + \Sigma_V^{-1}\mu_V)\big]^T (UU^T \Sigma_V^{-1})^{-1}$$

$$\big[V_j - (UU^T + \Sigma_V^{-1})^{-1}(UX_{\cdot j} + \Sigma_V^{-1}\mu_V)\big]\Big)$$

$$\sim MVN: \Big((UU^T + \Sigma_V^{-1})^{-1}(UX_{\cdot j} + \Sigma_V^{-1}\mu_V),\ (UU^T + \Sigma_V^{-1})^{-1} \Big)$$

$$\pi(\mu_u \mid \text{others}) \propto \exp\left(-\frac{1}{2}\left(\underbrace{\sum_{i=1}^{n} (u_i-\mu_u)^T \Sigma_u^{-1} (u_i-\mu_u) + \mu_u^T k_0 \Sigma_u^{-1} \mu_u}_{\star\star\star}\right)\right)$$

$$\star\star\star = n \cdot \mu_u^T \Sigma_u^{-1} \mu_u + k_0 \mu_u^T \Sigma_u^{-1} \mu_u$$

$$- \sum_{i=1}^{n} u_i^T \Sigma_u^{-1} \mu_u - \sum_{i=1}^{n} \mu_u^T \Sigma_u^{-1} u_i + C$$

$$= (n+k_0) \mu_u^T \Sigma_u^{-1} \mu_u - \mu_u^T \left(\Sigma_u^{-1} \sum_{i=1}^{n} u_i\right) - \left(\sum_{i=1}^{n} u_i^T \Sigma_u^{-1}\right) \mu_u + C$$

$$= (n+k_0)\left(\mu_u^T \Sigma_n^{-1} \mu_u - \mu_u^T \Sigma_n^{-1} \frac{\sum_{i=1}^{n} u_i}{n+k_0} - \frac{\sum_{i=1}^{n} u_i}{n+k_0} \Sigma_n^{-1} \mu_u + c\right)$$

$$= tr\left((n+k_0) \Sigma_n^{-1} \left[\left(\mu_u - \frac{\sum_{i=1}^{n} u_i}{n+k_0}\right)^T \left(\mu_u - \frac{\sum_{i=1}^{n} u_i}{n+k_0}\right)\right]\right)$$

$$\sim MVN\left(\frac{\sum_{i=1}^{n} u_i}{n+k_0}, \frac{\Sigma_u^{-1}}{n+k_0}\right)$$

$$\pi(\mu_v \mid \text{others}) \propto \exp\left[-\frac{1}{2}\left(\underbrace{\mu_v^T k_0 \Sigma_v^{-1} \mu_v + \sum_{j=1}^{p} (v_j-\mu_v)^T \Sigma_v^{-1} (v_j-\mu_v)}_{\star\star\star\star}\right)\right]$$

$$\star\star\star\star = (n+k_0) \mu_v^T \Sigma_v^{-1} \mu_v - \sum_{j=1}^{p} v_j^T \Sigma_v^{-1} \mu_v - \mu_v^T \Sigma_v^{-1} \sum_{j=1}^{p} v_j + C$$

$$\sim MVN\left(\frac{\sum_{j=1}^{p} v_j}{n+k_0}, \frac{\Sigma_v^{-1}}{n+k_0}\right)$$

$$\pi(\Sigma_u \mid \text{others}) \propto |\Sigma_u|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mu_u)^T k_0 \Sigma_u^{-1} (\mu_u)\right)$$

$$\prod_{i=1}^{n} \left(|\Sigma_u|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (u_i-\mu_u)^T \Sigma_u^{-1} (u_i-\mu_u)\right)\right)$$

$$|\Sigma_u|^{-\frac{p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr\left(\Sigma_u^{-1}\right)\right)$$

$$\propto |\Sigma_u|^{-\frac{1+n+p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr\left(\Sigma_u^{-1} \left(k_0 \mu_u^T \mu_u + \sum_{i=1}^{n} (\mu_i-\mu_u)^T (\mu_i-\mu_u) + I\right)\right)\right)$$

$\sim \text{Inv-Wishart} : \left(p+n+2,\ k_0 \mu_u^T \mu_u + \sum_{i=1}^{n} (u_i - \mu_u)^T (u_i - \mu_u) + I \right)$

$\pi(\Sigma_v \mid \text{others}) \propto |\Sigma_v|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\mu_v)^T k_0 \Sigma_v^{-1} (\mu_v) \right)$

$\prod_{j=1}^{p} \left(|\Sigma_v|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (v_j - \mu_v)^T \Sigma_v^{-1} (v_j - \mu_v) \right) \right)$

$|\Sigma_v|^{-\frac{p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr(\Sigma_v^{-1}) \right)$

$\propto |\Sigma_v|^{-\frac{1+p+p+1+k+1}{2}} \exp\left(-\frac{1}{2} tr \left(\Sigma_v^{-1} \left(k_0 \mu_v^T \mu_v + \sum_{j=1}^{p} (v_j - \mu_v)^T (v_j - \mu_v) + I \right) \right) \right)$

$\sim \text{Inv-Wishart} : \left(2p+2,\ k_0 \mu_v^T \mu_v + \sum_{j=1}^{p} (v_j - \mu_v)^T (v_j - \mu_v) + I \right)$

In all:

$u_i \mid \text{others} \sim MVN \left((VV^T + \Sigma_u^{-1})^{-1} (VX_{i\cdot}^T + \mu_u \Sigma_u^{-1}),\ (VV^T + \Sigma_u^{-1})^{-1} \right)$

$v_j \mid \text{others} \sim MVN : \left((uu^T + \Sigma_v^{-1})^{-1} (uX_{\cdot j} + \Sigma_v^{-1} \mu_v),\ (uu^T + \Sigma_v^{-1})^{-1} \right)$

$\mu_u \mid \text{others} \sim MVN \left(\frac{\sum_{i=1}^{n} u_i}{n+k_0},\ \frac{\Sigma_u^{-1}}{n+k_0} \right)$

$\mu_v \mid \text{others} \sim MVV \left(\frac{\sum_{j=1}^{p} v_j}{n+k_0},\ \frac{\Sigma_v^{-1}}{n+k_0} \right)$

$\Sigma_u \mid \text{others} \sim \text{Inv-Wishart} : \left(p+n+2,\ k_0 \mu_u^T \mu_u + \sum_{i=1}^{n} (u_i - \mu_u)^T (u_i - \mu_u) + I \right)$

$\Sigma_v \mid \text{others} \sim \text{Inv-Wishart} : \left(2p+2,\ k_0 \mu_v^T \mu_v + \sum_{j=1}^{p} (v_j - \mu_v)^T (v_j - \mu_v) + I \right)$

Tables are in another file.

6. In this question, we will analyze data on 10 power plant pumps using a Poisson gamma model. The number of failures Y_i is assumed to follow a Poisson distribution

$$Y_i|\theta_i \sim_{ind} Poisson(\theta_i t_i), \quad i = 1, ..., 10$$

where θ_i is the failure rate for pump i and t_i is the length of operation time of the pump (in 1000s of hours). The data is as follows:

pump	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
y_i	5	1	5	14	3	19	1	1	4	22

We consider the conjugate gamma prior distribution for the failure rate

$$\theta_i|\alpha, \beta \sim_{ind} Gamma(\alpha, \beta), \quad i = 1, ..., 10$$

with hyperpriors $\alpha \sim Exp(1)$ and $\beta \sim Gamma(0.1, 1.0)$. All Gamma distribution uses the shape and rate parameters.

(a) Write out the steps of a Metropolis-Hastings within Gibbs sampling algorithm to analyze these data.

(b) Apply the algorithm to the data and show histograms of the posterior marginal distributions for α and β, and a scatter plot of the bivariate posterior distribution.

(c) Analytically integrate θ_i from the posterior and derive (up to proportionality) the posterior $p(\alpha, \beta|y)$.

(d) Construct a Metropolis-Hastings algorithm to sample from the posterior $p(\alpha, \beta|y)$ without sampling θ_i.

(e) Repeat part (b) using this reduced sampler.

(f) Describe how you can draw samples from $p(\theta_i|y)$ from the reduced sampler

(a) We need to get the posterior first:

$\pi(\theta_i|\alpha,\beta) = \theta_i^{\alpha-1} \exp(-\beta\theta_i) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}$ $\quad \pi(\alpha) \propto \exp(-\alpha)$

$f(Y_i|\theta_i) = \frac{(\theta_i t_i)^{y_i}}{y_i!} e^{-\theta_i t_i}$ $\quad \pi(\beta) \propto \beta^{a-1} \exp(-b\beta)$

where $a = 0.1, \; b = 10$

$$\pi(\theta, \alpha, \beta \mid y_i, t_i) \propto \prod_{i=1}^{n} \theta_i^{\alpha-1} \exp(-\beta\theta_i)(\theta_i)^{y_i} \exp(-\theta_i t_i) \exp(-\alpha)\, \beta^{a-1} \exp(-b\beta) \cdot \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{n}$$

Full Conditional:

$$\pi(\theta_i \mid y_i, t_i) \propto \theta_i^{\alpha+y_i-1} \exp(-(\beta+t_i)\theta_i)$$

$$\sim \text{Gamma}(\alpha+y_i, \beta+t_i)$$

$$\pi(\alpha \mid \text{others}) \propto \left[\prod_{i=1}^{n} \theta_i\right]^{\alpha-1} \exp(-\alpha)\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{n}$$

$$\pi(\beta \mid \text{others}) \propto \exp\left(-\left(b+\sum_{i=1}^{n}\theta_i\right)\beta\right) \cdot \beta^{a+n\alpha-1}$$

Since $\alpha, \beta > 0$, we get $t_\alpha = \log\alpha$ when we use MH

$$t_\alpha = \log\alpha \Rightarrow \alpha = \exp(t_\alpha) \quad \frac{\partial\alpha}{\partial t_\alpha} = \exp(t_\alpha)$$

$$\pi(t_\alpha \mid \text{others}) \propto \left(\prod_{i=1}^{n}\theta_i\right)^{[\exp(t_\alpha)-1]} \exp(-\exp(t_\alpha)) \cdot \frac{\beta^{\exp(t_\alpha)}}{\Gamma(\exp(t_\alpha))} \cdot \exp(t_\alpha)$$

$$\propto \left(\prod_{i=1}^{n}\theta_i\right)^{(\exp(t_\alpha)-1)} \exp(t_\alpha - \exp(t_\alpha)) \cdot \frac{\beta^{\exp(t_\alpha)}}{\Gamma(\exp(t_\alpha))}$$

$$\pi(\beta \mid \text{others}) \sim \text{Gamma}\left(a+\alpha, b+\sum_{i=1}^{n}\theta_i\right)$$

Step 1: Initialize. $\alpha^{(1)}, \beta^{(1)}, \theta_i^{(1)}$

Step 2: Generate $\beta^{(t+1)} \sim \text{Gamma}\left(a+\alpha^{(t)}, b+\sum_{i=1}^{n}\theta_i^{(t)}\right)$

Step 3: Generate $\log(\alpha^{(t+1)})$ by MH steps:

Random Walk proposal:

Generate $t_\alpha^* \sim N(t_\alpha^{(t)}, \sigma_t^2)$

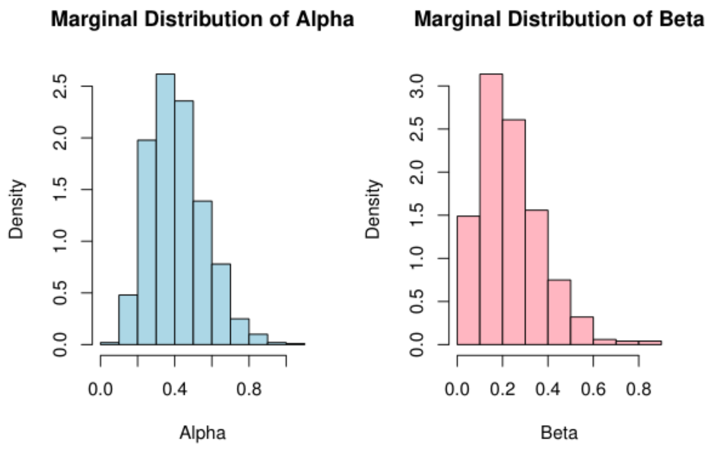
$$p = \min\left\{ \frac{\pi(t_\alpha^*)}{\pi(t_\alpha^{(t)})}, 1 \right\}$$

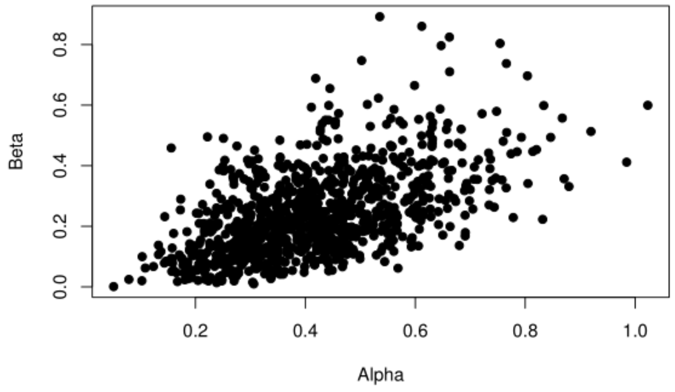
generate $\ell \sim \text{Unif}(0,1)$

if $\ell \le p$: $t_\alpha^* \to t_\alpha^{(t+1)}$, o.w. $t_\alpha^{(t)} \to t_\alpha^{(t+1)}$

Step 4: Generate $\theta_i \sim \text{Gamma}(\alpha^{(t+1)} + y_i, \beta^{(t+1)} + t_i)$

(b).





(c)

$$\pi(\theta_i|\alpha,\beta) = \theta_i^{\alpha-1}\exp(-\beta\theta_i)\cdot\frac{\beta^\alpha}{\Gamma(\alpha)} \quad \pi(\alpha)\propto\exp(-\alpha)$$

$$f(Y_i|\theta_i) = \frac{(\theta_i t_i)^{y_i}}{y_i!} e^{-(\theta_i t_i)} \quad \pi(\beta)\propto\beta^{a-1}\exp(-b\beta)$$

$$\pi(\alpha,\beta,\theta_i|Y,t)\propto\prod_{i=1}^n \theta_i^{\alpha+y_i-1}\exp(-\beta\sum_{i=1}^n\theta_i)\exp(\sum_{i=1}^n-\theta_i t_i)$$

$$\exp(-\alpha)\cdot\beta^{a-1}\exp(-b\beta)\cdot\left[\frac{\beta^\alpha}{\Gamma(\alpha)}\right]^n$$

$$\pi(\alpha,\beta|Y,t) = \int\int\cdots\int\pi(\alpha,\beta,\theta|Y,t)\,d\theta_1\cdots d\theta_n$$

$$= \beta^{a-1}\exp(-\alpha-b\beta)\left[\beta^\alpha/\Gamma(\alpha)\right]^n\cdot\prod_{i=1}^n\int\theta_i^{\alpha+y_i-1}\exp(-(\beta+t_i)\theta_i)\,d\theta_i$$

$$= \beta^{a-1}\exp(-\alpha-b\beta)\left[\frac{\beta^\alpha}{\Gamma(\alpha)}\right]^n\cdot\prod_{i=1}^n\left[\frac{\Gamma(\alpha+y_i)}{(\beta+t_i)^{\alpha+y_i}}\right]$$

$$= \exp(-\alpha-b\beta)\frac{\beta^{\alpha n+a-1}}{\prod_{i=1}^n(\beta+t_i)^{\alpha+y_i}}\cdot\frac{\prod_{i=1}^n\Gamma(\alpha+y_i)}{\Gamma(\alpha)^n}$$

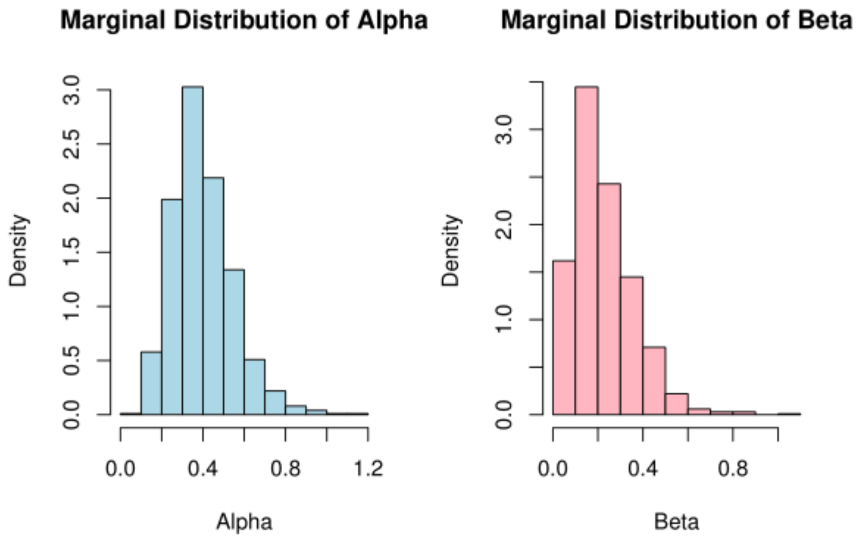
Let $t_\alpha = \log(\alpha)$, $t_\beta = \log\beta$: $\alpha = e^{t_\alpha}$ $\beta = e^{t_\beta}$

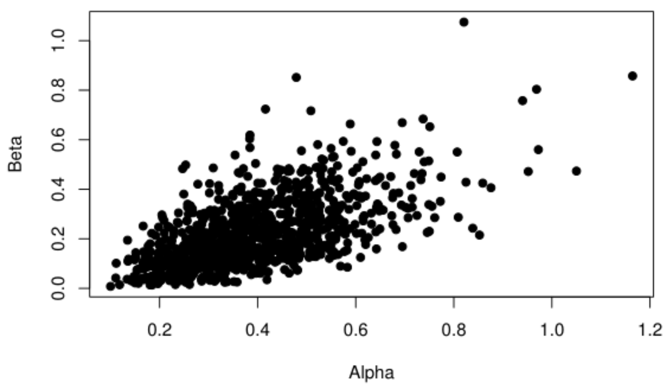
$$J = \begin{vmatrix} \partial\alpha/\partial t_\alpha & \partial\alpha/\partial t_\beta \\ \partial\beta/\partial t_\alpha & \partial\beta/\partial t_\beta \end{vmatrix} = \begin{vmatrix} e^{t_\alpha} & 0 \\ 0 & e^{t_\beta} \end{vmatrix} = e^{t_\alpha+t_\beta}$$

$$\pi(t_\alpha,t_\beta|Y,t)\propto\exp(-\exp(t_\alpha)-b\exp(t_\beta))\cdot\frac{\exp(t_\beta)^{n\cdot\exp(t_\alpha)+a-1}}{\prod_{i=1}^n(\exp(t_\beta)+t_i)^{[\exp(t_\alpha)+y_i]}}$$

$$\cdot\frac{\prod_{i=1}^n\Gamma(\exp(t_\alpha)+y_i)}{\Gamma(\exp(t_\alpha))^n}\cdot\exp(t_\alpha+t_\beta)$$

(d)





(e) After getting the sample of $(\alpha, \beta)^{(t)}$ we can generate $\theta_i \sim \text{Gamma}(\alpha^{(t)}, \beta^{(t)})$, then we get posterior θ_i.