Problem 7: Continuous mixture models.

Problem 7:

(a) Since $y|\theta \sim \text{Poisson}(\theta)$, and $\theta \sim \text{Gamma}(\alpha, \beta)$,

$$P(y|\theta) = P(y=k|\theta) = \frac{\theta^k e^{-\theta}}{k!}$$

$$P(\theta) = \frac{\theta^{\alpha-1} e^{-\beta\theta} \beta^\alpha}{\Gamma(\alpha)}$$

$$P(y=k) = \int p(y=k|\theta) P(\theta) d\theta$$

$$= \int \frac{\theta^k e^{-\theta}}{k!} \frac{\theta^{\alpha-1} e^{-\beta\theta}\beta^\alpha}{\Gamma(\alpha)} d\theta$$

$$= \frac{\beta^\alpha}{k!\,\Gamma(\alpha)} \int \theta^{\alpha+k-1} e^{-(\beta+1)\theta} d\theta$$

The mean:

$$E(y) = \sum_{k=1}^{n} k P(y=k)$$

$$= \sum_{k=1}^{n} \frac{k\beta^\alpha}{k!\,\Gamma(\alpha)} \int \theta^{\alpha+k-1} e^{-(\beta+1)\theta} d\theta$$

$$= \sum_{k=1}^{n} \frac{k\beta^\alpha}{k!\,\Gamma(\alpha)} \left[\int \frac{\theta^{\alpha+k-1} e^{-(\beta+1)\theta} (\beta+1)^{\alpha+k}}{\Gamma(\alpha+k)} d\theta\right] \frac{\Gamma(\alpha+k)}{(\beta+1)^{\alpha+k}}$$

$$= \sum_{k=1}^{n} k \cdot \frac{\beta^\alpha \Gamma(\alpha+k)}{k!\,\Gamma(\alpha)\,(\beta+1)^{\alpha+k}}$$

$$= \sum_{k=1}^{n} k \cdot \frac{\beta^\alpha (\alpha+k-1)!}{k!\,(\alpha-1)!\,(\beta+1)^{\alpha+k}}$$

$$= \sum_{k=1}^{n} k \cdot \binom{\alpha+k-1}{\alpha-1} \left(\frac{1}{\beta+1}\right)^k \left(\frac{\beta}{\beta+1}\right)^\alpha$$

$$\cancel{= (\alpha+k-1)\left(\frac{1}{\beta+1}\right)\frac{\beta}{\beta+1} = \frac{\beta(\alpha+k-1)}{(\beta+1)^2}}$$

$$= \frac{1}{\beta+1}\alpha \bigg/ \left(\frac{\beta}{\beta+1}\right) = \frac{\alpha}{\beta}$$

The variance:

$$E(y^2) = \sum_{k=1}^{n} k^2 p(y=k)$$

$$= \sum_{k=1}^{n} \frac{k^2\beta^{\alpha}}{k!\,\Gamma(\alpha)} \int \theta^{\alpha+k-1} e^{-(\beta+1)\theta}\, d\theta$$

$$= \sum_{k=1}^{n} k^2 \binom{\alpha+k-1}{\alpha-1}\left(\frac{1}{\beta+1}\right)^{k}\left(\frac{\beta}{\beta+1}\right)^{\alpha} = \left[\frac{1}{\beta+1}\alpha\Big/\left(\frac{\beta}{\beta+1}\right)^2 + \frac{\alpha^2}{\beta^2}\right]$$

$$\cancel{= \left[(\alpha+k-1)^2\left(\frac{1}{\beta+1}\right)^2 + (\alpha+k-1)\left(\frac{1}{\beta+1}\right)\left(\frac{\beta}{\beta+1}\right)\right]\frac{\beta}{\beta+1}}$$

$$\cancel{= \frac{\beta(\alpha+k-1)(\alpha+\beta+k-1)}{(\beta+1)^3}} \qquad = \frac{\alpha^2+\alpha(\beta+1)}{\beta^2}$$

So $Var(y) = E(y^2) - E^2(y) = \frac{\alpha^2+\alpha(\beta+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha(\beta+1)}{\beta^2}$.

$$\cancel{= \frac{\beta(\alpha+k-1)(\alpha+\beta+k-1)}{(\beta+1)^3} - \frac{\beta^2(\alpha+k-1)^2}{(\beta+1)^4}}$$

$$\cancel{= \frac{\beta(\beta+1)(\alpha+k-1)(\alpha+\beta+k-1) - \beta^2(\alpha+k-1)^2}{(\beta+1)^4}}$$

$$\cancel{= \frac{\beta(\alpha+k-1)(\beta^2+\alpha+\beta+k-1)}{(\beta+1)^4}}$$

(b) Since $y|(\mu,\sigma^2) \sim N(\mu,\sigma^2)$, $p(\mu,\sigma^2) \propto 1/\sigma^2$,

the posterior distribution of (μ,σ^2) is

$$p(\mu,\sigma^2|y) \propto p(y|\mu,\sigma^2)\,p(\mu,\sigma^2)$$

$$\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}\frac{1}{\sigma^2}$$

$$= (2\pi)^{-\frac{n}{2}} \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} \exp\left\{-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right\}$$

$$= (2\pi)^{-\frac{n}{2}} \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} \exp\left\{-\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

$$p(\mu|y) = \int p(\mu,\sigma^2|y)\, d\sigma^2$$

$$= \int (2\pi)^{-\frac{n}{2}} \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\} d\sigma^2$$

$$p(u|y) \propto (2\pi)^{-\frac{n}{2}} \int (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{(n-1)s^2+n(\bar{y}-u)^2}{2\sigma^2}} d\sigma^2$$

$$= (2\pi)^{-\frac{n}{2}} \left\{ \int \frac{\left(\frac{(n-1)s^2+n(\bar{y}-u)^2}{2}\right)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} (\sigma^2)^{-(\frac{n}{2}+1)} e^{-\frac{\frac{(n-1)s^2+n(\bar{y}-u)^2}{2}}{\sigma^2}} d\sigma^2 \right\} \frac{\Gamma(\frac{n}{2})}{\left(\frac{(n-1)s^2+n(\bar{y}-u)^2}{2}\right)^{\frac{n}{2}}}$$

$$= (2\pi)^{-\frac{n}{2}} \cdot \frac{\Gamma(\frac{n}{2})}{\left(\frac{(n-1)s^2+n(\bar{y}-u)^2}{2}\right)^{\frac{n}{2}}} = \pi^{-\frac{n}{2}} \cdot \frac{\Gamma(\frac{n}{2})}{((n-1)s^2+n(\bar{y}-u)^2)^{\frac{n}{2}}}$$

$$= \frac{\pi^{-\frac{n}{2}}\Gamma(\frac{n}{2})}{[(n-1)s^2]^{\frac{n}{2}}} \left(1+\frac{\left(\frac{\bar{y}-u}{s/\sqrt{n}}\right)^2}{n-1}\right)^{-\frac{n}{2}}$$

$\boxed{\cancel{u|y \sim t_{n-1}}}$

The mean is

$$\star \boxed{\frac{u-\bar{y}}{s/\sqrt{n}} \sim t_{n-1}} \Longrightarrow \begin{cases} E\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right) = 0 \\ Var\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right) = \frac{n}{n-2}, \ n>2. \end{cases}$$

$$E\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right) = \int \frac{u-\bar{y}}{s/\sqrt{n}} \cdot p(u|y)\, du$$

$$= \int \frac{u-\bar{y}}{s/\sqrt{n}} \cdot \frac{\pi^{-\frac{n}{2}}\Gamma(\frac{n}{2})}{[(n-1)s^2]^{\frac{n}{2}}} \left(1+\frac{\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)^2}{n-1}\right)^{-\frac{n}{2}} du$$

$$= \left\{ \int \frac{u-\bar{y}}{s/\sqrt{n}} \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi}\,\Gamma(\frac{n-1}{2})} \left(1+\frac{\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)^2}{n-1}\right)^{-\frac{n}{2}} d\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right) \right\} \frac{\Gamma(\frac{n-1}{2})}{[(n-1)\pi s^2]^{\frac{n-1}{2}} \cdot \sqrt{n}}$$

$$= 0$$

The variance is :

$$E\left(\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)^2\right) = \int \left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)^2 p(u|y)\, du$$

$$= \left\{ \int \left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)^2 \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi}\,\Gamma(\frac{n-1}{2})} \left(1+\frac{\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)^2}{n-1}\right)^{-\frac{n}{2}} d\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right) \right\} \frac{\Gamma(\frac{n-1}{2})}{[(n-1)\pi s^2]^{\frac{n-1}{2}} \sqrt{n}}$$

$$= \left(0 + \frac{n}{n-2}\right) \frac{\Gamma(\frac{n-1}{2})}{[(n-1)\pi s^2]^{\frac{n-1}{2}} \sqrt{n}} = \frac{\sqrt{n}\,\Gamma(\frac{n-1}{2})}{((n-1)\pi s^2)^{\frac{n-1}{2}} \cdot (n-2)} \text{ when } n>2.$$

$$Var\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right) = E\left(\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)^2\right) - E^2\left(\frac{u-\bar{y}}{s/\sqrt{n}}\right)$$

$$= \frac{\sqrt{n}\,\Gamma(\frac{n-1}{2})}{(n-2)\cdot((n-1)\pi s^2)^{\frac{n-1}{2}}}, \quad n>2.$$

Problem 8: Discrete mixture models.

Problem 8:

① Since $P_m(\theta)$, $m=1,2,\cdots,M$ are conjugate prior densities for $y|\theta$,

$P_m(\theta|y) \propto p(y|\theta)\cdot P_m(\theta)$, $m=1,2,\cdots,M$, has the same pdf family as $P_m(\theta)$

Then a class of finite mixture prior density, $P(\theta)=\sum_{m=1}^{M}\lambda_m P_m(\theta)$, where $\sum_{m=1}^{M}\lambda_m=1$

We have
$$
\begin{aligned}
P(\theta|y) &\propto P(y|\theta)P(\theta) \\
&= P(y|\theta)\sum_{m=1}^{M}\lambda_m P_m(\theta) \\
&= \sum_{m=1}^{M} P(y|\theta)\lambda_m P_m(\theta) \\
&\propto \sum_{m=1}^{M}\lambda_m P_m(\theta|y),
\end{aligned}
$$

Therefore, $P(\theta)=\sum_{m=1}^{M}\lambda_m P_m(\theta)$ where $\sum_{m=1}^{M}\lambda_m=1$ is also a conjugate class for $y|\theta$.

② A bimodal prior: $P_1(\theta)=\frac{1}{\sqrt{2\pi}\,0.5}e^{-\frac{(\theta-a)^2}{2\cdot 0.5^2}}$ $a\simeq 1$

$\theta_1 \sim N(a, 0.5^2)$, $a\simeq 1$

$P_2(\theta)=\frac{1}{\sqrt{2\pi}\,0.5}e^{-\frac{(\theta-b)^2}{2\cdot 0.5^2}}$ $b\simeq -1$

$\theta_2 \sim N(b, 0.5^2)$ $b\simeq -1$

$$
\begin{aligned}
P(\theta) &= \lambda_1 P_1(\theta)+\lambda_2 P_2(\theta) \\
&= \frac{1}{\sqrt{2\pi}\,0.5}\left[\lambda_1 e^{-\frac{(\theta-a)^2}{2\cdot 0.5^2}}+\lambda_2 e^{-\frac{(\theta-b)^2}{2\cdot 0.5^2}}\right]
\end{aligned}
$$
where $a\simeq 1, b\simeq -1$, $\lambda_1>\lambda_2$

Likelihood:
$$
\begin{aligned}
P(y|\theta) &= \prod_{i=1}^{10}\frac{1}{\sqrt{2\pi}\cdot 1}\exp\left\{-\frac{(y_i-\theta)^2}{2\cdot 1^2}\right\} \\
&= (2\pi)^{-5}\exp\left\{-\frac{\sum_{i=1}^{10}(y_i-\theta)^2}{2}\right\}
\end{aligned}
$$

Posterior:
$$
\begin{aligned}
P(\theta|y) &\propto P(y|\theta)P(\theta) \\
&= (2\pi)^{-5}\exp\left\{-\frac{\sum_{i=1}^{10}(y_i-\theta)^2}{2}\right\}\cdot\frac{1}{\sqrt{2\pi}\,0.5}\left[\lambda_1 e^{-\frac{(\theta-a)^2}{2\cdot 0.5^2}}+\lambda_2 e^{-\frac{(\theta-b)^2}{2\cdot 0.5^2}}\right]
\end{aligned}
$$

$$p(\theta|y) = 2\cdot(2\pi)^{-5.5}\exp\left\{-\frac{10\theta^2-20\bar{y}\theta+\sum_{i=1}^{10}y_i^2}{2}\right\}\left[\lambda_1 e^{-2(\theta-a)^2}+\lambda_2 e^{-2(\theta-b)^2}\right]$$

$$= 2\cdot(2\pi)^{-5.5} e^{-\frac{\sum_{i=1}^{10}y_i^2}{2}}\cdot e^{-\frac{10\theta^2+5\theta}{2}}\left[\lambda_1 e^{-2\theta^2+4a\theta-2a^2}+\lambda_2 e^{-2\theta^2+4b\theta-2b^2}\right]$$

$$= 2\cdot(2\pi)^{-5.5} e^{-\frac{\sum_{i=1}^{10}y_i^2}{2}}\left[\lambda_1 e^{-7\theta^2+(4a-2.5)\theta-2a^2}+\lambda_2 e^{-7\theta^2+(4b-2.5)\theta-2b^2}\right]$$

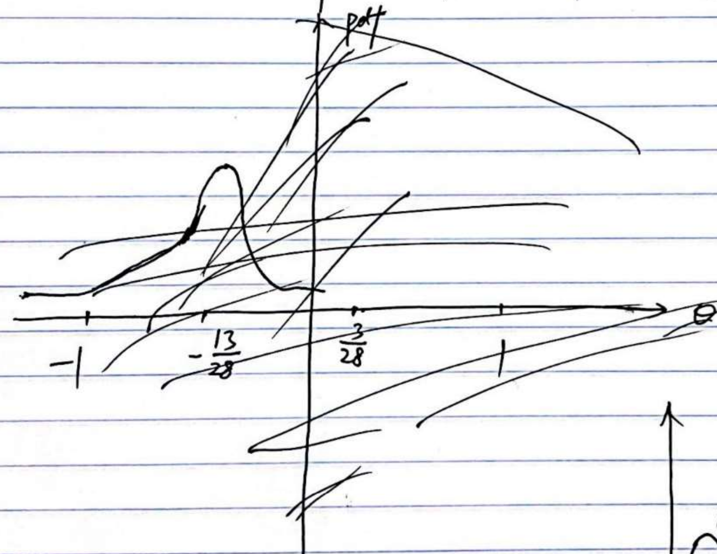
posterior distribution:

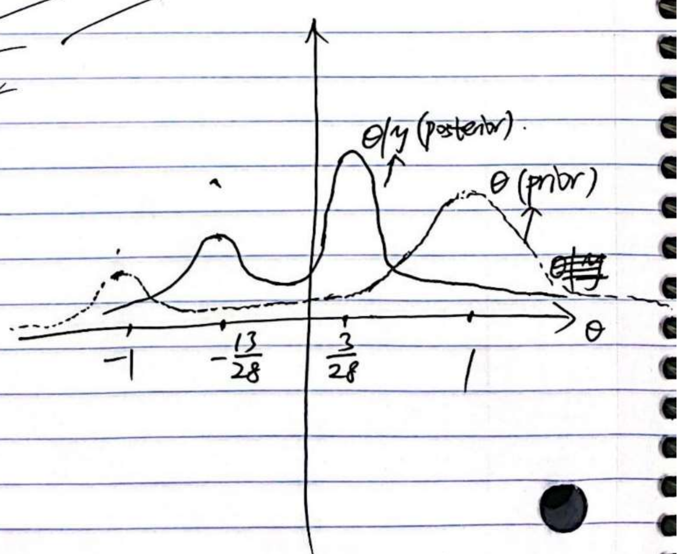
thus: $\theta|y \sim \lambda_1 N\left(\frac{2a}{7}-\frac{5}{28}, \frac{1}{14}\right)+\lambda_2 N\left(\frac{2b}{7}-\frac{5}{28}, \frac{1}{14}\right)$

Therefore:

prior density: $\theta \sim \lambda_1 N(a, \frac{1}{4})+\lambda_2 N(b, \frac{1}{4})$ where $\begin{cases} a\simeq 1 \\ b\simeq -1 \\ \lambda_1>\lambda_2 \\ \lambda_1+\lambda_2=1 \end{cases}$

So we can make a sketch of both prior and posterior densities as follows:





Problem 13: Hierarchical binomial model.

Problem 13:

(a) Our statistical model:

$$Y_j \overset{ind}{\sim} Bin(n_j, \theta_j)$$
$$\theta_j \overset{ind}{\sim} Beta(\alpha, \beta)$$
$$\alpha, \beta \sim p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}, \quad j = 1, 2, \cdots, 10$$

$\begin{cases} Y_j: \text{the observed number of bicycles at location } j \\ n_j: \text{the total number of vehicles at location } j \end{cases}$

The joint posterior distribution:

$$\begin{aligned} p(\theta, \alpha, \beta \mid y) &\propto p(y \mid \theta, \alpha, \beta)\, p(\theta, \alpha, \beta) \\ &= p(y \mid \theta) \cdot p(\theta \mid \alpha, \beta)\, p(\alpha, \beta) \\ &= \left[\prod_{j=1}^{10} p(y_j \mid \theta_j)\, p(\theta_j \mid \alpha, \beta)\right] p(\alpha, \beta) \\ &= p(\alpha, \beta) \cdot \prod_{j=1}^{10} \binom{n_j}{y_j} \theta_j^{y_j} (1-\theta_j)^{n_j - y_j} \cdot \frac{\theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}}{B(\alpha, \beta)} \\ &= p(\alpha, \beta) \prod_{j=1}^{10} \binom{n_j}{y_j} \frac{\theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}}{B(\alpha, \beta)} \\ &\propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{10} \binom{n_j}{y_j} \frac{\theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}}{B(\alpha, \beta)} \end{aligned}$$

(b) The marginal posterior density of the hyperparameters:

$$\begin{aligned} p(\alpha, \beta \mid y) &\propto p(y \mid \alpha, \beta)\, p(\alpha, \beta) \\ &= \prod_{j=1}^{10} p(y_j \mid \alpha, \beta)\, p(\alpha, \beta) \\ &= p(\alpha, \beta) \cdot \prod_{j=1}^{10} \int p(y_j \mid \theta_j)\, p(\theta_j \mid \alpha, \beta)\, d\theta_j \\ &= p(\alpha, \beta) \prod_{j=1}^{10} \int_0^1 \binom{n_j}{y_j} \frac{\theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}}{B(\alpha, \beta)}\, d\theta_j \\ &= p(\alpha, \beta) \frac{1}{B(\alpha, \beta)} \prod_{j=1}^{10} \binom{n_j}{y_j} \int_0^1 \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}\, d\theta_j \\ &= p(\alpha, \beta) \prod_{j=1}^{10} \binom{n_j}{y_j} \frac{B(\alpha+y_j, \beta+n_j-y_j)}{B(\alpha, \beta)} \\ &\propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{10} \binom{n_j}{y_j} \frac{B(\alpha+y_j, \beta+n_j-y_j)}{B(\alpha, \beta)}. \end{aligned}$$

b. (continued):

Drawing simulations from the joint posterior distribution of the parameters and hyperparameters is done through the following two steps:

(1) Draw simulations from the marginal distribution of the hyperparameters $p(\alpha, \beta|y)$ below:

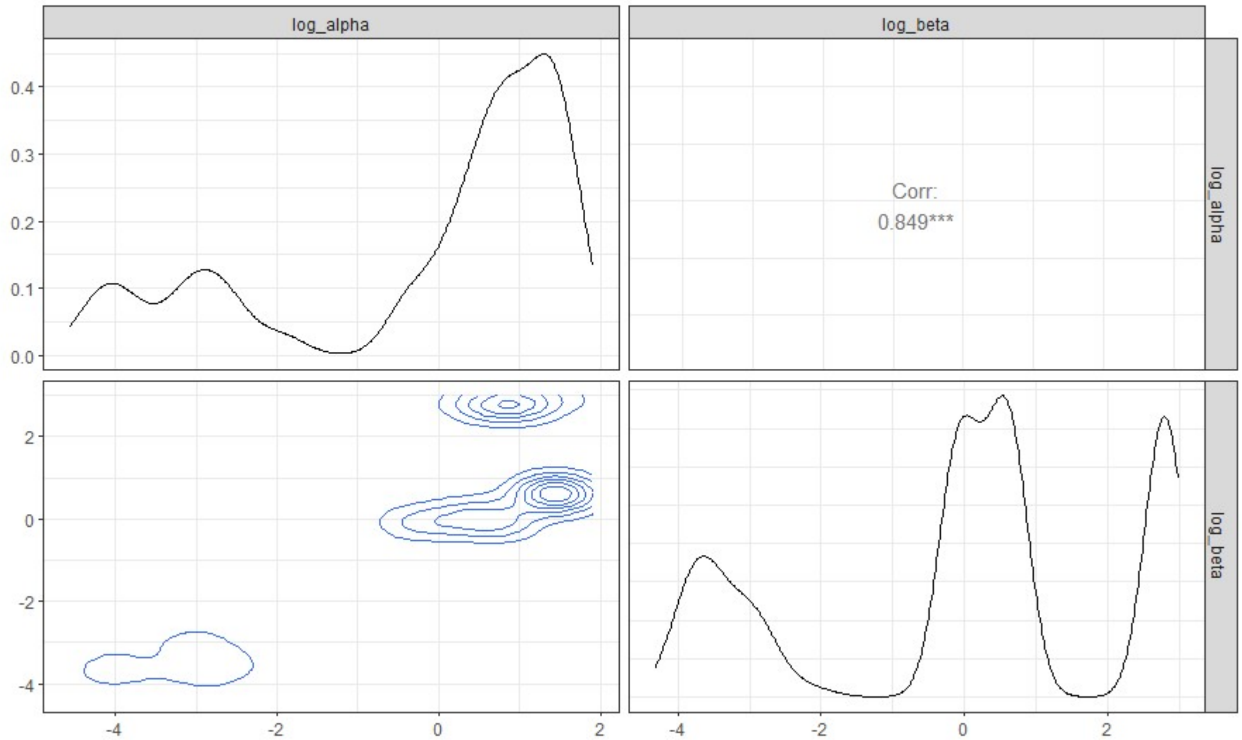


Figure 1: Sampling from the marginal distribution of the hyperparameters $p(\alpha, \beta|y)$

(2) Draw simulations from the conditional posterior density of the parameters $p(\theta|\alpha, \beta, y)$

$$p(\theta|\alpha, \beta, y) = \prod_{j=1}^{10} p(\theta_j|\alpha, \beta, y_j) = \prod_{j=1}^{10} Beta(\theta_j|\alpha + y_j, \beta + n_j - y_j)$$

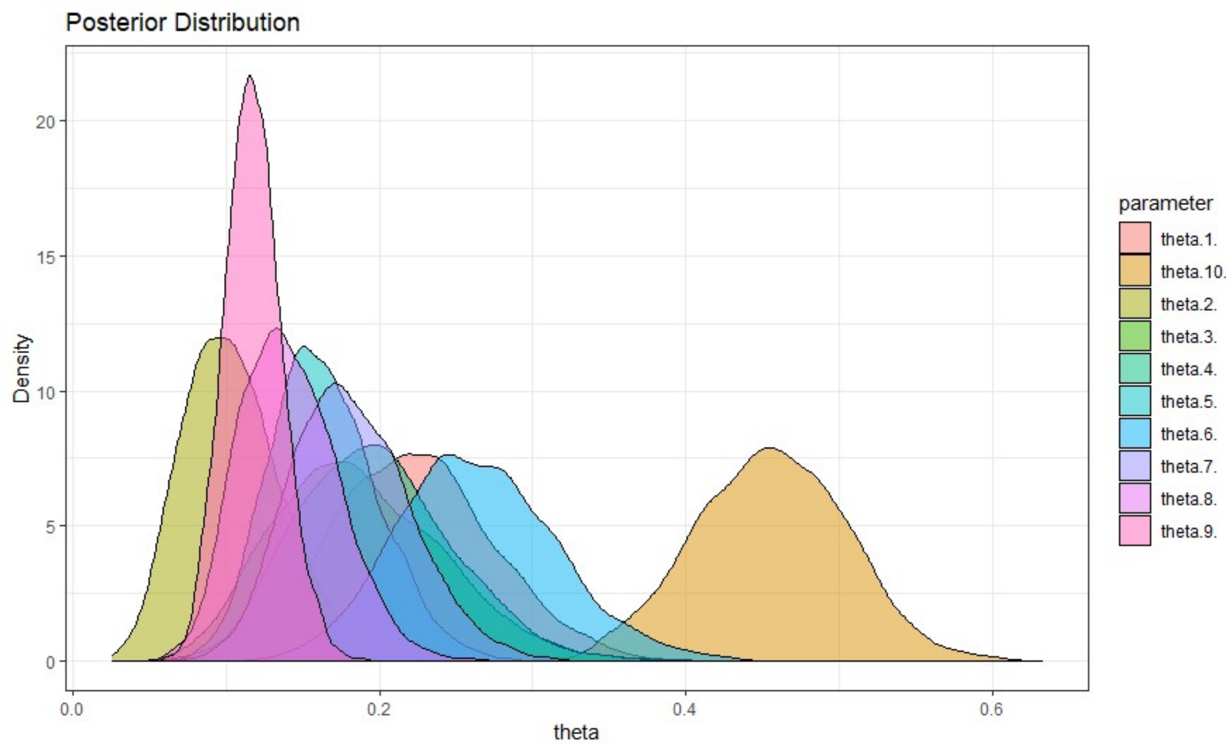


Figure 2: Sampling from the posterior density of the parameters $p(\theta|\alpha, \beta, y)$

c. Comparison

Figure 3 shows a plot of the comparison between the posterior distributions of the parameters θ_j and the raw proportions in location j. The black dots and solid lines separately denote the posterior means and 95% credible intervals of θ_j, and the corresponding x-axis values of the black dots and solid lines are raw proportions (the observed number of bicycles / total number of vehicles) in location j. Totally, the posterior means of θ_j are slightly higher than the observed raw proportions at location j, and the raw proportions of location j all fall into the 95% credible intervals of the posterior estimates.

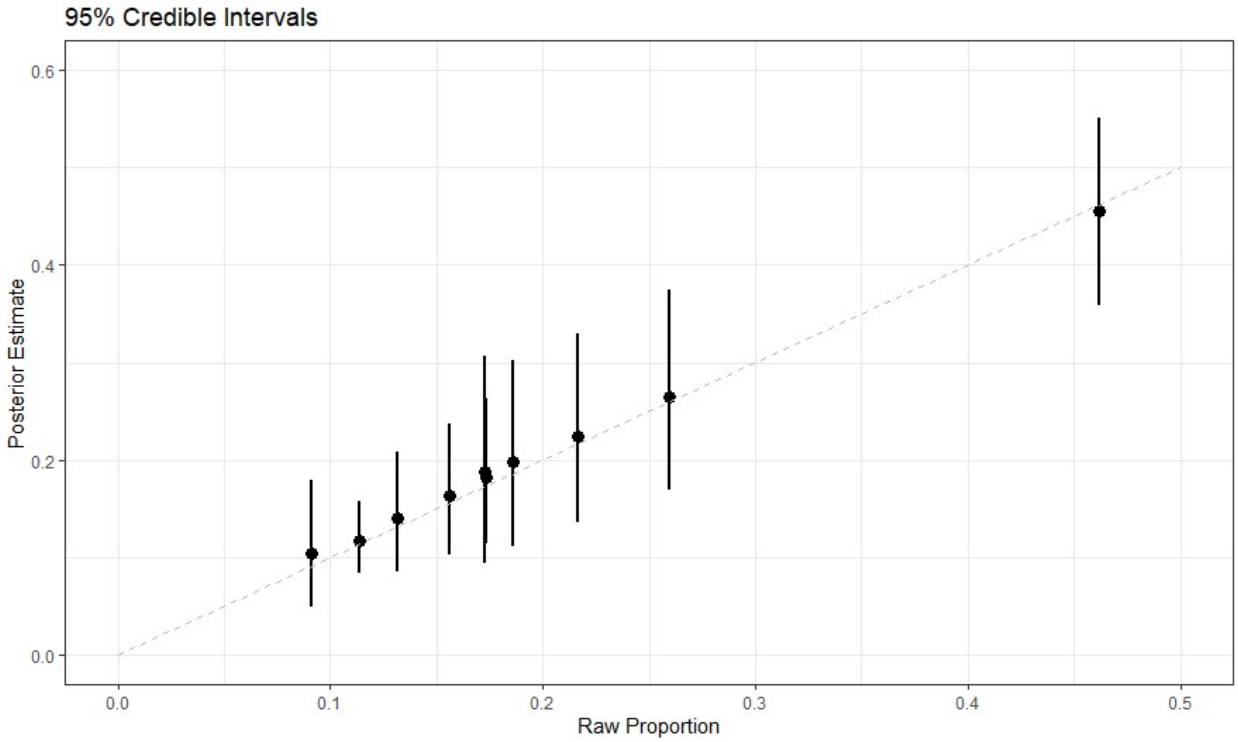


Figure 3: Posterior means and 95% credible intervals of the proporition of traffic that is bicycles θ_j (posterior estimates vs. raw proportions) based on simulations from the joint posterior distribution. The dashed line $y = x$ corresponds to the unpooled estimates $\theta_j = y_j/n_j$.

d. 95% posterior interval

Table 1 gives a 95% posterior interval for the proportions of traffic that is bicycles at location j and as a whole (all locations included). The 95% posterior interval for the average underlying proportion of traffic that is bicycles (all locations included) is [0.1613,0.2062]. These results are further reflected in Figure 4 where "theta[11]" represents the average proportion of traffic that is bicycles among all locations.

Table 1: Summary of bicycles and other vehicles in one hour in each of 10 blocks in Residential street labeled as “bike routes” and assocated statistics of posterior estimates (i.e., mean, median, and 95% credible intervals).

Block (j)	Observations			Posterior Estimates			
	Bicycles	Traffic	Proportion	mean	lcl	median	ucl
1	16	74	0.2162	0.2246	0.1361	0.2223	0.3302
2	9	99	0.0909	0.1049	0.0498	0.1018	0.1788
3	10	58	0.1724	0.1874	0.0948	0.1823	0.3060
4	13	70	0.1857	0.1976	0.1113	0.1947	0.3028
5	19	122	0.1557	0.1642	0.1028	0.1618	0.2374
6	20	77	0.2597	0.2643	0.1698	0.2614	0.3745
7	18	104	0.1731	0.1817	0.1139	0.1787	0.2625
8	17	129	0.1318	0.1411	0.0861	0.1385	0.2085
9	35	308	0.1136	0.1177	0.0836	0.1171	0.1570
10	55	119	0.4622	0.4555	0.3590	0.4561	0.5510
Total	212	1160	0.1828	0.1834	0.1613	0.1833	0.2062

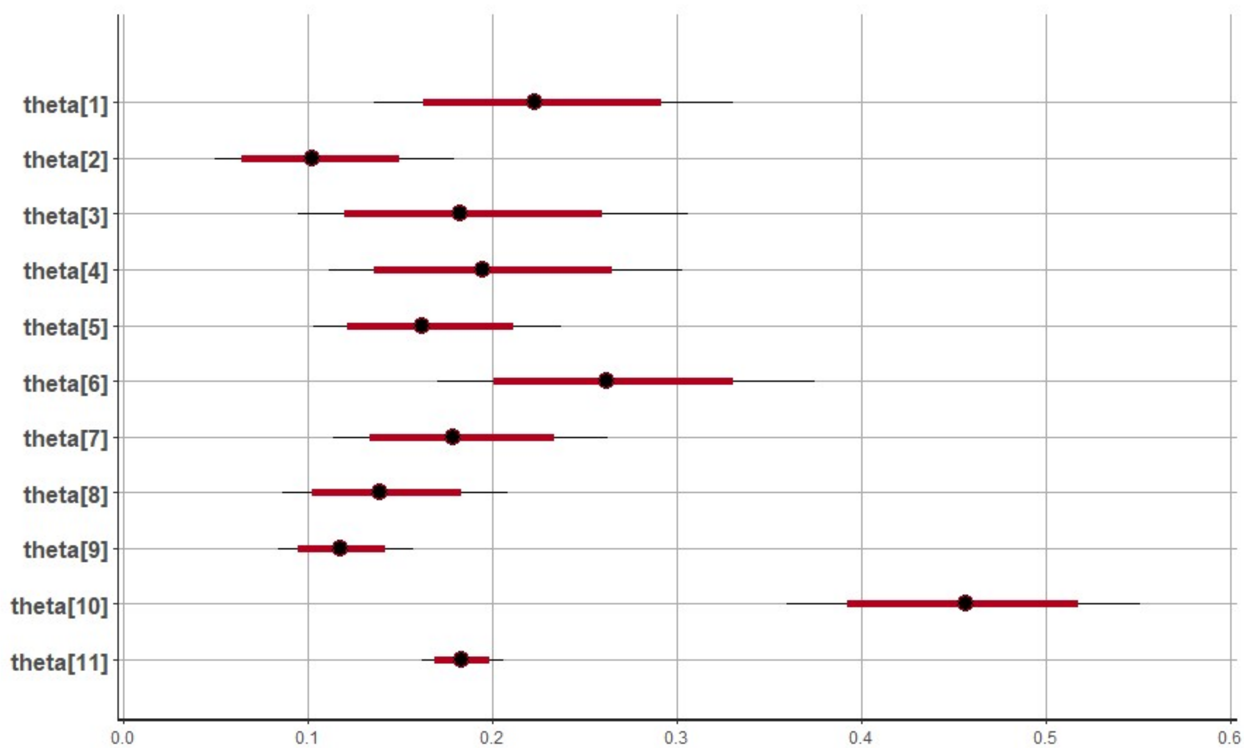


Figure 4: Mean, 80%, and 95% posterior intervals for the proportions of traffic that is bicycles at location $j = 1,2,\cdots,10$ and as a whole ($j = 11$).

e. Prediction

Since we have no idea or prior information of the exact location of a new city block and only know that the new city block is residential street with a bike route, it is more reasonable to use the posterior estimates of θ based on all locations (as a whole) than each location j. As a result, the 95% posterior interval of θ in the new city block is [0.1613,0.2062]. Therefore, when 100 vehicles of all kinds go by in an hour of observation, a 95% posterior interval for the number of those vehicles that are bicycles is [16, 21]. This is also the most possible and precise estimates of the number of bicycles going by that we can make based on the existing information.

f. Discussion

Although we cannot exactly know what the prior distribution of parameters θ is, we think that the beta distribution for the θ_j's is generally reasonable because (1) the posterior means of parameters θ_j at location j are very close to the raw proportions of traffic that is bicycles at each location. (2) The 95% posterior credible intervals of parameters θ_j at location j are not large and all contain the values of the raw proportions. Additionally, a better prior distribution may exist for this problem than the beta distirbution, however, the posterior estimates (i.e., means and 95% intervals) derived from the beta distribution are feasible, reasonable, and not bad. Therefore, we can conclude that the beta distribution for the θ_j's is reasonable.

Problem 14: Hierarchical Poisson model.

Problem 14:

(a) Our statistical model:

$y_j \overset{ind}{\sim} \text{Pois}(\theta_j)$

$\theta_j \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$

$\alpha, \beta \sim p(\alpha,\beta) \propto (\alpha+\beta)^{-5/2}$, $j = 1, 2, \cdots, 10$

$\left\{ \begin{array}{l} y_j: \text{the total number of vehicles observed at each Location } j \\ \theta_j: \text{the "true" rate of traffic per hour at Location } j. \end{array} \right.$

The joint posterior distribution:

$$
\begin{aligned}
p(\theta,\alpha,\beta|y) &\propto p(y|\theta,\alpha,\beta)\, p(\theta,\alpha,\beta) \\
&= p(y|\theta)\, p(\theta|\alpha,\beta) \cdot p(\alpha,\beta) \\
&= p(\alpha,\beta) \prod_{j=1}^{10} p(y_j|\theta_j)\, p(\theta_j|\alpha,\beta) \\
&= p(\alpha,\beta) \prod_{j=1}^{10} \frac{\theta_j^{y_j} e^{-\theta_j}}{y_j!} \frac{\beta^\alpha \theta_j^{\alpha-1} e^{-\beta\theta_j}}{\Gamma(\alpha)} \\
&= p(\alpha,\beta) \prod_{j=1}^{10} \frac{\beta^\alpha \theta_j^{\alpha+y_j-1} e^{-(\beta+1)\theta_j}}{y_j!\,\Gamma(\alpha)} \\
&\propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{10} \frac{\beta^\alpha \theta_j^{\alpha+y_j-1} e^{-(\beta+1)\theta_j}}{y_j!\,\Gamma(\alpha)}
\end{aligned}
$$

(b) The marginal posterior density of the hyperparameters:

$$
\begin{aligned}
p(\alpha,\beta|y) &\propto p(y|\alpha,\beta)\, p(\alpha,\beta) \\
&= p(\alpha,\beta) \prod_{j=1}^{10} p(y_j|\alpha,\beta) \\
&= p(\alpha,\beta) \prod_{j=1}^{10} \int p(y_j|\theta_j)\, p(\theta_j|\alpha,\beta)\, d\theta_j \\
&= p(\alpha,\beta) \prod_{j=1}^{10} \int \frac{\beta^\alpha \theta_j^{\alpha+y_j-1} e^{-(\beta+1)\theta_j}}{y_j!\,\Gamma(\alpha)}\, d\theta_j \\
&= p(\alpha,\beta) \prod_{j=1}^{10} \int \frac{\beta^\alpha}{y_j!\,\Gamma(\alpha)} \theta_j^{\alpha+y_j-1} e^{-(\beta+1)\theta_j}\, d\theta_j \\
&= p(\alpha,\beta) \prod_{j=1}^{10} \frac{\beta^\alpha \Gamma(\alpha+y_j)}{y_j!\,\Gamma(\alpha) \cdot (\beta+1)^{\alpha+y_j}} \int \frac{(\beta+1)^{\alpha+y_j} \theta_j^{\alpha+y_j-1} e^{-(\beta+1)\theta_j}}{\Gamma(\alpha+y_j)}\, d\theta_j \\
&= p(\alpha,\beta) \prod_{j=1}^{10} \frac{\beta^\alpha \Gamma(\alpha+y_j)}{y_j!\,\Gamma(\alpha)(\beta+1)^{\alpha+y_j}} \propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{10} \frac{\beta^\alpha \Gamma(\alpha+y_j)}{y_j!\,\Gamma(\alpha) \cdot (\beta+1)^{\alpha+y_j}} \\
&\propto (\alpha+\beta)^{-5/2} \prod_{j=1}^{10} \binom{\alpha+y_j-1}{\alpha-1} \left(\frac{1}{\beta+1}\right)^{y_j} \left(\frac{\beta}{\beta+1}\right)^{\alpha}, \quad \boxed{MB}
\end{aligned}
$$

b. (continued):

We simulate 2000 random draws from the marginal posterior distribution $(\alpha+\beta)^{-5/2}NB(\alpha,\frac{\beta}{\beta+1})$ of the hyperparameters $p(\alpha,\beta|y)$, and plot the contours and scatter plots of (α,β), separately as shown in Figure 5, 6, and 7. Specifically, Figure 5 presents the distributions of $log(\alpha)$, $\log(\beta)$, as well as their contour plots. Figure 6 shows the scatter plot of (α,β) based on the simuilation draws, which suggests that (α,β) have some correlations violating the independence assumption. Instead, when we plot the scatter plot of $(\log(\alpha/\beta), \log(\alpha+\beta))$, it can lead to a better result.

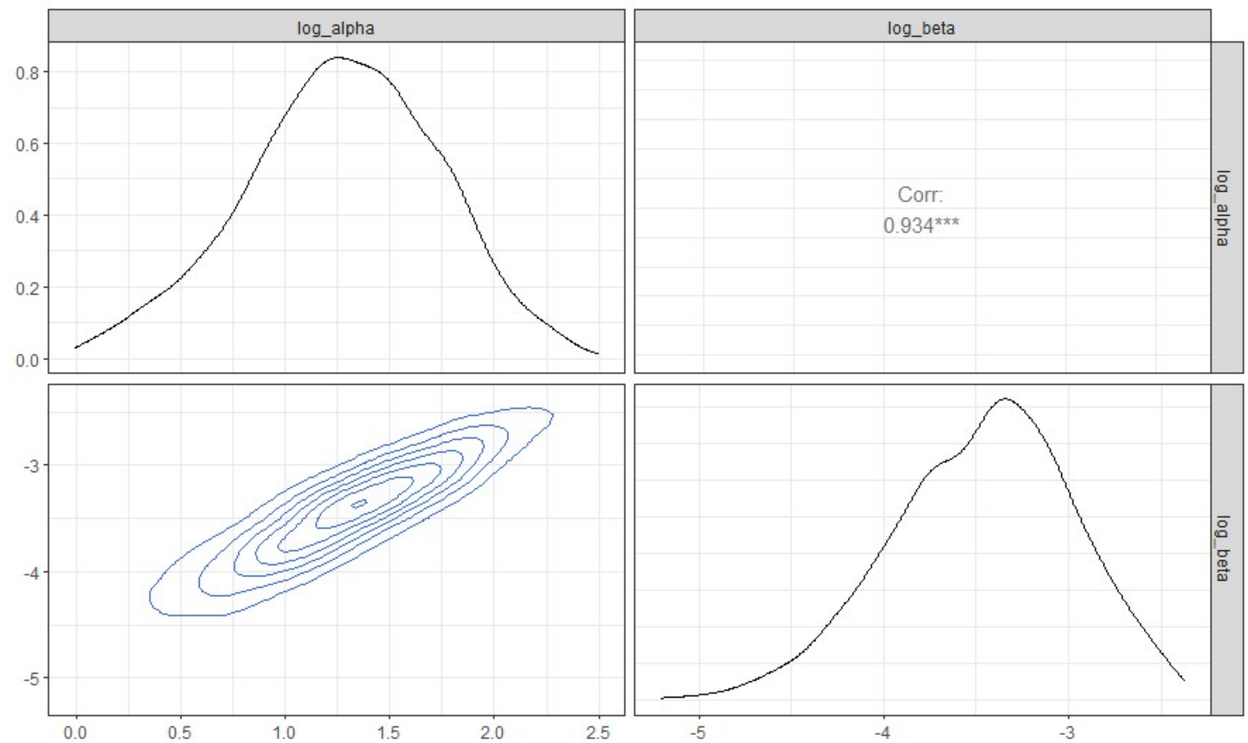


Figure 5: Sampling from the marginal distribution of the hyperparameters $p(\alpha,\beta|y)$

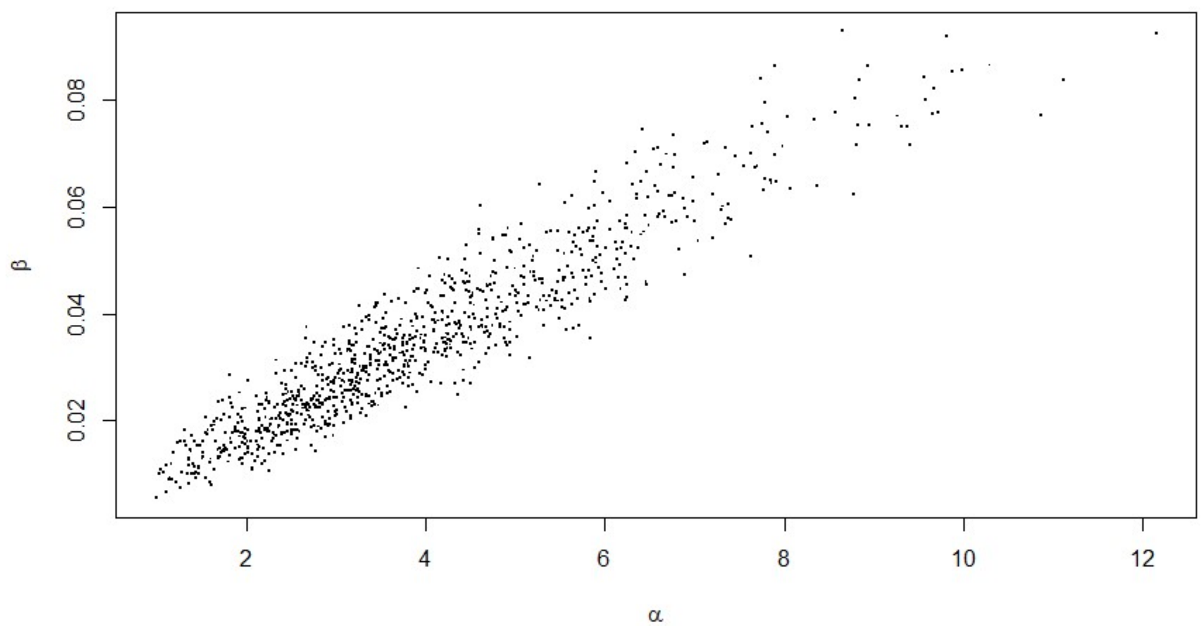


Figure 6: The scatter plot of (α, β) based on the simulation draws from the marginal distribution of the hyperparameters $p(\alpha, \beta|y)$.

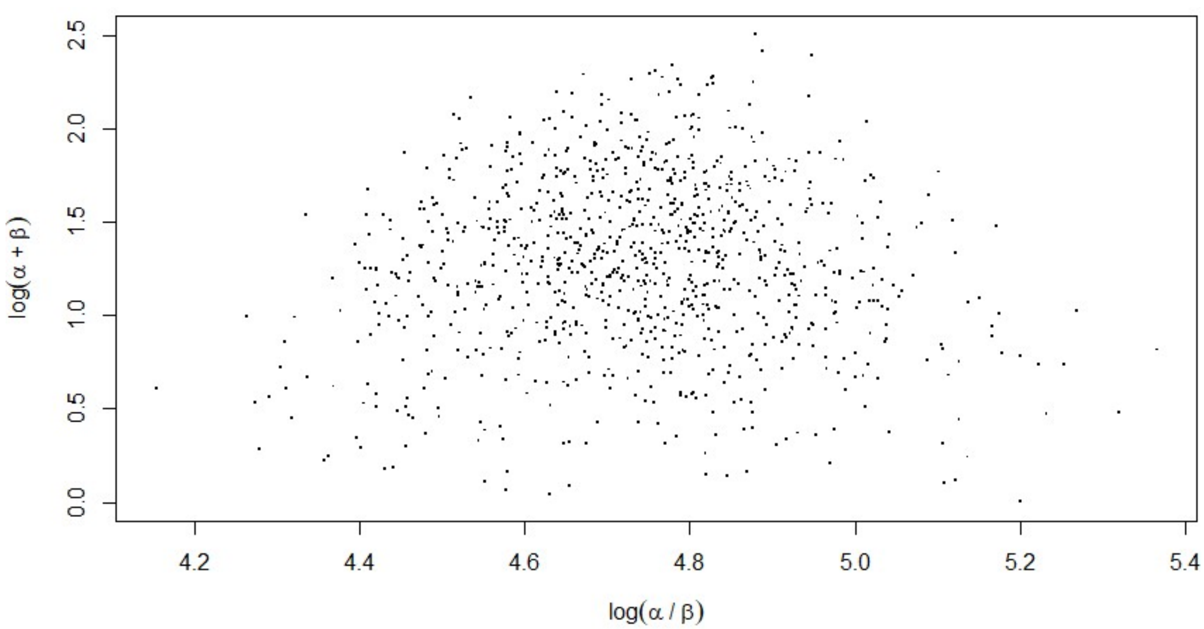


Figure 7: The scatter plot of $(\log(\alpha/\beta), \log(\alpha + \beta))$ based on the simulation draws from the marginal distribution of the hyperparameters $p(\alpha, \beta|y)$.

c-d. Integrable:

The marginal posterior density of the hyperparameters (α, β), as well as the joint posterior density of parameters, are both integrable. Specifically, as shown in Figure 5, the plots of the marginal posterior density of (α, β) suggest that the values of $log(\alpha)$, $\log(\beta)$ are both under finite curves. Furthermore, it can be easily observed from Figure 8-10 that the density plots of the sampled α, β and $\theta_j, j = 1,2, \cdots, 10$ are all finite and that their densities at the limits are all equal to zero. Thus, we conclude that the posterior density is integrable from an empirical perspective.

From an analytical perspective, we can easily derive the joint posterior density $p(\theta, \alpha, \beta|y)$ below:

$$p(\theta, \alpha, \beta|y) \propto (\alpha+\beta)^{-\frac{5}{2}} \prod_{j=1}^{10} \theta_j^{\alpha+y_j+1} e^{-(\beta+1)\theta_j} \propto (\alpha+\beta)^{-\frac{5}{2}} \prod_{j=1}^{10} Gamma\left(\theta_j|\alpha+y_j, \beta+1\right)$$

Thus, based on the proporties of Gamma Distribution and decaying characteristics of $(\alpha+\beta)^{-\frac{5}{2}}$, the joint posterior density $p(\theta, \alpha, \beta|y)$ is finite at the limits and therefore us integrable.

e. Simulations

Drawing simulations from the joint posterior distribution of the parameters and hyperparameters is done through the following two steps:

(1) Draw simulations from the marginal distribution of the hyperparameters $p(\alpha, \beta|y)$ below:

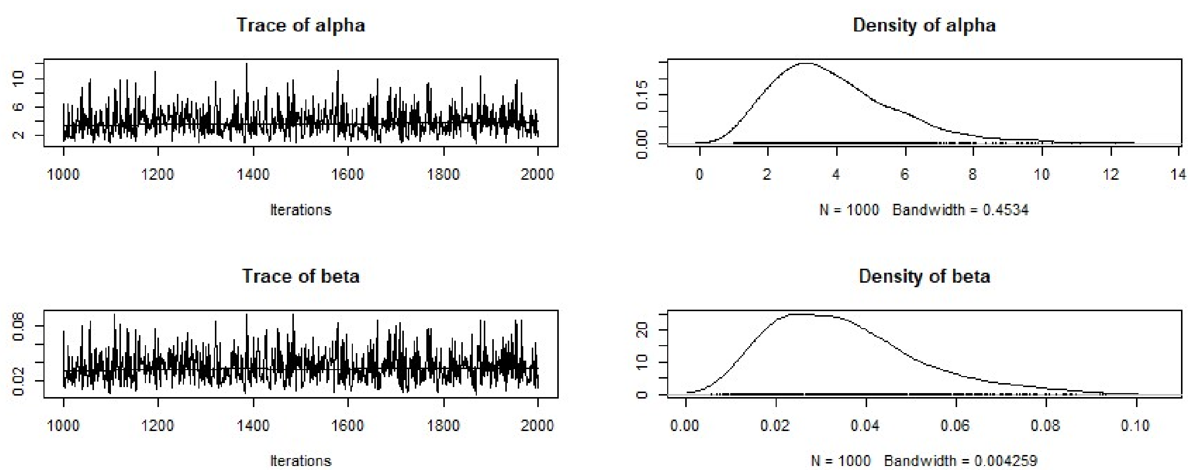
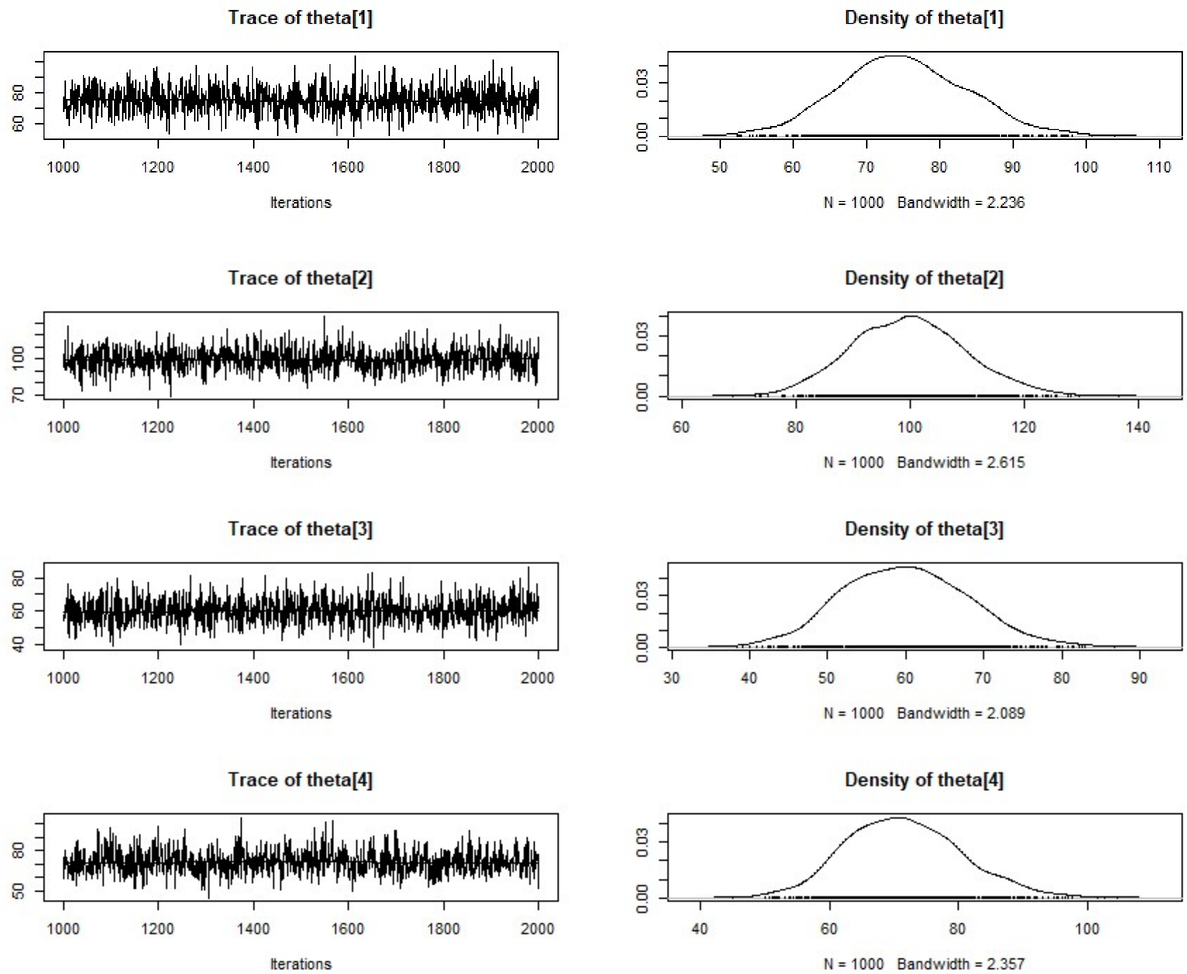


Figure 8: Trace and density of α, β based on simualtion samples from the marginal distribution of the hyperparameters $p(\alpha, \beta|y)$.

(2) Draw simulations from the conditional posterior density of the parameters $p(\theta|\alpha,\beta,y)$

$$p(\theta|\alpha,\beta,y) = \prod_{j=1}^{10} p(\theta_j|\alpha,\beta,y_j) = \prod_{j=1}^{10} Gamma(\theta_j|\alpha + y_j, \beta + 1)$$



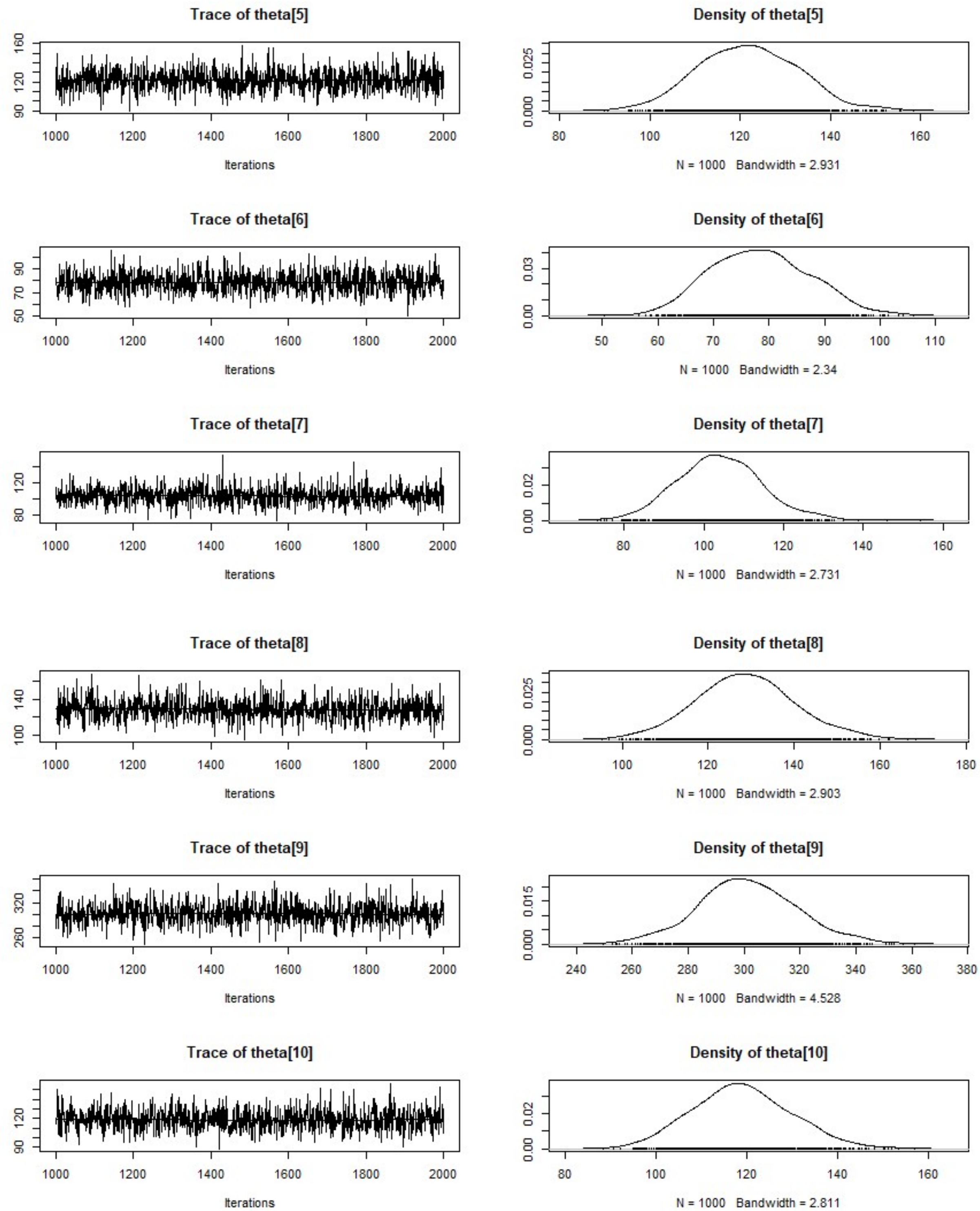


Figure 9: Trace and density of $\theta_j, j = 1,2,\cdots,10$ based on simualtion samples from the posterior density of the parameters $p(\theta|\alpha, \beta, y)$.

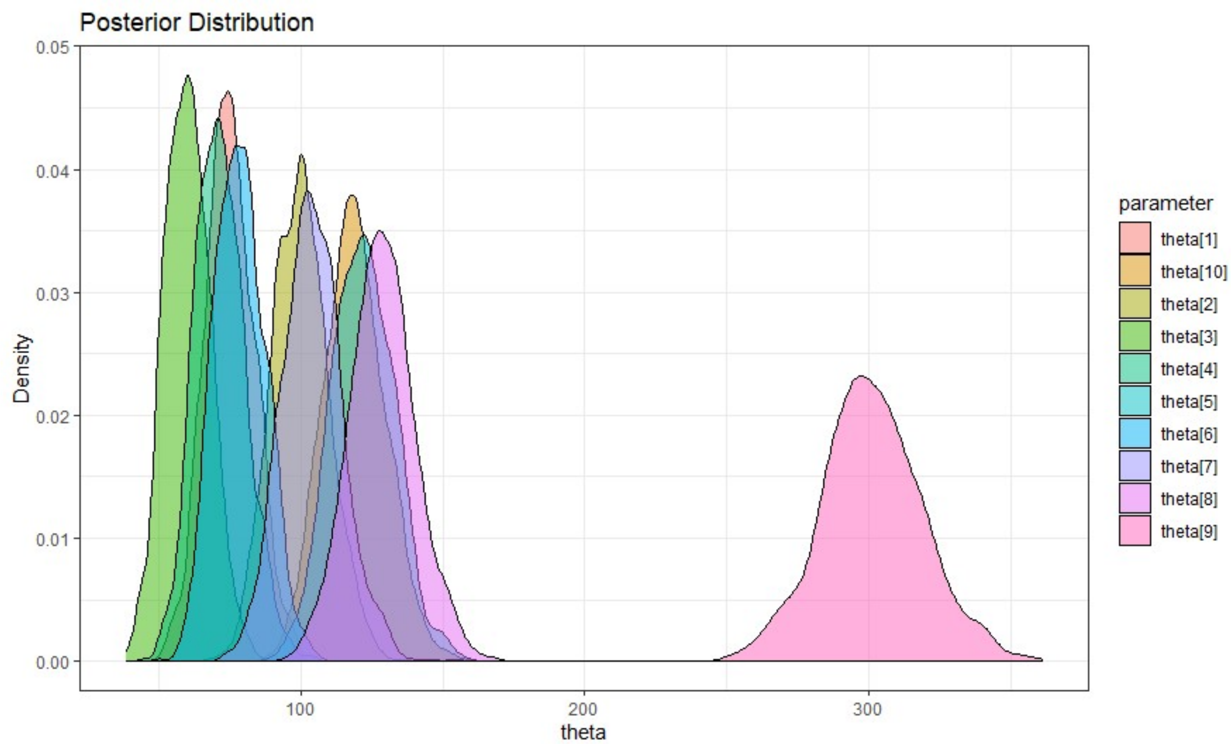


Figure 10: Plots of the posterior distributions of $\theta_j, j = 1,2,\cdots,10$ based on samplings from the posterior density of the parameters $p(\theta|\alpha,\beta,y)$.