1. Let $X_1, \ldots, X_n$ be an i.i.d. sample such that $X_i \mid \theta, \sigma^2 \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$, where σ^2 is known and θ is unknown. Also, let your prior for θ be a mixture of conjugate priors, i.e.,

$$\pi(\theta) = \sum_{\ell=1}^{K} w_\ell \phi(\theta \mid \mu_\ell, \tau^2)$$

where $\phi(\theta \mid \mu_\ell, \tau^2)$ denotes the Gaussian density with mean μ_ℓ and variance τ^2 and mixture weights $0 < w_\ell < 1$ for all $\ell = 1, \ldots, K$ with $\sum_{\ell=1}^{K} w_\ell = 1$.

Note: This questions is challenging. Use the results from the class example with $X_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$ and $\theta \sim \mathrm{N}(\mu, \tau^2)$.

(a) Find the posterior distribution for θ based on this prior.

(b) Find the posterior mean.

(c) Find the prior predictive distribution associated with this model (i.e., the marginal distribution of data).

(d) Find the posterior predictive distribution associated with this model.

(a) $\pi(\theta \mid x) \propto f(x \mid \theta) \cdot \pi(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \cdot \pi(\theta)$

$$\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta)^2\right) \cdot \sum_{\ell=1}^{K} w_\ell \cdot \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\theta - \mu_\ell)^2}{2\tau^2}\right)$$

$$\propto \sum_{\ell=1}^{k} w_\ell \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - \theta)^2\right)\right) \cdot \exp\left(-\frac{1}{2\tau^2}(\theta - \mu_l)^2\right)$$

constant given x

$$\propto \sum_{\ell=1}^{k} w_l \exp\left(-\frac{n(\bar{x} - \theta)^2}{2\sigma^2} - \frac{(\theta - \mu_l)^2}{2\tau^2}\right)$$

$$= \sum_{\ell=1}^{k} W_\ell \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\theta\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_l}{\tau^2}\right) + \frac{n\bar{x}^2}{\sigma^2} + \frac{\mu_l^2}{\tau^2}\right]\right\}$$

$$= \sum_{\ell=1}^{k} W_\ell \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)\left[\theta^2 - 2\theta \cdot \underbrace{\frac{n\bar{x}/\sigma^2 + \mu_l/\tau^2}{n/\sigma^2 + 1/\tau^2}}_{\text{Denote to } \tilde{\mu}_L} + \frac{n\bar{x}^2/\sigma^2 + \mu_l^2/\tau^2}{n/\sigma^2 + 1/\tau^2}\right]\right\}$$

$$= \sum_{\ell=1}^{k} W_\ell \exp\left\{-\frac{1}{2(n/\sigma^2 + 1/\tau^2)^{-1}}\left[(\theta - \tilde{\mu}_L)^2 + \underbrace{\frac{n\bar{x}^2/\sigma^2 + \mu_l^2/\tau^2}{n/\sigma^2 + 1/\tau^2} - \tilde{\mu}_L^2}_{\text{Denote to } \star}\right]\right\}$$

$$\phi = \frac{n/\sigma^2\, \bar{x}^2 + 1/\tau^2\, \mu_l^2}{n/\sigma^2 + 1/\tau^2} - \frac{(n/\sigma^2\, \bar{x} + 1/\tau^2 \mu_l^2)^2}{(n/\sigma^2 + 1/\tau^2)^2}$$

let $n/\sigma^2 = a$, $1/\tau^2 = b$

$$= \frac{1}{a+b}\left(\frac{(a\bar{x}^2 + b\mu_l^2)(a+b)}{a+b} - \frac{(a\bar{x} + b\mu_l)^2}{a+b}\right)$$

$$= \frac{1}{a+b} \cdot \frac{\cancel{a^2\bar{x}^2} + ab(\bar{x}^2 + \mu_l^2) + \cancel{b^2\mu_l^2} - \cancel{a^2\bar{x}^2} - 2ab\bar{x}\mu_l - \cancel{b^2\mu_l^2}}{a+b}$$

$$= \frac{1}{a+b} \cdot \frac{ab}{a+b}(\bar{x} - \mu_l)^2$$

Unnormalized New weight: $\tilde{W}_l$

$$\Rightarrow \pi(\theta|x) \propto \sum_{l=1}^{k} \exp\left(-\frac{n/\sigma^2 \cdot 1/\tau^2}{2(n/\sigma^2 + 1/\tau^2)}(\bar{x} - \mu_l)^2\right) \cdot W_l$$

$$\exp\left(-\frac{1}{2(n/\sigma^2 + 1/\tau^2)^{-1}}(\theta - \tilde{\mu}_l)^2\right)$$

$\tilde{\mu}_l \to$ New mean

$(n/\sigma^2 + 1/\tau^2)^{-1} \to$ New variance

But, the weight should be normalized, since

$$\sum_{l=1}^{k} \tilde{W}_l = 1$$

Still a Gaussian Mixture model, with weight proportional to $\tilde{W}_l$ (Accurate weight $= \tilde{W}_l / \sum_{l=1}^{k} \tilde{W}_l$).

new mean of each component is $\tilde{\mu}_l = \dfrac{n\bar{x}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2}$,

new variance: $(n/\sigma^2 + 1/\tau^2)^{-1} = \tilde{\tau}^2$

(b) Since it's still a Gaussian Mixture:

$$E(\theta|x) = \sum_{l=1}^{k} \frac{\tilde{w}_l}{\sum_{j=1} \tilde{w}_j} \cdot \tilde{\mu}_l$$

(c). $m(x) = \int_R f(x|\theta) \cdot \pi(\theta)\, d\theta$

By (a): $m(x) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}\right) \cdot \sum_{l=1}^{k} \tilde{w}_l$

$$\underbrace{\int_R \sum_{l=1}^{k} \frac{\tilde{w}_l}{\sum_{j=1}^{k} \tilde{w}_j} \exp\left(-\frac{(\theta-\tilde{\mu}_l)^2}{2\tilde{\tau}^2}\right) d\theta}$$

It's a Gaussian Mixture kernel. $\Rightarrow$ it's $\sqrt{2\pi\tilde{\tau}^2}$

in which, $\tilde{\tau}^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}$

$$\Rightarrow m(x) = (2\pi)^{-\frac{1}{2}} \frac{1}{\sqrt{\sigma^2\tau^2}} \cdot \frac{1}{\sqrt{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}} \cdot \sum_{l=1}^{k} \tilde{w}_l \cdot \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}\right)$$

$$= \left[2\pi(n\tau^2+\sigma^2)\right]^{-\frac{1}{2}} \cdot \sum_{l=1}^{k} \tilde{w}_l \cdot \exp\left(-\frac{\sum(x_i-\bar{x})^2}{2\sigma^2}\right)$$

$$= \left[2\pi\left(\sigma^2\tau^2\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\right)\right]^{-\frac{1}{2}} \cdot \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}\right) \cdot$$

$$\sum_{l=1}^{k} w_l \cdot \exp\left(-\frac{\frac{n}{\sigma^2} \cdot \frac{1}{\tau^2}}{2\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)} (\bar{x}-\mu_l)^2\right)$$

Gaussian Mixture:

Weight: $w_l \cdot \exp\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}\right)$

mean of each component: μ_l

Variance: $\frac{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}{\frac{n}{\sigma^2} \cdot \frac{1}{\tau^2}}$

(d) $f(y|x) = \int_{\mathbb{R}} f(y|\theta)\pi(\theta|x)\, d\theta$

$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\theta)^2}{2\sigma^2}\right) \cdot \pi(\theta|x)\, d\theta \longrightarrow I$

Denote normalized weight $\frac{\tilde{W}_\ell}{\sum_{\ell=1}^{k} W_\ell}$ to be W_{N_ℓ}

$\Rightarrow I = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp\left(-\frac{y^2 - 2\theta y + \theta^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi\tilde{\tau}^2}} \cdot \sum_{\ell=1}^{k} W_{N_\ell} \exp\left(-\frac{(\theta - \tilde{\mu}_\ell)^2}{2\tilde{\tau}^2}\right) d\theta$

$\propto \int_{\mathbb{R}} \sum_{\ell=1}^{k} W_{N_\ell} \exp\left(-\frac{1}{2} \cdot \left[\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta - \mu_\ell)^2}{\tilde{\tau}^2}\right]\right) d\theta$

$= \int_{\mathbb{R}} \sum_{\ell=1}^{k} W_{LN} \exp\left(-\frac{1}{2}\left[\theta^2\left(\frac{1}{\sigma^2} + \frac{1}{\tilde{\tau}^2}\right) - 2\theta\left(\frac{y}{\sigma^2} + \frac{\mu_\ell}{\tilde{\tau}^2}\right) + \frac{y^2}{\sigma^2} + \frac{\mu_\ell^2}{\tilde{\tau}^2}\right]\right) d\theta$

$= \sum_{\ell=1}^{k} W_{N_\ell} \exp\left(-\frac{\frac{y^2}{\sigma^2} + \frac{\mu_\ell^2}{\tilde{\tau}^2}}{2}\right) \int_{\mathbb{R}} \exp\left(-\frac{\frac{1}{\sigma^2} + \frac{1}{\tilde{\tau}^2}}{2}\left[\left(\theta - \frac{\frac{y}{\sigma^2} + \frac{\mu_\ell}{\tilde{\tau}^2}}{\frac{1}{\sigma^2} + \frac{1}{\tilde{\tau}^2}}\right)^2 - \left(\frac{\frac{y}{\sigma^2} + \frac{\mu_\ell}{\tilde{\tau}^2}}{\frac{1}{\sigma^2} + \frac{1}{\tilde{\tau}^2}}\right)^2\right]\right)$

$= \sum_{\ell=1}^{k} W_{N_\ell} \exp\left(-\frac{1}{2}\underbrace{\left(\frac{y^2}{\sigma^2} + \frac{\mu_\ell^2}{\tilde{\tau}^2} - \frac{\left(\frac{y}{\sigma^2} + \frac{\mu_\ell}{\tilde{\tau}^2}\right)^2}{\frac{1}{\sigma^2} + \frac{1}{\tilde{\tau}^2}}\right)}_{\star}\right) \cdot \sqrt{2\pi\left(\frac{1}{\sigma^2} + \frac{1}{\tilde{\tau}^2}\right)^{-1}}$

$\star = \frac{\cancel{\frac{y^2}{\sigma^4}} + \frac{y^2}{\sigma^2\tilde{\tau}^2} + \frac{\mu_\ell^2}{\sigma^2\tilde{\tau}^2} + \cancel{\frac{\mu_\ell^2}{\tilde{\tau}^4}} - \cancel{\frac{y^2}{\sigma^4}} - \frac{2y\mu_\ell}{\sigma^2\tilde{\tau}^2} - \cancel{\frac{\mu_\ell^2}{\tilde{\tau}^4}}}{\frac{1}{\sigma^2} + \frac{1}{\tilde{\tau}^2}}$

$= \frac{(y - \mu_\ell)^2}{\sigma^2 + \tilde{\tau}^2}$

$\Rightarrow f(y|x) \propto \sum_{\ell=1}^{k} W_{N_\ell} \cdot \exp\left(-\frac{1}{2(\sigma^2 + \tilde{\tau}^2)}(y - \mu_\ell)^2\right)$

⇒ Gaussian Mixture Model.

Weight : W_{N_L}

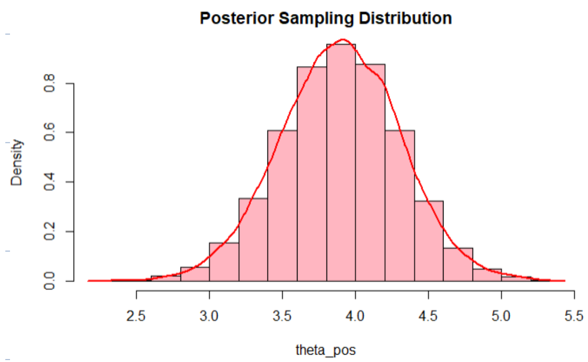
$\text{Mean}_L = \tilde{\mu}_L$

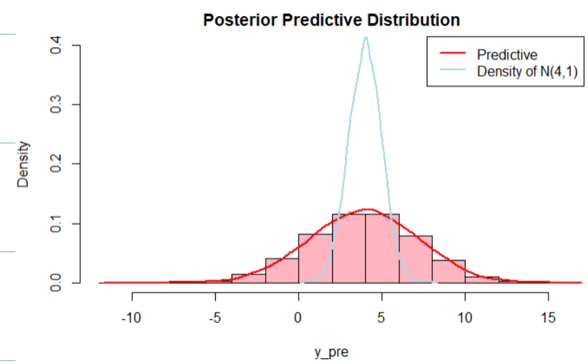
Var : $\sigma^2 \tau^2 \left(1/\sigma^2 + 1/\tilde{\tau}^2 \right)$

⇒ W_{N_L}, $\tilde{\mu}_L$, $\tilde{\tau}^2$ is in page 2, but W_{N_L} is normalized $\tilde{W}_L$.

(e) Simulate $n = 50$ i.i.d. observations from $N(4, 10)$. Let $K = 3$, $\tau^2 = 1$, $(\mu_1, \mu_2, \mu_3) = (-3, 0, 3)$ and $(\mu_1, \mu_2, \mu_3) = (1/3, 1/3, 1/3)$, and fit the above model to the simulated data. (i) draw samples of θ and make a histogram of the samples. (ii) draw samples of an unknown observable X from the posterior predictive distribution and make a histogram of the samples overlaid with the density of $N(4, 1)$. Choose option 'probability = TRUE' to make the histogram and the density plot comparable.

(i) Posterior distribution and Predictive Distribution.



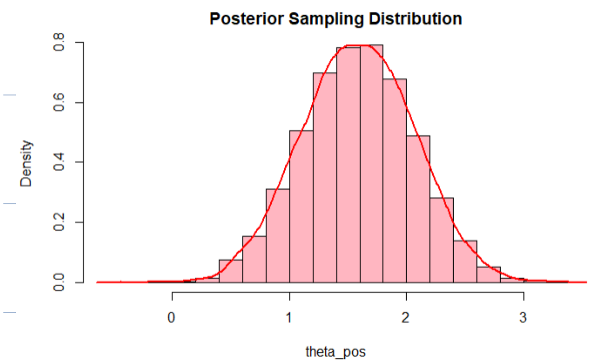


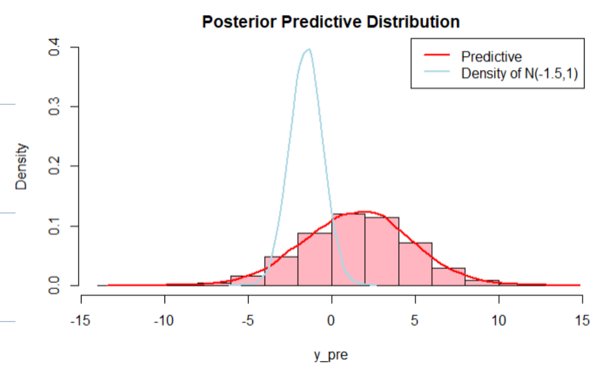
It seems not reasonable to compare with $N(4,1)$

since $Var(y|x)$ should be greater than $Var(x)$, i.e. 10.

(f) Simulate $n = 50$ i.i.d. observations from $\mathrm{N}(1.5, 10)$. Let $K = 3$, $\tau^2 = 1$, $(\mu_1, \mu_2, \mu_3) = (-3, 0, 3)$ and $(\mu_1, \mu_2, \mu_3) = (1/3, 1/3, 1/3)$, and fit the above model to the simulated data. (i) draw samples of θ and make a histogram of the samples. (ii) draw samples of an unknown observable X from the posterior predictive distribution and make a histogram of the samples overlaid with the density of $\mathrm{N}(-1.5, 1)$.

Note: The histograms are to approximate the posterior distribution of θ and the posterior predictive distribution of an unknown observable X.





2. Let X be $\mathrm{N}(0, \sigma^2)$. Assume that the unknown $1/\sigma^2$ has a gamma distribution with parameters $\alpha = r/2$ and $\beta = 2/r$ (i.e., $\sigma^2 \sim \mathrm{IG}(r/2, r/2)$), where r is a positive integer. Show that the marginal distribution of X is a $t-$distribution with r degrees of freedom.

$$\pi(\sigma^2) \propto (\sigma^2)^{-\frac{r}{2}-1} \exp\left(-\frac{1}{\sigma^2} \cdot \frac{r}{2}\right)$$

$$f(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$f(x,\sigma^2) \propto (\sigma^2)^{-\frac{r+1}{2}-1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{x^2+r}{2}\right)\right)$$

$$m(x) = \int_0^\infty f(x,\sigma^2)\, d\sigma^2$$

$$\propto \int_0^\infty (\sigma^2)^{-\frac{r+1}{2}-1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{x^2+v}{2}\right)\right) d\sigma^2$$

It's a kernel of $IG[\frac{r+1}{2}, \frac{x^2+V}{2})$

$$\Rightarrow \quad m(x) \propto \left(\frac{x^2+r}{2}\right)^{\frac{r+1}{2}}$$

$$\propto \left(1 + \frac{x^2}{r}\right)^{\frac{r+1}{2}}$$

It's a t-distribution with $df = r$.

3. Let (X_1, X_2, X_3) have trinomial distribution with density

$$f(x_1, x_2, x_3|\theta_1, \theta_2) \propto \theta_1^{x_1}\theta_2^{x_2}(1-\theta_1-\theta_2)^{x_3}.$$

Derive Jeffreys prior for (θ_1, θ_2).

exponential family, regular condition met

$$L(\theta) = \prod_{i=1}^n f(\vec{x}) = \theta_1^{\sum_{i=1}^n X_{1i}} \theta_2^{\sum_{i=1}^n X_{2i}} (1-\theta_1-\theta_2)^{\sum_{i=1}^n X_{3i}}$$

$$l(\theta) = \sum_{i=1}^n X_{1i} \log(\theta_1) + \sum_{i=1}^n X_{2i} \log(\theta_2) + \sum_{i=1}^n X_{3i} \log(1-\theta_1-\theta_2)$$

Also, $\sum_{j=1}^3 \sum_{i=1}^n X_{ji} = n$, let $\sum_{i=1}^n X_{1i} = A_1$, $\sum_{i=1}^n X_{2i} = A_2$

$$l(\theta) = A_1 \log(\theta_1) + A_2 \log(\theta_2) + (n-A_1-A_2)\log(1-\theta_1-\theta_2)$$

$$\frac{\partial l(\theta)}{\partial \theta_1} = \frac{A_1}{\theta_1} - \frac{n-A_1-A_2}{1-\theta_1-\theta_2} \qquad \frac{\partial^2 l(\theta)}{\partial\theta_1\partial\theta_2} = \frac{-(n-A_1-A_2)}{(1-\theta_1-\theta_2)^2}$$

$$\frac{\partial l(\theta)}{\partial \theta_2} = \frac{A_2}{\theta_2} - \frac{n-A_1-A_2}{1-\theta_1-\theta_2} \qquad \frac{\partial^2 l(\theta)}{\partial\theta_2\partial\theta_1} = \frac{-(n-A_1-A_2)}{(1-\theta_1-\theta_2)^2}$$

$$\frac{\partial^2 l(\theta)}{\partial \theta_1^2} = \frac{-A_1}{\theta_1^2} - \frac{n-A_1-A_2}{(1-\theta_1-\theta_2)^2} \qquad E(A_1) = n\theta_1 \quad E(A_2) = n\theta_2$$

$$\frac{\partial^2 l(\theta)}{\partial \theta_2^2} = \frac{-A_2}{\theta_2^2} - \frac{n-A_1-A_2}{(1-\theta_1-\theta_2)^2} \qquad E(n-A_1-A_2) = n\cdot(1-\theta_1-\theta_2)$$

$$\Rightarrow I(\theta) = \begin{bmatrix} \frac{n\theta_1}{\theta_1^2} + \frac{n(1-\theta_1-\theta_2)}{(1-\theta_1-\theta_2)^2} & \frac{n(1-\theta_1-\theta_2)}{(1-\theta_1-\theta_2)^2} \\ \frac{n(1-\theta_1-\theta_2)}{(1-\theta_1-\theta_2)^2} & \frac{n\theta_2}{\theta_2^2} + \frac{n(1-\theta_1-\theta_2)}{(1-\theta_1-\theta_2)^2} \end{bmatrix}$$

$$= \begin{bmatrix} n(1/\theta_1 + 1/(1-\theta_1-\theta_2)) & n/(1-\theta_1-\theta_2) \\ n/(1-\theta_1-\theta_2) & n(1/\theta_2 + 1/(1-\theta_1-\theta_2)) \end{bmatrix}$$

$$\det(I(\theta)) = n^2\left(\frac{(1-\theta_2)(1-\theta_1)}{\theta_1\theta_2(1-\theta_1-\theta_2)^2} - \frac{\theta_1\theta_2}{\theta_1\theta_2(1-\theta_1-\theta_2)^2}\right)$$

$$= \frac{1-\theta_1-\theta_2}{\theta_1\theta_2(1-\theta_1-\theta_2)^2} = \frac{1}{\theta_1\theta_2(1-\theta_1-\theta_2)}$$

$$\sqrt{\det(I(\theta))} = \theta_1^{-\frac{1}{2}}\theta_2^{-\frac{1}{2}}(1-\theta_1-\theta_2)^{-\frac{1}{2}}$$

It's dirichlet distibution $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$

4. (Robert Problem 3.9) Let $x \mid \theta \sim \text{Bin}(n, \theta)$ and $\theta \sim \text{Be}(\alpha, \beta)$. Determine whether there exits values of α, β such that $\pi(\theta \mid x)$ is the uniform posterior on [0, 1], even for a single value of x.

$$f(x|\theta) = \prod_{i=1}^{m}\binom{n}{x_i}\theta^{x_i}(1-\theta)^{n-x_i}$$

$$\propto \theta^{\sum_{i=1}^m x_i}(1-\theta)^{\sum_{i=1}^m (n-x_i)}$$

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\pi(\theta|x) \propto \theta^{\alpha+\sum_{i=1}^m x_i - 1}(1-\theta)^{\beta+\sum_{i=1}^m (n-x_i)-1}$$

$$\text{Beta}\left(\alpha+\textstyle\sum_{i=1}^m x_i,\ \beta+nm-\sum_{i=1}^m x_i\right)$$

If it's uniform: $\alpha + \sum_{i=1}^m x_i = 1 = \beta + nm - \sum_{i=1}^m x_i$

$\sum_{i=1}^m x_i \in \{0, 1, 2, \dots, mn\} \Rightarrow$ Iff $\alpha = 1 - \sum_{i=1}^m x_i$ and

$\beta = 1 - nm + \sum_{i=1}^m x_i$, the posterior will be uniform.

However! $\alpha, \beta > 0$. $\Rightarrow \begin{cases} \sum_{i=1}^m x_i < 1 \\ nm - \sum_{i=1}^m x_i < 1 \\ \sum_{i=1}^m x_i \in \{0, 1, 2, \cdots, mn\} \end{cases}$

There exists NO α and β or x satisfying this.

5. (Robert Problem 3.10) Let $x \mid \theta \sim \mathrm{Pa}(\alpha, \theta)$, a Pareto distribution, and $\theta \sim \mathrm{Be}(\mu, \nu)$. Show that if $\alpha < 1$ and $x > 1$, a particular choice of μ and ν gives $\pi(\theta \mid x)$ as the uniform posterior on [0,1].

$$f(x \mid \alpha, x_0) = \alpha \frac{x_0^\alpha}{x^{\alpha+1}}, \quad x \geq x_0.$$

$$f(x|\theta) = \prod_{i=1}^{n} \alpha \cdot \frac{\theta^\alpha}{x_i^{\alpha+1}} \cdot \mathbb{I}(x_i \geq \theta)$$

$$\pi(\theta) \propto \theta^{\mu-1}(1-\theta)^{\nu-1}$$

$$\pi(\theta|x) \propto \theta^{\alpha+\mu-1}(1-\theta)^{\nu-1} \cdot \alpha^n \cdot \left(\prod_{i=1}^{n} x_i\right)^{\alpha+1}$$

$$\propto \theta^{\alpha+\mu-1}(1-\theta)^{\nu-1}$$

Set : $\begin{cases} \alpha + \mu = 1 \\ \nu = 1 \end{cases}$ satisfying $\begin{cases} \mu > 0 \\ \nu > 0 \\ \alpha < 1 \end{cases}$

$\Rightarrow \mu = 1-\alpha, \ \nu = 1$

And here, $\pi(\theta|x) \sim \mathrm{Unif}[0,1]$

6. (Part of Robert Problem 3.31) Consider $x \mid \theta \sim \mathrm{N}(\theta, \theta)$ with $\theta > 0$. Determine the Jeffreys prior $\pi^J(\theta)$.

$$f(x|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \cdot \exp\left(-\frac{(x_i-\theta)^2}{2\theta}\right)$$

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

$$\propto \theta^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\theta} + \frac{n\theta}{2}\right)$$

$$l(L(\theta)) = -\frac{n}{2}\log(\theta) - \frac{\sum_{i=1}^n x_i^2}{2\theta} + \frac{n\theta}{2}$$

$$S(\theta) = \frac{\partial l(\theta)}{\partial(\theta)} = \frac{-n}{2\theta} + \frac{\sum_{i=1}^n x_i^2}{2\theta^2} + \frac{n}{2}$$

$$\frac{\partial S(\theta)}{\partial\theta} = \frac{n}{2\theta^2} - \frac{\sum_{i=1}^n x_i^2}{\theta^3}$$

$$I(\theta) = -E\left(\frac{\partial S(\theta)}{\partial(\theta)}\right) = -E\left(\frac{n}{2\theta^2} - \frac{\sum_{i=1}^n x_i^2}{\theta^3}\right) = \frac{n(\theta^2+\theta)}{\theta^2} - \frac{n}{2\theta^2}$$

$$= n + \frac{n}{\theta} - \frac{n}{2\theta^2} = n\left(\frac{\theta^2+\theta-\frac{1}{2}}{\theta^2}\right)$$

$$\sqrt{I(\theta)} = \sqrt{\frac{\theta^2+\theta-\frac{1}{2}}{\theta^2}} \Rightarrow \pi^J(\theta) \propto \sqrt{1+1/\theta-\frac{1}{2\theta^2}}$$

7. Consider a model of the form $x \mid \theta \sim \text{Bin}(n, \theta)$ and $\theta \sim \text{Be}(1/2, 1/2)$. Assume that you observe $n = 10$ and $x = 1$.

   (a) Report an exact 95% (symmetric) posterior credible interval for θ (for example, you can use the `qbeta` function in `R`).

   (b) Report an approximate credible interval for θ using the Laplace approximation.

   (c) Report an approximate credible interval for θ using Monte Carlo simulation.

   (d) Repeat the previous calculations with $n = 100, x = 10$ and with $n = 1000, x = 100$. Comment on the difference between all 9 situations.

(a) Posterior: $\text{Beta}(\frac{1}{2}+1, \frac{1}{2}+9) = \text{Beta}(1.5, 9.5)$

The 95% credible interval is $[0.011, 0.381]$

(b) $\pi(\theta|x) \propto \theta^{1.5-1}(1-\theta)^{9.5-1} = \theta^{0.5}(1-\theta)^{8.5}$

$$\frac{\partial \pi(\theta|x)}{\partial \theta} = \frac{1}{2}\theta^{-0.5}(1-\theta)^{8.5} - \theta^{0.5} \cdot 8.5(1-\theta)^{7.5}$$

Set this to be 0:

$$\frac{1}{2}\theta^{-0.5}(1-\theta)^{8.5} = 8.5\,\theta^{0.5}(1-\theta)^{7.5}$$

$$0.5(1-\theta) = 8.5\,\theta$$

$$1-\theta = 17\theta$$

$$\theta = 1/18$$

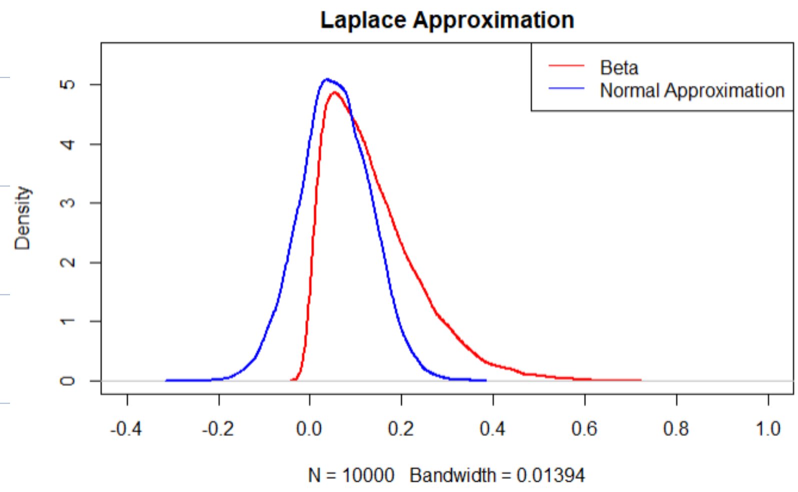
$$\log(\pi(\theta|x)) = c + 0.5\log(\theta) + 8.5\log(1-\theta)$$

$$\frac{d\log(\pi(\theta|x))}{d\theta} = \frac{0.5}{\theta} - \frac{8.5}{1-\theta}$$

$$\frac{d^2(\log(\pi(\theta|x)))}{d\theta^2} = -\frac{0.5}{\theta^2} - \frac{8.5}{(1-\theta)^2}$$

$$A = \left.\frac{0.5}{\theta^2} + \frac{8.5}{(1-\theta)^2}\right|_{\theta = 1/18}$$

$\Rightarrow \pi(\theta|x)$ Approximately $\sim N\left(1/18, \frac{1}{162+7.58179}\right)$



Therefore, the Laplace 95% HPD: $[-0.09485, 0.2061]$

(c) Monte Carlo: Generate 10000 samples from Beta(1.5, 9.5) and calculate 2.5% and 97.5% quantile: $[0.0106, 0.3819]$

Table 1: N=10, X=1

	2.5% Quantile	97.5 Quantile
True Beta	0.0110117	0.3813148
Laplace	-0.0949521	0.2060632
Monte Carlo	0.0113094	0.3833829

(d) for the case that $n=100$, $x=10$

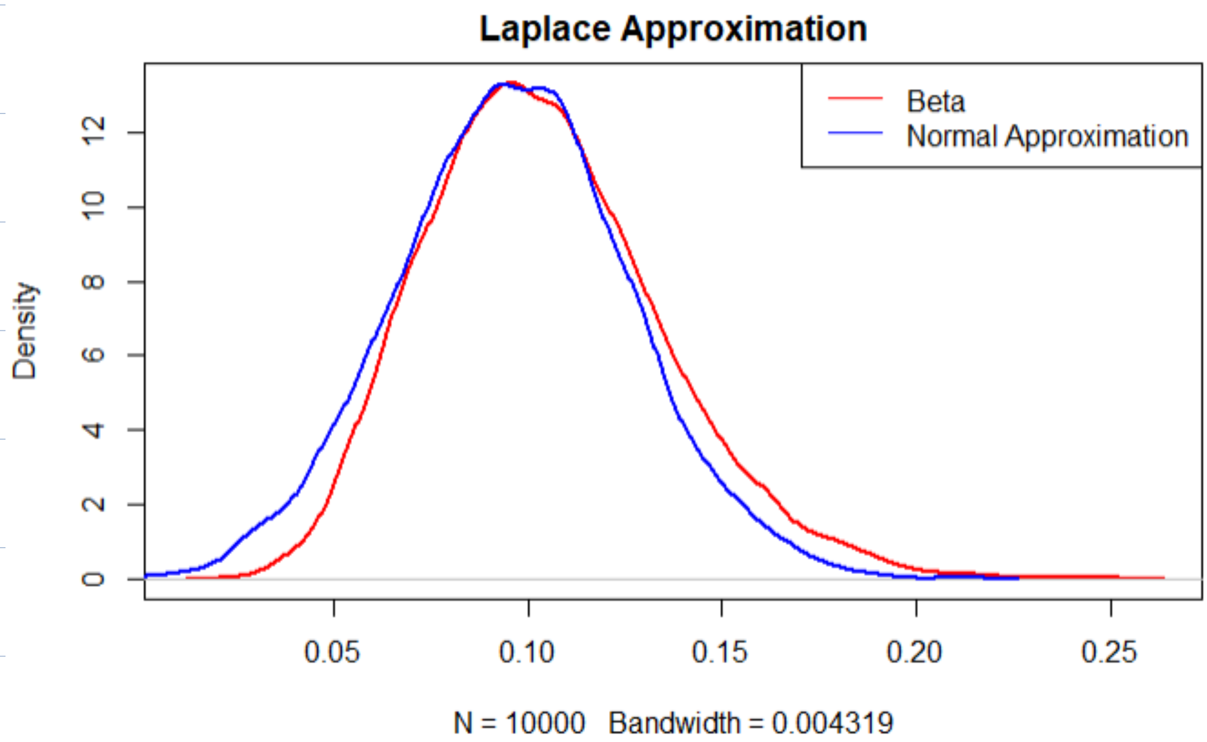


Table 2: N=100, X=10

	2.5% Quantile	97.5 Quantile
True Beta	0.0525847	0.1701239
Laplace	0.0379408	0.1539784
Monte Carlo	0.0521660	0.1697928

for the case that $N=1000$, $x=100$

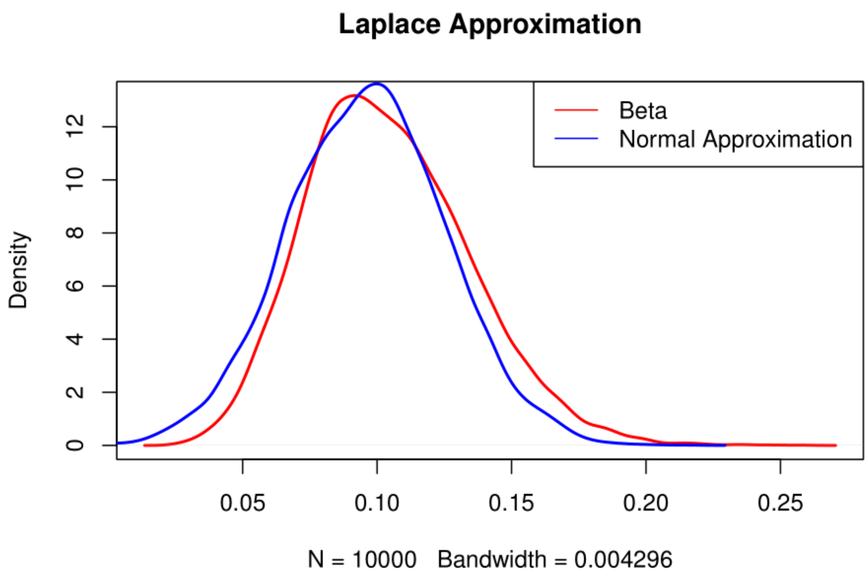


Table 3: N=100, X=10

	2.5% Quantile	97.5 Quantile
True Beta	0.0825627	0.1197483
Laplace	0.0810296	0.1181696
Monte Carlo	0.0826010	0.1196873

As N and x $\uparrow$, the Laplace approximation is becoming more and more accurate, for Monte Carlo method, as long as # of samples is large enough. by WLLN, we will converge in p to the true value.

8. Let $x_1, \ldots, x_n$ be an i.i.d. sample from a Gumbel type-II distribution with density

$$p(x \mid \alpha, \beta) = \alpha\beta x^{-\alpha-1}\exp(-\beta x^{-\alpha}), \qquad x > 0$$

with $\alpha, \beta > 0$. Let $\pi(\alpha, \beta) = 1$ for $\alpha, \beta > 0$ be the prior distribution. Simulate $n = 500$ i.i.d. observations from the Gumbel type-II distribution with $\alpha = \beta = 5$ (may use the inverse CDF method to generate random draws).

(a) Find the posterior $p(\alpha, \beta \mid x_1, \ldots, x_n)$. Use the simulated dataset and evaluate the posterior density in the grid of (α, β). You may find the normalizing constant numerically.

(b) Find the Laplace approximation to the posterior $p(\alpha, \beta \mid x_1, \ldots, x_n)$. Observe the mode can be found using some numerical method. Use the same simulated data and evaluate the approximated posterior on the grid of (α, β).

(c) Compare (a) and (b) and comment.

(a) $f(x|\alpha,\beta) = \prod_{i=1}^{n} \alpha\beta x_i^{-\alpha-1} \exp(-\beta x_i^{-\alpha})$

$= (\alpha\beta)^n \left(\prod_{i=1}^{n} x_i\right)^{-\alpha-1} \exp\left(-\beta \sum_{i=1}^{n} x_i^{-\alpha}\right)$

$\pi(\alpha,\beta|x) \propto f(x|\alpha,\beta) \cdot \pi(\alpha,\beta)$

$= (\alpha\beta)^n \left(\prod_{i=1}^{n} x_i\right)^{-\alpha-1} \exp\left(-\beta \sum_{i=1}^{n} x_i^{-\alpha}\right)$

$F(x|\alpha,\beta) = \int_0^x \alpha\beta t^{-\alpha-1} \exp(-\beta t^{-\alpha})\, dt$

Let $t^{-\alpha} = m \Rightarrow t = m^{-\frac{1}{\alpha}} \qquad \frac{dt}{dm} = +\frac{1}{\alpha} m^{-\frac{\alpha+1}{\alpha}}$

$= \int_{x^{-\alpha}}^{\infty} \cancel{\alpha}\beta\, \cancel{m} \cdot \cancel{m^{\frac{1}{\alpha}}} \exp(-\beta \cdot m)\left(+\frac{1}{\cancel{\alpha}}\right) \cancel{m^{-\frac{\alpha+1}{\alpha}}}\, dm$

$= \int_{x^{-\alpha}}^{\infty} \beta \exp(-\beta m)\, dm$

$= -\exp(-\beta m)\Big|_{x^{-\alpha}}^{\infty} = \exp(-\beta x^{-\alpha})$

Generate random number from $[0,1]$ uniformly and calculate the inverse CDF:

$$x = \exp(-\beta y^{-\alpha})$$

$$\log(x) = -\beta y^{-\alpha} \Rightarrow y^{-\alpha} = -\frac{\log(x)}{\beta} \Rightarrow y = \left[-\frac{\log(x)}{\beta}\right]^{-\frac{1}{\alpha}}$$

this is the inverse CDF, $x \in [0,1]$. $y > 0$

$$\pi(\alpha, \beta \mid x) \propto (\alpha\beta)^n (\Pi_{i=1}^n x_i)^{-\alpha-1} \exp(-\beta \Sigma_{i=1}^n x_i^{-\alpha})$$

$$\int \pi(\alpha, \beta \mid x)\, d\beta = \alpha^n (\Pi_{i=1}^n x_i)^{-\alpha-1} \int_0^\infty \beta^n \exp(-\beta \Sigma_{i=1}^n x_i^{-\alpha})\, d\beta$$

$$= \alpha^n (\Pi_{i=1}^n x_i)^{-\alpha-1} \frac{\Gamma(n+1)}{(\Sigma_{i=1}^n x_i^{-\alpha})^{n+1}}$$

This is a stange distribution, use importance sampling?

to get the constant?

I used "adaptIntegrate" function in R and get the

$$\int_0^\infty \int_0^\infty (\alpha\beta)^n (\Pi_{i=1}^n x_i)^{-\alpha-1} \exp(-\beta \Sigma_{i=1}^n x_i^{-\alpha})\, d\alpha\, d\beta$$

Also, I worked with $\exp(\log(\pi(\alpha, \beta \mid x)))$, it

says the $\int_0^\infty \int_0^\infty \pi(\alpha, \beta \mid x)\, d\alpha\, d\beta$ is $1.22\, e^{-78}$

the proportional posterior

$\Rightarrow$ True posterior:

$$(1.22\, e^{-78})^{-1} (\alpha\beta)^n (\Pi_{i=1}^n x_i)^{-\alpha-1} \exp(-\beta \Sigma_{i=1}^n x_i^{-\alpha})$$

(b) $\pi(\alpha.\beta|x) \propto (\alpha\beta)^n (\prod_{i=1}^n x_i)^{-\alpha-1} \exp(-\beta\sum_{i=1}^n x_i^{-\alpha})$

$= C\cdot(\alpha\beta)^n (\prod_{i=1}^n x_i)^{-\alpha-1} \exp(-\beta\sum_{i=1}^n x_i^{-\alpha})$

$\log(\pi(\alpha.\beta|x)) = \log(c) + n[\log(\alpha)+\log(\beta)] - (\alpha+1)\sum_{i=1}^n \log(x_i)$

$- \beta\sum_{i=1}^n x_i^{-\alpha}$

use "optim" function in R, I got:

$(\alpha_0, \beta_0) = (6.542\times10^{-37}, 0.0126)$ [Approximately]

$\partial L/\partial\alpha = \frac{n}{\alpha} + \sum_{i=1}^n \log(x_i) + \beta\sum_{i=1}^n \log(\alpha)\, x_i^{-\alpha}$

$\partial l/\partial\beta = \frac{n}{\beta} - \sum_{i=1}^n x_i^{-\alpha}$

$\partial^2 l/\partial\alpha^2 = -\frac{n}{\alpha} - \beta[\log(\alpha)]^2 \sum x_i^{-\alpha}$

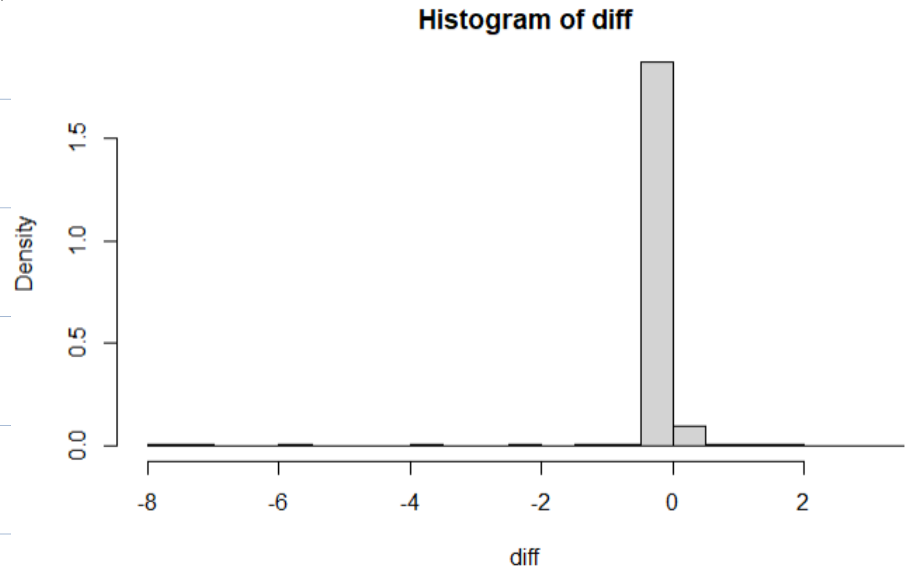
$\partial^2 l/\partial\alpha\partial\beta = \sum_{i=1}^n \log(\alpha)\cdot x_i^{-\alpha}$

$\partial^2 l/\partial\beta\partial\alpha = \sum_{i=1}^n \log(\alpha)\, x_i^{-\alpha}$

$\partial^2 l/\partial\beta^2 = -\frac{n}{\beta^2}$

$\Rightarrow \begin{pmatrix}\alpha_0\\ \beta_0\end{pmatrix} \sim N\left(\begin{pmatrix}5.5\\ 5\end{pmatrix}, \begin{pmatrix}0.0037 & -0.0266\\ -0.0266 & 0.2433\end{pmatrix}\right)$

(c). I choosed $(\alpha, \beta) \in [0,10] \times [0,10]$ and compared the difference:



Also, here is an overall summary:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-7.75824	0.00000	0.00000	-0.03802	0.00000	3.39968

We can see Laplace Approximation is performing good in our case.

Seems performing nice.

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.000000 0.000000 0.000000 0.009991 0.000000 3.399737
```

9. Let $x_1, \ldots, x_n$ be an i.i.d. sample such that $x_i \mid \theta, \sigma^2 \sim \mathrm{N}(\theta, \sigma^2)$ with θ and σ^2 unknown. Assume a conjugate normal-inverse-gamma prior on (θ, σ^2) such that $\theta \mid \sigma^2 \sim \mathrm{N}(\theta_0, \kappa_0\sigma^2)$ and $\sigma^2 \sim \mathrm{IG}(a, b)$ with θ_0, κ_0, a and b known.

Note: The lecture covered Parts (a)-(d). Also, see HW1 solution for Q11-1. You may repeat for practice or you use the results of the lecture directly.

(a) Find the joint posterior $p(\theta, \sigma^2 \mid \boldsymbol{x})$ (up to proportionality).

(b) Find $p(\theta \mid \sigma^2, \boldsymbol{x})$.

(c) Find $p(\sigma^2 \mid \boldsymbol{x})$.

(d) Find $p(\theta \mid \boldsymbol{x})$.

(a)

Prior: $\begin{cases} \sigma^2 : & \dfrac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp\left(-\beta \cdot \dfrac{1}{\sigma^2}\right) \\ \theta \mid \sigma^2 : & \dfrac{1}{\sqrt{2\pi k_0 \sigma^2}} \exp\left(-\dfrac{(\theta-\theta_0)^2}{2k_0\sigma^2}\right) \end{cases}$

Joint $\pi(\theta, \sigma^2) \propto (\sigma^2)^{-\alpha-1-\frac{1}{2}} \exp\left(-\beta/\sigma^2 - \dfrac{(\theta-\theta_0)^2}{2k_0\sigma^2}\right)$

$f(x \mid \theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\dfrac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}\right)$ $\quad \begin{cases} \bar{x} = \dfrac{\sum_{i=1}^n x_i}{n} \\ s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \end{cases}$

$\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\dfrac{s^2 + n(\bar{x}-\theta)^2}{2\sigma^2}\right)$

$\pi(\theta, \sigma^2 \mid x) \propto (\sigma^2)^{-\alpha-1-\frac{1}{2}-\frac{n}{2}} \exp\left(-\dfrac{s^2+n(\bar{x}-\theta)^2}{2\sigma^2} - \dfrac{\beta}{\sigma^2} - \dfrac{(\theta-\theta_0)^2}{2k_0\sigma^2}\right)$

(b) $\pi(\theta \mid \sigma^2, x) \propto \pi(\theta \mid \sigma^2) \cdot f(x \mid \sigma^2, \theta)$

$\propto \exp\left(-\dfrac{(\theta-\theta_0)^2}{2k_0\sigma^2} - \dfrac{s^2 + n(\bar{x}-\theta)^2}{2\sigma^2}\right)$

$\propto \exp\left(-\dfrac{1}{2k_0\sigma^2}\left[(\theta-\theta_0)^2 + k_0 n(\bar{x}-\theta)^2\right]\right)$

$\exp\left(-\dfrac{1}{2k_0\sigma^2}\left[\theta^2(1+k_0 n) - 2\theta(\theta_0 + k_0 n\bar{x})\right]\right)$

$\sim N\left(\dfrac{\theta_0 + k_0 n\bar{x}}{1+k_0 n}, \dfrac{k_0\sigma^2}{1+k_0 n}\right)$

(c) $p(\sigma^2|x) = \int \pi(\sigma^2, \theta | x)\, d\theta$

$$\propto \int (\sigma^2)^{-\alpha-1-\frac{n}{2}\cdot\frac{1}{2}} \exp\left(-\frac{\beta}{\sigma^2}\right) \exp\left(-\frac{S^2+n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2k_0\sigma^2}\right) d\theta$$

$$= (\sigma^2)^{-\alpha-\frac{1}{2}\cdot\frac{n}{2}-1} \exp\left(-\frac{\frac{S^2}{2}+\beta}{\sigma^2}\right) \int \exp\left(-\frac{1}{2k_0\sigma^2}\left(k_0 n\bar{x}^2 - 2k_0 n\bar{x}\theta + k_0 n\theta^2 + \theta^2 - 2\theta\theta_0 + \theta_0^2\right)\right) d\theta$$

$$= (\sigma^2)^{-\alpha-\frac{1}{2}-\frac{n}{2}-1} \exp\left(\frac{\frac{S^2}{2}+\beta}{\sigma^2}\right) \int \exp\Bigg(\underbrace{-\frac{1}{2k_0\sigma^2}\left((1+k_0 n)\theta^2 - 2\theta(k_0 n\bar{x}+\theta_0) + k_0 n\bar{x}^2 + \theta_0^2\right)}_{☆}\Bigg) d\theta$$

$$☆ = -\frac{1+k_0 n}{2k_0\sigma^2}\left(\theta^2 - 2\theta\cdot\frac{k_0 n\bar{x}+\theta_0}{1+k_0 n} + \frac{k_0 n\bar{x}^2+\theta_0^2}{1+k_0 n}\right)$$

$$= -\frac{1+k_0 n}{2k_0\sigma^2}\left[\left(\theta - \frac{k_0 n\bar{x}+\theta_0}{1+k_0 n}\right)^2 \underbrace{- \frac{(k_0 n\bar{x}+\theta_0)^2}{(1+k_0 n)^2} + \frac{k_0 n\bar{x}^2+\theta_0^2}{1+k_0 n}}_{☆☆}\right]$$

$$☆☆ = \frac{-\left[\cancel{(k_0 n)^2\bar{x}^2} + 2k_0 n\bar{x}\theta_0 + \cancel{\theta_0^2}\right] + k_0 n\bar{x}^2 + \cancel{\theta_0^2} + \cancel{(k_0 n)^2}\bar{x}^2 + k_0 n\theta_0^2}{(1+k_0 n)^2}$$

$$= \frac{k_0 n(\bar{x}^2 - 2\bar{x}\theta_0 + \theta_0^2)}{(1+k_0 n)^2}$$

$$p(\sigma^2|x) \propto (\sigma^2)^{-\alpha-\frac{1}{2}-\frac{n}{2}-1} \exp\left(\frac{\frac{S^2}{2}+\beta}{-\sigma^2}\right) \exp\left(-\frac{\cancel{k_0} n(\bar{x}-\theta_0)^2}{2\cancel{k_0}(1+k_0 n)\sigma^2}\right) \cdot (\sigma^2)^{\frac{1}{2}}$$

It's $IG\left(\alpha + \frac{n}{2},\ \beta + \frac{S^2}{2} + \frac{n(\bar{x}-\theta_0)^2}{2(1+k_0 n)}\right)$

(d) $p(\theta \mid x) = \int_0^\infty \pi(\theta, \sigma^2 \mid x)\, d\sigma^2$

$$\propto \int_0^\infty (\sigma^2)^{-\alpha-1-\frac{1}{2}-\frac{n}{2}} \exp\left(-\frac{s^2+n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{\beta}{\sigma^2} - \frac{(\theta-\theta_0)^2}{2k_0\sigma^2}\right) d\sigma^2$$

It's a IG kernel:

$$\propto \Bigg[\underbrace{\frac{s^2+n(\bar{x}-\theta)^2}{2} + \frac{(\theta-\theta_0)^2}{2k_0}}_{\star\star\star} \Bigg]^{-\frac{2\alpha+n+1}{2}}$$

$$\star\star\star \propto k_0 s^2 + k_0 n(\bar{x}-\theta)^2 + (\theta-\theta_0)^2$$

$$= k_0 s^2 + (k_0 n+1)\theta^2 - 2\theta(k_0 n\bar{x}+\theta_0) + k_0 n\bar{x}^2 + \theta_0^2$$

$$= k_0 s^2 + k_0 n\bar{x}^2 + \theta_0^2 + (k_0 n+1)\left(\theta - \frac{k_0 n\bar{x}+\theta_0}{k_0 n+1}\right)^2 - \frac{(k_0 n\bar{x}+\theta_0)^2}{k_0 n+1}$$

$$= k_0 s^2 + \frac{\cancel{(k_0 n)^2\bar{x}^2} + k_0 n\theta_0^2 + k_0 n\bar{x}^2 + \cancel{\theta_0^2}}{k_0 n+1} - \frac{\cancel{(k_0 n)^2\bar{x}^2} + 2k_0 n\theta_0\bar{x} + \cancel{\theta_0^2}}{k_0 n+1}$$

$$+ (k_0 n+1)\left(\theta - \frac{k_0 n\bar{x}+\theta_0}{k_0 n+1}\right)^2$$

$$= k_0 s^2 + \frac{k_0 n(\theta_0-\bar{x})^2}{k_0 n+1} + (k_0 n+1)\left(\theta - \frac{k_0 n\bar{x}+\theta_0}{k_0 n+1}\right)^2$$

$$\propto 1 + \frac{\left(\theta - \frac{k_0 n\bar{x}+\theta_0}{k_0 n+1}\right)^2}{\frac{(k_0 n+1)\left(k_0 s^2 + \frac{k_0 n(\theta_0-\bar{x})^2}{k_0 n+1}\right)}{2\alpha+n}(2\alpha+n)}$$

It follows a t-distribution:

$$\begin{cases} \text{Location: } \dfrac{k_0 n\bar{x} + \theta}{1 + k_0 n} \\ \text{Variance: } \dfrac{(k_0 n+1)\left(k_0 s^2 + \dfrac{k_0 n(\theta_0 - \bar{x})^2}{k_0 n + 1}\right)}{2\alpha + n} \\ \text{DF: } 2\alpha + n \end{cases}$$

(e) Simulate $n = 1000$ i.i.d. observations from a N$(5, 1)$. Fit the above model to these data assuming the following prior scenarios: (i) fairly informative priors around the true values of both parameters, (ii) informative prior on θ and vague on σ^2 (iii) informative prior on σ^2 and vague on θ (iv) vague on both parameters. Specify the form of your posteriors in each case.

(i) Prior: $\theta | \sigma^2 : N(5, \frac{1}{10}\sigma^2)$

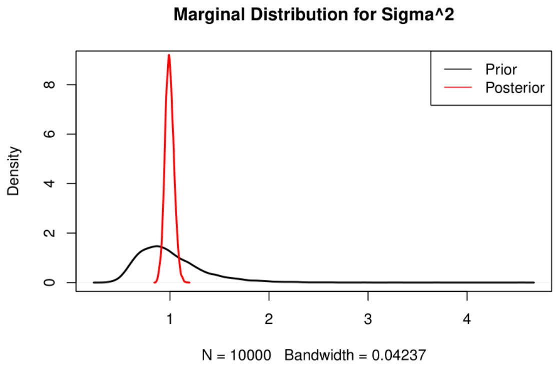
$\sigma^2 : IG(11, 10)$

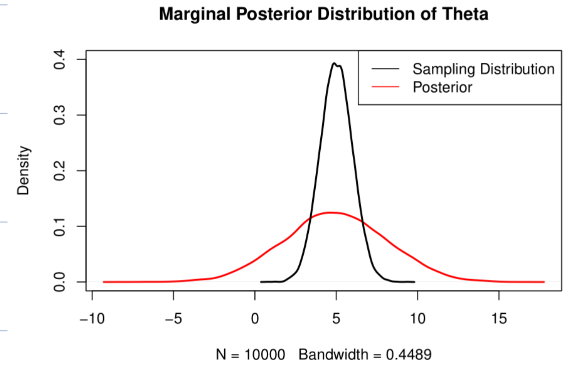
for $IG(\alpha, \beta)$: Mean: $\dfrac{\beta}{\alpha - 1}$ Var: $\dfrac{\beta^2}{(\alpha-1)^2(\alpha-2)} = (\text{mean})^2 \cdot \dfrac{1}{\alpha - 2}$

Therefore, as $\alpha \uparrow$, Var $\downarrow$, more informative.

Here is the form of posterior and comparisons with other distributions:

posterior: $\begin{cases} \sigma^2 | x : IG \\ \theta | x : t \end{cases}$

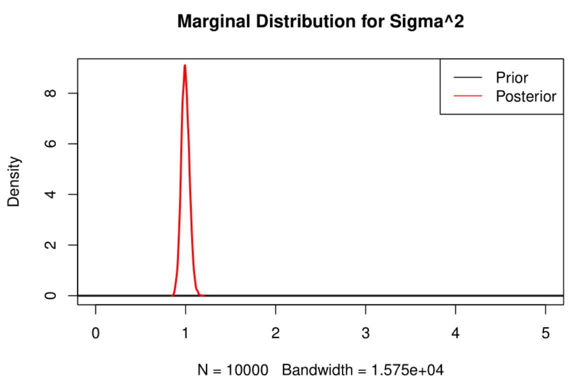


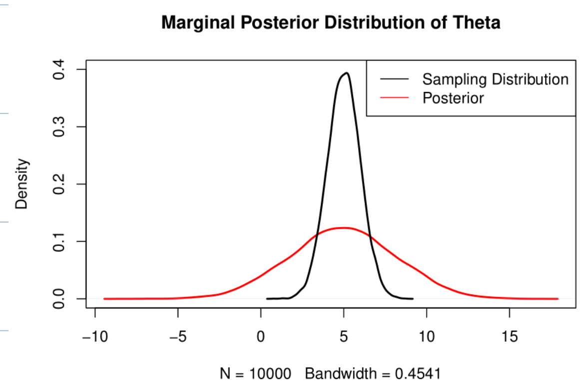


(ii) $\theta | \sigma^2 \sim N(5, \frac{1}{10}\sigma^2)$

$\sigma^2 \sim IG(0.1, 0.1)$

θ is informative and σ^2 is vague





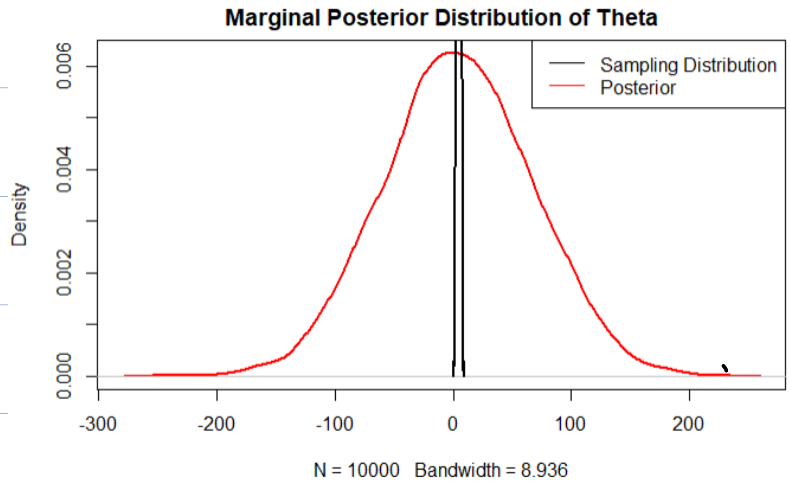
(iii) $\theta|\sigma^2 \sim N(5,\ 0.2\sigma^2)$

$\sigma^2 \sim IG(11, 10)$

In this case, I find something interesting.

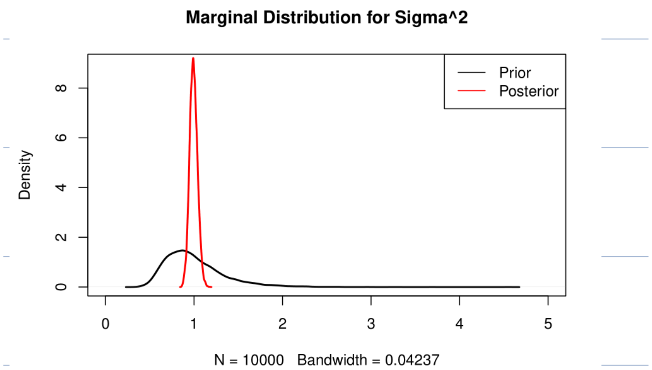
If k_0 is very big, then the posterior of theta seems flat

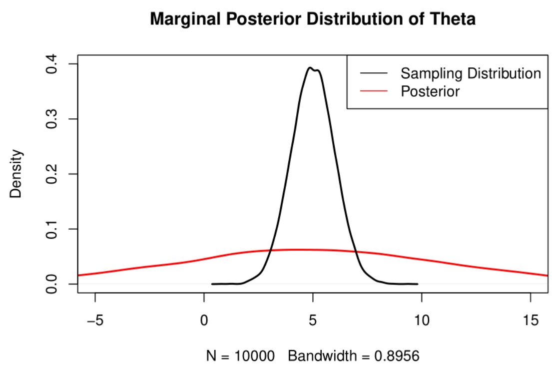
compared with the $N(5,1)$. it's still non-informative



This is the case that $k_0 = 2$

But for the case $k_0 = 0.2$, $\theta_0 = 5$:





However, if the θ_0 is wrongly specified:

for example: $k_0 = 0.2$ $\theta_0 = 15$

As $k_0 \uparrow$, the affect on $\sigma^2 | x$ is decreasing.

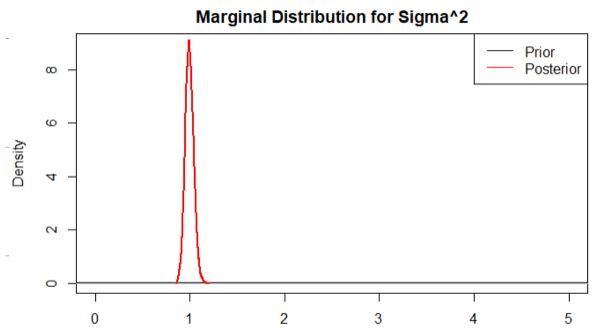
If k is small, the $\sigma^2 | x$ will also specify a wrong concentration. i.e. We should specify θ_0 correctly!

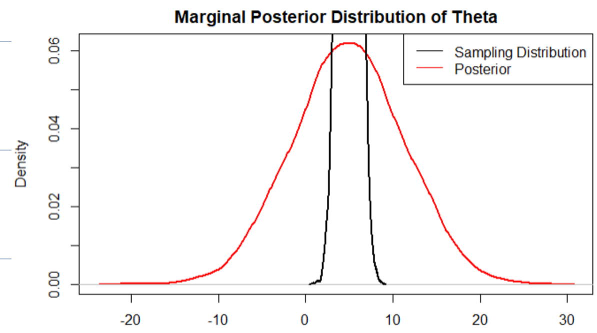
prior.

It's very important.

(iv). $\theta \mid \sigma^2 \sim N(5,\ 0.2 \cdot \sigma^2)$

$\sigma^2 \sim IG(0.1,\ 0.1)$





Still, if we specify θ_0 wrongly, if k_0 is big, then no effect on $\sigma^2 \mid x$, but if k_0 is not big enough, then $\sigma^2 \mid x$ will also be far from the true value.

(f) Assume that you are interested in estimating $\eta = \theta/\sigma$. Develop a Monte Carlo algorithm for computing the posterior mean and a 95% credible interval for η. Use the algorithm to compute such quantities under all the prior scenarios described above.

I will do in this way:

I. Sample $\sigma^{2(i)}$ from $\pi(\sigma^2 | x)$

II. Sample $\theta^{(i)}$ from $\pi(\theta | \sigma^{2(i)}, x)$

III. Calculate $\theta^{(i)} / \sqrt{\sigma^{2(i)}} = \eta^{(i)}$

IV. Use quantile of η to estimate 95% HPD.

$$\sigma^2 \sim IG\left(\alpha + \frac{n}{2}, \beta + \frac{S^2}{2} + \frac{n(\bar{x}-\theta_0)^2}{2(1+k_0 n)}\right)$$

$$\theta | \sigma^2 \sim N\left(\frac{\theta_0 + k_0 n \bar{x}}{1 + k_0 n}, \frac{k_0 \sigma^2}{1 + k_0 n}\right)$$

Still, I find that if θ_0 is far from the true mean, i.e. 5, then the ratio changes a lot!

But in my case, I set $\theta_0 = 5$ for all:

Table 4: Ratio of Mean to SD for 4 Cases

	2.5% Quantile	97.5% Quantile
Case 1	4.789701	5.208691
Case 2	4.768147	5.227301
Case 3	4.789512	5.208661
Case 4	4.767879	5.227299

If θ_0 is correct in prior of $\theta|\sigma^2$, then I have following results:
As long as σ^2 is vague, then this ratio varies more than the other cases. This ratio seems not to be sensitive to the change of vague prior of $\theta|\sigma^2$.

10. Consider the usual regression model, $y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$, covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$ and a coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$.

Fact: a sufficient statistic is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$, where $\boldsymbol{X}$ is a $n \times p$ matrix whose ith row is $\boldsymbol{x}_i$, and $\boldsymbol{y}$ is the $n-$dimensional column vector of $(y_1, \ldots, y_n)$. It is also the maximum likelihood estimator and the least-squares estimator of $\boldsymbol{\beta}$.

(a) Consider the following priors on $(\boldsymbol{\beta}, \sigma^2)$;

$$\boldsymbol{\beta} \mid \sigma^2 \sim \mathrm{N}_p\left(\boldsymbol{\beta}_0, \frac{\sigma^2}{n_0}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\right), \quad \text{and} \quad \sigma^2 \sim \mathrm{IG}(\nu/2, s_0^2/2),$$

where $\boldsymbol{\beta}_0$, n_0, ν and s_0^2 are fixed.

i. Find an expression of the joint posterior distribution as follows;

$$\pi(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) = \pi_1(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X})\pi_2(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}).$$

Also, identify $\pi_1(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X})$ and $\pi_2(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X})$

ii. Is the prior conjugate? Explain.

$f(\mathbf{x} \mid \boldsymbol{\theta}, \Sigma) = |\Sigma|^{-1/2}(2\pi)^{-p/2} \exp\left\{-(\mathbf{x}-\boldsymbol{\theta})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\theta})/2\right\}.$

(a) $\pi(\beta, \sigma^2) = \pi(\beta \mid \sigma^2) \cdot \pi(\sigma^2)$

$\propto \left[\frac{\sigma^2}{n_0}(X^TX)^{-1}\right]^{-\frac{1}{2}} \exp\left(-\frac{(\beta-\beta_0)^T\left(\frac{n_0}{\sigma^2}\cdot X^TX\right)(\beta-\beta_0)}{2}\right)$

$\cdot \; {\sigma^2}^{-\frac{\nu}{2}-1} \exp\left(\frac{-s_0^2}{2\sigma^2}\right)$

$f(x, y \mid \beta, \sigma^2) = \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right)$

$\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i - x_i\beta)^2}{2\sigma^2}\right)$

$\pi(\beta, \sigma^2 \mid x, y) \propto \pi(\beta, \sigma^2) \cdot f(x, y \mid \beta, \sigma^2)$

$\propto (\sigma^2)^{-\frac{1}{2}-\frac{\nu}{2}-\frac{n}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\left(s_0^2 + (\beta-\beta_0)^T \cdot n_0(X^TX)(\beta-\beta_0) + \sum_{i=1}^n (y_i - x_i\beta)^2\right)\right)$

$\pi(\beta \mid \sigma^2, x, y) = \dfrac{\pi(\beta, \sigma^2 \mid x, y)}{\int_0^\infty \pi(\beta, \sigma^2 \mid x, y)\, d\sigma^2} \propto \pi(\beta, \sigma^2 \mid x, y)$

$$\pi(\beta,\sigma^2 \mid x,y) \propto \pi(\beta,\sigma^2)\cdot f(x,y\mid \beta,\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{1}{2}-\frac{\nu}{2}-\frac{n}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\left(S_0^2+(\beta-\beta_0)^T\cdot n_0(x^Tx)(\beta-\beta_0)+\sum_{i=1}^n(y_i-x_i\beta)^2\right)\right)$$

$$\pi(\beta\mid\sigma^2,x,y) \propto \exp\left(-\frac{1}{2\sigma^2}\left[(\beta-\beta_0)^T n_0(x^Tx)(\beta-\beta_0)+(y-x\beta)^T(y-x\beta)\right]\right)$$

$\star$

$y: n\times 1 \quad x: n\times p \quad \beta: p\times 1 \quad \beta_0: p\times 1$

$$\star: \beta^T n_0(x^Tx)\beta - \beta^T n_0(x^Tx)\beta_0 - \beta_0^T n_0(x^Tx)\beta + \beta_0^T n(x^Tx)\beta_0$$

Constant

Constant

$$+y^Ty - y^Tx\beta - \beta^Tx^Ty + \beta^Tx^Tx\beta$$

$$= \beta^T[n_0(x^Tx)+(x^Tx)]\beta - \beta^T(n_0(x^Tx)\beta_0+x^Ty) - (\beta_0^Tn_0(x^Tx)+y^Tx)\beta + \text{constant}$$

$$\Sigma_{new} = \left(\frac{n_0+1}{\sigma^2}\right)(x^Tx)^{-1}$$

$$\mu_{new} = [(n_0+1)(x^Tx)]^{-1}(n_0(x^Tx)\beta_0+x^Ty)$$

It's a MVN

$$\pi_2 = \int_{R^p} \pi(\theta, \beta \mid X, y)\, d\beta$$

$$\propto \int_{R^p} (\sigma^2)^{-\frac{1}{2}-\frac{\nu}{2}-\frac{n}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\left(S_0^2 + (\beta-\beta_0)^T \cdot n_0(X^TX)(\beta-\beta_0) + \sum_{i=1}^{n}(y_i - x_i\beta)^2\right)\right) d\beta$$

$$\propto (\sigma^2)^{-\frac{1}{2}-\frac{\nu}{2}-\frac{n}{2}-1} \exp\left(-\frac{S_0^2 + \beta_0^T n(X^TX)\beta_0 + y^Ty}{2\sigma^2}\right)$$

$$\int_{R^p} \exp\left(-\frac{1}{2\sigma^2}\left(\beta^T n_0(X^TX)\beta - \beta^T n_0(X^TX)\beta_0 - \beta_0^T n_0(X^TX)\beta - y^TX\beta - \beta^TX^Ty + \beta^TX^TX\beta\right)\right) d\beta$$

$\star\star$

$\star\star =$

$$[(n_0+1)(X^TX)] \cdot \Big\{ \left[\beta - [(n_0+1)(X^TX)]^{-1}(n_0 X^TX\beta_0 + X^Ty)\right]^T \left[\beta - [(n_0+1)(X^TX)]^{-1}(n_0 X^TX\beta_0 + X^Ty)\right]$$

$$- \left\{[(n_0+1)(X^TX)]^{-1}(n_0 X^TX\beta_0 + X^Ty)\right\}^T [(n_0+1)(X^TX)]^{-1}(n_0 X^TX\beta_0 + X^Ty) \Big\}$$

MVN Kernel

$$\pi_2 \propto (\sigma^2)^{-\frac{1}{2}-\frac{\nu}{2}-\frac{n}{2}-1} \exp\left(-\frac{S_0^2 + \beta_0^T n(X^TX)\beta_0 + y^Ty}{2\sigma^2}\right) \cdot$$

$$\exp\left(-\frac{(n_0+1)(X^TX)}{2\sigma^2}\left[[(n_0+1)(X^TX)]^{-1}(n_0X^TX\beta_0 + X^Ty)\right]^T \left[[(n_0+1)(X^TX)]^{-1}(n_0X^TX\beta_0 + X^Ty)\right]\right)$$

$$\cdot (\sigma^2)^{\frac{1}{2}}$$

let this be β_1

$$\pi_2 = \propto (\sigma^2)^{-\frac{\nu}{2}-\frac{n}{2}-1} \quad \exp\left(-\frac{S_0^2 + y^Ty + \beta_0^T n x^T x \beta_0 + (n_0+1) x^T x \beta_1^T \beta_1}{2\sigma^2}\right)$$

It's IG

$$\left(\frac{\nu+n}{2}, -\frac{S_0^2 + y^Ty + \beta_0^T n x^T x \beta_0 + (n_0+1)(x^Tx)\beta_1^T\beta_1}{2}\right)$$

(b)

It's conjugate.

Since $\pi(\beta|\sigma^2)$ and $\pi(\beta|\sigma^2, x, y)$ both belong to MVN family, and $\pi(\sigma^2)$, $\pi(\sigma^2|x,y)$ both belong to IG family.