Assignment 2
Bayesian Inference FTN0548

Usama Zafar, Li Ju

March 13, 2024

1 Problem 4.1

(a) With the given $p(x|\theta)$ in the table and $\pi(\theta)$, the joint distribution and marginal distributions of x and θ are computed as follows and shown in Table 1.

$$
\begin{aligned}
p(x, \theta) &= p(x|\theta)\pi(\theta) \\
p(x) &= \sum_{\theta \in \{0,1,2\}} p(x, \theta)
\end{aligned}
$$

	$x = -1$	$x = 0$	$x = 1$	$\pi(\theta)$
$\theta = 0$	$0.5 - 0.5p$	$0.3 - 0.3p$	$0.2 - 0.2p$	$1 - p$
$\theta = 1$	0	$0.5p$	$0.5p$	p
$p(x)$	$0.5 - 0.5p$	$0.3 + 0.2p$	$0.2 + 0.3p$	

Table 1: Joint and marginal distributions of x and θ.

The posterior distribution of θ given x is computed as follows and shown in Table 2.

$$
\pi(\theta|x) = \frac{p(x, \theta)}{p(x)}
$$

	$x = -1$	$x = 0$	$x = 1$
$\theta = 0$	1	$\frac{3-3p}{3+2p}$	$\frac{2-2p}{2+3p}$
$\theta = 1$	0	$\frac{5p}{3+2p}$	$\frac{5p}{2+3p}$

Table 2: Posterior distribution of θ given x $\Pr(\theta|x)$.

(b) The Bayes estimator is defined as

$$
\begin{aligned}
\hat{\theta}^{\pi} &:= \arg\min_{\hat{\theta}} \mathbb{E}_{\pi(\theta|x)} L(\theta, \hat{\theta}) \\
&= \arg\min_{\hat{\theta}} \sum_{\theta_0 \in \{0,1\}} L(\theta_0, \hat{\theta}) \Pr(\theta = \theta_0 | x)
\end{aligned}
$$

The expected posterior loss is computed as follows:

$$\mathbb{E}_{\pi(\theta|x)} L(\theta, \hat{\theta}) = \begin{cases} L(0, \hat{\theta}) & x = -1 \\ L(0, \hat{\theta}) \frac{3-3p}{3+2p} + L(1, \hat{\theta}) \frac{5p}{3+2p} & x = 0 \\ L(0, \hat{\theta}) \frac{2-2p}{2+3p} + L(1, \hat{\theta}) \frac{5p}{2+3p} & x = 1 \end{cases}$$

Hence, with the $0-1$ loss function L, the Bayes estimator for θ is given in Table 3.

	$p > \frac{3}{8}$	$\frac{2}{7} \geq p > \frac{2}{7}$	$\frac{2}{7} \geq p$
$x = -1$	0	0	0
$x = 0$	1	0	0
$x = 1$	1	1	0

Table 3: The Bayes estimator for θ.

(c) The frequentist risk is defined by

$$\begin{aligned} R(\theta_0, \hat{\theta}) &:= \mathbb{E}_{p(x|\theta_0)} L(\theta_0, \hat{\theta}) \\ &= \sum_{x \in \{-1,0,1\}} L(\theta_0, \hat{\theta}) \Pr(X = x|\theta_0) \end{aligned}$$

Hence, the frequentist risks of the Bayes estimator of θ are given in Table 4.

	$p > \frac{3}{8}$	$\frac{2}{7} \geq p > \frac{2}{7}$	$\frac{2}{7} \geq p$
$\theta_0 = 0$	0.5	0.2	0
$\theta_0 = 1$	0	0.5	1

Table 4: The frequentist risk for the Bayes estimator $\hat{\theta}^{\pi}$.

(d) The Bayes risk is defined by

$$r(\pi, \delta) := \mathbb{E}^{\pi} R(\theta, \delta) = \sum_{\theta_0 \in \{0,1\}} R(\theta_0, \delta) \pi(\theta_0)$$

Hence, the Bayes risk of the Bayes estimator for θ is given by

$$r(\pi, \hat{\theta}^{\pi}) = \begin{cases} 0.5 - 0.5p & p > \frac{3}{8} \\ 0.2 + 0.3p & \frac{2}{7} \geq p > \frac{2}{7} \\ p & \frac{2}{7} \geq p \end{cases} \tag{1}$$

(e) The least favourable prior is defined by

$$\pi_0 := \arg\max_{\pi} r(\pi, \hat{\theta}^{\pi})$$

Given that $r(\pi, \hat{\theta}^{\pi(p)})$ is a piecewise linear function w.r.t p as shown in Equation 1, the suppremum of the $r(\pi, \hat{\theta}^{pi})$ is given when $p = \frac{3}{8}$. Hence, the least favourable prior is $\pi_0 = \Pr(\theta = 1) = \frac{3}{8}$.

2 Problem 4.2

(a) Recognizing that the prior is conjugate to the binomial likelihood function, with observation x, we have the posterior distribution for θ as

$$\pi(\theta|x) \sim \mathrm{Beta}(\alpha + x, \beta + n - x)$$

Then the expected posterior loss is given by

$$\begin{aligned}
\mathbb{E}_{p(\theta|x)} L(\theta, d) &= \int L(\theta, d)\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x+1} d\theta \cdot \frac{1}{\mathbf{B}(\alpha + x, \beta + n - x)} \\
&= k_1 \int_0^d (d-\theta)\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x+1} d\theta \cdot \frac{1}{\mathbf{B}(\alpha + x, \beta + n - x)} \\
&\quad + k_2 \int_d^1 (\theta-d)\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x+1} d\theta \cdot \frac{1}{\mathbf{B}(\alpha + x, \beta + n - x)} \\
&= k_1 \left(d\mathrm{CDF}_{\alpha+x,\beta+n-x}(d) - \mathrm{CDF}_{\alpha+x+1,\beta+n-x}(d) \frac{\mathbf{B}(\alpha + x + 1, \beta + n - x)}{\mathbf{B}(\alpha + x, \beta + n - x)} \right) \\
&\quad + k_2 \left((1 - \mathrm{CDF}_{\alpha+x+1,\beta+n-x}(d)) \frac{\mathbf{B}(\alpha + x + 1, \beta + n - x)}{\mathbf{B}(\alpha + x, \beta + n - x)} - d(1 - \mathrm{CDF}_{\alpha+x,\beta+n-x}(d)) \right) \\
&= k_1 \left(d\mathrm{CDF}_{\alpha+x,\beta+n-x}(d) - \mathrm{CDF}_{\alpha+x+1,\beta+n-x}(d) \frac{\alpha + x}{\alpha + \beta + n} \right) \\
&\quad + k_2 \left((1 - \mathrm{CDF}_{\alpha+x+1,\beta+n-x}(d)) \frac{\alpha + x}{\alpha + \beta + n} - d(1 - \mathrm{CDF}_{\alpha+x,\beta+n-x}(d)) \right)
\end{aligned}$$

where $\mathbf{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, and $\mathrm{CDF}_{a,b}$ denotes the cummulative distribution function of distribution $\mathrm{Beta}(a, b)$.

The derivative of $\mathbb{E}_{p(\theta|x)} L(\theta, d)$ w.r.t d is given by

$$\begin{aligned}
\frac{\partial \mathbb{E}_{p(x|\theta)} L(\theta, d)}{\partial d} &= k_1 \left(\mathrm{CDF}_{\alpha+x,\beta+n-x}(d) + d\mathrm{PDF}_{\alpha+x,\beta+n-x}(d) - \frac{\alpha + x}{\alpha + \beta + n} \mathrm{PDF}_{\alpha+x+1,\beta+n-x}(d) \right) \\
&\quad + k_2 \left(-\frac{\alpha + x}{\alpha + \beta + n} \mathrm{PDF}_{\alpha+x+1,\beta+n-x}(d) - 1 + \mathrm{CDF}_{\alpha+x,\beta+n-x}(d) + d\mathrm{PDF}_{\alpha+x,\beta+n-x}(d) \right) \\
&= -k_2 + (k_1 + k_2)\mathrm{CDF}(d)
\end{aligned}$$

where $\mathrm{PDF}_{a,b}$ denotes the probability density function of $\mathrm{Beta}(a, b)$.

The Bayes estimator is given when $\frac{\partial \mathbb{E}_{p(x|\theta)} L(\theta, d)}{\partial d}(\hat{\theta}^\pi) = 0$. Hence, we have

$$\mathrm{CDF}_{\alpha+x,\beta+n-x}(\hat{\theta}^\pi) = \frac{k_2}{k_1 + k_2}$$

Hence, the Bayes estimator for θ is given by the $\frac{k_2}{k_1+k_2}$ quantile of the posterior distribution $\mathrm{Beta}(\alpha + x, \beta + n - x)$.

(b) Let $k_1 = k_2 = k$. The Bayes estimator satisfies

$$\mathrm{CDF}_{\alpha+x,\beta+n-x}(\hat{\theta}^\pi) = \frac{1}{2}$$

Recall that the median of $\mathrm{Beta}(a, b)$ is approximately given by $\frac{a-\frac{1}{3}}{a+b-\frac{2}{3}}$, the Bayes estimator is given by

$$\hat{\theta}^\pi(x) = \frac{\alpha + x - \frac{1}{3}}{\alpha + \beta + n - \frac{2}{3}}$$

(c) The use of asymmetric loss is advantageous when the repercussions of overestimation differ significantly from those of underestimation. By allocating larger coefficients to the components that could lead to more serious consequences, the frequency of such adverse outcomes may be diminished. An real-life example is medical diagnosis, in which the use of an asymmetric loss function is crucial because the costs of false negatives and false positives are highly imbalanced.

(d) In Bayesian decision theory, achieving a constant risk means that the expected loss incurred by the decision rule is constant regardless of the true value of the parameter θ. This is desirable because it ensures a consistent performance of the decision rule across different parameter values.

In this particular case it is very difficult to prove or disprove existence of such choice of prior that would yield constant risk. To illustrate the problem faced, we begin by recalling that the risk for an estimator $\hat{\theta}$ given θ_0 is defined as

$$R(\theta_0, \hat{\theta}) := \mathbb{E}_{p(x|\theta_0)} L(\theta_0, \hat{\theta}) = \int_{\mathcal{X}} L(\theta_0, \hat{\theta}) p(x|\theta_0) dx$$

Since the loss function under consideration is asymmetric we split the above integral into two parts i.e.

$$\begin{aligned} R(\theta_0, \hat{\theta}) &= \int_{\substack{x \in \mathcal{X} \\ \theta_0 > \hat{\theta}}} L(\theta_0, \hat{\theta}) p(x|\theta_0) dx + \int_{\substack{x \in \mathcal{X} \\ \theta_0 < \hat{\theta}}} L(\theta_0, \hat{\theta}) p(x|\theta_0) dx \\ &= k_2 \int_{\substack{x \in \mathcal{X} \\ \theta_0 > \hat{\theta}}} (\theta_0 - \hat{\theta}) \binom{n}{x} \theta_0^x (1-\theta_0)^{n-x} dx - k_1 \int_{\substack{x \in \mathcal{X} \\ \theta_0 < \hat{\theta}}} (\theta_0 - \hat{\theta}) \binom{n}{x} \theta_0^x (1-\theta_0)^{n-x} dx \end{aligned}$$

To achieve constant risk is same as to say putting the derivative of above equal to 0. Computing the derivative for above expression is quite complicated (might not even have a closed form solution). We conjecture that it might not be possible to find a prior that results in constant risk but we do not have formal way to prove it.

3 Problem 5.1

Notation: $\Gamma(a, b)$ denotes a gamma distribution with parameters a and b, and $\Gamma(z)$ denotes a gamma function over z.

(a) Given the model $X|\theta \sim \text{Exp}(\theta)$ and prior $\theta \sim \Gamma(\alpha, \beta)$, we have the posterior distribution of θ given an i.i.d. sample $\boldsymbol{x}_{(n)} = (x_1, \ldots, x_n)$

$$\begin{aligned} \pi(\theta|\boldsymbol{x}_{(n)}) &\propto p(\boldsymbol{x}_{(n)}|\theta)\pi(\theta) \\ &= \prod \text{Exp}(\theta)\Gamma(\alpha, \beta) \\ &= \theta^n \exp\{-\theta \sum_{i=1}^n x_i\} \theta^{\alpha-1} \exp\{-\beta\theta\} \\ &= \theta^{n+\alpha-1} \exp\{-\theta(\beta + \sum_{i=1}^n x_i)\} \end{aligned}$$

Hence, the posterior distribution $\theta|\boldsymbol{x}_{(n)} \sim \Gamma(n + \alpha, \beta + \sum_{i=1}^n x_i)$.

(b) With L_2 loss, the Bayes estimator for θ is given by

$$\begin{aligned}
\delta(\boldsymbol{x}_{(n)}) &= \arg\min_{\hat{\theta}} \mathbb{E}_{\pi(\theta|\boldsymbol{x}_{(n)})} L_2(\theta, \hat{\theta}) \\
&= \arg\min_{\hat{\theta}} \int (\theta - \hat{\theta})^2 \theta^{n+\alpha-1} \exp\{-\theta(\beta + \sum_{i=1}^n x_i)\} d\theta \cdot \frac{(\beta + \sum_{i=1}^n x_i)^{\alpha+n}}{\Gamma(n+\alpha)} \\
&= \arg\min_{\hat{\theta}} \hat{\theta}^2 - 2\hat{\theta} \int \theta^{n+\alpha} \exp\{-\theta(\beta + \sum_{i=1}^n x_i)\} d\theta \cdot \frac{(\beta + \sum_{i=1}^n x_i)^{\alpha+n}}{\Gamma(n+\alpha)} + \text{const.} \\
&= \arg\min_{\hat{\theta}} \hat{\theta}^2 - 2\hat{\theta} \frac{\Gamma(n+\alpha+1)}{(\beta + \sum_{i=1}^n x_i)^{\alpha+n+1}} \frac{(\beta + \sum_{i=1}^n x_i)^{\alpha+n}}{\Gamma(n+\alpha)} + \text{const.} \\
&= \arg\min_{\hat{\theta}} \hat{\theta}^2 - 2\hat{\theta} \frac{n+\alpha}{\beta + \sum_{i=1}^n x_i} + \text{const.} \\
\delta(\boldsymbol{x}_{(n)}) &= \frac{n+\alpha}{\beta + \sum_{i=1}^n x_i}
\end{aligned}$$

(c)

Proof. First we show the strong consistency of the posterior $\pi(\theta|\boldsymbol{x}_n)$. With a fixed n, the expectation and variance of the posterior are given by

$$\begin{aligned}
\mathbb{E}_{\pi(\theta|\boldsymbol{x}_n)}[\theta] &= \frac{n+\alpha}{\beta + \sum_{i=1}^n x_i} \\
\mathbb{V}_{\pi(\theta|\boldsymbol{x}_n)}[\theta] &= \frac{n+\alpha}{(\beta + \sum_{i=1}^n x_i)^2}
\end{aligned}$$

The strong law of large numbers yields $\sum_i^n x_i \to n\mathbb{E}[X]$a.s. for $n \to \infty$. With $\mathbb{E}_{p(X|\theta)}[X] = \frac{1}{\theta}$ and $\mathbb{V}_{p(X|\theta)} = \frac{1}{\theta^2}$ then we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}_{\pi(\theta|\boldsymbol{x}_n)}[\theta] &= \lim_{n\to\infty} \frac{n+\alpha}{\beta + \sum_{i=1}^n x_i} \\
&= \lim_{n\to\infty} \frac{n+\alpha}{\beta + \frac{n}{\theta}} \\
&= \theta \\
\lim_{n\to\infty} \mathbb{V}_{\pi(\theta|\boldsymbol{x}_n)}[\theta] &= \lim_{n\to\infty} \frac{n+\alpha}{(\beta + \sum_{i=1}^n x_i)^2} \\
&= \lim_{n\to\infty} \frac{n+\alpha}{\beta^2 + 2\beta\sum_{i=1}^n x_i + (\sum_{i=1}^n x_i)^2} \\
&= \lim_{n\to\infty} \frac{n+\alpha}{\beta^2 + \frac{2n\beta}{\theta} + \frac{n^2}{\theta^2} + \frac{n^2}{\theta^2}} \\
&= 0
\end{aligned}$$

Then with a prior $\theta \sim \Gamma(\alpha, \beta)$, which has a support over all positive real numbers and an L_2 loss function, which has an unique minimum at θ, using Theorem 5.1, we ensure that the Bayes estimator $\delta(\boldsymbol{x}_{(n)}) = \frac{n+\alpha}{\beta+\sum_{i=1}^n x_i}$ is consistent.

□

4 Problem 5.2

(a) Assume $p(\boldsymbol{\theta})$ in exponential family in the form of $p(\boldsymbol{\theta}) = h(x)g(\boldsymbol{\theta})\exp(\boldsymbol{\eta}(\boldsymbol{\theta})^\top \boldsymbol{T}(x))$. The Kullback-Leibler divergence between $p(\boldsymbol{\theta}_0)$ and $p(\boldsymbol{\theta})$ is given as follows:

$$
\begin{aligned}
\mathrm{KL}\left(p(\boldsymbol{\theta}_0)\|p(\boldsymbol{\theta})\right) &= \int p_x(\boldsymbol{\theta}_0) \ln \frac{p_x(\boldsymbol{\theta}_0)}{p_x(\boldsymbol{\theta})} dx \\
&= \int p(\boldsymbol{\theta}_0) \ln \frac{h(x)g(\boldsymbol{\theta}_0)\exp(\boldsymbol{\eta}(\boldsymbol{\theta}_0)^\top \boldsymbol{T}(x))}{h(x)g(\boldsymbol{\theta})\exp(\boldsymbol{\eta}(\boldsymbol{\theta})^\top \boldsymbol{T}(x))} dx \\
&= \ln \frac{g(\boldsymbol{\theta}_0)}{g(\boldsymbol{\theta})} + \int p(\boldsymbol{\theta}_0) \left(\boldsymbol{\eta}(\boldsymbol{\theta}_0) - \boldsymbol{\eta}(\boldsymbol{\theta})\right)^\top \boldsymbol{T}(x) dx \\
&= \ln \frac{g(\boldsymbol{\theta}_0)}{g(\boldsymbol{\theta})} + \left(\boldsymbol{\eta}(\boldsymbol{\theta}_0) - \boldsymbol{\eta}(\boldsymbol{\theta})\right)^\top \mathbb{E}_{p(\boldsymbol{\theta}_0)}\left[\boldsymbol{T}(x)\right]
\end{aligned}
$$

(b) For Gamma distributions, first we reform $\Gamma(\alpha, \beta)$ in exponential form:

$$\Gamma_x(\alpha, \beta) = h(x)g(\alpha, \beta)\exp(\boldsymbol{\eta}(\alpha, \beta)\boldsymbol{T}(x))$$

where $h(x) = 1$, $g(\boldsymbol{\theta}) = \frac{\beta^\alpha}{\Gamma(\alpha)}$, $\boldsymbol{\eta}(\alpha, \beta) = (\alpha - 1, -\beta)^\top$ and $\boldsymbol{T}(x) = (\ln x, x)^\top$.

Then the Kullback-Leibler divergence between $\Gamma(\alpha_0, \beta_0)$ and $\Gamma(\alpha, \beta)$ is given by

$$
\begin{aligned}
\mathrm{KL}\left(\Gamma(\alpha_0, \beta_0)\|\Gamma(\alpha, \beta)\right) &= \ln \frac{\beta_0^{\alpha_0}\Gamma(\alpha)}{\beta^\alpha \Gamma(\alpha_0)} + (\alpha_0 - \alpha)\mathbb{E}_{\Gamma(\alpha_0,\beta_0)}[\ln x] - (\beta_0 - \beta)\mathbb{E}_{\Gamma(\alpha_0,\beta_0)}[x] \\
&= \alpha_0 \ln \beta_0 - \alpha \ln \beta + \ln \Gamma(\alpha) - \ln \Gamma(\alpha_0) + (\alpha_0 - \alpha)(\psi(\alpha_0) - \ln \beta_0) - (\beta_0 - \beta)\frac{\alpha_0}{\beta_0} \\
&= \alpha(\ln \beta_0 - \ln \beta) + \ln \Gamma(\alpha) - \ln \Gamma(\alpha_0) + (\alpha_0 - \alpha)\psi(\alpha_0) + \frac{\beta - \beta_0}{\beta_0}\alpha_0
\end{aligned}
$$

where ψ is the digamma function $\psi(z) = \frac{d\Gamma(z)}{dz}$.

5 Problem 5.4

Proof. With the conjugate prior $\theta \sim \Gamma(\alpha, \beta)$, the likelihood $X|\theta \sim \Gamma(\nu, \theta)$, and a set of i.i.d. observatios $\boldsymbol{X}_{(n)} = \{X_1, \ldots, X_n\}$, the posterior is given by

$$
\begin{aligned}
\pi(\theta|\boldsymbol{X}_{(n)}) &\propto \prod_{i=1}^n p(X_i|\theta)\pi(\theta) \\
&\propto \prod_{i=1}^n \theta^\nu X_i^{\nu-1} \exp\{-\theta X_i\} \cdot \theta^{\alpha-1} \exp\{-\beta\theta\} \\
&\propto \theta^{n\nu+\alpha-1} \exp\{-\theta(\sum_{i=1}^n X_i + \beta)\}
\end{aligned}
$$

It is recognized that the posterior distribution of $\theta|\boldsymbol{X}_{(n)} \sim \Gamma(n\nu + \alpha, \sum_{i=1}^n X_i + \beta)$.

For prior $\pi_1, \pi_2 : \theta \sim \Gamma(\alpha_1, \beta), \theta \sim \Gamma(\alpha_2, \beta)$, the Kullback-Leibler divergence of the corresponding posteriors is given by

$$
\begin{aligned}
\mathrm{KL}\left(\pi_1^n(\theta|\boldsymbol{X})\|\pi_2^n(\theta|\boldsymbol{X})\right) &= \int_0^\infty \pi_1^n(\theta|\boldsymbol{X})\ln\frac{\pi_1^n(\theta|\boldsymbol{X})}{\pi_2^n(\theta|\boldsymbol{X})}d\theta \\
&= \int_0^\infty \pi_1^n(\theta|\boldsymbol{X})\ln\frac{\Gamma(n\nu+\alpha_0,\sum_{i=1}^n X_i+\beta)}{\Gamma(n\nu+\alpha_1,\sum_{i=1}^n X_i+\beta)}d\theta \\
&= \int_0^\infty \pi_1^n(\theta|\boldsymbol{X})\ln\left((\sum_{i=1}^n X_i+\beta)^{\alpha_0-\alpha_1}\frac{\Gamma(n\nu+\alpha_1)}{\Gamma(n\nu+\alpha_0)}\theta^{\alpha_0-\alpha_1}\right)d\theta \\
&= (\alpha_0-\alpha_1)\ln(\sum_{i=1}^n X_i+\beta)+\ln\frac{\Gamma(n\nu+\alpha_1)}{\Gamma(n\nu+\alpha_0)}+(\alpha_0-\alpha_1)\mathbb{E}_{\pi_1^n}[\ln\theta] \\
&= (\alpha_0-\alpha_1)\ln(\sum_{i=1}^n X_i+\beta)+\ln\frac{\Gamma(n\nu+\alpha_1)}{\Gamma(n\nu+\alpha_0)}+(\alpha_0-\alpha_1)(\psi(n\nu+\alpha_0)-\ln(\sum_{i=1}^n X_i+\beta)) \\
&= \ln\frac{\Gamma(n\nu+\alpha_1)}{\Gamma(n\nu+\alpha_0)}+(\alpha_0-\alpha_1)\psi(n\nu+\alpha_0)
\end{aligned}
$$

where $\psi(x)$ is the digamma function.

The strong law of large numbers yields $\sum_i^n Y_i \to n\mathbb{E}[Y]$a.s. for $n\to\infty$. Also for digamma function $\psi(x)$, we have $\psi(x)\sim\ln x-\frac{1}{2x}$ for $n\to\infty$. Thus we have

$$
\begin{aligned}
&\lim_{n\to\infty}\mathrm{KL}\left(\pi_1^n(\theta|\boldsymbol{X})\|\pi_2^n(\theta|\boldsymbol{X})\right) \\
=& \lim_{n\to\infty}(\alpha_0-\alpha_1)\ln\frac{\Gamma(n\nu+\alpha_1)}{\Gamma(n\nu+\alpha_0)}+(\alpha_0-\alpha_1)\psi(n\nu+\alpha_0) \\
=& \lim_{n\to\infty}\ln\frac{\Gamma(n\nu+\alpha_0)(n\nu+\alpha_0)^{\alpha_1-\alpha_0}}{\Gamma(n\nu+\alpha_0)}+(\alpha_0-\alpha_1)\left\{\ln(n\nu+\alpha_0)-\frac{1}{2(n\nu+\alpha_0)}\right\} \\
=& \lim_{n\to\infty}(\alpha_1-\alpha_0)\ln(n\nu+\alpha_0)+(\alpha_0-\alpha_1)\ln(n\nu+\alpha_0)-\frac{\alpha_0-\alpha_1}{2n\nu+2\alpha_0} \\
=& \lim_{n\to\infty}-\frac{\alpha_0-\alpha_1}{2n\nu+2\alpha_0} \\
=& 0
\end{aligned}
$$

By applying Pinsker's inequality, we have

$$
\sqrt{\frac{1}{2}\mathrm{KL}\left(\pi_1^n(\theta|\boldsymbol{X})\|\pi_2^n(\theta|\boldsymbol{X})\right)} \geq \sup_A\left|\underset{n}{\overset{\pi_1}{\Pr}}(A|\boldsymbol{X})-\underset{n}{\overset{\pi_1}{\Pr}}(A|\boldsymbol{X})\right| \geq 0
$$

Take the limitation on $n\to\infty$, hence we have

$$
0\leq\lim_{n\to\infty}\sup_A\left|\underset{n}{\overset{\pi_1}{\Pr}}(A|\boldsymbol{X})-\underset{n}{\overset{\pi_1}{\Pr}}(A|\boldsymbol{X})\right| \leq \lim_{n\to\infty}\sqrt{\frac{1}{2}\mathrm{KL}\left(\pi_1^n(\theta|\boldsymbol{X})\|\pi_2^n(\theta|\boldsymbol{X})\right)}=0
$$

Then

$$
\lim_{n\to\infty}\sup_A\left|\underset{n}{\overset{\pi_1}{\Pr}}(A|\boldsymbol{X})-\underset{n}{\overset{\pi_1}{\Pr}}(A|\boldsymbol{X})\right|=0
$$

□

6 Problem 6.1

(a) Let $\boldsymbol{x}_i=(1,x_i,z_i)^\top$, $\boldsymbol{X}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^\top$ and $\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2)^\top$. The Bayes model $\{\mathcal{P},\pi\}$ in matrix form is given by

$$
\{\{\mathcal{N}_n(\boldsymbol{X}^\top\boldsymbol{\beta},\boldsymbol{I}_n):\boldsymbol{\beta}\in\mathbb{R}^3\},\mathcal{N}_3(\boldsymbol{\gamma},\boldsymbol{\Gamma})\}
$$

where

$$\boldsymbol{\gamma} = \begin{pmatrix}1\\1\\1\end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma} = \begin{pmatrix}1 & 0.5 & 0\\0.5 & 1 & 0\\0 & 0 & 1\end{pmatrix}$$

Let $\boldsymbol{y} = (y_1, \dots, y_n)^\top$, then we have the model equation as

$$\boldsymbol{y} = \boldsymbol{X}^\top \boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{I}_n)$ and $\boldsymbol{\beta} \sim \mathcal{N}_3(\boldsymbol{\gamma}, \boldsymbol{\Gamma})$. Additional the design matrix $\boldsymbol{X}$ is orthogonal, such that $\boldsymbol{X}^\top \boldsymbol{X} = n\boldsymbol{I}_3$.

(b) The posterior distribution of $\boldsymbol{\beta}$ is given as follows:

$$\boldsymbol{\theta}|\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\gamma}_{\boldsymbol{\beta}|\boldsymbol{y}}, \boldsymbol{\Gamma}_{\boldsymbol{\beta}|\boldsymbol{y}})$$

where

$$\begin{aligned}
\boldsymbol{\Gamma}_{\boldsymbol{\beta}|\boldsymbol{y}} &= (\boldsymbol{\Gamma}^{-1} + \boldsymbol{X}^\top \boldsymbol{I}_n^{-1} \boldsymbol{X})^{-1} \\
&= \left(\begin{pmatrix} \frac{4}{3} & -\frac{2}{3} & 0 \\ -\frac{2}{3} & \frac{4}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} + n\boldsymbol{I}_3 \right)^{-1} \\
&= \begin{pmatrix} \frac{4+7n+3n^2}{4+12n+11n^2+3n^3} & \frac{2+2n}{4+12n+11n^2+3n^3} & 0 \\ \frac{2+2n}{4+12n+11n^2+3n^3} & \frac{4+7n+3n^2}{4+12n+11n^2+3n^3} & 0 \\ 0 & 0 & \frac{4+7n+3n^2}{4+12n+11n^2+3n^3} \end{pmatrix} \\
\boldsymbol{\gamma}_{\boldsymbol{\beta}|\boldsymbol{y}} &= \boldsymbol{\Gamma}_{\boldsymbol{\beta}|\boldsymbol{y}} (\boldsymbol{X}^\top \boldsymbol{I}_3^{-1} \boldsymbol{y} + \boldsymbol{\Gamma}^{-1} \gamma) \\
&= \boldsymbol{\Gamma}_{\boldsymbol{\beta}|\boldsymbol{y}} \left(\boldsymbol{X}^\top \boldsymbol{y} + \left(\frac{2}{3}, \frac{2}{3}, 1 \right)^\top \right)
\end{aligned}$$

7 Problem 6.2

(a) Let $\boldsymbol{y} = (y_1, \dots, y_n)^\top$ and $\boldsymbol{x} = (x_1, \dots, x_n)^\top$. For $\theta = (\alpha, \beta, \sigma^2)$, we have

$$\begin{aligned}
\boldsymbol{y}|\theta &\sim \mathcal{N}_n \left(\boldsymbol{1}_n \alpha + \boldsymbol{x}\beta, \sigma^2 \boldsymbol{I}_n \right) \\
p(\boldsymbol{y}|\theta) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{1}{2} \left(\frac{y_i - \alpha - x_i\beta}{\sigma} \right)^2 \right\} \\
\ln p(\boldsymbol{y}|\theta)) &= \sum_{i=1}^n - \ln\sigma - \frac{1}{2} \left(\frac{y_i - \alpha - x_i\beta}{\sigma} \right)^2 + \text{const.} \\
&= -\frac{n}{2} \ln\sigma^2 - \frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \alpha - x_i\beta}{\sigma} \right)^2 + \text{const.}
\end{aligned}$$

The score function is then

$$\begin{aligned}
V(\theta|\boldsymbol{y}) &= \frac{\partial \ln p(\boldsymbol{y}|\theta)}{\partial\theta} \\
&= \begin{pmatrix} \frac{\partial \ln p(\boldsymbol{y}|\theta)}{\partial\alpha} \\ \frac{\partial \ln p(\boldsymbol{y}|\theta)}{\partial\beta} \\ \frac{\partial \ln p(\boldsymbol{y}|\theta)}{\partial\sigma^2} \end{pmatrix} \\
&= \begin{pmatrix} -\frac{n\alpha + \sum_{i=1}^n (x_i\beta - y_i)}{\sigma^2} \\ -\frac{\beta \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i(\alpha - y_i)}{\sigma^2} \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \alpha - x_i\beta)^2 \end{pmatrix}
\end{aligned}$$

Further the Jacobian of $\ln p(\theta|\boldsymbol{y})$ is

$$J(\theta|\boldsymbol{y}) = \begin{pmatrix} \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial^2 \alpha} & \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial \alpha \partial \beta} & \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial \alpha \partial \sigma^2} \\ \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial \beta \partial \alpha} & \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial^2 \beta} & \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial \sigma^2 \partial \alpha} & \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial^2 \sigma^2} \end{pmatrix}$$
$$= \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{\sum_{i=1}^n x_i}{\sigma^2} & \frac{n\alpha + \sum_{i=1}^n (x_i\beta - y_i)}{\sigma^4} \\ -\frac{\sum_{i=1}^n x_i}{\sigma^2} & -\frac{\sum_{i=1}^n x_i^2}{\sigma^2} & \frac{\beta \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i(\alpha - y_i)}{\sigma^4} \\ \frac{n\alpha + \sum_{i=1}^n (x_i\beta - y_i)}{\sigma^4} & \frac{\beta \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i(\alpha - y_i)}{\sigma^4} & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (y_i - \alpha - x_i\beta)^2 \end{pmatrix}$$

We have $\mathbb{E}_{p(y|\boldsymbol{\theta})} y = \alpha + x\beta$. Then the Fisher information matrix following $I(\theta) = -\mathbb{E}_{p(y|\boldsymbol{\theta})} J(\theta|\boldsymbol{y})$ is then

$$I(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{\sum_{i=1}^n x_i}{\sigma^2} & 0 \\ \frac{\sum_{i=1}^n x_i}{\sigma^2} & \frac{\sum_{i=1}^n x_i^2}{\sigma^2} & 0 \\ 0 & 0 & -\frac{n}{2\sigma^4} \end{pmatrix}$$

The Jeffreys prior is given by

$$\pi_{\text{Jeff}}(\alpha, \beta, \sigma^2) \propto \sqrt{\det(I(\theta))} \propto \sqrt{\frac{1}{\sigma^8}} = \frac{1}{\sigma^4}$$

Further, we assume the independence of (α, β) and σ^2 as $\pi^{\text{ind}}(\alpha, \beta, \sigma^2) = \pi^{\text{ind}}(\alpha, \beta)\pi^{\text{ind}}(\sigma^2)$. Then we have

$$I(\alpha, \beta) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{\sum_{i=1}^n x_i}{\sigma^2} \\ \frac{\sum_{i=1}^n x_i}{\sigma^2} & \frac{\sum_{i=1}^n x_i^2}{\sigma^2} \end{pmatrix}$$

and $\pi_{\text{Jeff}}^{\text{ind}}(\alpha, \beta) \propto \sqrt{\det(I(\alpha, \beta))} \propto 1$. Also we have $I(\sigma^2) = \frac{n}{2\sigma^4}$ and $\pi_{\text{Jeff}}^{\text{ind}}(\sigma^2) \propto \sqrt{\det(I(\sigma^2))} \propto \frac{1}{\sigma^2}$. Then we have

$$\pi_{\text{Jeff}}^{\text{ind}}(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}$$

(b) (i). Using $\pi(\alpha, \beta) \propto 1$ and $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ under the assumption of the independence of (α, β) and σ^2, we have

$$\begin{aligned} \pi(\alpha, \beta|\boldsymbol{y}, \sigma^2) &\propto \pi(\alpha, \beta) p(\boldsymbol{y}|\alpha, \beta, \sigma^2) \\ &\propto \exp\left\{ -\frac{1}{2}\sum_{i=1}^n \left(\frac{y_i - \alpha - x_i\beta}{\sigma^2} \right)^2 \right\} \\ &\propto \exp\left\{ -\frac{1}{2}\left((\alpha, \beta) - (\hat{\alpha}, \hat{\beta}) \right) \Sigma^{-1} \left((\alpha, \beta) - (\hat{\alpha}, \hat{\beta}) \right)^\top \right\} \end{aligned}$$

Hence, we have

$$\begin{aligned} &\alpha, \beta|\boldsymbol{y}, \sigma^2 \sim \mathcal{N}_2\left((\hat{\alpha}_{\text{MLE}}, \hat{\beta}_{\text{MLE}})^\top, \Sigma^2 \right) \\ &\text{where } \hat{\beta}_{\text{MLE}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \ \hat{\alpha}_{\text{MLE}} = \bar{y} - \hat{\beta}\bar{x}, \\ &\Sigma^{-1} = \sigma^2 \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} \\ &\text{and } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \ \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \end{aligned}$$

(ii). For $\pi(\beta|\boldsymbol{y},\sigma^2)$, we have

$$\pi(\beta|\boldsymbol{y},\sigma^2) = \int \pi(\alpha,\beta|\boldsymbol{y},\sigma^2)d\alpha$$

Then we have

$$\beta|\boldsymbol{y},\sigma^2 \sim \mathcal{N}(\hat{\beta},\frac{\sigma^2}{n})$$

(iii). For $\pi(\beta|\boldsymbol{y},\alpha,\sigma^2)$, we have

$$\begin{aligned}
\pi(\beta|\boldsymbol{y},\alpha,\sigma^2) &\propto \pi(\beta)p(\boldsymbol{y}|\alpha,\beta,\sigma^2) \\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\alpha-x_i\beta}{\sigma}\right)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\frac{\beta^2\sum_{i=1}^{n}x_i^2+2\beta\sum_{i=1}^{n}x_i(\alpha-y_i)}{\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\frac{(\beta-\frac{\sum_{i=1}^{n}x_i(y_i-\alpha)}{\sum_{i=1}^{n}x_i^2})^2}{\frac{\sigma^2}{\sum_{i=1}^{n}x_i^2}}\right\}
\end{aligned}$$

Then we have

$$\beta|\boldsymbol{y},\alpha,\sigma^2 \sim \mathcal{N}(\frac{\sum_{i=1}^{n}x_i(y_i-\alpha)}{\sum_{i=1}^{n}x_i^2},\frac{\sigma^2}{\sum_{i=1}^{n}x_i^2})$$

(iv). For $\pi(\alpha,\beta,\sigma^2|\boldsymbol{y})$, we have

$$\pi(\alpha,\beta,\sigma^2|\boldsymbol{y}) = p(\boldsymbol{y}|\alpha,\beta,\sigma^2)\pi(\alpha,\beta,\sigma^2)$$

With Theorem 6.7, we have

$$\alpha,\beta,\sigma^2|\boldsymbol{y} \sim \mathcal{NIG}(n-2,\sum_{i=1}^{n}\left(y_i-\hat{\alpha}_{\text{MLE}}-x_i\hat{\beta}_{\text{MLE}}\right)^2,(\hat{\alpha}_{\text{MLE}},\hat{\beta}_{\text{MLE}})^\top,\Sigma^{-1})$$

With the properties of normal inverse gamma distribution, we have

$$\sigma^2|\boldsymbol{y} \sim \mathcal{IG}\left(\frac{n}{2}-1,\frac{1}{2}\sum_{i=1}^{n}\left(y_i-\hat{\alpha}_{\text{MLE}}-x_i\hat{\beta}_{\text{MLE}}\right)^2\right)$$

(v). Similarly for $\pi(\alpha,\beta|\boldsymbol{y})$, we have

$$\alpha,\beta|\boldsymbol{y} \sim \boldsymbol{t}_2(n-2,(\hat{\alpha}_{\text{MLE}},\hat{\beta}_{\text{MLE}})^\top,\frac{1}{n-2}\sum_{i=1}^{n}\left(y_i-\hat{\alpha}_{\text{MLE}}-x_i\hat{\beta}_{\text{MLE}}\right)^2\Sigma^{-1})$$

where $\boldsymbol{t}_2$ is a bivariate t-distribution.

8 Problem 6.3

(a) With Jeffreys prior $\pi(\sigma^2)\propto\sigma^{-2}$, we have

$$\begin{aligned}
\pi(\sigma^2|\boldsymbol{y},\alpha,\beta) &\propto \pi(\sigma^2)p(\boldsymbol{y}|\alpha,\beta,\sigma^2) \\
&\propto \frac{1}{\sigma^n}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i-\alpha-x_i\beta}{\sigma}\right)^2\right\}\frac{1}{\sigma^2} \\
&\propto (\sigma^2)^{-\frac{n}{2}-1}\exp\left(-\frac{1}{\sigma^2}\frac{\sum_{i=1}^{n}(y_i-\alpha-x_i\beta)^2}{2}\right)
\end{aligned}$$

Recognizing that $\sigma^2|\boldsymbol{y},\alpha,\beta$ is an inverse gamma random variable, we have $\sigma^2|\boldsymbol{y},\alpha,\beta \sim \mathcal{IG}\left(\frac{n}{2},\frac{1}{2}\sum_{i=1}^{n}(y_i-\alpha-x_i\beta)^2\right)$.

(b) Both conditional and marginal posterior distribution for σ^2 follows inverse gamma distribution. The scale parameter of the conditional posterior distribution $\pi(\sigma^2|\boldsymbol{y}, \alpha, \beta)$ depends on the given α and β, while its counterpart is computed based on the maximum likelihood estimations of α and β.