Stats 370 Homework 3 Solution

Problem 1

Part (a): By definition,

$$Y = \mu + \beta X + \epsilon$$

we have

$$\mathbb{P}(Y) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(y-\mu-\beta X)^2}{2\tau}}$$

Then,

$$\log \mathbb{P}(Y) = C - \frac{1}{2}\log\tau - \frac{(y-\mu-\beta X)^2}{2\tau}$$

Considering the additivity of Fisher Information between independent trials, the corresponding Informaiton Matrix can be computed as

$$\mathbb{I}_{(\mu,\beta,\tau)} = \sum_{i=1}^{n} \mathbb{E} \begin{pmatrix} \frac{1}{\tau} & \frac{x_i}{\tau} & \frac{y_i-\mu-\beta x_i}{\tau^2} \\ \frac{x_i}{\tau} & \frac{x_i^2}{\tau} & \frac{x_i(y_i-\mu-\beta x_i)}{\tau^2} \\ \frac{y_i-\mu-\beta x_i}{\tau^2} & \frac{x_i(y_i-\mu-\beta x_i)}{\tau^2} & -\frac{1}{2\tau^2} + \frac{(y-\mu-\beta x_i)^2}{\tau^3} \end{pmatrix}$$

Note that $\mathbb{E}[y_i - \mu - \beta x_i] = 0$ and $\mathbb{E}[(y_i - \mu - \beta x_i)^2] = \tau$, we have

$$\mathbb{I}_{(\mu,\beta,\tau)} = \sum_{i=1}^{n} \begin{pmatrix} \frac{1}{\tau} & \frac{x_i}{\tau} & 0 \\ \frac{x_i}{\tau} & \frac{x_i^2}{\tau} & 0 \\ 0 & 0 & \frac{1}{2\tau^2} \end{pmatrix}$$

Thus, the Jeffreys' prior can be computed as

$$\mathcal{J}(\mu,\beta,\tau) \propto \sqrt{\det \mathbb{I}_{\mu,\beta,\tau}} = \frac{[\frac{1}{2}n^2\sum_{i=1}^{n} x_i^2 - \frac{1}{2}n(\sum_{i=1}^{n} x_i)^2]^{1/2}}{\tau^2} \propto \frac{1}{\tau^2}.$$

Part (b): Let X be the $n \times 2$ matrix with first column consisting of 1's and the second column consisting of the x_i's. Let Y be the $n \times 1$ matrix of the y_i's. Then $(X^T X)^{-1} X^T Y$ is the least squares solution for $(\hat{\mu}, \hat{\beta})^T$. For easier notation let α be the vector $(\mu, \beta)^T$ and let $\hat{\alpha}$ be the OLS solution $(X^T X)^{-1} X^T Y$.

The joint posterior distribution is proportional to

$$
\begin{aligned}
\mathbb{P}(\tau|Y) &\propto \frac{1}{\tau^2}\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\tau}}\exp\left(-\frac{(y_i-\mu-x_i\beta)^2}{2\tau}\right)\\
&= \left(\frac{1}{\tau}\right)^{n/2+2}\exp\left(-\frac{1}{2\tau}\|y-X\alpha\|^2\right)\\
&= \frac{1}{\tau^{n/2+2}}\exp\left(-\frac{1}{2\tau}\|y-X\hat{\alpha}+X\hat{\alpha}-X\alpha\|^2\right)\\
&= \frac{1}{\tau^{n/2+2}}\exp\left(-\frac{1}{2\tau}\|y-X\hat{\alpha}\|^2\right)\exp\left(-\frac{1}{2\tau}\|X\hat{\alpha}-X\alpha\|^2\right)\\
&= \frac{1}{\tau^{n/2+2}}\exp\left(-\frac{1}{2\tau}\|y-X\hat{\alpha}\|^2\right)\exp\left(\frac{1}{2\tau}(\alpha-\hat{\alpha})X^TX(\alpha-\hat{\alpha})\right)|X^TX/\tau|^{1/2}|X^TX/\tau|^{-1/2}.
\end{aligned}
$$

Integrating the last line over α gives

$$
\frac{1}{\tau^{n/2+2}}|X^TX/\tau|^{-1/2}\exp\left(-\frac{1}{2\tau}\|y-X\hat{\alpha}\|^2\right) = \frac{1}{\tau^{n/2+1}}\exp\left(-\frac{1}{2\tau}\|y-X\hat{\alpha}\|^2\right).
$$

Therefore, τ is distributed according to an inverse gamma distribution with shape $\alpha = n/2$ and scale $\beta = \|y-X\hat{\alpha}\|^2/2$.

Part (c): Next we want to integrate out τ to get the marginal posterior distribution for $\alpha = (\mu,\beta)$. Let $a = n/2+1$ and $b = \frac{1}{2}\|y-X\alpha\|^2$.

$$
\begin{aligned}
\mathbb{P}(\mu,\beta|Y) &= \int_0^\infty \tau^{-n/2-2}\exp\left(-\frac{1}{2\tau}\|y-X\alpha\|^2\right)d\tau\\
&= \left(\int_0^\infty \frac{x^a}{\Gamma(a)}\tau^{-a-1}\exp(-\beta/\tau)\,d\tau\right)\frac{\Gamma(a)}{b^a}\\
&= \frac{\Gamma(n/2+1)}{\left(\frac{1}{2}\|y-X\alpha\|^2\right)^{n/2+1}}\\
&\propto \frac{\Gamma(\frac{n+2}{2})}{(\|y-X\hat{\alpha}\|^2+\|X\alpha-X\hat{\alpha}\|^2)^{(n+2)/2}}\\
&\propto \left(\frac{1}{1+(\alpha-\hat{\alpha})^TX^TX(\alpha-\hat{\alpha})/\|y-X\hat{\alpha}\|^2}\right)^{(n+2)/2}\\
&\propto \left(1+(\alpha-\hat{\alpha})^T\frac{X^TX}{\|y-X\hat{\alpha}\|^2/n}(\alpha-\hat{\alpha})\frac{1}{n}\right)^{-(n+2)/2}.
\end{aligned}
$$

This is the form of a two-dimensional t distribution with location $\hat{\alpha}$, scale $(X^TX)^{-1}\|y-X\hat{\alpha}\|^2/n$, and n degrees of freedom.

Problem 2

Given λ, the censoring probability for the ith observation is $\exp(-\lambda c_i)$. Hence the likelihood can be written

$$p(z_i, \delta_i|\lambda) = [\lambda \exp(-\lambda z_i)]^{\delta_i} [\exp(-\lambda c_i)]^{1-\delta_i} = \lambda^{\delta_i} \exp(-\lambda z_i),$$

since $z_i = c_i$ when $\delta_i = 0$.

Part (a): Using the above expression for the likelihood, the posterior is proportional to

$$\lambda^{\alpha + \sum_i \delta_i - 1} \exp(-(\beta + \sum_i z_i)\lambda)$$

which is the form of a Gamma distribution with shape $\alpha + \sum_i \delta_i$ and rate $\beta + \sum_i z_i$.

Part (b): The second derivative of the log-likelihood is

$$\frac{\partial^2}{\partial \lambda^2} \sum_{i=1}^n \log p(z_i|\lambda) = -\frac{\sum_i \delta_i}{\lambda^2},$$

so the Fisher information is

$$\frac{\mathbb{E} \sum_{i=1}^n \delta_i}{\lambda^2} = \lambda^{-2} \sum_{i=1}^n (1 - \exp(-\lambda c_i)).$$

Note this is the usual (uncensored) information for the exponential parameter but with n replaced by the expected number of uncensored observations. The Jeffreys' prior is thus

$$\pi(\lambda) \propto \lambda^{-1} \sqrt{\sum_{i=1}^n (1 - \exp(-\lambda c_i))}.$$

Under this noninformative prior, the posterior has the following (unnormalized) density:

$$g(\lambda|z_1, \ldots, z_n) = \lambda^{\sum_i \delta_i - 1} \exp(-\lambda \sum_i z_i) \sqrt{\sum_{i=1}^n (1 - \exp(-\lambda c_i))}.$$

Note: for large c_i this is approximately Gamma($\sum_i \delta_i, \sum_i z_i$) [cf. part (a)].

To check whether the posterior is proper, we check that $g(\lambda|z_1, \ldots, z_n)$ can be normalized to a probability. Note that as $\lambda \to \infty$ the square root term tends to a constant, so by the finiteness of the gamma integral we have

$$\int_a^\infty g(\lambda|z_1, \ldots, z_n) d\lambda < \infty.$$

for any $a > 0$. As $\lambda \to 0$ it can be checked that the square root term is

$$\sqrt{(c_1 + \cdots + c_n)\lambda} + o(\sqrt{\lambda}).$$

Since

$$\int_0^a \lambda^{\sum_i \delta_i - 1} \exp(-\lambda \sum_i z_i) \sqrt{C\lambda} d\lambda = \sqrt{C} \int_0^a \lambda^{(\sum_i \delta_i + 1/2) - 1} \exp(-\lambda \sum_i z_i) d\lambda < \infty$$

(gamma integral again), we know the posterior is proper.

Part (c): Note that we can just find a posterior interval for λ and then transform:

$$P(\lambda \in (a, b)|z) = 0.95 \Rightarrow P(\exp(-\lambda t_0) \in (\exp(-bt_0), \exp(-at_0))|z) = 0.95.$$

In part (a), we could use R or similar to compute the .025 and .975 quantiles of the gamma posterior, which would take the the places of a and b as above. For part (b), we could numerically integrate $g(\lambda|z_1, \ldots, z_n)$ over a suitably large range to find the normalizing constant, after which the posterior cdf can be computed and the relevant quantiles computed.