Joseph High
Hopkins ID: 9E1FDC

EN.553.732 Homework 2

Problem 1.) Suppose $x_1, \ldots, x_n$ is a random sample from an exponential distribution with mean $1/\theta$.

a.) Derive the Jeffrey's prior for θ.

Solution: We are given that $x|\theta \sim Exp(\theta)$ with mean $1/\theta$. then,

$$p(x|\theta) = \theta e^{-\theta x}, \quad \theta \geq 0$$

Moreover, Jeffrey's prior is such that $\pi(\theta) \propto \sqrt{I(\theta)}$

where $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}[\log p(x|\theta)] \,\middle|\, \theta\right]$.

then, $\log p(x|\theta) = \log\theta - \theta x$; $\frac{\partial}{\partial\theta}[\log p(x|\theta)] = \frac{1}{\theta} - x$

$$\Rightarrow \frac{\partial^2}{\partial\theta^2}[\log p(x|\theta)] = \frac{-1}{\theta^2}$$

$$\Rightarrow I(\theta) = -E[-1/\theta^2 | \theta] = 1/\theta^2$$

$$\Rightarrow \text{Jeffrey's prior } \pi(\theta) \propto \sqrt{1/\theta^2} = 1/\theta, \quad \theta \geq 0$$

(b.) Derive the posterior distribution of θ using Jeffrey's prior.

Solution: $p(\theta|x_1, \ldots, x_n) \propto \pi(\theta) p(x_1, \ldots, x_n|\theta)$

where $\pi(\theta) = 1/\theta$ and $p(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta) =$

$$= \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

$\Longrightarrow$

$$\Rightarrow p(\theta|x_1,\dots,x_n) \propto \frac{1}{\theta}\left(\theta^n e^{-\theta\sum_{i=1}^n x_i}\right)$$

$$= \theta^{n-1} e^{-\theta\sum_{i=1}^n x_i}$$

and $\theta^{n-1} e^{-\theta\sum_{i=1}^n x_i} \longrightarrow \text{Gamma}(n, \sum_{i=1}^n x_i)$

i.e. $\theta|x_1,\dots,x_n \sim \text{Gamma}(n, \sum_{i=1}^n x_i)$

$$\Rightarrow p(\theta|x_1,\dots,x_n) = \frac{(\sum x_i)^n}{\Gamma(n)} \cdot \theta^{n-1} e^{-\theta\sum_{i=1}^n x_i} \quad \text{(By def. of the Gamma dist.)}$$

$$, \theta \geq 0$$

(c) Derive the predictive distribution of a future observation z.

Solution: Clearly $z \sim Exp(\theta)$ since z is from the same population as the $x_1,\dots,x_n$. Now, using $p(\theta|x_1,\dots,x_n)$ from part (b) and assuming the $x_1,\dots,x_n$ iid, the predictive distribution is

$$p(z|x_1,\dots,x_n) = \int_0^\infty p(z,\theta|x_1,\dots,x_n)\,d\theta$$

$$= \int_0^\infty p(z|\theta,x_1,\dots,x_n)\,p(\theta|x_1,\dots,x_n)\,d\theta \quad \text{(By Bayes' Rule)}$$

$$= \int_0^\infty p(z|\theta)\,p(\theta|x_1,\dots,x_n)\,d\theta$$ Assuming independence of z and the $x_1,\dots,x_n$ (Lecture 4)

$$= \int_0^\infty \theta e^{-\theta z} \cdot \frac{(\sum_{i=1}^n x_i)^n}{\Gamma(n)} \theta^{n-1} e^{-\theta\sum x_i}\,d\theta \quad \text{(from part (b))}$$

$$= \frac{(\sum x_i)^n}{\Gamma(n)} \int_0^\infty \theta^n e^{-\theta(z+\sum x_i)}\,d\theta ;$$

using exact density here, not proportional density

$$\Rightarrow \text{let } u = \theta(z + \sum_{i=1}^n x_i) \Rightarrow du = (z + \sum x_i)\,d\theta$$

$$\Rightarrow d\theta = \frac{du}{z+\sum x_i} \quad \text{and} \quad \theta = \frac{u}{z+\sum x_i} \Longrightarrow$$

Joseph High

Hopkins ID: 9E1FDC

$\Rightarrow$ $p(z \mid x_1, \dots, x_n)$ is then

$$= \frac{(\sum x_i)^n}{\Gamma(n)} \int_0^\infty \left(\frac{u}{z + \sum x_i}\right)^n e^{-u} \cdot \frac{du}{z + \sum x_i} =$$

$$= \frac{(\sum x_i)^n}{\Gamma(n)(z + \sum x_i)^{n+1}} \int_0^\infty u^n e^{-u}\, du = \frac{(\sum x_i)^n \Gamma(n+1)}{\Gamma(n)\,(z + \sum x_i)^{n+1}}$$

$$= \frac{n!\,(\sum x_i)^n}{(n-1)!\,(z + \sum x_i)^{n+1}} = \frac{n(\sum x_i)^n}{(z + \sum x_i)^{n+1}}$$

Thus, $p(z \mid x_1, \dots, x_n) = \dfrac{n(\sum x_i)^n}{(z + \sum x_i)^{n+1}}$ //

(d.) Derive a 95% Credible interval for z.

Solution: We are interested an c_1, c_2 such that

$$P(c_1 \le z \le c_2) = 0.95 \quad \text{i.e.}$$

$$= \int_{c_1}^{c_2} P(z \mid x_1, \dots, x_n)\, dz = 0.95$$, by part (c) we have

$$= \int_{c_1}^{c_2} \frac{n(\sum x_i)^n}{(z + \sum x_i)^{n+1}}\, dz = n(\sum x_i)^n \underbrace{\int_{c_1}^{c_2} \frac{dz}{(z + \sum x_i)^{n+1}}}_{(*)}$$

let $u = z + \sum x_i$ in $(*)$: $\displaystyle\int_{c_1 + \sum x_i}^{c_2 + \sum x_i} \frac{du}{u^{n+1}} = -\frac{1}{nu^n}\Bigg|_{c_1 + \sum x_i}^{c_2 + \sum x_i}$

$du = dz$

$$= -\frac{1}{n}\left[\frac{1}{(c_2+\sum x_i)^n} - \frac{1}{(c_1+\sum x_i)^n}\right] = \int_{c_1}^{c_2} \frac{dz}{(z+\sum x_i)^{n+1}}$$

$$\Rightarrow \quad 0.95 = n\left(\sum x_i\right)^n \int_{c_1}^{c_2} \frac{dz}{(z+\sum x_i)^{n+1}} =$$

$$= -\left(\sum x_i\right)^n \left[\frac{1}{(c_2+\sum x_i)^n} - \frac{1}{(c_1+\sum x_i)^n}\right] = 0.95$$

Letting $c_1 = 0$:

$$= \frac{-(\sum x_i)^n}{(c_2+\sum x_i)^n} + 1 = 0.95 \Rightarrow$$

$$\Rightarrow \frac{(\sum x_i)^n}{(c_2+\sum x_i)^n} = 0.05 \Rightarrow (c_2+\sum x_i)^n = \frac{(\sum x_i)^n}{0.05}$$

$$\Rightarrow \quad c_2 = \frac{\sum x_i}{(0.05)^{1/n}} - \sum x_i = \frac{\sum x_i\left(1-(0.05)^{1/n}\right)}{(0.05)^{1/n}}$$

Thus, a 95% Credible Interval for $z \mid x_1, \dots, x_n$

is $\left(0, \dfrac{\sum_{i=1}^{n} x_i\left(1-(0.05)^{1/n}\right)}{(0.05)^{1/n}}\right)$

$\square$

Joseph High

Problem 2

(a.) Let $Y \sim \text{binomial}(n, \theta)$. Obtain Jeffrey's prior distribution $p_J(\theta)$ for this model.

Solution: Since $Y \sim \text{binomial}(n, \theta)$, then letting $p(y|\theta)$ be the sampling model distribution,

$$p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad \theta \in [0,1].$$

Moreover, $p_J(\theta) \propto \sqrt{I(\theta)}$.

where $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}[\log p(y|\theta)] \,\Big|\, \theta\right]$

$$\log p(y|\theta) = \log \binom{n}{y}\theta^y(1-\theta)^{n-y} = \log\binom{n}{y} + y\log\theta + (n-y)\log(1-\theta)$$

then, $\frac{\partial}{\partial\theta}[\log p(y|\theta)] = \frac{y}{\theta} - \frac{n-y}{1-\theta}$;

$$\frac{\partial^2}{\partial\theta^2}[\log p(y|\theta)] = \frac{-y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

$$\Rightarrow I(\theta) = -E\left[-\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2} \,\Big|\, \theta\right] = E\left[\frac{y}{\theta^2}\Big|\theta\right] + E\left[\frac{n-y}{(1-\theta)^2}\Big|\theta\right]$$

$$= \frac{1}{\theta^2}E[y|\theta] + \frac{1}{(1-\theta)^2}(n - E[y|\theta])$$

since $y|\theta \sim \text{Bin}(n,\theta) \Rightarrow E[y|\theta] = n\theta$

$$\Rightarrow I(\theta) = \frac{n\theta}{\theta^2} + \frac{(n-n\theta)}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n(1-\theta)}{(1-\theta)^2} =$$

$$= \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n - n\theta + n\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$$

$\Longrightarrow$

Thus, $p_J(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\theta(1-\theta)}}$

(b.) Reparameterize the binomial sampling model with $\phi = \log\left[\frac{\theta}{1-\theta}\right]$, so that $p(y|\phi) = \binom{n}{y} e^{\phi y}(1+e^{\phi})^{-n}$.

Obtain Jeffrey's prior distribution $p_J(\phi)$ for this model.

Solution: We have $p(y|\phi) = \binom{n}{y} e^{\phi y}(1+e^{\phi})^{-n}$

where $\phi = \log\left(\frac{\theta}{1-\theta}\right)$.

$$\log p(y|\phi) = \log\binom{n}{y} + \phi y - n\left(\log(1+e^{\phi})\right)$$

$$\frac{\partial}{\partial\phi}\left[\log p(y|\phi)\right] = y - \frac{n}{1+e^{\phi}}e^{\phi}$$

$$\Rightarrow \frac{\partial^2}{\partial\phi^2}\left[\log p(y|\phi)\right] = \frac{ne^{\phi}}{(1+e^{\phi})^2}e^{\phi} - \frac{ne^{\phi}}{1+e^{\phi}} =$$

$$= \frac{ne^{2\phi} - ne^{\phi}(1+e^{\phi})}{(1+e^{\phi})^2} = \frac{-ne^{\phi}}{(1+e^{\phi})^2}$$

$$\Rightarrow I(\theta) = -E\left[\frac{-ne^{\phi}}{(1+e^{\phi})^2}\Big|\phi\right] = nE\left[\frac{e^{\phi}}{(1+e^{\phi})^2}\Big|\phi\right] =$$

$$= \frac{ne^{\phi}}{(1+e^{\phi})^2} \Longrightarrow p_J(\phi) \propto \sqrt{\frac{ne^{\phi}}{(1+e^{\phi})^2}} = \frac{\sqrt{ne^{\phi}}}{1+e^{\phi}}$$

Joseph High

(c) <u>Solution</u>: From the supposition in part (b) we know that

$$\phi = \log \frac{\theta}{1-\theta} = g(\theta).$$

The change of variables formula: $p_J(\phi) = p_J(\theta) \left| \frac{d\theta}{d\phi} \right|$

From (a) we have that $p_J(\theta) \propto \sqrt{\frac{n\theta}{1-\theta}}$

and, $\log \frac{\theta}{1-\theta} = \phi \Rightarrow \frac{\theta}{1-\theta} = e^{\phi} \Rightarrow \frac{1-\theta}{\theta} = \frac{1}{e^{\phi}}$

$$\Rightarrow \frac{1}{\theta} = \frac{1}{e^{\phi}} + 1 = \frac{1+e^{\phi}}{e^{\phi}} \Rightarrow \boxed{\theta = \frac{e^{\phi}}{1+e^{\phi}}}$$

Then, $\left| \frac{d\theta}{d\phi} \right| = \left| -e^{\phi}(1+e^{\phi})^{-2} e^{\phi} + e^{\phi}(1+e^{\phi})^{-1} \right| =$

$$= \left| \frac{-e^{2\phi}}{(1+e^{\phi})^2} + \frac{e^{\phi}(1+e^{\phi})}{(1+e^{\phi})^2} \right| = \left| \frac{e^{\phi}}{(1+e^{\phi})^2} \right| = \frac{e^{\phi}}{(1+e^{\phi})^2}$$

Then $p_J(\phi) \propto \sqrt{\frac{n}{\theta(1-\theta)}} \cdot \frac{e^{\phi}}{(1+e^{\phi})^2} =$

$$= \sqrt{\frac{n}{\frac{e^{\phi}}{1+e^{\phi}}\left(1-\frac{e^{\phi}}{1+e^{\phi}}\right)}} \cdot \frac{e^{\phi}}{(1+e^{\phi})^2} = \sqrt{\frac{n}{\left(\frac{e^{\phi}}{1+e^{\phi}}\right)^2\left(\frac{1+e^{\phi}}{e^{\phi}}-1\right)}} \cdot \frac{e^{\phi}}{(1+e^{\phi})^2} =$$

$$= \frac{1+e^{\phi}}{e^{\phi}} \sqrt{\frac{n}{\frac{1}{e^{\phi}}}} \cdot \frac{e^{\phi}}{(1+e^{\phi})^2} = \sqrt{ne^{\phi}} \cdot \frac{1}{(1+e^{\phi})}$$

$\Rightarrow p_J(\phi) \propto \frac{\sqrt{ne^{\phi}}}{(1+e^{\phi})}$. This is the same density found in part (b). Hence, consistency under reparameterization. $\square$

Problem 3) Suppose that $f(x|\theta, \sigma)$ is normal with mean θ and standard deviation σ.

(a) Suppose that σ is known, and show that the Jeffrey's prior for θ is given by $p(\theta) = 1$, $\theta \in \mathbb{R}$.

Proof: Since σ is known, without loss of generality let $\sigma = 1$ so that $X|\theta \sim N(\theta, 1)$.

we have, $f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$, $-\infty < x < \infty$, $-\infty < \theta < \infty$

Then,

$$\log f(x|\theta) = \log \frac{1}{\sqrt{2\pi}} - \frac{(x-\theta)^2}{2}$$

$$\Rightarrow \frac{\partial \log f(x|\theta)}{\partial \theta} = x - \theta \quad ; \quad \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} = -1$$

$$\Rightarrow I(\theta) = -E\left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \Big| \theta\right] = -E[-1|\theta] = 1$$

$$\Rightarrow p_J(\theta) = \sqrt{I(\theta)} = \sqrt{1} = 1$$

If instead we let $\sigma = \sigma_c$ for some known σ_c, then $I(\theta) = 1/\sigma_c \propto 1 \Rightarrow p_J(\theta) \propto 1$

(so generality is not lost by choosing a σ) $\square$

(b.) Next, suppose that θ is known, and show that the Jeffrey's prior for σ is given by $p(\sigma) = 1/\sigma$, $\sigma > 0$.

Pf: Without loss of generality let $\theta = 0$ (since θ known) so that $X|\sigma \sim N(0, \sigma^2)$

Then,

$$f(x|\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad \sigma > 0$$

$\Longrightarrow$

Joseph High

$$\Rightarrow \log f(x|\sigma) = \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{x^2}{2\sigma^2}$$

$$\frac{\partial \log f(x|\sigma)}{\partial \sigma} = -\sqrt{2\pi\sigma^2}\left(\tfrac{1}{2}\right)(2\pi\sigma^2)^{-3/2}(4\pi\sigma) + \frac{x^2}{\sigma^3}$$

$$= -\frac{4\pi\sigma}{2\pi\sigma^2} + \frac{x^2}{\sigma^3} = \frac{-2}{\sigma} + \frac{x^2}{\sigma^3}$$

$$\frac{\partial^2 \log f(x|\sigma)}{(\partial\sigma)^2} = \frac{2}{\sigma^2} - \frac{3x^2}{\sigma^4}$$

$$\Rightarrow -E\left[\frac{\partial^2 \log f(x|\sigma)}{(\partial\sigma)^2}\middle|\sigma\right] = -E\left[\frac{2}{\sigma^2} - \frac{3x^2}{\sigma^4}\middle|\sigma\right] =$$

$$= \frac{-2}{\sigma^2} + \frac{3}{\sigma^4}E[x^2|\sigma] = \frac{-2}{\sigma^2} + \frac{3}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}$$

$$\Rightarrow p_J(\sigma) \propto \sqrt{I(\sigma)} = \sqrt{1/\sigma^2} = \frac{1}{\sigma}$$

If we had used some arbitrary $\theta = \mu$, then

$$\frac{\partial^2 \log f(x|\sigma)}{(\partial\sigma)^2} = \frac{2}{\sigma^2} - \frac{3(x-\mu)^2}{\sigma^4}$$

$$\Rightarrow p_J(\sigma) \propto \sqrt{I(\sigma)} = \sqrt{-\frac{2}{\sigma^2} + \frac{3}{\sigma^4}E[(x-\mu)^2|\sigma]}$$

$$= \sqrt{-\frac{2}{\sigma^2} + \frac{3}{\sigma^4}\sigma^2} = \sqrt{\frac{1}{\sigma^2}} = 1/\sigma$$

(Just to illustrate that we indeed do not lose generality by choosing a θ.)

$$\Rightarrow p_J(\sigma) \propto \frac{1}{\sigma}$$

$\square$

(c) Finally, assume that θ and σ are both unknown, and show that the bivariate Jeffrey's prior is given by $p(\theta, \sigma) = 1/\sigma^2$; $\theta \in \mathbb{R}$, $\sigma > 0$.

Pf: We have that $f(x \mid \theta, \sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$; $x, \theta \in \mathbb{R}$, $\sigma > 0$

Let $l(\theta, \sigma) = \log f(x \mid \theta, \sigma)$ (for convenience).

Then, $l(\theta, \sigma) = \log \dfrac{1}{\sigma\sqrt{2\pi}} - \dfrac{(x-\theta)^2}{2\sigma^2}$

and $p_J(\theta, \sigma) \propto \sqrt{I(\theta, \sigma)} \; \underline{=} \; -E\left[|\nabla^2 l(\theta, \sigma)| \mid \theta, \sigma \right]$

where $\nabla^2 l(\theta, \sigma)$ is the Hessian of $l(\theta, \sigma)$ and $|\cdot|$ its determinant.

i.e. $\nabla^2 l(\theta, \sigma) = \begin{bmatrix} \dfrac{\partial^2 l}{(\partial\theta)^2} & \dfrac{\partial^2 l}{\partial\theta\,\partial\sigma} \\ \dfrac{\partial^2 l}{\partial\theta\,\partial\sigma} & \dfrac{\partial^2 l}{(\partial\sigma)^2} \end{bmatrix}$

Then, $\dfrac{\partial l}{\partial \theta} = \dfrac{(x-\theta)}{\sigma^2}$; $\dfrac{\partial^2 l}{(\partial\theta)^2} = \dfrac{-1}{\sigma^2}$;

$\dfrac{\partial^2 l}{\partial\theta\,\partial\sigma} = \dfrac{-2(x-\theta)}{\sigma^3}$; $\dfrac{\partial l}{\partial\sigma} = \dfrac{-\sigma\sqrt{2\pi}}{\sigma^2\sqrt{2\pi}} + \dfrac{(x-\theta)^2}{\sigma^3}$

$= -\dfrac{1}{\sigma} + \dfrac{(x-\theta)^2}{\sigma^3}$

$\Rightarrow \dfrac{\partial^2 l}{(\partial\sigma)^2} = \dfrac{1}{\sigma^2} - \dfrac{3(x-\theta)^2}{\sigma^4}$

$\Rightarrow \quad \nabla^2 l(\theta, \mu) = \begin{bmatrix} -\dfrac{1}{\sigma^2} & \dfrac{-2(x-\theta)}{\sigma^3} \\ \dfrac{-2(x-\theta)}{\sigma^3} & \dfrac{1}{\sigma^2} - \dfrac{3(x-\theta)^2}{\sigma^4} \end{bmatrix} \quad \Longrightarrow$

Joseph High

Now, $|\nabla^2 l(\theta,\sigma)| = -\frac{1}{\sigma^2}\left(\frac{1}{\sigma^2} - \frac{3(x-\theta)^2}{\sigma^4}\right) - \left(\frac{-2(x-\theta)}{\sigma^3}\right)^2$

$$= \frac{-1}{\sigma^4} + \frac{3(x-\theta)^2}{\sigma^6} - \frac{4(x-\theta)^2}{\sigma^6}$$

$$= \frac{-1}{\sigma^4} - \frac{(x-\theta)^2}{\sigma^6}$$

Then,

$$-E\left[|\nabla^2 l(\theta,\sigma)| \,\middle|\, \theta,\sigma\right] = \frac{1}{\sigma^4} + \frac{1}{\sigma^6} E[(x-\theta)^2 \mid \mu,\sigma]$$

$$= \frac{1}{\sigma^4} + \frac{1}{\sigma^6}\sigma^2 = \frac{2}{\sigma^4}$$

Then,

$$P_J(\theta,\sigma) \propto \sqrt{I(\theta,\sigma)} = \sqrt{\frac{2}{\sigma^4}} = \frac{\sqrt{2}}{\sigma^2} \propto \frac{1}{\sigma^2}$$

Problem 4: Solution: Let $\tilde{y} = (y_1, y_2, \ldots, y_n)$

To compute $p(\theta|\tilde{y})$ consider the following relationships

(1) $p(\theta|\tilde{y}) = \int p(\theta,\sigma^2|\tilde{y})\,d\sigma^2 \propto \int p(\tilde{y}|\theta,\sigma^2)\,p(\theta,\sigma^2)\,d\sigma^2$

(2.) $p(\theta,\sigma^2) = p(\theta|\sigma^2)\,p(\sigma^2)$

(3.) $\tilde{y}|\theta,\sigma^2 \sim N(\theta, \Sigma)$, since $\{y_i\}_{i=1}^n \sim$ i.i.d. $N(\theta,\sigma^2)$

(4.) $\frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right)$ and $\theta|\sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$ (from Chapter 5, pg. 74-75, Hoff)

We can then proceed as follows $\Longrightarrow$ (continued... $\Rightarrow$)

From (2) and (4) we have

$$p(\theta,\sigma^2) = p(\theta|\sigma^2)p(\sigma^2) = \left(\frac{1}{\sqrt{2\pi\frac{\sigma^2}{\kappa_0}}} e^{-\frac{(\theta-\theta_0)^2}{2\sigma^2/\kappa_0}}\right)\cdot\left(\frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\nu_0/2}\left(\frac{1}{\sigma^2}\right)^{\frac{\nu_0}{2}-1} e^{-\frac{\nu_0\sigma_0^2}{2\sigma^2}}}{\Gamma(\nu_0/2)}\right)$$

$$= \frac{\sqrt{\frac{\kappa_0}{2\pi}}\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\nu_0/2}}{\Gamma(\nu_0/2)}\left(\frac{1}{\sigma^2}\right)^{1/2}\left(\frac{1}{\sigma^2}\right)^{\frac{\nu_0}{2}-1} e^{-\frac{1}{2\sigma^2}\left[(\theta-\theta_0)^2\kappa_0+\nu_0\sigma_0^2\right]}$$

(5) $$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{\nu_0-3}{2}} e^{-\frac{\kappa_0}{2\sigma^2}\left[(\theta-\theta_0)^2+\frac{\nu_0\sigma_0^2}{\kappa_0}\right]} \propto p(\theta,\sigma^2)$$

By (3) we have

$$p(\tilde{y}|\theta,\sigma^2) = p(y_1,\dots,y_n|\theta,\sigma^2) = \prod_{i=1}^{n}(2\pi\sigma^2)^{-1/2} e^{-\frac{(y_i-\theta)^2}{2\sigma^2}}$$

(6) $$= (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2}$$

By (1), (5), and (6) we have that

$$p(\theta|\tilde{y}) = \int_0^\infty p(\theta,\sigma^2|\tilde{y})\,d\sigma^2 \propto \int_0^\infty p(\tilde{y}|\theta,\sigma^2)\,p(\theta,\sigma^2)\,d\sigma^2$$

$$= \int_0^\infty (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2}\left(\frac{1}{\sigma^2}\right)^{\frac{\nu_0-3}{2}} e^{-\frac{\kappa_0}{2\sigma^2}\left[(\theta-\theta_0)^2+\frac{\nu_0\sigma_0^2}{\kappa_0}\right]}\,d\sigma^2$$

$$\propto \int_0^\infty (\sigma^2)^{-\left(\frac{\nu_0+n-3}{2}\right)} e^{-\frac{\kappa_0}{2\sigma^2}\left[\frac{1}{\kappa_0}\sum(y_i-\theta)^2+(\theta-\theta_0)^2+\frac{\nu_0\sigma_0^2}{\kappa_0}\right]}\,d\sigma^2$$

Using the change of variables presented in Lecture 4 slides and on page 74-75 in Hoff:

Let

$$\kappa_n = \kappa_0 + n,\quad \theta_n = \frac{\kappa_0\theta_0 + n\bar{y}}{\kappa_n},\quad \nu_n = \nu_0 + n$$

and $$SS_n^2 = \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + \sum(y_i-\bar{y})^2 + \frac{\kappa_0 n}{\kappa_n}(\bar{y}-\theta_0)^2\right] \Rightarrow$$

$$\Rightarrow \propto \int_0^\infty (\sigma^2)^{-\left(\frac{\nu_n+1}{2}+1\right)} \exp\left\{-\frac{\kappa_n}{2\sigma^2}\left(\frac{SS_n^2}{\kappa_n}+(\theta-\theta_n)^2\right)\right\} d\sigma^2$$

(from Lecture Slides and Hoff)

$$\left(\text{Let } A = \kappa_n\left[\frac{SS_n^2}{\kappa_n}+(\theta-\theta_n)^2\right] \text{ and } z = \frac{A}{2\sigma^2}\right)$$

$$\Rightarrow \propto A^{-\frac{\nu_n+1}{2}} \int_0^\infty z^{\frac{(\nu_n-1)}{2}} e^{-z}\, dz \qquad \text{(Gamma Integral)}$$

$$\propto \left[1+\frac{1}{\nu_n}\left(\frac{\theta-\theta_n}{SS_n/\sqrt{\kappa_n}}\right)^2\right]^{-\frac{(\nu_n+1)}{2}}$$

Thus, $\theta \mid \tilde{y} \sim t\left(\nu_n, \theta_n, \frac{SS_n^2}{\kappa_n \nu_n}\right)$

Similarly, $p(\sigma^2 \mid \tilde{y}) = \int_{-\infty}^{\infty} p(\theta, \sigma^2 \mid \tilde{y})\, d\theta \propto \int_{-\infty}^{\infty} p(\tilde{y} \mid \theta, \sigma^2) p(\sigma^2)\, d\theta$

$$\propto \int_{-\infty}^{\infty} (\sigma^2)^{-\left(\frac{\nu_n+1}{2}+1\right)} \exp\left\{-\frac{\kappa_n}{2\sigma^2}\left[\frac{SS_n^2}{\kappa_n}+(\theta-\theta_n)^2\right]\right\} d\theta$$

$$= (\sigma^2)^{-\left(\frac{\nu_n}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2} SS_n^2\right\} \cdot \underbrace{\frac{1}{\sqrt{\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{\kappa_n}{2\sigma^2}(\theta-\theta_n)^2\right\} d\theta}_{\propto 1}$$

$$\propto (\sigma^2)^{-\left(\frac{\nu_n}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2} SS_n^2\right\} \longrightarrow \text{Inv-Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n}{2} SS_n^2\right)$$

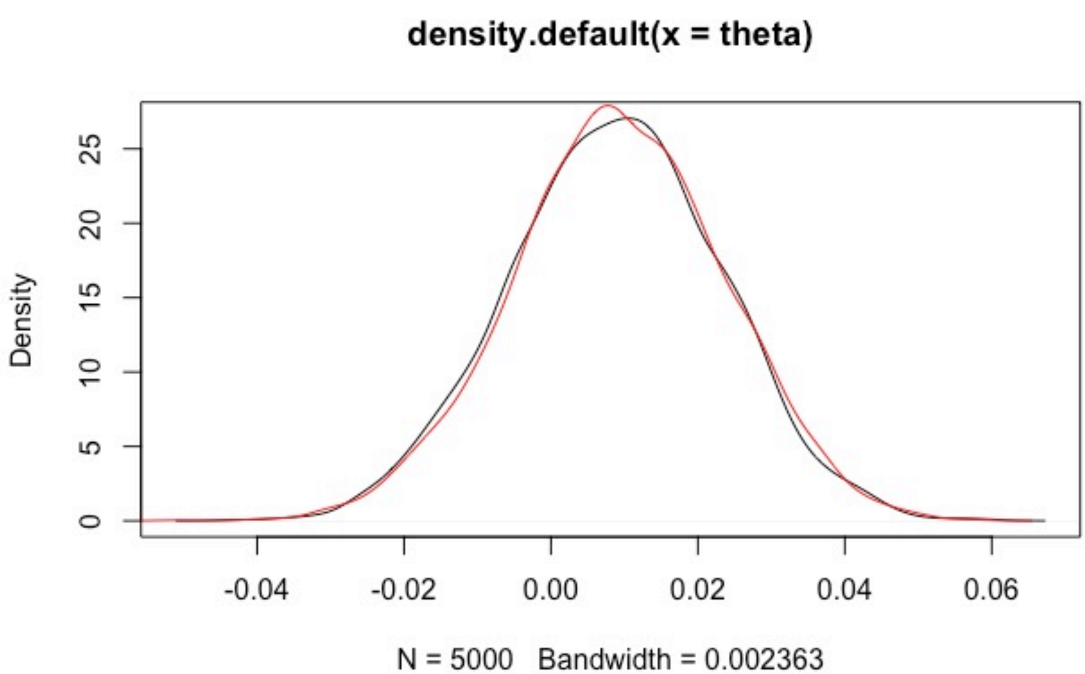
$$\Rightarrow \frac{1}{\sigma^2} \sim \text{Inv-Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n SS_n^2}{2}\right)$$

See R Code for Monte Carlo Simulation.

Joseph High
Hopkins ID: 9E1FDC

Problem 4 R Code and Plot

```
> #Problem 4
>
> n<-5000
>
> #Chosen values and Prior Parameters:
> theta0 = 1
> sig0 = 0.5
> v0 = 1
> k0 = 1
> kn<-k0+n
> vn<-v0+n
>
> y<-rnorm(n, mean=0, sd = 1)
> yb<-mean(y)
> SS<-sum((y-yb)^2)
> theta_n<-(k0*theta0+n*yb)/kn
> SSn<-(v0*sig0+SS+(k0*n)*(y-theta0)^2/kn)/vn
> sig <- 1/rgamma(5000,vn/2,vn*SSn/2)
> theta<-rnorm(5000, theta_n, sqrt(sig/kn))
> t<-rt(5000, df=vn)*sqrt(SSn/kn)+theta_n
> theta_density<-density(theta)
> t_dist<-density(t)
> plot(theta_density)
> lines(t_dist, col="red")
```



Joseph High
Hopkins ID: 9E1FDC

EN.553.732; Homework 2
Problems 5 and 6

Problem 5

Part (a):
R Code

```
#Problem 5
#Part (a)

mu0=5
sigma0=4
v0=2
k0=1
schooldata=list()

schooldata[1]<-read.table("school1.txt")
schooldata[2]<-read.table("school2.txt")
schooldata[3]<-read.table("school3.txt")

n = sapply(schooldata, length)
ybar=sapply(schooldata, mean)
s=sapply(schooldata, var)

kn=k0+n
vn=v0+n
mun=(k0*mu0+n*ybar)/kn
sigman=(v0*sigma0+(n-1)*s+k0*n*(ybar-mu0)^2/kn)/(vn)
sigma=mu=matrix(0, 10000, 3, dimnames = list(NULL, c("school1",
"school2", "school3")))
for (i in c(1, 2, 3)){
  sigma[,i]=1/rgamma(10000, vn[i]/2, vn[i]*sigman[i]/2)
  mu[,i]=rnorm(10000, mun[i], (sigma[,i]/kn[i])^0.5)
}

#Computing posterior means and 95% confidence interval for mu

colMeans(mu)
apply(mu, 2, function(x) {
  quantile(x, c(0.025, 0.975))
})

#Computing posterior means and 95% confidence interval for standard
deviation

colMeans(sqrt(sigma))
apply(sqrt(sigma), 2, function(x) {
  quantile(x, c(0.025, 0.975))
})
```

Joseph High
Hopkins ID: 9E1FDC

Results:

Posterior Means:

```
school1  school2  school3
9.290606 6.963136 7.814114
```

95% CI for mean:

```
       school1  school2  school3
2.5%   7.75762 5.150658 6.172948
97.5% 10.84183 8.787480 9.427163
```

Posterior Means for standard deviation:

```
school1  school2  school3
3.905729 4.402176 3.741269
```

95% CI for standard deviation:

```
       school1  school2  school3
2.5%  3.000531 3.349973 2.800034
97.5% 5.157399 5.889208 5.110928
```

Problem 5, Part b

R Code

```
#Part b

#combinat package installed for permn function use. Used to generate
all 6 permutations of {1,2,3}.

mu_ranks= t(apply(mu, 1, rank))
prob_ranks= list()
for (p in permn(3)) {
  index= apply(mu_ranks, 1, function(row) {
    all(row == p)
  })
  prob_ranks[[paste(p, collapse = ",")]] = length(mu_ranks[index,
1])/10000
}

prob_ranks[["1,2,3"]]
prob_ranks[["1,3,2"]]
prob_ranks[["2,1,3"]]
prob_ranks[["3,1,2"]]
prob_ranks[["2,3,1"]]
prob_ranks[["3,2,1"]]
```

Results:

```
> prob_ranks[["1,2,3"]]
[1] 0.0066

> prob_ranks[["1,3,2"]]
```

```
[1] 0.0042

> prob_ranks[["2,1,3"]]
[1] 0.0846

> prob_ranks[["3,1,2"]]
[1] 0.6639

> prob_ranks[["2,3,1"]]
[1] 0.0154

> prob_ranks[["3,2,1"]]
[1] 0.2253
```

Problem 5, Part c

R Code

```
#Part c

#Posterior predictive distribution
predict = matrix(0, 10000, 3, dimnames = list(NULL,
c("school1","school2", "school3")))
for (i in c(1, 2, 3)) {
  predict[, i]= rnorm(10000, mun[i], sqrt(sigma[,i]*((kn[i]+1)/kn[i])))
}

#Computing ranks and probabilities

pred_rank= t(apply(predict, 1, rank))
pred_probrank = list()
for (p in permn(3)) {
  index = apply(pred_rank, 1, function(row) {all(row == p)
  })
  pred_probrank[[paste(p, collapse = ",")]]= length(pred_rank[index,
1])/10000
}

pred_probrank[["1,2,3"]]
pred_probrank[["1,3,2"]]
pred_probrank[["3,1,2"]]
pred_probrank[["2,1,3"]]
pred_probrank[["2,3,1"]]
pred_probrank[["3,2,1"]]
```

Results:

```
> pred_probrank[["1,2,3"]]
[1] 0.1092

> pred_probrank[["1,3,2"]]
[1] 0.1041

> pred_probrank[["3,1,2"]]
[1] 0.2699

> pred_probrank[["2,1,3"]]
[1] 0.1828
```

```
> pred_probrank[["2,3,1"]]
[1] 0.1402

> pred_probrank[["3,2,1"]]
[1] 0.1938
```

Problem 5, part d

R Code and Results

```
> #Part d
> prob_ranks[["2,3,1"]]+prob_ranks[["3,2,1"]]
[1] 0.2407

> pred_probrank[["2,3,1"]]+pred_probrank[["3,2,1"]]
[1] 0.334
```

Joseph High

Problem 6

Part (a) : Our sampling model
$$y_1, y_2, \dots, y_J \mid \theta_1, \dots, \theta_J \sim \text{Multinomial}(\theta_1, \dots, \theta_J)$$

and the prior $\theta_1, \dots, \theta_J \sim \text{Dirichlet}(\alpha)$.
Then
$$\pi(\theta) \propto \prod_{j=1}^{J} \theta_j^{\alpha_j - 1} \quad \text{where } \theta = (\theta_1, \dots, \theta_J)$$

Then, by Lecture 5, slide 20 since $\theta \sim \text{Dirichlet}(\alpha)$ we have that the posterior is

$$p(\theta \mid y_1, \dots, y_J) \propto \prod_{j=1}^{J} \theta_j^{y_j + \alpha_j - 1}$$

i.e. $\theta \mid y_1, \dots, y_J \sim \text{Dirichlet}(\alpha + y)$

where $y = (y_1, \dots, y_J)$.
It is also well known that the marginal distribution of a subvector of $\theta = (\theta_1, \dots, \theta_J)$ is also Dirichlet i.e. for components θ_i, θ_j in θ

$$(\theta_i, \theta_j, 1 - \theta_i - \theta_j) \sim \text{Dirichlet}(\alpha_i, \alpha_j, \alpha_0 - \alpha_i - \alpha_j)$$

Then, $(\theta_1, \theta_2, 1 - \theta_1 - \theta_2) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_0 - \alpha_1 - \alpha_2)$
and so,
$$\mathbb{P}(\theta_1, \theta_2 \mid y_1, \dots, y_J) \propto \theta_1^{y_1 + \alpha_1 - 1} \theta_2^{y_2 + \alpha_2 - 1} (1 - \theta_1 - \theta_2)^{\tilde{y} + \tilde{\alpha} - 1}$$

where $\tilde{y} = \sum_{i=3}^{J} y_i$ and $\tilde{\alpha} = \sum_{i=3}^{J} \alpha_i$

It is given that $\alpha = \dfrac{\theta_1}{\theta_1 + \theta_2}$ and letting $u = \theta_1 + \theta_2$

$\Longrightarrow$ we get

Then,

$$1 = \iint p(\theta_1, \theta_2 | y_1, \dots, y_J)\, d\theta_1 d\theta_2 = \iint p(\alpha, u | y_1, \dots y_J) \, |J| \, d\alpha \, du$$

where $|J| = \begin{vmatrix} \frac{\partial \theta_1}{\partial \alpha} & \frac{\partial \theta_1}{\partial u} \\ \frac{\partial \theta_2}{\partial \alpha} & \frac{\partial \theta_2}{\partial u} \end{vmatrix}$, (the Jacobian)

w/ $\alpha = \frac{\theta_1}{\theta_1 + \theta_2}$, $u = \theta_1 + \theta_2$

$\Rightarrow \theta_1 = \frac{\theta_2 \alpha}{1-\alpha}$, $\theta_1 = u - \theta_2$

$\theta_2 = \frac{\theta_1 \alpha}{1-\alpha}$, $\theta_2 = u - \theta_1$

$\Rightarrow \theta_1 = \frac{(u-\theta_1)\alpha}{1-\alpha} \Rightarrow \theta_1 \left(\frac{1}{1-\alpha} \right) = \frac{u}{1-\alpha} \alpha$

$\Rightarrow \boxed{\theta_1 = u\alpha}$

$\Rightarrow \theta_2 = u - \theta_1 = u(1-\alpha)$

Then, $\frac{\partial \theta_1}{\partial \alpha} = u$ $\quad \frac{\partial \theta_1}{\partial u} = \alpha$ $\quad \frac{\partial \theta_2}{\partial \alpha} = -u$ $\quad \frac{\partial \theta_2}{\partial u} = 1 - \alpha$

Then, $J = \begin{vmatrix} u & \alpha \\ -u & 1-\alpha \end{vmatrix} = u(1-\alpha) + \alpha u = u$

and

$$p(\theta_1, \theta_2 | y_1 \dots y_J) = \left[\frac{\theta_1}{\theta_1 + \theta_2} (\theta_1 + \theta_2) \right]^{y_1 + \alpha_1 - 1} \left[\left(1 - \frac{\theta_1}{\theta_1 + \theta_2}\right)(\theta_1 + \theta_2) \right]^{y_2 + \alpha_2 - 1}$$

$$\cdot \left(1 - (\theta_1 + \theta_2)\right)^{\tilde{y} + \tilde{\alpha} - 1}$$

$\Longrightarrow$

$$= (\alpha u)^{y_1+\alpha_1-1}\,[(1-\alpha)u]^{y_2+\alpha_2-1}\,(1-u)^{\tilde{y}+\tilde{\alpha}-1}$$

$$\Rightarrow p(\alpha, u \mid y_1, \dots y_J) = u\left[\alpha^{y_1+\alpha_1-1}(1-\alpha)^{y_2+\alpha_2-1} \cdot u^{y_1+y_2+\alpha_1+\alpha_2-2}(1-u)^{\tilde{y}+\tilde{\alpha}-1}\right]$$

$$= \underbrace{\alpha^{y_1+\alpha_1-1}(1-\alpha)^{y_2+\alpha_2-1}}_{Beta(y_1+\alpha_1,\ y_2+\alpha_2)} \cdot \underbrace{u^{y_1+y_2+\alpha_1+\alpha_2-1}(1-u)^{\tilde{y}+\tilde{\alpha}-1}}_{Beta(y_1+y_2+\alpha_1+\alpha_2,\ \tilde{y}+\tilde{\alpha})}$$

Marginal distribution of alpha

$$\Rightarrow \alpha \mid y_1, \dots, y_J \sim Beta(y_1+\alpha_1,\ y_2+\alpha_2)$$

Binomial distribution with probability α and $n = y_1 + y_2$ with $Beta(\alpha_1, \alpha_2)$ prior

is

$$p(\alpha \mid y_1 \dots y_J) \propto \underbrace{\alpha^{y_1}(1-\alpha)^{y_1+y_2-y_1}}_{\text{Binomial w/ } n=y_1+y_2,\ y_1 \text{ observation.}}\ \underbrace{\alpha^{\alpha_1-1}(1-\alpha)^{\alpha_2-1}}_{Beta(\alpha_1,\ \alpha_2)}$$

$$= \alpha^{y_1+\alpha_1-1}(1-\alpha)^{y_2+\alpha_2-1}$$

Identical

$\square$

Part (b): (Referencing Lecture 5, Slide 21)

Pre-Debate: $p(\theta|y) = \theta^{294}\theta^{307}\theta^{38}$, $\theta \sim D(1,1,1)$

$\hookrightarrow \theta|y \sim D(295, 308, 39)$

Post Debate: and $p(\beta|y) = \beta^{288}\beta^{332}\beta^{19}$, $\beta \sim D(1,1,1)$

$\beta|y \sim D(289, 333, 20)$

Then, with $\alpha_1 = \dfrac{\theta_1}{\theta_1+\theta_2}$, $\alpha_2 = \dfrac{\beta_1}{\beta_1+\beta_2}$

Then by (a) we have that

$\alpha_1|y \sim Beta(295, 308)$

$\alpha_2|y \sim Beta(289, 333)$

The R Code for the histogram of $\alpha_2 - \alpha_1$ density is attached.

Indeed, since α_2 corresponds to the proportion of voters preferring Bush post-debate and α_1 the proportion voters preferring Bush pre-debate, then the posterior probability that there was a shift toward Bush corresponds to $\alpha_2 - \alpha_1$ being positive since this implies an increase in people tending toward Bush post debate.

The posterior probability that there was a shift in Bush's favor was found to be 0.1934.

Joseph High
Hopkins ID: 9E1FDC

Problem 6b

R Code

```
> #Problem 6 part b:
>
> N=5000
> a1= rbeta(N, 295, 308)
> a2= rbeta(N, 289, 333)
> x<-a2-a1
> hist(x, main='Problem 6(b) Histogram', xlab='a2-a1')
>
> ##As written in my HW, a2-a1>0 corresponds to a shift
toward Bush post debate, so this is why
> #this is taken here.
> mean(x>0)

[1] 0.1934
> #Thus, the posterior probability that there was a shift
in Bush's favor is 0.1934.
>
```

Histogram Plot