Winter 18 – AMS206B Homework 6

Due Friday March 3rd.

1. (a) **Note:** We treat y_1 as a constant since we are conditioning on it. In other words, no distribution for y_1.

$$\begin{aligned} f(y_2,\ldots,y_n|y_1,\rho,\sigma^2) &= \prod_{i=2}^{n} f(y_i \mid y_1,\ldots,y_{i-1},\rho,\sigma 2) \\ &= \prod_{i=2}^{n} f(y_i \mid y_{i-1},\rho,\sigma 2) \\ &= (\frac{1}{\sqrt{2\pi\sigma^2}})^{n-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n}(y_i-\rho y_{i-1})^2\right\}. \end{aligned}$$

(b) i.

$$\begin{aligned} p(\rho,\sigma^2 \mid y_1,\ldots,y_n) &\propto \underbrace{f(y_2,\ldots,y_n \mid y_1,\rho,\sigma^2)}_{\text{from part (a)}}\pi(\rho,\sigma^2) \\ &\propto (\frac{1}{\sqrt{2\pi\sigma^2}})^{n-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n}(y_i-\rho y_{i-1})^2\right\}\frac{1}{\sigma^2} \\ &\propto (\frac{1}{\sigma^2})^{\frac{n-1}{2}+1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n}(y_i-\rho y_{i-1})^2\right\} \end{aligned}$$

ii. when we find $p(\rho \mid \sigma^2, y_1,\ldots,y_n)$, we can treat σ^2 as a constant since we are conditioning on σ^2. We may drop any factor having σ^2 only.

$$\begin{aligned} p(\rho \mid \sigma^2, y_1,...,y_n) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n}(y_i-\rho y_{i-1})^2\right\} \\ &\propto \exp\left\{-\frac{\sum_{i=2}^{n} y_{i-1}^2}{2\sigma^2}(\rho^2 - 2\rho\frac{\sum_{i=2}^{n} y_i y_{i-1}}{\sum_{i=2}^{n} y_{i-1}^2})\right\} \\ &\propto \exp\left\{-\frac{1}{2\frac{\sigma^2}{\sum_{i=2}^{n} y_{i-1}^2}}(\rho - \frac{\sum_{i=2}^{n} y_i y_{i-1}}{\sum_{i=2}^{n} y_{i-1}^2})^2\right\} \end{aligned}$$

Thus, $\pi(\rho \mid \sigma^2, y_1,...,y_n)$ is

$$\mathrm{N}\left(\frac{\sum_{i=2}^{n} y_i y_{i-1}}{\sum_{i=2}^{n} y_{i-1}^2}, \frac{\sigma^2}{\sum_{i=2}^{n} y_{i-1}^2}\right).$$

Now, we need to be careful. By completing the normal density for ρ, we get more terms having σ^2 that need to be kept for the marginal posterior of σ^2.

$$\begin{aligned}
p(\sigma^2|y_1,\ldots,y_n) &\propto (\frac{1}{\sigma^2})^{\frac{n-1}{2}+1}\exp\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n} y_i^2\} \\
&\quad \times \int \exp\left\{-\frac{\sum_{i=2}^{n} y_{i-1}^2}{2\sigma^2}(\rho^2 - 2\rho\frac{\sum_{i=2}^{n} y_i y_{i-1}}{\sum_{i=2}^{n} y_{i-1}^2})\right\} d\rho \\
&\propto (\frac{1}{\sigma^2})^{\frac{n-1}{2}+1}\exp\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n} y_{i-1}^2\} \underbrace{\sqrt{\sigma^2}\exp\{\frac{1}{2\sigma^2}\frac{(\sum_{i=2}^{n} y_i y_{i-1})^2}{\sum_{i=1}^{n} y_{i-1}^2}\}}_{\text{From completing the normal density for } \rho} \\
&\propto (\frac{1}{\sigma^2})^{\frac{n}{2}-1+1}\exp\left\{-\frac{1}{2\sigma^2}\{\sum_{i=2}^{n} y_i^2 - \frac{(\sum_{i=2}^{n} y_i y_{i-1})^2}{\sum_{i=1}^{n} y_{i-1}^2}\}\right\}
\end{aligned}$$

Thus, $\pi(\sigma^2 \mid y_1, ..., y_n)$ is

$$\mathrm{IG}\left(\frac{n}{2}-1, \frac{1}{2}\sum_{i=2}^{n} y_i^2 - \frac{1}{2}\frac{(\sum_{i=2}^{n} y_i y_{i-1})^2}{\sum_{i=1}^{n} y_{i-1}^2}\right).$$

2. (a) To calculate the posterior mean and variance we must first find the posterior.

$$\begin{aligned}
\pi(\theta|X) &\propto f(X|\theta)\pi(\theta) \\
&\propto \binom{n}{x}\theta^x(1-\theta)^{n-x}\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&\propto \theta^x(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&\propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}
\end{aligned}$$

Which is the kernel of a $\mathrm{Be}(\alpha+x, \beta+n-x-1)$ distribution. Therefore,

$$\begin{aligned}
\mathrm{E}[\theta|x] &= \frac{x+\alpha}{x+\alpha+\beta+n-x} \\
&= \frac{x+\alpha}{\alpha+\beta+n} \\
\mathrm{Var}(\theta|x) &= \frac{(x+\alpha)(\beta+n-x)}{(\alpha+\beta+n)^2(\beta+n-x+1)}.
\end{aligned}$$

$\star\star\star$ Let's find the MAP as an estimate of θ and $\alpha+x>1, n-x+\beta>1$.

$$\begin{aligned}
\pi(\theta|X) &\propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1} \\
\log(\pi(\theta|X)) &\propto (x+\alpha-1)\log\theta + (\beta+n-x-1)\log(1-\theta) \\
\frac{\partial}{\partial\theta}(\log[\pi(\theta|X)]) &= \frac{x+\alpha-1}{\theta} - \frac{\beta+n-x-1}{1-\theta} \\
\frac{\partial^2}{\partial\theta^2}(\log[\pi(\theta|X)]) &= -\frac{x+\alpha-1}{\theta^2} - \frac{\beta+n-x-1}{(1-\theta)^2} < 0 \quad \forall\theta \in (0,1)
\end{aligned}$$

Then, we find $\hat{\theta}_{\text{MAP}}$ by equating $\frac{\partial}{\partial\theta}(\log[\pi(\theta|X)])$=0 for $\theta > 0$.

$$\frac{x+\alpha-1}{\hat{\theta}_{\text{MAP}}} - \frac{\beta+n-x-1}{1-\hat{\theta}_{\text{MAP}}} = 0 \Rightarrow \hat{\theta}_{\text{MAP}} = \frac{x+\alpha-1}{\alpha+\beta+n-2}$$

If $\alpha+x<1$ then $\pi(\theta|X)\to\infty$ as $\theta\to 0$ and if $n-x+\beta<1$ then $\pi(\theta|X)\to\infty$ as $\theta\to 1$. From the lecture, we know $\text{Var}(\hat{\theta}_{\text{MAP}}) = (\hat{\theta}-\hat{\theta}_{\text{MAP}})^2 + \text{Var}(\hat{\theta})$ where $\hat{\theta}$ is the posterior mean. So,

$$\text{Var}(\hat{\theta}_{\text{MAP}}) = \left(\frac{x+\alpha-1}{\alpha+\beta+n-2} - \frac{x+\alpha}{\alpha+\beta+n}\right)^2 + \frac{(x+\alpha)(\beta+n-x)}{(\alpha+\beta+n)^2(\beta+n-x+1)}$$

(b) To calculate the posterior mean and variance we must first find the posterior.

$$\begin{aligned}
\pi(\theta|X) &\propto f(X|\theta)\pi(\theta) \\
&\propto \left(\frac{1}{\theta}\right)^n \exp\left(-\frac{\sum\limits_{i=1}^n x_i}{\theta}\right) \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-\alpha-1}\exp\left(-\frac{\beta}{\theta}\right) \\
&\propto \left(\frac{1}{\theta}\right)^n \exp\left(-\frac{\sum\limits_{i=1}^n x_i}{\theta}\right) \theta^{-\alpha-1}\exp\left(-\frac{\beta}{\theta}\right) \\
&\propto \theta^{-\alpha-n-1}\exp\left[-\frac{1}{\theta}\left(\sum_{i=1}^n x_i+\beta\right)\right]
\end{aligned}$$

Which is the kernel of a IG $\left(\alpha+n, \sum\limits_{i=1}^n x_i+\beta\right)$ distribution. Therefore,

$$\begin{aligned}
\text{E}[\theta|x] &= \frac{\sum\limits_{i=1}^n x_i+\beta}{\alpha+n-1} \\
\text{Var}(\theta|x) &= \frac{\sum\limits_{i=1}^n x_i+\beta}{(\alpha+n-1)^2(\alpha+n-2)} \qquad \alpha+n>2.
\end{aligned}$$

$\star\star\star$ Let's find the MAP as an estimate of θ. Here the maximum was obtained by $\hat{\theta}_{\text{MAP}}$ found below:

$$\begin{aligned}
\pi(\theta|X) &\propto \theta^{-\alpha-n-1}\exp\left[-\frac{1}{\theta}\left(\sum_{i=1}^n x_i+\beta\right)\right] \\
\log(\pi(\theta|X)) &\propto (-\alpha-n-1)\log\theta - \frac{\sum_{i=1}^n x_i+\beta}{\theta} \\
\frac{\partial}{\partial\theta}(\log[\pi(\theta|X)]) &= -\frac{\alpha+n+1}{\theta} + \frac{\sum_{i=1}^n x_i+\beta}{\theta^2} \\
\frac{\partial^2}{\partial\theta^2}(\log[\pi(\theta|X)]) &= \frac{1}{\theta^2}\left(\alpha+n+1-2\frac{\sum_{i=1}^n x_i+\beta}{\theta}\right)
\end{aligned}$$

Then, we find $\hat{\theta}_{\text{MAP}}$ by equating $\frac{\partial}{\partial\theta}\left(\log[\pi(\theta|X)]\right)$=0 for $\theta > 0$.

$$-\frac{\alpha+n+1}{\hat{\theta}_{\text{MAP}}} + \frac{\sum_{i=1}^n x_i + \beta}{\hat{\theta}^2_{\text{MAP}}} = 0 \Rightarrow \hat{\theta}_{\text{MAP}} = \frac{\sum_{i=1}^n x_i + \beta}{\alpha+n+1}.$$

Now

$$\frac{\partial^2}{\partial\theta^2}\left(\log[\pi(\theta|X)]\right)\Bigg|_{\hat{\theta}_{\text{MAP}}} = -\frac{1}{\hat{\theta}^2_{\text{MAP}}}\left(\alpha+n+1\right) < 0$$

which confirms that $\hat{\theta}_{\text{MAP}}$ is indeed a maximum.
Since we know $\text{Var}(\hat{\theta}_{\text{MAP}}) = (\hat{\theta} - \hat{\theta}_{\text{MAP}})^2 + \text{Var}(\hat{\theta})$ where $\hat{\theta}$ is the posterior mean. So,

$$\text{Var}(\hat{\theta}_{\text{MAP}}) = \left(\frac{\sum_{i=1}^n x_i + \beta}{\alpha+n+1} - \frac{\sum_{i=1}^n x_i + \beta}{\alpha+n-1}\right)^2 + \frac{\sum_{i=1}^n x_i + \beta}{(\alpha+n-1)^2(\alpha+n-2)}.$$

(c) To calculate the posterior mean and variance we must first find the posterior.

$$\begin{aligned}
\pi(\theta|X) &\propto f(X|\theta)\pi(\theta)\\
&\propto \frac{x^{n/2-1}}{(2\theta)^{n/2}\Gamma(n/2)} exp\left\{-\frac{x}{2\theta}\right\}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-\alpha-1} exp\left(-\frac{\beta}{\theta}\right)\\
&\propto \left(\frac{1}{(2)^{n/2}}\right)\left(\frac{1}{(\theta)^{n/2}}\right) exp\left\{-\frac{x}{2\theta}\right\}\theta^{-\alpha-1} exp\left(-\frac{\beta}{\theta}\right)\\
&\propto \theta^{-\alpha-n/2-1} exp\left\{-\frac{1}{\theta}\left(x/2+\beta\right)\right\}.
\end{aligned}$$

Which is the kernel of a IG $(\alpha+n/2, x/2+\beta)$ distribution. Therefore,

$$\begin{aligned}
\text{E}[\theta|X] &= \frac{x/2+\beta}{\alpha+n/2-1} \quad \alpha+n/2>1\\
\text{Var}(\theta|X) &= \frac{x/2+\beta}{(\alpha+n/2-1)^2(\alpha+n/2-2)} \quad \alpha+n/2>2
\end{aligned}$$

$\star\star\star$ Let's find the MAP as an estimate of θ. Here the maximum was obtained by $\hat{\theta}_{\text{MAP}}$ found below:

$$\begin{aligned}
\pi(\theta|X) &\propto \theta^{-\alpha-n/2-1} exp\left\{-\frac{1}{\theta}\left(x/2+\beta\right)\right\}\\
\log(\pi(\theta|X)) &= -\frac{x}{2\theta} - (\alpha+n/2+1)\log\theta - \frac{\beta}{\theta}\\
\frac{\partial}{\partial\theta}\left(\log[\pi(\theta|X)]\right) &= \frac{x+2\beta}{2\theta^2} - \frac{\alpha+n/2+1}{\theta}\\
\frac{\partial^2}{\partial\theta^2}\left(\log[\pi(\theta|X)]\right) &= \frac{1}{\theta^2}\left(\alpha+n/2+1-\frac{x+2\beta}{\theta}\right).
\end{aligned}$$

Then, we find $\hat{\theta}_{\text{MAP}}$ by equating $\frac{\partial}{\partial\theta}$ $(\log[\pi(\theta|X)])$=0 for $\theta > 0$.

$$\frac{x+2\beta}{2\hat{\theta}^2_{\text{MAP}}} - \frac{\alpha+n/2+1}{\hat{\theta}_{MAP}} = 0 \Rightarrow \hat{\theta}_{\text{MAP}} = \frac{x/2+\beta}{\alpha+n/2+1}.$$

And

$$\frac{\partial^2}{\partial\theta^2}(\log[\pi(\theta|X)])\bigg|_{\hat{\theta}_{\text{MAP}}} = -\frac{1}{\hat{\theta}^2_{\text{MAP}}}(\alpha+n/2+1) < 0$$

which confirms that $\hat{\theta}_{\text{MAP}}$ is indeed a maximum.
Since we know $\text{Var}(\hat{\theta}_{\text{MAP}}) = (\hat{\theta} - \hat{\theta}_{\text{MAP}})^2 + \text{Var}(\hat{\theta})$ where $\hat{\theta}$ is the posterior mean. So,

$$\text{Var}(\hat{\theta}_{\text{MAP}}) = \left(\frac{x+2\beta}{2\alpha+m+2} - \frac{x/2+\beta}{\alpha+n/2-1}\right)^2 + \frac{x/2+\beta}{(\alpha+n/2-1)^2(\alpha+n/2-2)}.$$

(d) We already did this above.

3. (a) (Find the distribution of Y_n) Since Y is the n^{th} order statistic, and there is a general formula for order statistics then the formula can be applied. However, since the formula requires the CDF, the CDF must be found first.

$$f(x|\theta) = \frac{1}{\theta}I(0 < x < \theta)$$
$$F(x|\theta) = \int_0^x \frac{1}{\theta}d\theta$$
$$F(x|\theta) = \frac{x}{\theta}$$

So $F(x|\theta) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{\theta} & 0 < x < \theta \\ 1 & x > \theta \end{cases}$

Then $f_Y(y) = \binom{n-1}{n-1}\left(\frac{n}{\theta}\right)\left(\frac{1}{\theta}y\right)^{n-1} I(0 < x < 1)$ or $f(y) = \left(\frac{n}{\theta}\right)\left(\frac{y}{\theta}\right)^{n-1} I(0 < x < 1)$.

(b) (Find the Bayes estimate) It is known that the posterior mean minimizes least square error or is the "Bayes' point estimate" for least square error . Therefore, it is necessary to compute the posterior and its expectation.
The posterior is $\pi(\theta|Y) = \frac{f(Y|\theta)\pi(\theta)}{m(Y)}$ and $m(Y) = \int f(Y|\theta)\pi(\theta)d\theta$. Since $0 < y < \theta$, then $\theta > y$.

$$\begin{aligned} m(Y) &= \int_y^\infty \left(\frac{n}{\theta}\right)\left(\frac{y}{\theta}\right)^{n-1}\frac{\beta\alpha^\beta}{\theta^{\beta+1}}d\theta \\ &= \beta\alpha^\beta n y^{n-1}\int_y^\infty \left(\frac{1}{\theta}\right)^{n+\beta+1} d\theta \\ &= \beta\alpha^\beta n y^{n-1}\left[-\frac{1}{n+\beta}\left(\frac{1}{\theta^{n+\beta}}\right)\right]_y^\infty \\ &= \beta\alpha^\beta n y^{n-1}\left(\frac{1}{y^{n+\beta}}\right) \end{aligned}$$

Now, the posterior can be computed.

$$\pi(\theta|Y) = \frac{f(Y|\theta)\pi(\theta)}{m(Y)} = \frac{\left(\frac{n}{\theta}\right)\left(\frac{y}{\theta}\right)^{n-1}\frac{\beta\alpha^\beta}{\theta^{\beta+1}}}{\beta\alpha^\beta n y^{n-1}\left(\frac{1}{y^{n+\beta}}\right)} = (n+\beta)\left(\frac{1}{\theta}\right)^{n+\beta+1} y^{n+\beta}$$

NOTE: This is a Pareto distribution.
From this, the posterior mean or the "Bayes' point estimator" for mean squared loss can be computed.

$$\begin{aligned}
E[\theta|Y] &= \int_y^\infty (n+\beta)y^{n+\beta}\left(\frac{1}{\theta}\right)^{n+\beta} d\theta \\
&= (n+\beta)y^{n+\beta}\left[-\left(\frac{1}{n+\beta-1}\right)\left(\frac{1}{\theta}\right)^{n+\beta-1}\right]_y^\infty \\
&= \frac{(n+\beta)y^{n+\beta}}{(n+\beta-1)y^{n+\beta-1}} \\
&= \frac{(n+\beta)y}{(n+\beta-1)}
\end{aligned}$$

So the Bayes' point estimate for θ given mean squared error loss is $y\frac{n+\beta}{n+\beta-1}$.

4. $\frac{1}{\sigma^2} \sim \mathrm{Ga}(\frac{r}{2}, \frac{2}{r})$ implies $\sigma^2 \sim \mathrm{IG}(\frac{r}{2}, \frac{2}{r})$

$$\begin{aligned}
m(x) &= \int_0^\infty f(x|\sigma^2)\pi(\sigma^2)\, d\sigma^2 \\
&\propto \int_0^\infty \underbrace{\frac{1}{\sqrt{\sigma^2}}\exp\left\{\frac{-x^2}{2\sigma^2}\right\}}_{\text{from the likelihood}} \underbrace{\left(\frac{1}{\sigma^2}\right)^{\frac{r}{2}+1}\exp\left\{-\frac{1}{\sigma^2\left(\frac{2}{r}\right)}\right\}}_{\text{from the prior}} d\sigma^2
\end{aligned}$$

Recognize a kernel for $\mathrm{IG}(\frac{r}{2}+\frac{1}{2}, (\frac{x^2}{2}+\frac{r}{2})^{-1})$ and so obtain

$$\begin{aligned}
m(x) &\propto \left(\frac{x^2}{2}+\frac{r}{2}\right)^{-(r/2+1/2)} \\
&\propto \left(1+\frac{x^2}{r}\right)^{-\frac{r+1}{2}}
\end{aligned}$$

We recognize a kernel for a $t-$distribution and look for a parameter. We find that this is the pdf for a t-distribution with r degrees of freedom