Homework 3 Responses

Chris Crabtree

Oct. 28, 2021

1.a

> **Prompt**: *Show that if $\boldsymbol{y} \sim EFD(\boldsymbol{\theta})$ with EFD being an Exponential Family Distribution, then*
>
> $$\mathbb{E}[\boldsymbol{y}|\boldsymbol{\theta}] = \frac{\partial\psi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

For notational convenience I will write $\boldsymbol{y}|\boldsymbol{\theta}$ as just $\boldsymbol{y}$.

$$\begin{aligned}
1 &= \int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y} - \psi(\boldsymbol{\theta}))d\boldsymbol{y} && \text{(exp. family dist'n)} \\
\tfrac{\partial}{\partial\boldsymbol{\theta}} 1 &= \tfrac{\partial}{\partial\boldsymbol{\theta}} \int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y} - \psi(\boldsymbol{\theta}))d\boldsymbol{y} \\
0 &= \tfrac{\partial}{\partial\boldsymbol{\theta}} \exp(-\psi(\boldsymbol{\theta})) \int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y})d\boldsymbol{y} \\
0 &= \tfrac{\partial}{\partial\boldsymbol{\theta}} \Big[\exp(-\psi(\boldsymbol{\theta}))\Big] \int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y})d\boldsymbol{y} \\
&\quad + \exp(-\psi(\boldsymbol{\theta})) \tfrac{\partial}{\partial\boldsymbol{\theta}} \Big[\int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y})d\boldsymbol{y}\Big] && \text{(product rule)} \\
0 &= \tfrac{\partial}{\partial\boldsymbol{\theta}} \Big[\exp(-\psi(\boldsymbol{\theta}))\Big] \int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y})d\boldsymbol{y} \\
&\quad + \exp(-\psi(\boldsymbol{\theta})) \int_{\boldsymbol{y}} \tfrac{\partial}{\partial\boldsymbol{\theta}} \Big[h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y})\Big] d\boldsymbol{y} && \text{(Leibnitz rule and limits const. w.r.t. } \boldsymbol{\theta}) \\
0 &= -\tfrac{\partial}{\partial\boldsymbol{\theta}} \Big[\psi(\boldsymbol{\theta})\Big] \exp(-\psi(\boldsymbol{\theta})) \int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y})d\boldsymbol{y} \\
&\quad + \exp(-\psi(\boldsymbol{\theta})) \int_{\boldsymbol{y}} h(\boldsymbol{y}) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{y})\boldsymbol{y}d\boldsymbol{y} \\
0 &= -\tfrac{\partial}{\partial\boldsymbol{\theta}} \Big[\psi(\boldsymbol{\theta})\Big] \cdot 1 + \mathbb{E}[\boldsymbol{y}] \\
\mathbb{E}[\boldsymbol{y}] &= \tfrac{\partial}{\partial\boldsymbol{\theta}} \Big[\psi(\boldsymbol{\theta})\Big]
\end{aligned}$$

$\blacksquare$

1.b

> **Prompt**: *Show that $\mathbb{E}[\xi(\boldsymbol{\theta})] = \frac{\boldsymbol{y}_0}{\lambda} + c$ and $\mathbb{E}[\xi(\boldsymbol{\theta})|\boldsymbol{y}_{1:n}] = \frac{\boldsymbol{y}_0 + n\bar{\boldsymbol{y}}}{\lambda + n} + c$ for some constant c*

Using the hint and the result from part a, we have:

$$\begin{aligned}
\mathbb{E}[\xi(\boldsymbol{\theta})] = \mathbb{E}[\tfrac{\partial\psi(\theta)}{\partial\boldsymbol{\theta}}] = \mathbb{E}[\tfrac{\partial\psi(\theta)}{\partial\boldsymbol{\theta}}] \\
&= \mathbb{E}[\frac{1}{\lambda}\Big(\boldsymbol{y}_0 - \Big(\boldsymbol{y}_0 - \lambda\tfrac{\partial\psi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big)\Big)] \\
&= \mathbb{E}[\frac{1}{\lambda}\Big(\boldsymbol{y}_0 - \frac{\frac{\partial}{\partial\boldsymbol{\theta}}\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))}{\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))}\Big)] \\
&= \frac{1}{\lambda}\Big(\boldsymbol{y}_0 - \mathbb{E}[\frac{\frac{\partial}{\partial\boldsymbol{\theta}}\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))}{\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))}]\Big) \\
&= \frac{1}{\lambda}\Big(\boldsymbol{y}_0 - \int\frac{\frac{\partial}{\partial\boldsymbol{\theta}}\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))}{\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))}h(\boldsymbol{y}_0,\lambda)\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))d\boldsymbol{\theta}\Big) \\
&= \frac{1}{\lambda}\Big(\boldsymbol{y}_0 - h(\boldsymbol{y}_0,\lambda)\int\frac{\partial}{\partial\boldsymbol{\theta}}\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))d\boldsymbol{\theta}\Big) \\
&= \frac{1}{\lambda}\boldsymbol{y}_0 - \frac{h(\boldsymbol{y}_0,\lambda)}{\lambda}\int\frac{\partial}{\partial\boldsymbol{\theta}}\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))d\boldsymbol{\theta}
\end{aligned}$$

Since $\boldsymbol{y}|\boldsymbol{\theta}$ is distributed according to an exponential family distribution, $\mathbb{E}[\boldsymbol{y}|\boldsymbol{\theta}] = \mathbb{E}[\xi(\boldsymbol{\theta})]$ must exist. Therefore, $\int\frac{\partial}{\partial\boldsymbol{\theta}}\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))d\boldsymbol{\theta}$ must converge. Thus,

$$\frac{1}{\lambda}\boldsymbol{y}_0 - \frac{h(\boldsymbol{y}_0,\lambda)}{\lambda}\int\frac{\partial}{\partial\boldsymbol{\theta}}\exp(\boldsymbol{\theta}^\intercal\boldsymbol{y}_0 - \lambda\psi(\boldsymbol{\theta}))d\boldsymbol{\theta} = \frac{1}{\lambda}\boldsymbol{y}_0 + \textit{constant}$$

2.a

Prompt: *Show that the Binomial belongs to exponential family distributions.*

Using the notation of the factorization of exponential family distributions (EFD) introduced in lecture:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \equiv \text{EFD}(\boldsymbol{\theta}) \text{ iff } p(\boldsymbol{y}|\boldsymbol{\theta}) = g(\boldsymbol{\theta})h(\boldsymbol{y})\exp(\phi(\boldsymbol{\theta})^T u(\boldsymbol{y}))) \tag{1}$$

$$\tag{2}$$

The Binomial PDF can be manipulated as follows:

$$\begin{aligned}
\binom{n}{k}\theta^k(1-\theta)^{n-k} &= \binom{n}{k}(1-\theta)^n\Big(\frac{\theta}{(1-\theta)}\Big)^k \\
&= \binom{n}{k}(1-\theta)^n\exp(k\log\Big(\frac{\theta}{(1-\theta)}\Big))
\end{aligned}$$

Since the Binomial is a single parameter family (is not typically considered a parameter) I can a let equation 1 have scalar quantities and let:

$$\begin{aligned}
h(k) &= \binom{n}{k} \\
g(\theta) &= (1-\theta)^n \\
u(k) &= k \\
\phi(k) &= \log\Big(\frac{\theta}{(1-\theta)}\Big)
\end{aligned}$$

2.b

Prompt: *Show that the Negative Binomial belongs to exponential family distributions.*

The Negative Binomial PDF can be manipulated in a similar manner as above, however we must assume that r is known:

$$\binom{r+y-1}{y}\theta^y(1-\theta)^r = \binom{r+y-1}{y}(1-\theta)^r \exp(y\log(\theta)$$

Since the Binomial is a single parameter family (is not typically considered a parameter) I can a let equation 1 have scalar quantities and let:

$$\begin{aligned}
h(k) &= \binom{n}{k} \\
g(\theta) &= (1-\theta)^n \\
u(k) &= k \\
\phi(k) &= \log\left(\frac{\theta}{(1-\theta)}\right)
\end{aligned}$$

3.a

Prompt: *Show that the posterior $p(\theta|y)$ has the form:*

$$\frac{\frac{Beta(13,27)}{Beta(10,20)}\frac{1}{Beta(13,27)}\theta^{12}(1-\theta)^{26} + \frac{Beta(23,17)}{Beta(10,20)}\frac{1}{Beta(23,17)}\theta^{22}(1-\theta)^{16}}{\frac{Beta(13,27)}{Beta(10,20)} + \frac{Beta(23,17)}{Beta(10,20)}}$$

Firstly, note that the posterior above can alternatively be simplified by multiplying the numerator and denominator by $\text{Beta}(10, 20)$, cancelling the like Beta factors in the numerator, and writing the exponents in the standard form of a Beta pdf:

$$\frac{\theta^{13-1}(1-\theta)^{27-1} + \theta^{23-1}(1-\theta)^{17-1}}{\text{Beta}(13,27) + \text{Beta}(23,17)} \tag{3}$$

Now we can simply note that the posterior is proportional to the likelihood times the prior and the kernel of that quantity is of the form of another mixture of Betas:

$$\begin{aligned}
p(\theta|y) &\propto p(y|\theta)\cdot p(\theta) \\
&= \binom{10}{3}\theta^3(1-\theta)^7 \cdot \left(\frac{1}{2\text{Beta}(10,20)}\theta^{10-1}(1-\theta)^{20-1} + \frac{1}{2\text{Beta}(20,10)}\theta^{20-1}(1-\theta)^{10-1}\right) \\
&= \binom{10}{3}\frac{1}{2\text{Beta}(10,20)}\left(\theta^3(1-\theta)^7\theta^{10-1}(1-\theta)^{20-1} + \theta^3(1-\theta)^7\theta^{20-1}(1-\theta)^{10-1}\right) \\
&\propto \theta^{13-1}(1-\theta)^{27-1} + \theta^{23-1}(1-\theta)^{17-1}
\end{aligned}$$

I recognize the first term as a Beta((, 1)3, 27) and the second as a Beta((, 2)3, 17). I can now immediately write down the normalizing constant for the mixture:

$$\int \theta^{13-1}(1-\theta)^{27-1} + \theta^{23-1}(1-\theta)^{17-1} = \text{Beta}(13, 27) + \text{Beta}(23, 17)$$

The posterior can therefore be written:

$$p(\theta|y) = \frac{\theta^{13-1}(1-\theta)^{27-1} + \theta^{23-1}(1-\theta)^{17-1}}{\text{Beta}(13, 27) + \text{Beta}(23, 17)}$$

This is the same as equation 3 so we are finished.

3.b

I ran out of time to show this.

4

Prompt: *Show that Jeffrey's priors satisfy the invariance principle*

In Jeffrey's view, a prior is non-informative if the prior chosen for $p_\psi(\psi)$ is proportional to the pdf that appears with the transformation $p_\theta(\psi = g(\theta)) = p_\psi(\psi)$. The Jeffrey's prior for $p(\psi)$ would be $\left|I(\psi)\right|^{1/2}$ and the pdf of the transformation $p_\theta(\psi = g(\theta)) = p_\psi(\psi)$ is a well-known result if g is monotone, i.e. $p_\psi(\psi) = p_\theta(g^{-1}(\psi))\left|\frac{\partial}{\partial\psi}g^{-1}(\psi)\right|$. Therefore, I need to show that $p_\theta(\psi = g(\theta)) \propto \left|I(\psi)\right|^{1/2}$. Starting from the pdf of the transformation $p_\theta(\psi = g(\theta))$:

$$\begin{aligned}
p_\theta(\psi = g(\theta)) &= p_\theta(g^{-1}(\psi))\left|\tfrac{\partial}{\partial\psi}g^{-1}(\psi)\right| \\
&= \left|\tfrac{\partial^2}{\partial\theta^2}\log(p_\theta(y|\theta = g^{-1}(\psi)))\right|^{1/2}\left|\tfrac{\partial}{\partial\psi}g^{-1}(\psi)\right| \\
&\Rightarrow \\
p_\theta(\psi = g(\theta))^2 &= \left|\tfrac{\partial^2}{\partial\theta^2}\log(p_\theta(y|\theta = g^{-1}(\psi)))\right|\left|\tfrac{\partial}{\partial\psi}g^{-1}(\psi)\right|^2 \\
&= I(g^{-1}(\psi))\left|\tfrac{\partial}{\partial\psi}g^{-1}(\psi)\right|^2
\end{aligned}$$

From here I need to use a special property of Fisher information. This property relates the information of paramater to the fisher information of a function of a parameter. It states that if α and $\beta = h(\alpha)$ are scalar parameters and if h is a continuously differentiable function, then the Fisher information of the two parameterizations are related by the following equation:

$$I(\alpha) = I(h(\alpha))(\frac{d}{d\alpha}h(\alpha))^2$$

This is tricky for me to prove, but it seems to be a well known result. Using this property and letting $h = g^{-1}$ I get:

$$
\begin{aligned}
p_\theta(\psi = g(\theta))^2 &= I(g^{-1}(\psi))\left|\tfrac{\partial}{\partial\psi}g^{-1}(\psi)\right|^2 \\
&= I(\psi)^2 \\
&\Rightarrow \\
p_\theta(\psi = g(\theta)) &= I(\psi)^{1/2}
\end{aligned}
$$

Which is the Jeffrey's prior for $p(\psi)$ that I needed to prove.

5

I ran out of time :(

6.a

Prompt: *Given $y_{1:n} = y_1, y_2, ..., y_n \sim \mathcal{N}(\mu, \sigma^2)$ with known μ, compute the Jeffrey's prior for σ^2.*

First, the log-likelihood for y is:

$$
\begin{aligned}
\log(p(y|\sigma^2)) &= \log(\frac{1}{\sqrt{2\pi}\sigma}^n \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2)) \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2
\end{aligned}
\tag{4}
$$

Jeffrey's prior for θ is proportional to:

$$
\begin{aligned}
\left|\mathcal{I}(\theta)\right|^{1/2} &= \left|\mathop{\mathbb{E}}_y[\tfrac{\partial^2}{\partial(\theta\theta^T)^2}\mathcal{L}(\theta)]\right|^{1/2} \\
&= \left|\mathop{\mathbb{E}}_y[\tfrac{\partial^2}{\partial(\theta\theta^T)^2}\log(p(y|\theta))]\right|^{1/2}
\end{aligned}
\tag{5}
$$

Since μ is known in this problem, $\theta = \sigma^2$ is a scalar. This means I can write equation 5 as:

$$
\left|\mathcal{I}(\sigma^2)\right|^{1/2} = \left|\mathop{\mathbb{E}}_y[\tfrac{\partial^2}{\partial(\sigma^2)^2}\log(p(y|\sigma^2))]\right|^{1/2}
\tag{6}
$$

Now I plug the log-likelihood from equation 4 into equation 6 to obtain:

$$\begin{aligned}
\left|\mathcal{I}(\sigma^2)\right|^{1/2} &= \left|\mathop{\mathbb{E}}_y[\tfrac{\partial^2}{\partial(\sigma^2)^2}[-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2]]\right|^{1/2} \\
&= \left|\mathop{\mathbb{E}}_y[\tfrac{\partial}{\partial(\sigma^2)}[-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\mu)^2]]\right|^{1/2} \\
&= \left|\mathop{\mathbb{E}}_y[\frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(y_i-\mu)^2]\right|^{1/2} \\
&= \left|\frac{n}{2}\mathop{\mathbb{E}}_y[\frac{1}{(\sigma^2)^2}] - \frac{1}{(\sigma^2)^3}\mathop{\mathbb{E}}_y[\sum_{i=1}^{n}(y_i-\mu)^2]\right|^{1/2} \\
&= \left|\frac{n}{2(\sigma^2)^2} - \frac{(n-1)\sigma^2}{(\sigma^2)^3}\right|^{1/2} \\
&= \left|\frac{n}{2(\sigma^2)^2} - \frac{(n-1)}{(\sigma^2)^2}\right|^{1/2} \\
&= \left|\frac{n-2n+2}{2(\sigma^2)^2}\right|^{1/2} \\
&= \left|\frac{-(n+1)}{2}\frac{1}{(\sigma^2)^2}\right|^{1/2} \\
&\propto \frac{1}{\sigma^2}
\end{aligned}$$

This will not have a finite integral so this is an improper prior.

6.b

Prompt: *Computet the posterior from part a*

The prior is improper, so I will compute the posterior up to a constant and identify the kernel.

$$\begin{aligned}
p(\sigma^2|y) &= p(y|\sigma^2)\cdot p(\sigma^2) \\
p(\sigma^2|y) &= \frac{1}{2\pi\sigma^2}^{n/2}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2)\cdot p(\sigma^2) \\
p(\sigma^2|y) &\propto \frac{1}{\sigma^2}^{n/2}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2)\cdot\frac{1}{\sigma^2} \\
p(\sigma^2|y) &= \sigma^{-n/2-1}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2)\cdot
\end{aligned}$$

This is an Inverse Gamma distribution with $\alpha = n/2$ and $\beta = \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2)$.

6.c

I ran out of time to show this :(

6.d

I ran out of time to show this :(

6.e

These homeworks take forever!!!