Homework 3

Juliette Franqueville

October 20, 2022

(1) (a) Show that $E(y|\theta) = \xi(\theta) = \partial\psi(\theta)/\partial\theta$

Find the MGF of y:

$$\begin{aligned}
E(e^{ty}) &= \int e^{ty} P(y|\theta) dy \\
&= \int h(y) e^{ty+\theta^T y - \psi(\theta)} dy \\
&= \int h(y) e^{(t+\theta)^T y - \psi(\theta) + \psi(t+\theta) - \psi(\theta)} dy \\
&= e^{\psi(t+\theta)-\psi(\theta)} \int h(y) e^{(t+\theta)^T y - \psi(t+\theta)} dy \\
&= e^{\psi(t+\theta)-\psi(\theta)}
\end{aligned}$$

$$\frac{\partial M_y(t)}{\partial t}|_{t=0} = \psi'(\theta)$$

(1) (b) Show that $E(\xi(\theta)) = y_0/\lambda + c$ and $E(\xi(\theta)|y) = (y_0 + n\bar{y})/(\lambda + n)$ + c

$$\begin{aligned}
E(\psi'(\theta)) &= \int P(\theta)\psi'(\theta) d\theta \\
&= \int h(y_0, \lambda)\exp[\theta^T y_0 - \lambda\psi(\theta)]\psi'(\theta) d\theta \\
&= \int h(y_0, \lambda)\exp[\theta^T y_0 - \lambda\psi(\theta)] \frac{1}{\lambda}[y_0 - (y_0 - \lambda\psi'(\theta))] d\theta \\
&= 1/\lambda \int y_0 h(y_0, \lambda)\exp[\theta^T y_0 - \lambda\psi(\theta)] - [y_0 - \lambda\psi'(\theta)]\exp[\theta^T y_0 - \lambda\psi(\theta)] d\theta \\
&= y_0/\lambda - 1/\lambda \int h(y_0, \lambda)[y_0 - \lambda\psi'(\theta)]\exp[\theta^T y_0 - \lambda\psi(\theta)] d\theta \\
&= y_0/\lambda - 1/\lambda \int h(y_0, \lambda)\frac{\partial}{\partial\theta}\exp[\theta^T y_0 - \lambda\psi(\theta)] d\theta \\
&= y_0/\lambda - 1/\lambda \frac{\partial}{\partial\theta} \int h(y_0, \lambda)\exp[\theta^T y_0 - \lambda\psi(\theta)] d\theta \\
&= y_0/\lambda
\end{aligned}$$

We've shown that for the prior $P(\theta) = h(y_0, \lambda)\exp(\theta^T y_0 - \lambda\psi(\theta))$, $E(\xi(\theta)) = \frac{y_0}{\lambda}$. We know that the posterior for θ has form $P(\theta|y, \lambda, y_0) = h(y_0, \lambda)\exp(\theta^T(y_0 + n\bar{y}) - (\lambda + n)\psi(\theta))$. So as before, $E(\xi(\theta)|y)$ will be the ratio of the two coefficients, $\frac{y_0+n\bar{y}}{\lambda+n}$. Also, I could not find where that c integration constant came from since we are integrating from $-\infty$ to $+\infty$. I spoke to Dr Sarkar about it and he was not sure either, so I did not include it.

(2) Show that binomial and negative binomial distributions belong to exponential families

For the binomial distribution:

$$\begin{aligned}
P(y) &= \binom{n}{y} p^y (1-p)^{n-y} \\
&= \binom{n}{y} \exp\{\log[p^y(1-p)^{n-y}]\} \\
&= \binom{n}{y} \exp\{y\log p + (n-y)\log(1-p)\} \\
&= \binom{n}{y} \exp\left\{y\log\frac{p}{1-p} + n\log(1-p)\right\}
\end{aligned}$$

We have

$$\begin{aligned}
\theta &= \log\frac{p}{1-p} \\
\exp\theta &= \frac{p}{1-p} \\
\exp\theta(1-p) &= p \\
\exp\theta(1-p) &= p \\
\exp\theta - p\exp\theta &= p \\
\exp &= p(1+\exp\theta) \\
p &= \exp\theta/(1+\exp\theta) \\
1-p &= 1-\exp\theta/(1+\exp\theta) \\
&= 1/(1+\exp\theta)
\end{aligned}$$

So:

$$\begin{aligned}
P(y) &= \binom{n}{y} \exp\left\{y\log\frac{p}{1-p} + n\log(1-p)\right\} \\
&= \binom{n}{y} \exp\left\{y\log\frac{p}{1-p} - n\log[1+\exp\theta]\right\}
\end{aligned}$$

So we have $h(y) = \binom{n}{y}$, $\theta = \log\frac{p}{1-p}$ and $\psi(\theta) = n\log[1+\exp\theta]$
For the negative binomial distribution:

$$\begin{aligned}
P(y) &= \binom{y+r-1}{y} (1-p)^r p^y \\
&= \binom{y+r-1}{y} \exp\{\log(1-p)^r p^y\} \\
&= \binom{y+r-1}{y} \exp\{r\log(1-p) + y\log p\}
\end{aligned}$$

We have:

$$\theta = \log p$$
$$p = \exp\theta$$
$$1 - p = 1 - \exp\theta$$

So we have $h(y) = \binom{y+r-1}{y}$, $\theta = \log p$ and $\psi(\theta) = -r\log(1 - \exp\theta)$

(3) Let $y \sim Bin(10, \theta)$. Also, let the observed value of $y = 3$. The prior is a mixture of Betas

(a) Find the posterior

Dropping constants, we have:

$$\begin{aligned}
P(\theta|) &\propto P(y|\theta)P(\theta) \\
&\propto \theta^3(1-\theta)^7\left[\frac{\theta^9(1-\theta)^{19}}{B(10,20)} + \frac{\theta^{19}(1-\theta)^9}{B(20,10)}\right] \\
&\propto \left[\frac{\theta^{12}(1-\theta)^{26}}{B(10,20)} + \frac{\theta^{22}(1-\theta)^{16}}{B(20,10)}\right] \\
&\propto \left[\frac{B(13,27)}{B(13,27)}\frac{\theta^{12}(1-\theta)^{26}}{B(10,20)} + \frac{B(23,17)}{B(23,17)}\frac{\theta^{22}(1-\theta)^{16}}{B(20,10)}\right] \\
&\propto \left[B(13,27)\frac{Beta(13,27)}{B(10,20)} + B(23,\frac{Beta(23,17)}{B(20,10)}\right] \\
&\propto \pi_1 Beta(13,27) + \pi_2 Beta(23,17)
\end{aligned}$$

With $\pi_1 = \frac{\frac{B(13,27)}{B(10,20)}}{\frac{B(13,27)}{B(10,20)} + \frac{B(23,17)}{B(20,10)}}$ and $\pi_2 = \frac{\frac{B(23,17)}{B(20,10)}}{\frac{B(13,27)}{B(10,20)} + \frac{B(23,17)}{B(20,10)}}$ since we need $\pi_1 + \pi_2 = 1$.

(b) Plot the posterior superimposed on the prior

(c) Compute a 90% posterior credible interval for θ

To obtain the prior and posteriors, we sum the pdfs of the relevant betas. To find the 90% posterior credible interval, we find the value of θ corresponding to the location where the area under the pdf curve for the posterior is .05 and .95.

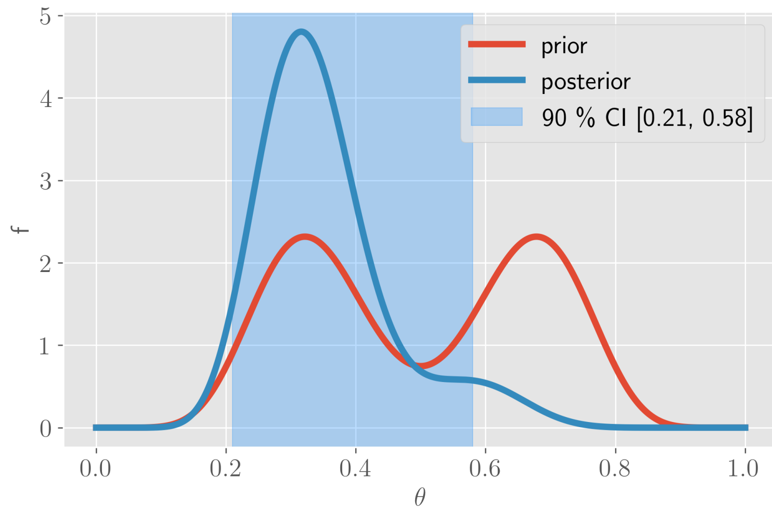

Figure 1: Prior / Posterior and CI

(4) Prove that Jeffreys' priors satisfy the invariance principle: starting with $p(\theta) \propto [\mathbf{det} I(\theta)]^{1/2}$, show that the induced prior on $\psi = g(\theta)$, where g is one-one, is $p(\psi) \propto [\mathbf{det} I(\psi)]^{1/2}$

Using the chain rule:

$$
\begin{aligned}
I(\psi) &= -E\left(\frac{\partial^2 \mathcal{L}(\psi)}{\partial\psi\partial\psi^T}\right) \\
&= -E\left(\frac{\partial^2 \mathcal{L}(\psi)}{\partial\theta\partial\theta^T}\left[\frac{\partial\theta}{\partial\psi}\right]^2 + \frac{\partial \mathcal{L}(\psi)}{\partial\theta}\frac{\partial^2\theta}{\partial\psi\psi\theta^T}\right) \\
&= -E\left(\frac{\partial^2 \mathcal{L}(\psi)}{\partial\theta\partial\theta^T}\right)\left[\frac{\partial\theta}{\partial\psi}\right]^2 - E\left(\frac{\partial \mathcal{L}(\psi)}{\partial\theta}\right)\frac{\partial^2\psi}{\partial\theta\partial\theta^T}
\end{aligned}
$$

$E\left(\frac{\partial \mathcal{L}(\psi)}{\partial\theta}\right)$ is the score and its expectation is 0:

$$
\begin{aligned}
E\left(\frac{\partial \mathcal{L}(\psi)}{\partial\theta}\right) &= \int_{-\infty}^{\infty} f(y|\psi)\frac{\partial \mathcal{L}(\psi)}{\partial\theta}dy \\
&= \int_{-\infty}^{\infty} f(y|\psi)\frac{\partial \log f(y|\psi)}{\partial\theta}dy \\
&= \int_{-\infty}^{\infty} f(y|\psi)\frac{1}{f(y|\psi)}\frac{\partial f(y|\psi)}{\partial\theta}dy \\
&= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(y|\psi)dy \\
&= \frac{\partial}{\partial\theta}(1) = 0
\end{aligned}
$$

So we have:

$$
\begin{aligned}
I(\psi) &= -E\left(\frac{\partial^2 \mathcal{L}(\psi)}{\partial\theta\partial\theta^T}\right)\left[\frac{\partial\theta}{\partial\psi}\right]^2 \\
&= I(\theta)\left[\frac{\partial\theta}{\partial\psi}\right]^2
\end{aligned}
$$

Then,

$$
\begin{aligned}
P(\psi) &\propto P(\theta)\left|\det\frac{\partial\psi}{\partial\theta}\right|^{-1} \\
&\propto \det[I(\theta)]^{1/2}\left|\det\frac{\partial\psi}{\partial\theta}\right|^{-1} \\
&\propto \det\left[I(\psi)\left[\frac{\partial\psi}{\partial\theta}\right]^2\right]^{1/2}\left|\det\frac{\partial\psi}{\partial\theta}\right|^{-1} \\
&\propto \det[I(\psi)]^{1/2}
\end{aligned}
$$

(5) For the Poisson likelihood model $y_1 \dots y_n \sim Poisson(\lambda)$, Jeffrey's (improper) prior was derived in class. Compute the corresponding posterior. Is it proper?

Jeffreys' prior for the Poisson distribution is $\lambda \propto \lambda^{-1/2}$

We have:

$$
\begin{aligned}
P(\lambda|y) &\propto P(y|\lambda)P(\lambda) \\
&\propto \prod \lambda^{y}\exp\{-\lambda\}\lambda^{-1/2} \\
&\propto \lambda^{1/2+\sum y_i - 1}\exp\{-\lambda n\} \\
&= Ga\left(1/2 + \sum y_i, n\right)
\end{aligned}
$$

Since the posterior is a Gamma distribution, it is proper (integrates to 1) as long as $n > 0$.

(6) Consider the likelihood model $y_1, \dots, y_n \sim N(\mu, \sigma)$, μ known. (a) Compute the Jeffreys' prior for σ^2

Note that we already derived the Fisher information matrix for the normal distribution in the previous homework.

$$
\begin{aligned}
P(\sigma^2) &\propto |I(\sigma^2)|^{1/2} \\
&\propto [n/(2\sigma^4)]^{1/2} \\
&\propto 1/\sigma^2
\end{aligned}
$$

(b) Compute also the corresponding posterior

$$
\begin{aligned}
P(\sigma^2|\mu, y) &\propto P(\mu, y|\sigma^2)P(\sigma^2) \\
&\propto \sigma^{-n}\text{exp}\left\{\frac{\sum(x-\mu)^2}{\sigma^2}\right\}\frac{1}{\sigma^2} \\
&\propto \sigma^{-n-2}\text{exp}\left\{\frac{\sum(x-\mu)^2}{\sigma^2}\right\} \\
&\propto \sigma^{2\frac{1}{2}(-n-2)}\text{exp}\left\{\frac{\sum(x-\mu)^2}{\sigma^2}\right\} \\
&= IG\left(n/2, \frac{1}{2}\sum(x-\mu)^2\right)
\end{aligned}
$$

(c) Draw a random sample of size 20 from a Normal(0, 1) distribution. Using these sampled values as data points and assuming the variance to now be unknown, plot the posterior superimposed on the general shape of the prior. (d) Compute a 90% centered quantile based credible interval for σ^2. (e) Compute also a 90% HPD interval for σ^2.

The plot below shows the prior and posterior and the CIs. Note that the HPD interval was found by iteratively looking for the shortest 90% interval.

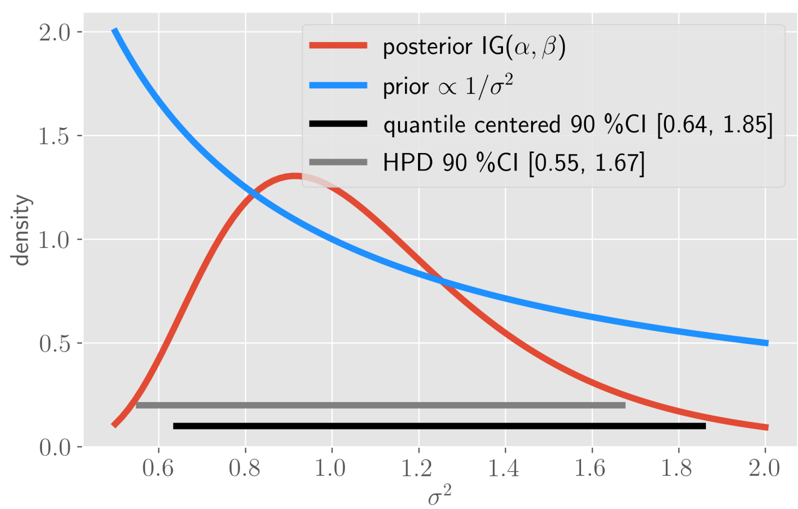

Figure 2: Prior / Posterior and CIs