Homework 5

Frederick J. Boehm

November 3, 2020

Question 1

Q1, part i

We see that

$$
\begin{aligned}
f^*(y) &= \sup_x (xy - |x|) \\
&= 1_{|y|>1}\infty
\end{aligned}
$$

To see that this is the case consider, for $|y| \leq 1$ and $x > 0$:

$$xy - |x| = xy - x \quad \leq 0$$

Additionally, we see that the supremum must be 0, rather than something smaller than zero, since, when $y = 1$,

$$xy - x = x - x = 0$$

Q1, part ii

$$\begin{aligned}
f^*(y) &= \sup_x \left(xy - f(x)\right) \\
&= \sup_x \left(xy - |x|^p\right) \\
&\implies \frac{d}{dx}\left(xy - |x|^p\right) = y - p|x|^{p-1} = 0 \\
&\implies \frac{y}{p} = |x|^{p-1} \\
&\implies \hat{x} = \left(\frac{y}{p}\right)^{\frac{1}{p-1}} \\
&\implies f^*(y) = y\left(\frac{y}{p}\right)^{\frac{1}{p-1}} - \left(\frac{y}{p}\right)^{\frac{p}{p-1}}
\end{aligned}$$

Q1, part iii

$$\begin{aligned}
f^*(y) &= \sup_x \left(xy - f(x)\right) \\
&= \sup_x \left(xy - x\log x\right) \\
&\implies \frac{d}{dx}\left(xy - \log x\right) = y - \log x - 1 = 0 \\
&\implies y = \log x + 1 \\
&\implies \hat{x} = e^{y-1} \\
&\implies f^*(y) = ye^{y-1} - (y-1)e^{y-1} = e^{y-1}
\end{aligned}$$

Question 2

2, part a

This follows from the definition of conjugate function.

See that

$$P(Y \geq \mu > \mathbb{E}Y) \leq \exp \theta Y -$$

2, part b

We start with the relationship $A(\theta) + A^*(\mu) = \theta\mu$, which follows from the definition of conjugate function.

We then see that $\frac{dA}{d\theta} = \mu$ and $\frac{dA^*}{d\mu} = \theta$, leading us to see that $\frac{d^2A}{d\theta^2} = \frac{d\mu}{d\theta}$ and $\frac{d^2A^*}{d\mu^2} = \frac{d\theta}{d\mu}$. Putting these expressions together, after recognizing that $\left(\frac{d\mu}{d\theta}\right)^{-1} = \frac{d\theta}{d\mu}$, we get the result.

2, part c

Nonnegative second derivative We first show that $\frac{d^2A^*}{d\mu^2}(\mu) \geq 0$.

$$\begin{aligned}\frac{d^2A}{d\theta^2} &= \frac{d^2}{d\theta^2}\left(\log \mathbb{E}e^{\theta Y}\right)\\ &= \frac{\mathbb{E}(Y^2 e^{Y\theta})}{\mathbb{E}e^{Y\theta}} - \frac{[\mathbb{E}(Ye^{Y\theta})]^2}{[\mathbb{E}(e^{Y\theta})]^2} \geq 0\end{aligned}$$

We get the remaining results from the definition of the cumulant generating function and the results in part b.

Question 3

Q3, part i

we first calculate the marginal density of the x's.

$$\begin{aligned}p(x_1,\ldots,x_n) &= \int p(x_1,\ldots,x_n)p(\theta)d\theta\\ &= \int \theta^{a-1+\sum_{i=1}^n x_i} e^{-n\theta - b\theta}\left(\frac{b^a}{\Gamma(b)}\right)\left(\frac{1}{\prod_{i=1}^n} x_i!\right)d\theta\\ &= \left(\frac{b^a}{\Gamma(b)}\right)\left(\frac{1}{\prod_{i=1}^n x_i!}\right)\int \theta^{a-1+\sum_{i=1}^n x_i} e^{-\theta(n+b)}\left(\frac{(n+b)^{a+\sum x_i}}{(n+b)^{a+\sum x_i}}\right)\left(\frac{\Gamma(n+b)}{\Gamma(n+b)}\right)d\theta\\ &= \left(\frac{b^a}{\Gamma(b)}\right)\left(\frac{\Gamma(n+b)}{(n+b)^{a+\sum x_i}}\right)\left(\frac{1}{\prod x_i!}\right)\end{aligned}$$

Note that the last equality follows from the observation that the integrand is a multiple of a gamma density.

We then calculate the posterior density.

$$
\begin{aligned}
p(\theta|x_1,\dots,x_n) &= \frac{p(x_1,\dots,x_n|\theta)p(\theta)}{p(x_1,\dots,x_n)} \\
&= \left(\frac{\theta^{\sum x_i} e^{-n\theta}}{\prod x_i!}\right)\left(\frac{b^a\theta^{a-1}e^{-b\theta}}{\Gamma(b)}\right)\left(\frac{\Gamma(b)}{b^a}\right)\left(\frac{(n+b)^{a+\sum x_i}}{\Gamma(n+b)}\right)\left(\prod x_i!\right) \\
&= \frac{\theta^{a-1+\sum x_i} e^{-\theta(n+b)}(n+b)^{a+\sum x_i}}{\Gamma(n+b)}
\end{aligned}
$$

Q3, part ii

We let T denote the sufficient statistic and θ the canonical parameter.

We then have

$$p(x|\theta) = h(x)\exp\left(\theta T - A(\theta)\right)$$

and

$$p(\theta|\eta,\zeta) = \exp\left(\eta\theta - \zeta A(\theta) - B(\eta,\zeta)\right)$$

We compute the posterior distribution:

$$
\begin{aligned}
p(\theta|x,\eta,\zeta) &= \frac{p(\theta|\eta,\zeta)p(x|\theta)}{\int p(\theta|\eta,\zeta)p(x|\theta)} \\
&= \frac{h(x)\exp\left(\theta T + \eta T - \zeta A(\theta) - B(\eta,\zeta) - A(\theta)\right)}{\int p(\theta|\eta,\zeta)p(x|\theta)} \\
&= \frac{\exp\left((\theta+\eta)T - (\zeta+1)A(\theta) - B(\eta,\zeta)\right)}{\int \exp\left((\theta+\eta)T - (\zeta+1)A(\theta) - B(\eta,\zeta)\right)} \\
&= \frac{\exp\left((\theta+\eta)T - (\zeta+1)A(\theta)\right)}{\int \exp\left((\theta+\eta)T - (\zeta+1)A(\theta)\right)}
\end{aligned}
$$

We then observe that the posterior has a form similar to that of the prior.

Q3, part iii

Question 4

Q4, part i

To get the first equality, we interchange the order of integration and differentiation in the definition of A_θ before applying the (differentiation) chain rule.

In symbols, we write:

$$\begin{aligned}
(A_\theta)_{kl} &= \frac{\partial}{\partial \theta_k} \int f_l p d\nu \\
&= \int \frac{\partial}{\partial \theta_k} (f_l p) d\nu \\
&= \int f_l \frac{\partial p}{\partial \theta_k} d\nu + \int \frac{\partial f_l}{\partial \theta_k} p d\nu \\
&= \int f_l \frac{\partial p}{\partial \theta_k} d\nu \\
&= \int f_l \frac{\partial p}{\partial \theta_k} p(\frac{1}{p}) d\nu \\
&= \int f_l \frac{\partial \log p}{\partial \theta_k} p d\nu \\
&= \int f_l W_k p d\nu \\
&= \mathbb{E}\left(f_l W_k\right)
\end{aligned}$$

Q4, part ii

We note that $I_\theta = \mathbb{E}\left([W - \mathbb{E}W]^2\right)$ and $V(f_l) = \mathbb{E}\left([f - \mathbb{E}f]^2\right)$ and $|A_\theta|^2_{kl} = \left(\mathbb{E}\left[(f - \mathbb{E}f)(W - \mathbb{E}W)\right]\right)^2$. Assembling these together and applying Cauchy-Schwarz Theorem, we see that

$$|(A_\theta)_{kl}|^2 \leq \mathbb{V} f_l (I_\theta)_{kk}$$

Q4, part iii

First we establish the equality

$$A_\theta = \mathbb{E}\left([f - \mathbb{E}f][T - \mathbb{E}T]\right)$$

where T is the sufficient statistic and we're working in an exponential family with natural parameter η and cgf B.

We start at Equation 2.166:

$$
\begin{aligned}
A_\theta &= \mathbb{E}\left([f - \mathbb{E}f][W - \mathbb{E}W]\right) \\
&= \mathbb{E}\left([f - \mathbb{E}f][(T - \frac{\partial B}{\partial \eta}) - \mathbb{E}(T - \frac{\partial B}{\partial \eta})]\right) \\
&= \mathbb{E}\left([f - \mathbb{E}f][T - \frac{\partial B}{\partial \eta}]\right) \\
&= \mathbb{E}\left([f - \mathbb{E}f][T - \mathbb{E}T]\right)
\end{aligned}
$$

We now justify the above equalities. First, observe that $W = T - \frac{\partial B}{\partial \eta}$. This follows from definition of our exponential family.

Second, observe that $\mathbb{E}T = \frac{\partial B}{\partial \eta}$. This means that $\mathbb{E}(T - \frac{\partial B}{\partial \eta})$ is zero in the above equalities. Finally, we apply this relationship to get the third line equaling the fourth line.

We then argue that $f = T$ implies that $A_\theta = I_\theta$, the fisher information. This follows from the definition of Fisher information.

Q4, part iv

First, we calculate the expected value of $\widehat{(A_\theta)_{kl}}$. Second, we verify the assumptions of the strong law of large numbers before reaching the conclusion.

Calculating the expected value of $\widehat{(A_\theta)_{kl}}$ We leverage the iid nature of the random variables throughout our argument. Superscript bars denote sample means.

$$
\begin{aligned}
\mathbb{E}\left([f_l(X_i) - f_l(\bar{X})][T(X_i) - T(\bar{X})]\right) &= \mathbb{E}\left(f_l(X_1)T(X_1) - f_l(\bar{X})T(X_1) - f_l(X_1)T(\bar{X}) + f_l(\bar{X})T(\bar{X})\right) \\
&= \mathbb{E}(f_l(X_1)T(X_1))\left(1 - \frac{1}{n}\right) - \mathbb{E}\left(f_l(X_1)T(X_2)\right)\left(\frac{n-1}{n}\right) \\
&= \frac{n-1}{n}(A_\theta)_{kl}
\end{aligned}
$$

Since we can apply Slutsky's theorem, we have, by the strong law of large numbers, the desired result. That is, $\frac{n-1}{n} \to 1$ as $n \to \infty$.

Verifying assumptions of SLLN We verify the assumptions of the strong law of large numbers.

We've already assumed finite second moments, and that is a sufficient condition for SLLN application, so we can use the Kolmogorov's SLLN without problems.

Q4, part v