STA 602. HW06

Yicheng Shen

10/10/2022

1. 3.12

(a) The binomial sampling model is

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

Then we derive the Jeffery's prior by starting with Fisher's information

$$\begin{aligned}
\text{By Definition } I(\theta) &= -\mathrm{E}(\frac{\partial^2 \ell(y|\theta)}{\partial\theta^2}) \\
\ell(y|\theta) &= \log p(y \mid \theta) \\
&= \log\left[\binom{n}{y}\theta^y(1-\theta)^{n-y}\right] \\
\text{First term is constant } &= \log\left(\binom{n}{y}\right) + y\log(\theta) + (n-y)\log(1-\theta) \\
\text{Take 1st derivative } \frac{\partial\ell(y|\theta)}{\partial\theta} &= \frac{y}{\theta} - \frac{n-y}{1-\theta} \\
\text{Take 2nd derivative } \frac{\partial^2\ell(y|\theta)}{\partial\theta^2} &= -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2} \\
I(\theta) &= -\mathrm{E}\left(-\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}\right) \\
&= -\left(-\frac{1}{\theta^2}\mathrm{E}(y) - \frac{1}{(1-\theta)^2}\mathrm{E}(n-y)\right) \\
&= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} \qquad \text{Since } \mathrm{E}(y) = n\theta \\
&= \frac{n}{\theta} + \frac{n}{1-\theta} \\
&= \frac{n-n\theta+n\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}
\end{aligned}$$

Now we can formulate the Jeffreys' prior as

$$\begin{aligned}
p_J(\theta) \propto I(\theta)^{1/2} &= \sqrt{\frac{n}{\theta(1-\theta)}} \\
&\sim Beta(0.5, 0.5)
\end{aligned}$$

(b) We again start by looking at the likelihood for Fisher's information

$$\begin{aligned}
\ell(y \mid \psi) &= \log p(y \mid \psi) \\
&= \log\left[\binom{n}{y} e^{\psi y}(1+e^{\psi})^{-n}\right] \\
&= \log\binom{n}{y} + \psi y - n\log\left(1+e^{\psi}\right) \\
\text{Take 1st derivative } \frac{\partial \ell(y|\psi)}{\partial \psi} &= y - \frac{ne^{\psi}}{e^{\psi}+1} \\
\text{Take 2nd derivative } \frac{\partial^2 \ell(y|\psi)}{\partial \psi^2} &= -\frac{ne^{\psi}}{\left(e^{\psi}+1\right)^2} \\
I(\psi) &= -\mathrm{E}\left[-\frac{ne^{\psi}}{\left(e^{\psi}+1\right)^2}\right] \\
\text{Since there is no y } &= \frac{ne^{\psi}}{\left(e^{\psi}+1\right)^2} \\
p_J(\psi) &\propto \sqrt{\frac{ne^{\psi}}{\left(e^{\psi}+1\right)^2}} = \frac{\sqrt{ne^{\psi}}}{e^{\psi}+1}
\end{aligned}$$

(c) Let $\psi = g(\theta) = \log\frac{\theta}{1-\theta}$ and $\theta = h(\psi) = \frac{e^{\psi}}{1+e^{\psi}}$. So we know that $\left|\frac{dh}{d\psi}\right| = \frac{e^{\psi}}{(e^{\psi}+1)^2}$

$$\begin{aligned}
p_{\psi}(\psi) &\propto p_{\theta}(h(\psi)) \times \left|\frac{dh}{d\psi}\right| \\
\text{Because } p_J(\theta) \propto \sqrt{\frac{n}{\theta(1-\theta)}} &\text{ and } \theta = \frac{e^{\psi}}{1+e^{\psi}} \\
p_{\psi}(\psi) &\propto \sqrt{\frac{n}{\frac{e^{\psi}}{1+e^{\psi}}\left(1-\frac{e^{\psi}}{1+e^{\psi}}\right)}} \times \frac{e^{\psi}}{(e^{\psi}+1)^2} \\
&\propto \sqrt{\frac{n}{\frac{e^{\psi}}{1+e^{\psi}}\frac{1}{1+e^{\psi}}}} \times \frac{e^{\psi}}{(e^{\psi}+1)^2} \\
&\propto \sqrt{\frac{n(e^{\psi}+1)^2}{e^{\psi}}} \times \frac{e^{\psi}}{(e^{\psi}+1)^2} \\
&\propto \frac{\sqrt{n}}{\sqrt{e^{\psi}}} \times (e^{\psi}+1) \times \frac{e^{\psi}}{(e^{\psi}+1)^2} \\
&\propto \frac{\sqrt{ne^{\psi}}}{e^{\psi}+1}
\end{aligned}$$

It is thus shown that the Jeffreys' prior is invariant with change of variables.

2. 3.13

(a) The Poisson density is specified as $p(y) = \frac{\theta^y e^{-\theta}}{y!}$.

$$
\begin{aligned}
\ell(y \mid \theta) &= \log p(y \mid \theta) \\
&= \log\left(\frac{\theta^y e^{-\theta}}{y!}\right) \\
&= -\log(y!) + y\log(\theta) - \theta \\
\text{Take 1st derivative } \frac{\partial \ell(y|\theta)}{\partial\theta} &= \frac{y}{\theta} - 1 \\
\text{Take 2nd derivative } \frac{\partial^2 \ell(y|\theta)}{\partial\theta^2} &= -\frac{y}{\theta^2} \\
p_J(\theta) &\propto \sqrt{\frac{1}{\theta}}
\end{aligned}
$$

By a closer look, it is noticed that $\int_0^\infty \frac{1}{\sqrt{\theta}}\, d\theta$ diverge and $p_J(\theta)$ cannot be proportional to an actual probability density for $\theta \in (0, \infty)$. This makes the above prior an improper prior.

(b) Now we are looking at the joint probability of θ, y.

$$
\begin{aligned}
f(\theta, y) &= \sqrt{I(\theta)} \times p(y \mid \theta) \\
&= \sqrt{\frac{1}{\theta}} \times \frac{\theta^y e^{-\theta}}{y!} \\
&= \theta^{-\frac{1}{2}} \theta^y \frac{e^{-\theta}}{\Gamma(y+1)} \\
&= \frac{\theta^{y-\frac{1}{2}} e^{-\theta}}{\Gamma(y+1)} \\
\text{y comes from data and is constant } &\propto \theta^{y-\frac{1}{2}} e^{-\theta} \\
&\sim \text{Gamma}(y + \frac{1}{2}, 1) \quad \text{ for } y \geq 0
\end{aligned}
$$

Now $\int f(\theta, y) d\theta$ could serve as the normalizing constant that makes sure this posterior distribution of θ is Gamma density and thus proper.

3. 3.14

(a) First we obtain the MLE:

$$
\begin{aligned}
\sum_{i=1}^{n} \log p(y_i|\theta) &= \sum_{i=1}^{n} \log(\theta^{y_i}(1-\theta)^{1-y_i}) \\
&= (\sum_{i=1}^{n} y_i)\log(\theta) + (\sum_{i=1}^{n} 1-y_i)\log(1-\theta) \\
\text{Take 1st Derivative as } 0 &= \frac{\sum_{i=1}^{n} y_i}{\hat{\theta}} - \frac{\sum_{i=1}^{n}(1-y_i)}{1-\hat{\theta}} \\
&= \frac{\sum_{i=1}^{n} y_i - \hat{\theta}\sum_{i=1}^{n} y_i - \hat{\theta}\sum_{i=1}^{n}(1-y_i)}{\theta(1-\hat{\theta})} \\
n\hat{\theta} &= \sum_{i=1}^{n} y_i \\
\text{MLE } \hat{\theta} &= \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} \\
\text{Check 2nd Derivative } &= -\frac{\sum_{i=1}^{n} y_i}{\theta^2} - \frac{\sum_{i=1}^{n}(1-y_i)}{(1-\theta)^2} < 0
\end{aligned}
$$

Then we get

$$
\begin{aligned}
J(\theta) &= -\frac{\partial^2 \ell(y|\theta)}{\partial\theta^2} \\
&= -[-\frac{\sum_{i=1}^{n} y_i}{\theta^2} - \frac{\sum_{i=1}^{n}(1-y_i)}{(1-\theta)^2}] \\
J(\hat{\theta})/n &= [\frac{\sum_{i=1}^{n} y_i}{\hat{\theta}^2} + \frac{\sum_{i=1}^{n}(1-y_i)}{(1-\hat{\theta})^2}]/n \\
&= \frac{\sum_{i=1}^{n} y_i}{\hat{\theta}^2 n} + \frac{\sum_{i=1}^{n}(1-y_i)}{(1-\hat{\theta})^2 n} \\
&= \frac{1}{\hat{\theta}^2}\bar{y} + \frac{1}{(1-\hat{\theta})^2}(1-\bar{y}) \\
\text{Bacuase } \hat{\theta} &= \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} \\
J(\hat{\theta})/n &= \frac{1}{\bar{y}} + \frac{1}{1-\bar{y}}
\end{aligned}
$$

(b) First we know for probability density, it has to be true that $\int p_U(\theta)d\theta = 1$.

$$\begin{aligned}
\log p_U(\theta) &= \ell(\theta|\boldsymbol{y})/n + c \\
p_U(\theta) &= e^{\ell(\theta|\boldsymbol{y})/n+c} \\
\int_0^1 p_U(\theta)d\theta &= 1 \\
\int_0^1 e^{\ell(\theta|\boldsymbol{y})/n+c}d\theta &= 1 \\
\int_0^1 e^{(\sum_{i=1}^n y_i)\log(\theta)/n+(\sum_{i=1}^n 1-y_i)\log(1-\theta)/n} \times e^c d\theta &= 1 \\
\int_0^1 \theta^{\frac{\sum_{i=1}^n y_i}{n}}(1-\theta)^{1-\frac{\sum_{i=1}^n y_i}{n}} \times e^c d\theta &= 1 \\
\int_0^1 \theta^{\bar{y}}(1-\theta)^{1-\bar{y}} \times e^c d\theta &= 1 \\
\text{Remove the constant and recognize the kernel of Beta} \qquad p_U(\theta) &\sim Beta(\bar{y}+1, 2-\bar{y})
\end{aligned}$$

Then compute the information

$$\begin{aligned}
\log p_U(\theta) &= \ell(\theta|\boldsymbol{y})/n + c \\
&= \frac{\sum_{i=1}^n \log(\theta^{y_i}(1-\theta)^{1-y_i})}{n} + c \\
&= (\sum_{i=1}^n y_i)\log(\theta)/n + (\sum_{i=1}^n 1-y_i)\log(1-\theta)/n + c \\
\partial \log p_U(\theta)/\partial\theta &= \frac{\sum_{i=1}^n y_i}{\theta n} - \frac{\sum_{i=1}^n(1-y_i)}{(1-\theta)n} \\
-\partial^2 \log p_U(\theta)/\partial\theta^2 &= \frac{\sum_{i=1}^n y_i}{\theta^2 n} + \frac{\sum_{i=1}^n(1-y_i)}{(1-\theta)^2 n} \\
&= \frac{\bar{y}}{\theta^2} + \frac{1-\bar{y}}{(1-\theta)^2}
\end{aligned}$$

(c) The posterior is a Beta distribution as the following

$$\begin{aligned}
p(\theta|\boldsymbol{y}) &\propto p_U(\theta) \times p(y_1, y_2, \dots y_n|\theta) \\
&\propto \theta^{\bar{y}}(1-\theta)^{1-\bar{y}} \times \prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i} \\
&\propto \theta^{\bar{y}}(1-\theta)^{1-\bar{y}} \times \theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} \\
&\propto \theta^{\bar{y}+n\bar{y}}(1-\theta)^{1-\bar{y}+n(1-\bar{y})} = \theta^{(n+1)\bar{y}}(1-\theta)^{(n+1)(1-\bar{y})} \\
\text{Recognize kernel of Beta} \qquad &\sim Beta((n+1)\bar{y}+1, (n+1)(1-\bar{y})+1)
\end{aligned}$$

(d) First we obtain the MLE and $J(\hat{\theta})/n$:

$$\begin{aligned}
p(y|\theta) &= \frac{\theta^y e^{-\theta}}{y!} \\
\sum_{i=1}^n \log p(y_i|\theta) &= -\log \sum_{i=1}^n y_i! + \log(\theta) \times \sum_{i=1}^n y_i - n\theta \\
\text{Take 1st Derivative as } 0 &= \frac{\sum_{i=1}^n y_i}{\hat{\theta}} - n \\
n &= \frac{\sum_{i=1}^n y_i}{\hat{\theta}} \\
\text{MLE } \hat{\theta} &= \frac{\sum_{i=1}^n y_i}{n} = \bar{y} \\
\text{Check 2nd Derivative } &= -\frac{\sum_{i=1}^n y_i}{\theta^2} < 0 \\
J(\hat{\theta})/n &= -[-\frac{\sum_{i=1}^n y_i}{\hat{\theta}^2}]/n \\
&= \frac{\sum_{i=1}^n y_i}{\hat{\theta}^2 n} \\
&= \frac{1}{\bar{y}}
\end{aligned}$$

Then find $p_U(\theta)$:

$$\begin{aligned}
1 &= \int_0^1 p_U(\theta) d\theta \\
&= \int_0^1 e^{\ell(\theta|\boldsymbol{y})/n+c} d\theta \\
&= \int_0^1 e^{[-\log \sum_{i=1}^n y_i! + \log(\theta) \times \sum_{i=1}^n y_i - n\theta]/n} \times e^c d\theta \\
&= \int_0^1 e^{\log(\theta) \times (\sum_{i=1}^n y_i)/n - \theta} \times e^c d\theta \\
&= \int_0^1 \theta^{\bar{y}} \times e^{-\theta} \times e^c d\theta \\
\text{Recognize kernel of gamma} \quad & p_U(\theta) \sim Gamma(\bar{y}+1, 1)
\end{aligned}$$

Then compute the information

$$\begin{aligned}
\log p_U(\theta) &= \ell(\theta|\boldsymbol{y})/n + c \\
&= \frac{-\log \sum_{i=1}^n y_i! + \log(\theta) \times \sum_{i=1}^n y_i - n\theta}{n} + c \\
\partial \log p_U(\theta)/\partial\theta &= \frac{\bar{y}}{\theta} - 1 \\
-\partial^2 \log p_U(\theta)/\partial\theta^2 &= -\frac{\bar{y}}{\theta^2}
\end{aligned}$$

The posterior is a Beta distribution as the following

$$
\begin{aligned}
p(\theta|\boldsymbol{y}) &\propto p_U(\theta) \times p(y_1, y_2, \dots y_n|\theta) \\
&\propto \theta^{\bar{y}} \times e^{-\theta} \times \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} \\
&\propto \theta^{\bar{y}} \times e^{-\theta} \times \theta^{\sum y_i} \times e^{-n\theta} \\
&\propto \theta^{\bar{y}+\sum y_i} \times e^{-(n+1)\theta} \\
\text{Recognize kernel of Gamma} \qquad &\sim Gamma(1 + \bar{y} + \sum_{i=1}^{n} y_i, n+1)
\end{aligned}
$$

4. 4.7