# Stats 579 – Intermediate Bayesian Modeling

## Assignment # 3 Solutions

1. An hypothetical study considers the lifespan of fluorescent light bulbs. Let $y_1, ..., y_n$ be the duration (in years) it takes for each of $n$ light bulbs to fail. Assume that all tests are performed under laboratory conditions and observations are *iid*. Researchers are interested in whether bulb lifespan is better modeled with model $M_1$, an $\text{Exp}(\lambda)$ distribution, or with model $M_2$, a $\text{Weibull}(3, \lambda)$ distribution. For both models, the researchers assume $p(\lambda) = e^{-\lambda}$.

   For this problem, please use the Exponential and Weibull parameterizations from your textbook, which give

$$\begin{aligned} y_E &\sim \text{Exp}(\lambda) \\ f(y_E \mid \lambda) &= \lambda \exp(-\lambda y) I_{(0,\infty)}(y) \\ y_W &\sim \text{Weibull}(\alpha, \lambda) \\ f(y_W \mid \alpha, \lambda) &= \lambda \alpha y^{\alpha-1} \exp(-\lambda y^{\alpha}) I_{(0,\infty)}(y) \end{aligned}$$

   (a) Obtain the marginal density for these data under each model. (HINT: Take advantage of conjugacy.)

   The marginal density under the exponential model is given by

$$\begin{aligned} f(y_1, ..., y_n \mid M_1) &= \int \left[\prod_{i=1}^{n} f(y_E \mid \lambda)\right] p(\lambda) d\lambda \\ &= \int \left[\prod_{i=1}^{n} \lambda \exp(-\lambda y_i) I_{(0,\infty)}(y_i)\right] \exp(-\lambda) d\lambda \\ &= \left[\int \lambda^n \exp\left(-\lambda\left[1 + \sum_{i=1}^{n} y_i\right]\right) d\lambda\right] \prod_{i=1}^{n} I_{(0,\infty)}(y_i) \\ &= \frac{\Gamma(n+1)}{\left(1 + \sum_{i=1}^{n} y_i\right)^{n+1}} \prod_{i=1}^{n} I_{(0,\infty)}(y_i) \\ &\quad \times \int \frac{\left(1 + \sum_{i=1}^{n} y_i\right)^{n+1}}{\Gamma(n+1)} \lambda^n \exp\left(-\lambda\left[1 + \sum_{i=1}^{n} y_i\right]\right) d\lambda \\ &= \frac{\Gamma(n+1)}{\left(1 + \sum_{i=1}^{n} y_i\right)^{n+1}} \prod_{i=1}^{n} I_{(0,\infty)}(y_i) \end{aligned}$$

The marginal density under the Weibull model is given by

$$\begin{aligned}
f(y_1,...,y_n \mid M_2) &= \int \left[\prod_{i=1}^{n} f\left(y_W \mid \lambda\right)\right] p(\lambda) d\lambda \\
&= \int \left[\prod_{i=1}^{n} 3\lambda y_i^2 \exp\left(-\lambda y_i^3\right) I_{(0,\infty)}(y_i)\right] \exp(-\lambda) d\lambda \\
&= 3^n \left[\int \lambda^n \exp\left(-\lambda\left[1+\sum_{i=1}^{n} y_i^3\right]\right) d\lambda\right] \prod_{i=1}^{n} y_i^2 I_{(0,\infty)}(y_i) \\
&= 3^n \frac{\Gamma(n+1)}{\left(1+\sum_{i=1}^{n} y_i^3\right)^{n+1}} \prod_{i=1}^{n} y_i^2 I_{(0,\infty)}(y_i) \\
&\quad \times \int \frac{\left(1+\sum_{i=1}^{n} y_i^3\right)^{n+1}}{\Gamma(n+1)} \lambda^n \exp\left(-\lambda\left[1+\sum_{i=1}^{n} y_i^3\right]\right) d\lambda \\
&= \frac{3^n\Gamma(n+1)}{\left(1+\sum_{i=1}^{n} y_i^3\right)^{n+1}} \prod_{i=1}^{n} y_i^2 I_{(0,\infty)}(y_i)
\end{aligned}$$

(b) Obtain an expression for the Bayes factor comparing $M_1$ to $M_2$.

This is straightforward from part (a). The Bayes factor comparing $M_1$ to $M_2$ is given by

$$\begin{aligned}
BF_{(M_1:\,M_2)} &= \frac{f(y_1,...,y_n \mid M_1)}{f(y_1,...,y_n \mid M_2)} \\
&= \frac{\frac{\Gamma(n+1)}{\left(1+\sum_{i=1}^{n} y_i\right)^{n+1}} \prod_{i=1}^{n} I_{(0,\infty)}(y_i)}{\frac{3^n\Gamma(n+1)}{\left(1+\sum_{i=1}^{n} y_i^3\right)^{n+1}} \prod_{i=1}^{n} y_i^2 I_{(0,\infty)}(y_i)} \\
&= \frac{\left(1+\sum_{i=1}^{n} y_i^3\right)^{n+1}}{3^n\left(1+\sum_{i=1}^{n} y_i\right)^{n+1} \prod_{i=1}^{n} y_i^2}
\end{aligned}$$

Note that this quantity is defined only for $y_1,...,y_n$ satisfying $\prod_{i=1}^{n} I_{(0,\infty)}(y_i) > 0$, otherwise the Bayes factor is undefined.

(c) Evaluate the Bayes factor when the data are: $\{8.05, 6.56, 3.20, 6.85, 5.67\}$.

Using R to evaluate the value, I obtain 0.5885679.

(d) Explain which model seems preferable based on the Bayes factor. Explain which model would be preferable if you had a prior belief that $M_1$ were nine times more likely to be correct than $M_2$.

According to the Bayes factor, these data appear more likely to have arisen from $M_2$ than from $M_1$, but not extraordinarily more likely to have done so. If our prior odds were $^9/_1$, then the posterior odds for $M_1$ over $M_2$ would be about 5.297, still strongly in favor of the first (exponential) model.

2. Using the same set-up as in Problem 1, the researchers want to compare these models in terms of AIC.

   (a) Find the MLE for $\lambda$ under $M_1$ and $M_2$. Note that although both models have a parameter named $\lambda$, these parameters are not the same and may maximize at different values for each model. It may be helpful to write the MLEs as $\hat{\lambda}_1$ and $\hat{\lambda}_2$ to help distinguish them.

For $M_1$, we have the likelihood

$$L_1(\lambda \mid y_1, ..., y_n) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} y_i\right),$$

and the log-likelihood

$$l_1(\lambda \mid y_1, ..., y_n) = n \log \lambda - \lambda \sum_{i=1}^{n} y_i.$$

Taking the derivative gives us

$$l_1'(\lambda \mid y_1, ..., y_n) = \frac{n}{\lambda} - \sum_{i=1}^{n} y_i,$$

and the second derivative yields

$$l_1^{(2)}(\lambda \mid y_1, ..., y_n) = -\frac{n}{\lambda^2}.$$

Since the second derivative is non-negative, any root of the first derivative must be a maximum. The first-derivative maximizes at

$$\hat{\lambda}_1 = \frac{n}{\sum_{i=1}^{n} y_i},$$

which is thus the maximum likelihood estimate for $\lambda$ under $M_1$.

Under $M_2$, our likelihood is

$$L_2(\lambda \mid y_1, ..., y_n) = \prod_{i=1}^{n} 3^n \lambda^n \left(\prod_{i=1}^{n} y_i^2\right) \exp\left(-\lambda \sum_{i=1}^{n} y_i^3\right),$$

with log-likelihood

$$l_2(\lambda \mid y_1, ..., y_n) = n \log(3\lambda) + 2 \sum_{i=1}^{n} \log y_i - \lambda \sum_{i=1}^{n} y_i^3.$$

The derivative for this log-likelihood is

$$l_2'(\lambda \mid y_1, ..., y_n) = \frac{n}{\lambda} - \sum_{i=1}^{n} y_i^3,$$

and the second derivative again yields

$$l_2^{(2)}(\lambda \mid y_1, ..., y_n) = -\frac{n}{\lambda^2}.$$

Since the second derivative is again non-negative, any root of the first derivative must be a maximum. The first-derivative maximizes at

$$\hat{\lambda}_2 = \frac{n}{\sum_{i=1}^{n} y_i^3},$$

which is the maximum likelihood estimate for $\lambda$ under $M_2$.

(b) Calculate AIC for each model using the data provided above. Which model seems preferable based on AIC?

In our scenario, the AIC for model $j$ is given by the equation

$$AIC_j = -2 \sum_{i=1}^{n} \log f_j \left( y_i \mid \hat{\lambda}_j \right) + 2k_j,$$

where $f_j$ specifies the density function for model $j$ and $k_j$ is the number of parameters in the model. Note that $k_j$ is one for both $M_1$ and $M_2$ since $\lambda$ is the only unknown parameter in both models.

Using the data from Question 1, we find that the MLE for $M_1$ is 0.165 and for $M_2$ it is 0.00373. Plugging these into the model equations, we find

$$\begin{aligned}
AIC_1 &= -2 \sum_{i=1}^{5} \left( \log \hat{\lambda}_1 - \hat{\lambda}_1 y_i \right) + 2 \\
&= -10 \log \hat{\lambda}_1 + 2\hat{\lambda}_1 \sum_{i=1}^{5} y_i + 2 \\
&= -10 \log \hat{\lambda}_1 + 12 \\
&= 30.027 \\
AIC_2 &= -2 \sum_{i=1}^{5} \left( \log \left( 3\hat{\lambda}_2 \right) + 2 \log y_i - \hat{\lambda}_2 y_i^3 \right) + 2 \\
&= -10 \log \left( 3\hat{\lambda}_2 \right) - 4 \sum_{i=1}^{n} \log y_i + 2\hat{\lambda}_2 \sum_{i=1}^{n} y_i^3 + 2 \\
&= -10 \log \left( 3\hat{\lambda}_2 \right) - 4 \sum_{i=1}^{n} \log y_i + 12 \\
&= 21.770
\end{aligned}$$

Thus by AIC, we also find $M_2$ preferable to $M_1$.

3. Let $y_1, ..., y_n$ be independent conditional on some model parameter $\theta$. Let $y_i \sim f(y_i \mid \theta)$ and let $\theta$ have prior $p(\theta)$. Consider the conditional predictive ordinate for an observation $y_j$,

$$CPO_j = f\left(y_j \mid y_{(j)}\right),$$

where $y_{(j)}$ denotes the set $\{y_1, ..., y_{j-1}, y_{j+1}, ..., y_n\}$.

(a) Show that

$$CPO_j = \frac{\int \prod_{i=1}^n f(y_i \mid \theta)p(\theta)}{\int \prod_{i \neq j} f(y_i \mid \theta)p(\theta)}.$$

$$\begin{aligned}
CPO_j &= f\left(y_j \mid y_{(j)}\right) \\
&= \int f(y_j \mid \theta)p(\theta \mid y_{(j)})d\theta \\
&= \int f(y_j \mid \theta)\left(\frac{\prod_{i \neq j} f(y_i \mid \theta)p(\theta)}{f(y_{(j)})}\right) d\theta \\
&= \frac{1}{f(y_{(j)})}\int \prod_{i=1}^{n} f(y_i \mid \theta)p(\theta)d\theta \\
&= \frac{\int \prod_{i=1}^n f(y_i \mid \theta)p(\theta)d\theta}{\int \prod_{i \neq j} f(y_i \mid \theta)p(\theta)d\theta}
\end{aligned}$$

(b) Now show that

$$CPO_j^{-1} = \int \left[\frac{1}{f(y_j \mid \theta)}\right] p(\theta \mid y)d\theta,$$

where $y$ denotes the entire collection of data.

$$\begin{aligned}
CPO_j^{-1} &= \frac{1}{f\left(y_j \mid y_{(j)}\right)} \\
&= \frac{\int \prod_{i \neq j} f(y_i \mid \theta)p(\theta)d\theta}{\int \prod_{i=1}^n f(y_i \mid \theta)p(\theta)d\theta} \\
&= \frac{\int \prod_{i \neq j} f(y_i \mid \theta)p(\theta)d\theta}{f(y)} \\
&= \int \frac{1}{f(y_j \mid \theta)}\left(\frac{\prod_{i=1}^n f(y_i \mid \theta)p(\theta)}{f(y)}\right) d\theta \\
&= \int \frac{1}{f(y_j \mid \theta)} p(\theta \mid y)d\theta
\end{aligned}$$

Thus the inverse CPO, $CPO_j^{-1}$, is equal to the posterior expectation of the inverse of the density for $y_j$.