Name: ____________________

Intermediate Bayesian Modeling
Midterm Exam

Instructions: Attempt as many parts of as many problems as possible. Show enough work to convince me that you know what you are doing. Write your answer to each problem on the page for that problem—you may use the back of the page if necessary. Do well!

1. (40 pts.) Let $y = \{y_1, ..., y_n\}$ be a collection of exchangeable random variables, and consider the conditional predictive ordinate for an observation y_j,

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

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

$$CPO_j^{-1} = \mathrm{E}_{\theta|y}\left[\frac{1}{g_j(y_j \mid \theta)}\right]$$

for some function g_j and some parameter θ. (HINT: You don't need to prove the results for exchangeability; it's sufficient to state what they allow you to conclude in context.)

Note that if $y_1, ..., y_n$ are exchangeable, then by de Finetti's Theorem they must be conditionally independent relative to some latent parameter. Let θ be this latent parameter and let, for $j \in \{1, ..., n\}$, g_j be the density of observation y_j conditional on θ. Then the joint distribution of all the data conditional on θ can be written

$$g(y_1, ..., y_n \mid \theta) = \prod_{i=1}^{n} g_i(y_i \mid \theta).$$

Further, the marginal distribution of the data can be written

$$f(y_1, ..., y_n) = \int \prod_{i=1}^{n} g_i(y_i \mid \theta) p(\theta) d\theta,$$

for some prior distribution $p(\theta)$.

Then we have the following:

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

2. (40 pts.) Astronomers are interested in the number of supernovae appearing in a particular region of the night sky each year. Let $y = \{y_1, ..., y_n\}$ be a collection of such data. It is reasonable to assume that these data are *iid*, and that as count data they can be well modeled with a Poisson distribution.

An astronomer wants to compare the performance of two Poisson models for these data using a Bayes Factor. In both models, it is assumed that $y_i \overset{\text{iid}}{\sim} \text{Pois}(\lambda)$ for $i \in \{1, ..., n\}$. In the first model, M_1, a Jeffreys prior is used for λ, $p(\lambda) \propto \lambda^{-1/2}$. In the second model, M_2, an informative prior is used. The astronomer thinks the supernova rate in a region should be about 20 with a 95% probability interval from 14 to 26. This prior belief can be roughly approximated with a Gamma$(50, 2.5)$ prior.

- Obtain an analytical expression for the Bayes factor comparing these models. (25 pts.)

To obtain the Bayes factor, we need to find the marginal pdf for both models. For the first model, we have

$$p(y \mid M_1) = \int \left(\prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \right) \lambda^{-1/2} d\lambda \qquad =$$

The marginal pdf for the second model is

$$p(y \mid M_2) \propto \int \left(\prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \right) \lambda^{49} e^{-2.5\lambda} d\lambda$$

- Explain what would happen to this Bayes factor if the number of observations, n, was increased substantially? (15 pts.)

3. (40 pts.) Let $y \mid \theta \sim \text{NegBin}(r, \theta)$, and let θ have a $\text{Beta}(3,5)$ prior.

- Find the marginal distribution for y. (15 pts.)

The marginal distribution is given by

$$\begin{aligned}
f(y) &= \int_0^1 \binom{y}{r} \theta^{y-r}(1-\theta)^r \frac{\Gamma(5+3)}{\Gamma(5)\Gamma(3)} \theta^{3-1}(1-\theta)^{5-1} d\theta \\
&= \frac{y!}{r!(y-r)!} \frac{\Gamma(5+3)}{\Gamma(5)\Gamma(3)} \int_0^1 \theta^{2+y-r}(1-\theta)^{4+r} d\theta \\
&= \frac{\Gamma(1+y)}{\Gamma(1+r)\Gamma(1+y-r)} \frac{\Gamma(8)}{\Gamma(5)\Gamma(3)} \frac{\Gamma(5+r)\Gamma(3+y-r)}{\Gamma(8+y)} \\
&\quad \times \int_0^1 \frac{\Gamma(8+y)}{\Gamma(5+r)\Gamma(3+y-r)} \theta^{2+y-r}(1-\theta)^{4+r} d\theta \\
&= \frac{\Gamma(1+y)\Gamma(8)\Gamma(5+r)\Gamma(3+y-r)}{\Gamma(1+r)\Gamma(1+y-r)\Gamma(5)\Gamma(3)\Gamma(8+y)} \\
&= 105 \frac{\Gamma(1+y)\Gamma(5+r)\Gamma(3+y-r)}{\Gamma(1+r)\Gamma(1+y-r)\Gamma(8+y)}
\end{aligned}$$

- Let $\tilde{y}$ be a new *binomial* observation on 5 trials that are *iid* with the original trials. Obtain an analytical expression for the predictive probability that $\tilde{y} = 5$. (25 pts.)

The predictive distribution of $\tilde{y}$ is given by

$$\begin{aligned}
f(\tilde{y} = 5 \mid y) &= \int_0^1 f(\tilde{y} \mid \theta) p(\theta \mid y) d\theta \\
&= \int_0^1 \binom{5}{5} \theta^5 (1-\theta)^0 \frac{\binom{y}{r}\theta^{y-r}(1-\theta)^r \frac{\Gamma(5+3)}{\Gamma(5)\Gamma(3)} \theta^{3-1}(1-\theta)^{5-1}}{f(y)} d\theta \\
&= \frac{\frac{y!}{r!(y-r)!} \frac{\Gamma(5+3)}{\Gamma(5)\Gamma(3)}}{f(y)} \int_0^1 \theta^{3+(y-r)+4}(1-\theta)^{0+r+4} d\theta \\
&= \frac{105 \frac{\Gamma(1+y)}{\Gamma(1+r)\Gamma(1+y-r)}}{105 \frac{\Gamma(1+y)\Gamma(3+r)\Gamma(5+y-r)}{\Gamma(1+r)\Gamma(1+y-r)\Gamma(8+y)}} \\
&\quad \times \frac{\Gamma(5+r)\Gamma(8+y-r)}{\Gamma(13+y)} \int_0^1 \frac{\Gamma(13+y)}{\Gamma(5+r)\Gamma(8+y-r)} \theta^{7+y-r}(1-\theta)^{4+r} d\theta \\
&= \frac{1}{\frac{\Gamma(5+r)\Gamma(3+y-r)}{\Gamma(8+y)}} \frac{\Gamma(5+r)\Gamma(8+y-r)}{\Gamma(13+y)} \\
&= \frac{\Gamma(8+y)\Gamma(8+y-r)}{\Gamma(13+y)\Gamma(3+y-r)}
\end{aligned}$$

4. (40 pts.) A certain machine in nuclear power plants is known to often blow fuses, so these machines include a main fuse and three identical redundant backup fuses. When the first fuse is blown, the machine will switch to using the second fuse. This continues until the fourth fuse is blown, at which point the machine fails.

   Researchers are studying data on failure times for a collection of these machines. Let $y = \{y_1, ..., y_n\}$ be these failure time data, and assume that these data are *iid*. They want to perform an hypothesis test to decide which of two failure time models is better for their data. The first failure time model, M_1, is a simple $\text{Exp}(\lambda_1)$ model. The second model, M_2, is an $\text{Erlang}(4, \lambda_2)$ model. (The Erlang distribution is a special case of the Gamma distribution where the α parameter is a positive integer, and has an interpretation as the sum of α *iid* $\text{Exp}(\beta)$ random variables.) If failure of the machine does in fact happen because of blown fuses where the lifespan of fuses are *iid*, then the Erlang should be a good model for machine lifespan. If failure of the machine happens because of other factors, the Exponential may be a better model.

   The researchers have a prior belief that failure is 4 times more likely to occur because of blown fuses than because of other factors. In other words, the researchers believe $\Pr[M_1] = 0.2$ and $\Pr[M_2] = 0.8$. The researchers also believe, given the previously observed failure times machines of this sort, that a $\text{Gamma}(1, 4)$ prior should be used for λ_1 and a $\text{Gamma}(1, 1)$ prior should be used for λ_2.

   Obtain an analytical expression for the posterior probability that the data are better modeled by the Exponential, $\Pr[M_1 \mid y]$.

   One simple way to do this problem is to calculate the posterior odds for the two models, which is simply the prior odds times the Bayes factor. If we have the posterior odds, we can convert this into probabilities regarding the models. Since we already know the prior odds to be 1/4, we proceed to finding the Bayes factor.

$$
\begin{aligned}
BF_{(1:\,2)} &= \frac{f(y \mid M_1)}{f(y \mid M_2)} \\
&= \frac{\int_0^\infty \left(\prod_{i=1}^n \lambda_1 e^{-\lambda_1 y_i}\right) 4e^{-4\lambda_1} d\lambda_1}{\int_0^\infty \left(\prod_{i=1}^n \frac{\lambda_2^4}{\Gamma(4)} y_i^3 e^{-\lambda_2 y_i}\right) e^{-\lambda_2} d\lambda_2} \\
&= \frac{\int_0^\infty \left(\lambda_1^n e^{-\lambda_1 \sum_{i=1}^n y_i}\right) 4e^{-4\lambda_1} d\lambda_1}{\int_0^\infty \left(\frac{\lambda_2^{4n}}{6^n} \left(\prod_{i=1}^n y_i\right)^3 e^{-\lambda_2 \sum_{i=1}^n y_i}\right) e^{-\lambda_2} d\lambda_2} \\
&= \frac{\int_0^\infty 4\lambda_1^n e^{-\lambda_1\left(4+\sum_{i=1}^n y_i\right)} d\lambda_1}{\int_0^\infty \frac{\lambda_2^{4n}}{6^n} \left(\prod_{i=1}^n y_i\right)^3 e^{-\lambda_2\left(1+\sum_{i=1}^n y_i\right)} d\lambda_2} \\
&= \frac{4\frac{\Gamma(n+1)}{\left(4+\sum_{i=1}^n y_i\right)^{n+1}} \int_0^\infty \frac{\left(4+\sum_{i=1}^n y_i\right)^{n+1}}{\Gamma(n+1)} \lambda_1^n e^{-\lambda_1\left(4+\sum_{i=1}^n y_i\right)} d\lambda_1}{\frac{\left(\prod_{i=1}^n y_i\right)^3}{6^n} \frac{\Gamma(4n+1)}{\left(1+\sum_{i=1}^n y_i\right)^{4n+1}} \int_0^\infty \frac{\left(1+\sum_{i=1}^n y_i\right)^{4n+1}}{\Gamma(4n+1)} \lambda_2^{4n} e^{-\lambda_2\left(1+\sum_{i=1}^n y_i\right)} d\lambda_2} \\
&= \frac{4(6^n)\Gamma(n+1)\left(1+\sum_{i=1}^n y_i\right)^{4n+1}}{\left(\prod_{i=1}^n y_i\right)^3 \Gamma(4n+1)\left(4+\sum_{i=1}^n y_i\right)^{n+1}}
\end{aligned}
$$

Then, incorporating the prior odds, the posterior odds are

$$
\frac{p(M_1 \mid y)}{p(M_2 \mid y)} = \frac{6^n\Gamma(n+1)\left(1+\sum_{i=1}^n y_i\right)^{4n+1}}{\left(\prod_{i=1}^n y_i\right)^3 \Gamma(4n+1)\left(4+\sum_{i=1}^n y_i\right)^{n+1}},
$$

and the posterior probability for M_1 is

$$
\Pr\left[M_1 \mid y\right] = \frac{6^n\Gamma(n+1)\left(1+\sum_{i=1}^n y_i\right)^{4n+1}}{6^n\Gamma(n+1)\left(1+\sum_{i=1}^n y_i\right)^{4n+1} + \left(\prod_{i=1}^n y_i\right)^3 \Gamma(4n+1)\left(4+\sum_{i=1}^n y_i\right)^{n+1}}.
$$

Table 1: Common distributions and densities.

Distribution	Notation	Density
Bernoulli	$\text{Bern}\,(\theta)$	$f(y\|\theta) = \theta^y(1-\theta)^{1-y}$; $y = 0, 1$
Binomial	$\text{Bin}\,(n, \theta)$	$f(y\|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$; $y = 0, 1, \ldots, n$
Negative Binomial	$\text{NegBin}\,(r, \theta)$	$f(y\|\theta) = \binom{y}{r}\theta^{y-r}(1-\theta)^r$; $y = r, r+1, \ldots$
Beta	$\text{Beta}\,(a, b)$	$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}I_{(0,1)}(\theta)$
Poisson	$\text{Pois}\,(\theta)$	$f(y\|\theta) = \theta^y e^{-\theta}/y!$; $y = 0, 1, 2, \ldots$
Exponential	$\text{Exp}\,(\theta)$	$f(y\|\theta) = \theta e^{-\theta y} I_{(0,\infty)}(y)$
Gamma / Erlang	$\text{Gamma}\,(a, b)$	$p(\theta) = [b^a/\Gamma(a)]\theta^{a-1}e^{-b\theta}I_{(0,\infty)}(\theta)$
Weibull	$\text{Weibull}\,(\alpha, \lambda)$	$f(y\|\alpha, \lambda) = \lambda\alpha y^{\alpha-1}\exp(-\lambda y^\alpha)\, I_{(0,\infty)}(y)$

Table 2: Some conjugate families.

$f(y\|\theta)$	$p(\theta)$	$p(\theta\|y)$
$\text{Bin}\,(n, \theta)$	$\text{Beta}\,(a, b)$	$\text{Beta}\,(a+y, b+n-y)$
$\text{NegBin}\,(r, \theta)$	$\text{Beta}\,(a, b)$	$\text{Beta}\,(a+y-r, b+r)$
$\text{Pois}\,(\theta)$	$\text{Gamma}\,(a, b)$	$\text{Gamma}\,(a+\sum y_i, b+n)$
$\text{Exp}\,(\theta)$	$\text{Gamma}\,(a, b)$	$\text{Gamma}\,(a+n, b+\sum y_i)$
$\text{Gamma}\,(k, \theta)$	$\text{Gamma}\,(a, b)$	$\text{Gamma}\,(a+nk, b+\sum y_i)$