The exam is open book and open notes. No laptops are permitted but you may use a calculator. Giving or receiving assistance from other students is not allowed. Show work to receive full credit! Partial credit will be given, but only for work written on the exam.

Assume that \( x \sim \text{Gamma}(a, b) \), \( y \sim \text{InvGamma}(a, b) \), and \( z \sim \text{Exponential}(\lambda) \) implies that

\[
p(x|a, b) = \frac{b^a e^{-ax}}{\Gamma(a)}, \quad p(y|a, b) = \frac{b^a e^{-by}}{\Gamma(a)y^{a+1}}, \quad p(z|\lambda) = \frac{1}{\lambda} e^{-z/\lambda}, \quad \text{and} \quad \text{Prob}(z < c) = 1 - e^{-c/\lambda}.
\]

1. Let \( y_1, \ldots, y_n \overset{iid}{\sim} \text{Uniform}(0, \theta) \). We select the prior \( \theta \sim \text{Pareto}(a, b) \) with prior density

\[
p(\theta) = \frac{ab^a}{\theta^{a+1}} I(\theta > b),
\]

where \( I(\theta > b) = 1 \) if \( \theta > b \) and 0 otherwise.

(a) (10 points) Calculate the posterior of \( \theta \) (give the parametric family and its parameters, e.g., \( \theta|y \sim N(a, b) \)).

\[
p(\theta|y) \propto \prod_{i=1}^{n} p(y_i|\theta) p(\theta) \propto \frac{1}{\theta^a} \prod_{i=1}^{n} I(y_i < b) I(\theta > b)
\]

\[
= \frac{1}{\theta^{a+n-1}} I(\theta > \max \{y_1, \ldots, y_n, b\})
\]

\[
\Rightarrow \theta \sim \text{Pareto}(n+g, \max \{y_1, \ldots, y_n, b\})
\]

(b) (5 points) Is the estimate \( \hat{\theta} = E(\theta|y_1, \ldots, y_n) \) consistent for all values of \( a, b, \) and \( \theta_0 \) (the true value of \( \theta \))? Justify your conclusion.

If \( \theta_0 < b \), then \( E(\theta|y) > b \) for all \( y > \theta_0 \) if it can't be consistent.

I
2. Let \( y_1, \ldots, y_n \overset{\text{iid}}{\sim} \text{Gamma}(a, \theta) \) where \( a \) is known and \( \theta > 0 \).

(a) (10 points) Compute the Jeffreys prior for \( \theta \).

\[
\begin{align*}
\ell(\theta) &= \sum_{i=1}^{n} \left[ a \log(\theta y_i) + (1-a) \log y_i - 2\theta y_i - a \log \Gamma(\theta) \right] \\
\ell'(\theta) &= \frac{a y_i^2}{\theta} - a \frac{y_i}{\theta} \\
\ell''(\theta) &= -\frac{a y_i^2}{\theta^2}
\end{align*}
\]

So \( p(\theta) = \sqrt{-\ell''(\theta)} \propto \frac{1}{\theta} \)

(b) (10 points) Identify a conjugate prior and resulting posterior for \( \theta \).

\[
\begin{align*}
\theta &\sim \text{Beta}(c_1, c_2) \\
p(\theta | y) &\propto \theta^{n_\theta + c_1 - 1} (\theta y)^{c_2 - 1} e^{-c_2 \theta} \\
\theta | y &\sim \text{Beta}(n_\theta + c_1, c_2 + n_\theta y)
\end{align*}
\]

(c) (5 points) Now assume that \( a \) is not known, but it is to be fixed using an empirical Bayes approach. Define and justify a value for \( a \) in terms of the sample mean and variance of \( y_1, \ldots, y_n \), denoted \( \bar{y} \) and \( s^2 \), respectively.

\[
\begin{align*}
E(Y) &= \frac{a}{\theta} = \frac{1}{\theta} \\
V(Y) &= \frac{1}{a^2 \theta^2} = \frac{1}{a^2 \theta^2} = \left( \frac{1}{\theta^2} \right) E(Y)^2 \\
\text{So } a &= \frac{E(Y)^2}{V(Y)} \\
\text{So } a &= \frac{\bar{y}^2}{s^2} \text{ seems like a justifiable method of moments estimate.}
\end{align*}
\]
3. (10 points) The data for this problem come from a smoking cessation study. Smokers were given hypnosis theory to help them quit smoking. We call the subjects every morning to determine if they had a cigarette. For $n$ subjects, they answer the phone every day until they report a relapse. Let $y_1, \ldots, y_n$ be the relapse times for these subjects. For $m$ subjects, they stop answering the phone and we never know their relapse time. For these censored subjects, let $y_{n+1}, \ldots, y_{n+m}$ be the number of days until they stopped answering the phone. We model the times until relapse as exponential with mean $\lambda$, pdf $f(y)$, and CDF $F(y)$. The likelihood for subject $i$ is

$$
\begin{cases}
    f(y_i) & i \leq n \\
    1 - F(y_i) & i > n
\end{cases}
$$

Assuming an inverse gamma prior for $\lambda$, compute $\lambda$'s posterior (give the parametric distribution).

\[ \lambda \sim \text{Inv}(a, b) \]

\[
\begin{align*}
    p(\lambda | y) & \propto \prod_{i=1}^{n} f(y_i) \left[ \prod_{i=n+1}^{m} 1 - f(y_i) \right] \rho(\lambda) \\
    & \propto \left[ \lambda^{-n-1} e^{-\lambda \sum_{i=1}^{n} y_i} \right] \left[ e^{-\lambda \sum_{i=n+1}^{m} y_i} \right] \lambda^{-a-1} e^{-\frac{\lambda}{\alpha}} \\
    & \propto \lambda^{-n+a-1} e^{-\frac{1}{\lambda} \left[ \sum_{i=1}^{m} y_i + b \right]} \\
    \lambda & \sim \text{Inv}(n+a, \sum_{i=1}^{m} y_i + b)
\end{align*}
\]
4. Referring again to the survival analysis problem (3), assume that \( y_i \overset{iid}{\sim} \text{Expo}(\lambda) \), and that there is no censoring, i.e., \( m = 0 \). Compare the following two estimators of the mean survival time, \( \lambda \): \( \hat{\lambda}_1 = (y_1 + \ldots + y_n)/n \) and \( \hat{\lambda}_2 = E(\lambda | y_1, \ldots, y_n) \), where \( \lambda \) has an improper flat prior on \([0, \infty)\).

(a) (10 points) Derive \( \hat{\lambda}_2 \). For which values of \( n \) is it well-defined?

\[
\hat{\lambda}_2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda} e^{-\frac{y_i}{\lambda}} > \frac{1}{n-1} \sum_{i=1}^{n} \frac{1}{\lambda} e^{-\frac{y_i}{\lambda}}
\]

\( \lambda \sim \text{Inv}(a-1, b) \) which is valid if \( n > 1 \).

\( E(\lambda) = \frac{a}{a-2} \bar{y} \) which is valid if \( n > 2 \).

(b) (10 points) Compute and compare the bias and variance for \( \hat{\lambda}_1 \) and \( \hat{\lambda}_3 = c\hat{\lambda}_1 \) for some \( c > 0 \).

\[
\text{Bias}(\hat{\lambda}_1) = 0 \quad \text{Var}(\hat{\lambda}_1) = \frac{\lambda^2}{n}
\]

\[
\text{Bias}(\hat{\lambda}_2) = \left(1 - \frac{c}{\lambda} \right) \lambda \quad \text{Var}(\hat{\lambda}_2) = c^2 \frac{\lambda^2}{n} = c^2 \text{Var}(\hat{\lambda}_1)
\]

So \( |\text{Bias}(\hat{\lambda}_2)| \geq |\text{Bias}(\hat{\lambda}_1)| \) for all cases.

\( \text{Var}(\hat{\lambda}_2) > \text{Var}(\hat{\lambda}_1) \) if \( c > 1 \) or vice versa.

(c) (10 points) In terms of \( n \) and the true value of \( \lambda \), when is \( \hat{\lambda}_2 \) preferred over \( \hat{\lambda}_1 \)?

\( \hat{\lambda}_2 = c \hat{\lambda}_1 \) where \( c = \frac{a}{a-2} > 1 \).

So it is worse than \( \hat{\lambda}_1 \) in terms of both bias and variance for all \( n \geq 1 \).