1. Non-small cell lung cancer (NSCLC) is the most common type of lung cancer in humans. A recent study in *Japanese Journal of Clinical Oncology* examined a small group of NSCLC patients treated with gefitinib and erlotinib (two cancer drugs). Here were the times until treatment failure (TTF, in months) for $n = 14$ patients:

0.8 7.5 13.4 1.4 0.5 68.9 16.1 20.4 15.6 4.2 2.4 8.2 5.3 14.0

Consider modeling TTF by using an exponential distribution; specifically, suppose the $n = 14$ patient times $Y_1, Y_2, ..., Y_{14}$ are iid from the population-level pdf

$$f_Y(y|\theta) = \begin{cases} \theta e^{-\theta y}, & y > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where θ is modeled *a priori* as $\theta \sim \text{gamma}(a, b)$, where $a > 0$ and $b > 0$ are known. That is, the prior pdf is

$$g(\theta) = \begin{cases} \dfrac{1}{\Gamma(a)b^a}\, \theta^{a-1} e^{-\theta/b}, & \theta > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show the posterior distribution $g(\theta|\mathbf{y})$ is gamma with (updated) shape parameter $n + a$ and (updated) scale parameter $b/(1 + bu)$, where $u = \sum_{i=1}^{n} y_i$.
(b) Suppose $a = 1/2$ and $b = 1/5$. Graph the posterior distribution for θ using the data above. Report posterior mean, median, and mode estimates for θ.
(c) Derive Jeffreys' prior distribution for θ. Is this prior distribution proper? What is the posterior distribution $g(\theta|\mathbf{y})$ when using Jeffreys' prior?

2. Suppose Y, the time to be seen by a medical professional at the Thomson Student Health Center, follows a uniform distribution, specifically, $Y \sim \mathcal{U}(0, \theta)$, where $\theta > 0$. Recall the population pdf of Y is

$$f_Y(y|\theta) = \begin{cases} \dfrac{1}{\theta}, & 0 < y < \theta \\ 0, & \text{otherwise.} \end{cases}$$

In turn, the parameter θ is best regarded as random with a Pareto prior distribution; i.e., $\theta \sim g(\theta)$, where

$$g(\theta) = \begin{cases} \dfrac{\beta\alpha^\beta}{\theta^{\beta+1}}, & \theta > \alpha \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$ are known. We want to do a Bayesian analysis for θ based on a random sample of student waiting times $Y_1, Y_2, ..., Y_n$.
(a) We proved in STAT 512 that the maximum order statistic $T = T(Y_1, Y_2, ..., Y_n) = Y_{(n)}$ is a sufficient statistic for θ. Find the pdf of T. Denote the pdf by $f_{T|\theta}(t|\theta)$.
(b) Derive the posterior distribution $g(\theta|t)$ and find the posterior mean $\widehat{\theta}_B$.

(c) We proved in STAT 512 that $T = T(Y_1, Y_2, ..., Y_n) = Y_{(n)}$ is also the maximum likelihood estimator (MLE) for θ. Can the posterior mean $\widehat{\theta}_B$ in part (b) be written as a linear combination of the prior mean and T? If so, prove it. If not, show this can not be done.

3. Insurance payments are typically positively skewed with a long upper tail. A reasonable parametric model for this type of data is the Weibull distribution. Let $Y_1, Y_2, ..., Y_n$ denote a sample of n payments modeled as iid observations with common pdf

$$f_Y(y|\theta) = \begin{cases} \dfrac{m}{\theta} y^{m-1} e^{-y^m/\theta}, & y > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where $m > 0$ is known. Suppose θ is best regarded as random with an IG(α, β) prior; i.e., the pdf of θ is

$$g(\theta) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^\alpha} \dfrac{1}{\theta^{\alpha+1}} e^{-1/\beta\theta}, & \theta > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show the posterior $g(\theta|\mathbf{y})$ is also inverted gamma and determine the updated parameters.
(b) Suppose $m = 2$ so that the population-level model is Rayleigh(θ), and assume $\theta \sim$ IG$(\alpha = 0.5, \beta = 2)$. An actuary records a random sample of $n = 10$ payments (in \$10,000s) from the most recent period:

0.269 0.071 0.469 0.819 3.970 0.268 0.245 2.831 0.085 0.118

Graph and prior pdf $g(\theta)$ and the posterior pdf $g(\theta|\mathbf{y})$ on the same graph and calculate a 95 percent equal-tail credible interval for θ.

4. The following table gives the number of goals scored per game in the 2013-2014 English Premier League season:

Goals	0	1	2	3	4	5	6	7	8	9	10+
Frequency	27	73	80	72	65	39	17	4	1	2	0

For example, 27 games ended in a 0-0 draw, 73 games ended 1-0, and so on. There were $n = 380$ games total. Let's model the number of goals scored in these $n = 380$ games as iid Poisson(λ), where λ is modeled noninformatively using Jeffreys' prior

$$g(\lambda) \propto \frac{1}{\sqrt{\lambda}}$$

as derived in the notes.

(a) Derive the posterior distribution $g(\lambda|\mathbf{y})$. Then, using the English Premier League data, calculate an equal-tail 99 percent credible interval for λ. Interpret the interval.

(b) When $Y_1, Y_2, ..., Y_n$ are iid Poisson(λ), a (non-Bayesian) confidence interval for λ can be derived by arguing

$$Z = \frac{\overline{Y} - \lambda}{\sqrt{\frac{\overline{Y}}{n}}} \sim \mathcal{AN}(0,1)$$

for large n. First, use the CLT and Slutsky's Theorem to make this argument. Then, using Z as a large-sample pivot, show this leads to a large-sample $100(1-\alpha)\%$ confidence interval for λ of the form

$$\overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\overline{Y}}{n}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile from a $\mathcal{N}(0,1)$ distribution. Using the English Premier League data, calculate a 99 percent confidence interval for λ using this formula. Interpret the interval. Are your credible and confidence intervals about the same? Why do you think this is?

(c) Using the English Premier League data, perform a Bayesian hypothesis test of

$$H_0 : \lambda < 3$$
$$\text{versus}$$
$$H_a : \lambda \geq 3.$$

What is the probability H_0 is true? (Note this question would make no sense to a non-Bayesian).