Homework 3

Q1

Marietta Traffic Authority is concerned about the repeated accidents at the intersection of Canton and Piedmont Roads. Bayes-inclined city-engineer would like to estimate the accident rate, even better, find a credible set.

A well-known model for modeling the number of road accidents in a particular location/time window is the Poisson distribution. Assume that X represents the number of accidents in a 3 month period at the intersection of Canton and Piedmont Roads.

Assume that $[X \mid \theta] \sim \mathcal{P}oi(\theta)$. Nothing is known a priori about θ, so it is reasonable to assume the Jeffreys prior

$$\pi(\theta) = \frac{1}{\sqrt{\theta}}\mathbf{1}(0 < \theta < \infty)$$

In the four most recent three-month periods the following realizations for X are observed: $1, 2, 0,$ and 2 .

(a) Compare the Bayes estimator for θ with the MLE (For Poisson, recall, $\hat{\theta}_{MLE} = \bar{X}$).

For X ~Poisson(θ), the likelihood is:

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \left(\prod_{i=1}^{n} \frac{1}{x_i!}\right) e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}$$

Since Posterior $\propto$ Likelihood $\times$ Prior:

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto f(\mathbf{x}|\theta)\pi(\theta) \\
&= \left(\prod_{i=1}^{n} \frac{1}{x_i!}\right) e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}\frac{1}{\sqrt{\theta}} \\
&\propto e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i - \frac{1}{2}}
\end{aligned}$$

Which we can see as the kernel of a Gamma($\sum_{i=1}^{n} x_i + \frac{1}{2}$, n). Don't forget to explicitly compute the parameters using the observations. After doing this you should notice that the posterior is a $\boxed{\text{Gamma}(5.5, 4)}$ distribution. We use the mean of this gamma distribution, along with the observed values provided as a Bayes estimator for θ:

$$\mathbb{E}_{\theta|x}[\theta] = \frac{\sum_{i=1}^{n} x_i + \frac{1}{2}}{n} \Rightarrow \hat{\theta}_{\text{Bayes}} = \frac{[1+2+0+2]+\frac{1}{2}}{4} \boxed{= 1.375}$$

While $\hat{\theta}_{\text{MLE}} = \frac{1+2+0+2}{4} = \frac{5}{4}$

(b) Compute the 95% equitailed credible set.

In this case, the resulting posterior is of a form that is well-known and widely available in statistical software. Simply use the inverse CDF for a Gamma(5.5,4) at quantiles 0.025 and 0.975. On Python this might look like:

```
from scipy.stats import gamma
gamma(a=5.5,scale=1/4).ppf([0.025,1-0.025])
>[0.477, 2.74]
```

Notice how `scipy.stats` parametrizes the gamma distribution in comparison to how it is shown in your part (a) solution (reciprocals of each other). This occurs frequently throughout the course. Please read the documentation carefully for your chosen statistical software.

(c) Compute (numerically) the 95% HPD credible set. Use an optimization method, not a sampling method.

Recall that the HPD credible set is the smallest interval $[a, b]$ such that $\int_a^b \pi(\theta|\mathbf{x})d\theta = 0.95$. In the case of a Gamma distribution, a closed form solution for arbitrary parameters does not exist. A numerical method must be used to estimate a value of c that satisfies:

$$\int_{\theta:\pi(\theta|\mathbf{x})\geq c} \pi(\theta|\mathbf{x})d\theta = 0.95$$

Section 4.11 of the Python repository shows an example of this. Focusing just on the numerical component, this should look something like:

```
from scipy.optimize import fsolve
from scipy.stats import gamma
guess_lwr, guess_upr = 0.2, 3
def conditions (x) :
    a, b =5.5, 4
    alpha = 0.05
    lwr, upr = x
    cond_1 = gamma.pdf (upr, a, scale=1 / b) - gamma.pdf (lwr, a, scale=1 / b)
    cond_2 = (gamma.cdf (upr, a, scale=1 / b) - gamma.cdf(lwr, a, scale=1 / b) - (1 - alpha))
    return cond_1, cond_2

fsolve(conditions, (guess_lwr, guess_upr))
>[0.36915183, 2.53811837]
```

(d) Numerically find the mode of the posterior, that is, MAP estimator of θ. Make sure it matches the result of the known equation for the posterior mode.

The MAP is the posterior mode. A gamma distribution with shape parameter α and rate parameter β has a mode $\frac{\alpha-1}{\beta}$. In this case, $\boxed{\hat{\theta}_{\text{MAP}} = \frac{4.5}{4} = 1.125}$. Assuming that x, α, and β are all greater than zero (properties of a gamma distribution), analytically deriving the mode would look something like this:

$$\begin{aligned}\frac{d}{dx}\left[\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp\{-\beta x\}\right] &= \frac{\beta^\alpha}{\Gamma(\alpha)}\left[(\alpha-1)x^{\alpha-2}\exp\{-\beta x\} - \beta x^{\alpha-1}\exp\{-\beta x\}\right]\\ 0 &= \exp\{-\beta x\}x^{\alpha-2}[(\alpha-1)-\beta x]\end{aligned}$$

Solving for x here requires some deductive reasoning and the zero property, along with the properties of x, α, β. Note $\exp\{-\beta x\}$ and $x^{\alpha-2}$ will always be positive because of $\alpha, \beta, x > 0$. The only part of the expression above that could result in $x = 0$ lies in the factor $[(\alpha-1)-\beta x]$. Setting equal to zero and solving gives a critical point on$f$ at $x = \frac{\alpha-1}{\beta}$.A second derivative test would show this corresponds to a maximum value of f.

Fortunately, we don't need to do that. We're using a numeric method, such as the following:

```
from scipy.optimize import minimize
from scipy.stats import gamma

posterior = gamma(5.5, scale=1 / 4)

bounds = [(0, 5)]

minimize_dict = minimize(
    lambda p:  -posterior.pdf(p), x0=1, method="L-BFGS-B", bounds=bounds
)

print(f"MAP={minimize_dict['x'][0]:.3f}")
> MAP=1.125
```

(e) If you test the hypotheses

$$H_0 : \theta \geq 1 \quad \text{vs} \quad H_1 : \theta < 1$$

based on the posterior, which hypothesis will be favored?

Calculate the posterior probabilities using a Gamma(5.5,4) CDF:

$$\begin{aligned}&P(\theta \geq 1) = 1 - P(\theta < 1) \approx 1 - 0.2867 = 0.7133\\ &P(\theta < 1) \approx 0.2867\end{aligned}$$

Since $P(H_0) > P(H_1)$, H_0 is the favored hypothesis.

(f) Derive the posterior predictive distribution. Based on this, how many accidents do you predict for the next year?

We begin by defining some notations and recalling what we know about what we've seen so far. Going forward, the data used to create our posterior above, $(x_1, \cdots, x_n)$, will be denoted by vector $\mathbf{x}$. Then x_{n+1} is the next observation. There will be several instances of $\sum_i x_i$ as well: it's important to remember that this is the sum of the *first* n observations.

The **gamma function** $\Gamma(z) = \int_0^\infty t^{z-1}\exp\{-t\}dt$ will also make a few cameo appearances in our derivation below. When z is an integer, $\Gamma(z) = (z-1)!$. From Lecture 4.13 Bayesian Prediction, the posterior predictive distribution has the form:

$$f(x_{n+1}|\mathbf{x}) = \int_\Theta f(x_{n+1}|\theta)\pi(\theta|\mathbf{x})$$

Where f is the Poisson(θ) likelihood with a single observation x_{n+1}, and π is the posterior distribution we calculated above. Making those substitutions:

$$\begin{aligned} f(x_{n+1}|\mathbf{x}) &= \int_0^\infty \frac{\exp\{-\theta\} \cdot \theta^{x_{n+1}}}{x_{n+1}!} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \times \theta^{\alpha-1} \times \exp\{-\theta\beta\} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha) \cdot (x_{n+1})!} \cdot \int_0^\infty \theta^{(x_{n+1}+\alpha)-1} \exp\{-\theta \cdot (1+\beta)\} d\theta \end{aligned} \tag{2}$$

Note that $\frac{(1+\beta)^{x_{n+1}+\alpha}}{\Gamma(x_{n+1}+\alpha)} \int_0^\infty \theta^{(x_{n+1}+\alpha)-1} \exp\{-(1+\beta)\theta\} d\theta = 1$ (Why? Hint: PDF of a Gamma random variable).

$$\begin{aligned} f(x_{n+1}|\mathbf{x}) &= \frac{\beta^\alpha}{\Gamma(\alpha) \cdot (x_{n+1})!} \cdot \frac{\Gamma(x_{n+1}+\alpha)}{(1+\beta)^{x_{n+1}+\alpha}} \\ &= \frac{\Gamma(x_{n+1}+\alpha)}{\Gamma(\alpha) \cdot (x_{n+1})!} \times \left(\frac{\beta}{1+\beta}\right)^\alpha \times \left(\frac{1}{1+\beta}\right)^{x_{n+1}} \end{aligned}$$

Substituting $p = \frac{\beta}{1+\beta}$ and using the Gamma representation of $(x_{n+1})!$:

$$f(x_{n+1}|\mathbf{x}) = \frac{\Gamma(x_{n+1}+\alpha)}{\Gamma(\alpha) \cdot \Gamma(x_{n+1}+1)} \cdot p^\alpha \cdot (1-p)^{x_{n+1}}$$

We recognize the Negative Binomial Distribution's Probability Mass Function with parameters $r = n + \alpha$ and p as defined before. A more familiar representation includes a binomial coefficient as the normalizing constant, having PMF $\binom{k+r-1}{k} \cdot (1-p)^k \cdot p^r$. That form is less applicable here, since α is not necessarily an integer. We leave it for the reader to verify that $\binom{k+r-1}{k} = \frac{(k+r-1)!}{k! \cdot (r-1)!} = \frac{\Gamma(k+r)}{\Gamma(k+1)\Gamma(r)}$. The extension of the Negative Binomial distribution to non-integer values of r is known as the Pólya Distribution.

Our solution is NegBin$(r = 5.5, p = 0.8)$.

Lastly, we can estimate the mean of the posterior predictive using the Negative Binomial distribution mean $\frac{r(1-p)}{p} = 1.375$. To extrapolate over the next year, we multiply 1.375 by 4 for a prediction of 5.5 (or 6, rounding up).

Q2

Waiting time. The waiting time for a bus at a given corner at a certain time of day is known to have a $U(0, \theta)$ distribution. It is desired to test $H_0 : 0 \leq \theta \leq 15$ versus $H_1 : \theta > 15$. From other similar routes, it is known that θ has a Pareto $(5, 3)$ distribution. If waiting times of $10, 8, 10, 5,$ and 14 are observed at the given corner, calculate the posterior odds ratio and the Bayes factor. Which hypothesis would you favor?

Note: the density of a Pareto distribution with parameters (c, α) is given by

$$\frac{\alpha c^\alpha}{\theta^{\alpha+1}} \mathbf{1}(\theta > c)$$

Suppose that

$$X_1, \ldots, X_5 \overset{\text{iid}}{\sim} U(0, \theta)$$

Then the likelihood function is

$$\begin{aligned} L(\theta \mid X) &= \prod_{i=1}^{5} \frac{1}{\theta} \mathbf{1}(0 x_i \leq \theta) \\ &= \frac{1}{\theta^5} \mathbf{1}\Big(\theta \geq \max\{10, 8, 10, 5, 14\}\Big) \\ &= \frac{1}{\theta^5} \mathbf{1}(\theta \geq 14) \end{aligned}$$

A Pareto prior for θ:

$$\pi(\theta) = \frac{\alpha c^\alpha}{\theta^{\alpha+1}} \mathbf{1}(\theta > c)$$

The posterior with values subbed in:

$$\begin{aligned} \pi(\theta \mid x) &\propto L(\theta \mid X)\, \pi(\theta) \\ &\propto \frac{1}{\theta^5} \cdot \frac{3 \cdot 5^3}{\theta^4} \mathbf{1}(\theta \geq 14) \\ &\propto \frac{1}{\theta^{8+1}} \mathbf{1}(\theta \geq 14). \end{aligned}$$

Recognize the Pareto posterior:

$$\theta \mid x \sim \text{Pareto}\Big(c = 14, \alpha = 8\Big)$$

Now, testing the following hypotheses:

$$H_0 : \ \theta \leq 15 \quad \text{versus} \quad H_1 : \ \theta > 15.$$

Under the posterior Pareto$(14, 8)$ distribution the cumulative distribution function is

$$F(\theta) = 1 - \left(\frac{14}{\theta}\right)^8, \quad \theta \geq 14.$$

Thus, the posterior probability of H_0 is

$$P(H_0 \mid x) = P(\theta \leq 15 \mid x) = 1 - \left(\frac{14}{15}\right)^8 \approx .424$$

and that of H_1 is

$$P(H_1 \mid x) = \left(\frac{14}{15}\right)^8 \approx 0.576$$

Hence the posterior odds in favor of H_0 (over H_1) are

$$\frac{P(H_0 \mid x)}{P(H_1 \mid x)} \approx \frac{0.424}{0.576} \approx 0.737.$$

Next, we compute the prior probabilities. Under the Pareto$(5, 3)$ prior, the CDF is

$$F(\theta) = 1 - \left(\frac{5}{\theta}\right)^3, \quad \theta \geq 5.$$

Thus,

$$P(H_0) = P(\theta \leq 15) = 1 - \left(\frac{5}{15}\right)^3 = 1 - \frac{1}{27} \approx 0.963$$

and

$$P(H_1) = P(\theta > 15) = \frac{1}{27} \approx 0.037$$

The prior odds in favor of H_0 are then

$$\frac{P(H_0)}{P(H_1)} = \frac{26/27}{1/27} = 26$$

The Bayes factor in favor of H_0 (relative to H_1) is the ratio of the posterior odds to the prior odds:

$$B_{01} = \frac{\frac{P(H_0 \mid x)}{P(H_1 \mid x)}}{\frac{P(H_0)}{P(H_1)}} \approx \frac{0.737}{26} \approx 0.028.$$

Equivalently, the Bayes factor in favor of H_1 is

$$B_{10} = \frac{1}{B_{01}} \approx 35.30$$

Since $B_{10} \approx 35.30$, the data provide strong evidence in favor of H_1, that is, in favor of $\theta > 15$.

Q3

The Maxwell distribution with parameter $\alpha > 0$, has a probability density function for $x > 0$ given by

$$p(x \mid \alpha) = \sqrt{\frac{2}{\pi}} \alpha^{3/2} x^2 \exp\left(-\frac{1}{2}\alpha x^2\right)$$

(a) Find the Jeffreys prior for α.

The Maxwell distribution with parameter $\alpha > 0$ has the probability density function

$$p(x \mid \alpha) = \sqrt{\frac{2}{\pi}}\, \alpha^{3/2} x^2 \exp\Bigl(-\frac{1}{2}\alpha x^2\Bigr), \quad x > 0.$$

First, compute the log-density:

$$\log p(x \mid \alpha) = \frac{1}{2}\log\left(\frac{2}{\pi}\right) + \frac{3}{2}\log\alpha + 2\log x - \frac{1}{2}\alpha x^2.$$

Differentiate with respect to α:

$$\frac{\partial}{\partial\alpha}\log p(x \mid \alpha) = \frac{3}{2}\frac{1}{\alpha} - \frac{1}{2}x^2.$$

Taking the second derivative gives:

$$\frac{\partial^2}{\partial\alpha^2}\log p(x \mid \alpha) = -\frac{3}{2}\frac{1}{\alpha^2}.$$

The Fisher information is then

$$I(\alpha) = -E\left[\frac{\partial^2}{\partial\alpha^2}\log p(x \mid \alpha)\right] = \frac{3}{2}\frac{1}{\alpha^2},$$

since the second derivative does not depend on x. Thus, the Jeffreys prior is

$$\pi_J(\alpha) \propto \sqrt{I(\alpha)} \propto \frac{1}{\alpha}.$$

(b) Find a transformation of this parameter in which the corresponding prior is uniform.

We wish to find a one-to-one transformation $\phi = g(\alpha)$ so that the induced prior on ϕ is uniform. Recall that if $\pi_\alpha(\alpha)$ is the prior on α and $\phi = g(\alpha)$ then the induced prior is

$$\pi_\phi(\phi) = \pi_\alpha(\alpha)\left|\frac{d\alpha}{d\phi}\right|.$$

Since we have found

$$\pi_\alpha(\alpha) \propto \frac{1}{\alpha},$$

we want

$$\frac{1}{\alpha}\left|\frac{d\alpha}{d\phi}\right| \propto 1.$$

A natural choice is to let

$$\phi = \log\alpha.$$

Then

$$\frac{d\alpha}{d\phi} = e^\phi = \alpha,$$

so that

$$\pi_\phi(\phi) \propto \frac{1}{\alpha}\cdot\alpha = 1.$$

That is, the induced prior on ϕ is uniform.

(c) Find the posterior distribution for n independent and identically distributed datapoints $x_1, \ldots, x_n$ from the Maxwell distribution, assuming the Jeffreys prior on α from part (a).

Assume we observe n independent data points

$$x_1, x_2, \ldots, x_n \sim \text{Maxwell}(\alpha)$$

with density

$$p(x \mid \alpha) = \sqrt{\frac{2}{\pi}}\,\alpha^{3/2}x^2\exp\Big(-\frac{1}{2}\alpha x^2\Big), \quad x > 0.$$

The likelihood function is:

$$
\begin{aligned}
L(\alpha \mid x_1, \ldots, x_n) &= \prod_{i=1}^{n} p(x_i \mid \alpha) \\
&= \prod_{i=1}^{n} \left[\sqrt{\frac{2}{\pi}}\, \alpha^{3/2} x_i^2 \exp\Big(-\frac{1}{2}\alpha x_i^2\Big) \right] \\
&= \left(\sqrt{\frac{2}{\pi}} \right)^{n} \alpha^{\frac{3n}{2}} \left(\prod_{i=1}^{n} x_i^2 \right) \exp\left(-\frac{1}{2}\alpha \sum_{i=1}^{n} x_i^2 \right).
\end{aligned}
$$

Using the Jeffreys prior for α from part (a),

$$
\pi_J(\alpha) \propto \frac{1}{\alpha},
$$

the unnormalized posterior is given by

$$
\begin{aligned}
\pi(\alpha \mid x_1, \ldots, x_n) &\propto L(\alpha \mid x_1, \ldots, x_n)\, \pi_J(\alpha) \\
&\propto \alpha^{\frac{3n}{2}} \cdot \frac{1}{\alpha} \exp\left(-\frac{1}{2}\alpha \sum_{i=1}^{n} x_i^2 \right) \\
&= \alpha^{\frac{3n}{2}-1} \exp\left(-\frac{1}{2}\alpha \sum_{i=1}^{n} x_i^2 \right).
\end{aligned}
$$

Recognize that this is the kernel of a Gamma density. Comparing exponents, we identify

$$
a = \frac{3n}{2} \quad \text{and} \quad b = \frac{1}{2}\sum_{i=1}^{n} x_i^2.
$$

The posterior distribution for α is then

$$
\alpha \mid x_1, \ldots, x_n \sim \text{Gamma}\left(\frac{3n}{2}, \frac{1}{2}\sum_{i=1}^{n} x_i^2 \right).
$$