Exercise: We would like to study and compare the popularity of two competing restaurant chains, X and Y, with n and m branches, respectively. We develop two different models, A and B:

A. The number of daily clients each branch receives is assumed to be iid for both chains, $x_i|\theta \sim \mathcal{P}(\theta)$, $i \in [n]$ and $y_i|\theta \sim \mathcal{P}(\theta)$, $i \in [m]$. The assumed prior distribution is $\theta \sim \mathcal{G}(p, q)$.

B. The number of daily clients each branch receives is assumed to be iid for each chain, $x_i|\theta_1 \sim \mathcal{P}(\theta_1)$, $i \in [n]$ and $y_i|\theta_2 \sim \mathcal{P}(\theta_2)$, $i \in [m]$, with $\mathbf{x}|\theta_1 \perp\!\!\!\perp \mathbf{y}|\theta_2$. The assumed prior distributions are $\theta_1 \sim \mathcal{G}(p_1, q_1)$ and $\theta_2 \sim \mathcal{G}(p_2, q_2)$, $\theta_1 \perp\!\!\!\perp \theta_2$.

Perform the following analysis:

1. For model A, derive the posterior distribution of $\theta|(\mathbf{x}, \mathbf{y})$. For model B, derive the joint posterior distribution of $(\theta_1, \theta_2)|(\mathbf{x}, \mathbf{y})$. Show the conjugacy of the prior in both cases.

2. Assuming known prior model probabilities $P(A)$ and $P(B)$, perform a model comparison and derive the posterior model probabilities.

3. The generalized beta prime distribution $\beta'(\alpha, \beta, p, q)$ is a continuous distribution with pdf

$$f(x|\alpha, \beta, p, q) = \frac{pq^{\beta p}}{B(\alpha, \beta)} x^{\alpha p - 1}(q^p + x^p)^{-(\alpha+\beta)}, \quad x > 0.$$

   Show that, in model B, the posterior distribution of $u := \theta_1/\theta_2$ is the beta prime $\beta'(P_1, P_2, 1, Q)$, with $P_1 := \sum_{i=1}^n x_i + p_1$, $P_2 := \sum_{i=1}^m y_i + p_2$ and $Q := (q_2 + m)/(q_1 + n)$. Derive a (1-a)% equal-tailed credible region for $u|(\mathbf{x}, \mathbf{y})$.

4. The following observations are available for the two chains:

$$\mathbf{x} = (15, 10, 15, 7, 12, 7, 5, 8, 16, 10),$$
$$\mathbf{y} = (14, 20, 20, 14, 9, 13, 16, 15, 12, 18, 19, 11)$$

   i. A general review on restaurant traffic indicates that $\mathrm{E}(\theta) = 9$. A thorough study on each chain indicates that $\mathrm{E}(\theta_1) = 6$ and $\mathrm{E}(\theta_2) = 12$. Choose appropriate hyperparameters p, q, p_1, q_1, p_2, q_2 for the two models A and B that follow the literature guidelines but impose a (relatively large) variance, $\mathrm{Var}(\theta) = \mathrm{Var}(\theta_1) = \mathrm{Var}(\theta_2) = 60$.

   ii. Derive the posterior distributions for each model. What is the posterior expectation in each case?
   Bonus: Knowing that the data were in fact simulated from $x_i \sim \mathcal{P}(10)$ and $y_i \sim \mathcal{P}(15)$, are the posterior expectations closer to the true values of θ_1 and θ_2?

   iii. Assuming prior model probabilities $P(A) = P(B) = 0.5$, compare the two models.

   iv. Create a 95% equal-tailed credible region for $u|(\mathbf{x}, \mathbf{y})$.