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 $u|(\mathbf{x}, \mathbf{y})$.

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.

   Solution:

   For model A, the posterior is:

$$\begin{aligned} f(\theta|\mathbf{x},\mathbf{y}) &\propto f(\mathbf{x},\mathbf{y}|\theta)f(\theta) \\ &\propto f(\mathbf{x}|\theta)f(\mathbf{y}|\theta)f(\theta) \\ &\propto \theta^{p+\sum_{i=1}^n x_i+\sum_{i=1}^n y_i-1}e^{-(q+n+m)\theta}. \end{aligned}$$

   Therefore,

$$\theta|(\mathbf{x},\mathbf{y}) \sim \mathcal{G}(P,Q), \quad P = p+\sum_{i=1}^n x_i+\sum_{i=1}^n y_i,\ Q = q+n+m.$$

   For model B, the joint posterior is:

$$\begin{aligned} f(\theta_1,\theta_2|\mathbf{x},\mathbf{y}) &\propto f(\mathbf{x},\mathbf{y}|\theta_1,\theta_2)f(\theta_1,\theta_2) \\ &\propto f(\mathbf{x}|\theta_1)f(\mathbf{y}|\theta_2)f(\theta_1)f(\theta_2) \\ &\propto \theta_1^{p_1+\sum_{i=1}^n x_i-1}e^{-(q_1+n)\theta_1}\theta_2^{p_2+\sum_{i=1}^m y_i-1}e^{-(q_2+m)\theta_2}. \end{aligned}$$

   Therefore,

$$\begin{aligned} \theta_1|\mathbf{x} &\sim \mathcal{G}(P_1,Q_1), \quad P_1 = p_1+\sum_{i=1}^n x_i,\ Q_1 = q_1+n, \\ \theta_2|\mathbf{y} &\sim \mathcal{G}(P_2,Q_2), \quad P_2 = p_2+\sum_{i=1}^m y_i,\ Q_2 = q_2+m. \end{aligned}$$

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

   Solution:

$$P(H_0|\mathbf{x},\mathbf{y}) = \frac{f(\mathbf{x},\mathbf{y}|H_0)P(H_0)}{f(\mathbf{x},\mathbf{y}|H_0)P(H_0)+f(\mathbf{x},\mathbf{y}|H_1)P(H_1)}.$$

   For model A:

$$\begin{aligned} f(\mathbf{x},\mathbf{y}|H_0) &= \int f(\mathbf{x},\mathbf{y},\theta|H_0)d\theta \\ &= \int f(\mathbf{x},\mathbf{y}|\theta,H_0)f(\theta|H_0)d\theta \\ &= \frac{1}{\prod_{i=1}^n x_i!\prod_{i=1}^m y_i!}\frac{q^p}{\Gamma(p)}\frac{Q^P}{\Gamma(P)}. \end{aligned}$$

For model B:

$$\begin{aligned}
f(\mathbf{x},\mathbf{y}|H_1) &= \iint f(\mathbf{x},\mathbf{y},\theta_1,\theta_2|H_1)d\theta_1 d\theta_2 \\
&= \iint f(\mathbf{x},\mathbf{y}|\theta_1,\theta_2,H_1)f(\theta_1\theta_2|H_1)d\theta_1 d\theta_2 \\
&= \iint f(\mathbf{x}|\theta_1,H_1)f(\theta_1|H_1)f(\mathbf{y}|\theta_2,H_2)f(\theta_2|H_2)d\theta_1 d\theta_2 \\
&= \int f(\mathbf{x}|\theta_1,H_1)f(\theta_1|H_1)d\theta_1 \int f(\mathbf{y}|\theta_2,H_1)f(\theta_2|H_1)d\theta_2 \\
&= \frac{1}{\prod_{i=1}^n x_i!}\frac{q_1^{p_1}}{\Gamma(p_1)}\frac{Q_1^{P_1}}{\Gamma(P_1)}\frac{1}{\prod_{i=1}^m y_i!}\frac{q_2^{p_2}}{\Gamma(p_2)}\frac{Q_2^{P_2}}{\Gamma(P_2)}.
\end{aligned}$$

By denoting

$$A = \frac{q^p}{\Gamma(p)}\frac{Q^P}{\Gamma(P)}, \quad B = \frac{q_1^{p_1}}{\Gamma(p_1)}\frac{Q_1^{P_1}}{\Gamma(P_1)}\frac{q_2^{p_2}}{\Gamma(p_2)}\frac{Q_2^{P_2}}{\Gamma(P_2)},$$

we get the posterior of the hypothesis:

$$P(H_0|\mathbf{x},\mathbf{y}) = \frac{AP(H_0)}{AP(H_0)+BP(H_1)}.$$

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 $P_1 := p_1 + \sum_{i=1}^n x_i$, $P_2 := p_2 + \sum_{i=1}^m y_i$ and $Q := (q_2+m)/(q_1+n)$. Derive a (1-a)% equal-tailed credible region for $u|(\mathbf{x},\mathbf{y})$.

Solution:

We will transform $(u_1,u_2) := g(\theta_1,\theta_2) := (\theta_1/\theta_2,\theta_2)$, $\theta_1,\theta_2,u_1,u_2>0$. Note that $g^{-1}(u_1,u_2) = (u_1u_2,u_2)$. Then:

$$\begin{aligned}
f_u(u_1,u_2) &= f_\theta(u_1u_2,u_2)\,\big|\big|J_{g^{-1}}(u_1,u_2)\big|\big| \\
&\propto (u_1u_2)^{P_1-1}e^{-Q_1(u_1u_2)}u_2^{P_2-1}e^{-Q_2u_2}u_2 \\
&= u_1^{P_1-1}u_2^{P_1+P_2-1}e^{-(Q_1u_1+Q_2)u_2}.
\end{aligned}$$

We compute the marginal of u_1:

$$f(u_1) = \int_0^{+\infty} f_u(u_1,u_2)du_2 \propto \frac{u_1^{P_1-1}}{\left(u_1+\frac{Q_2}{Q_1}\right)^{P_1+P_2}}.$$

Therefore, $u_1 \sim \beta'(P_1,P_2,1,Q)$, where $Q := Q_2/Q_1$.