

# MA40189 - Solution Sheet Three

Simon Shaw, [s.shaw@bath.ac.uk](mailto:s.shaw@bath.ac.uk)  
<https://people.bath.ac.uk/mass/ma40189.html>

2021/22 Semester II

1. Let  $X_1, \dots, X_n$  be conditionally independent given  $\theta$ , so  $f(x|\theta) = \prod_{i=1}^n f(x_i|\theta)$  where  $x = (x_1, \dots, x_n)$ , with each  $X_i|\theta \sim Po(\theta)$ . Suppose we judge that  $\theta \sim Gamma(\alpha, \beta)$ .

- (a) Find the distribution of  $\theta|x$ .

As  $\theta \sim Gamma(\alpha, \beta)$  then

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \quad (1)$$

whilst as  $X_i|\theta \sim Po(\theta)$

$$f(x|\theta) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^n x_i!} \quad (2)$$

From (1) and (2) the posterior density

$$\begin{aligned} f(\theta|x) &\propto \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^n x_i!} \times \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\ &\propto \theta^{(\alpha+n\bar{x})-1} e^{-(\beta+n)\theta} \end{aligned}$$

which is a kernel of a  $Gamma(\alpha + n\bar{x}, \beta + n)$  density, so  $\theta|x \sim Gamma(\alpha + n\bar{x}, \beta + n)$ .

- (b) Show that the posterior mean can be written as a weighted average of the prior mean, denoted  $\theta_0$ , and the maximum likelihood estimate of  $\theta$ ,  $\bar{x}$ .

As  $\theta|x \sim Gamma(\alpha + n\bar{x}, \beta + n)$ ,

$$\begin{aligned} E(\theta|x) &= \frac{\alpha + n\bar{x}}{\beta + n} = \frac{\beta \left(\frac{\alpha}{\beta}\right) + n\bar{x}}{\beta + n} \\ &= \lambda\theta_0 + (1 - \lambda)\bar{x} \end{aligned}$$

where  $\lambda = \frac{\beta}{\beta+n}$  and  $\theta_0 = \frac{\alpha}{\beta} = E(\theta)$  as  $\theta \sim Gamma(\alpha, \beta)$ . Thus, the posterior mean is a weighted average of the prior mean and the maximum likelihood estimate.

- (c) Let  $Z$  be a future (unobserved) observation. Find the mean and variance of the predictive distribution  $Z|x$ . You may assume that  $X$  and  $Z$  are conditionally independent given  $\theta$  and that  $Z|\theta \sim Po(\theta)$ .

We do not need to find the full predictive distribution. As  $X$  and  $Z$  are conditionally independent given  $\theta$  we can use expectation and variance the formulas given in lectures so

$$\begin{aligned} E(Z|X) &= E(E(Z|\theta)|X) \\ &= E(\theta|X) \end{aligned} \tag{3}$$

$$= \frac{\alpha + n\bar{x}}{\beta + n} \tag{4}$$

where (3) follows as  $Z|\theta \sim Po(\theta)$  and (4) as  $\theta|x \sim Gamma(\alpha + n\bar{x}, \beta + n)$ . Similarly,

$$\begin{aligned} Var(Z|X) &= Var(E(Z|\theta)|X) + E(Var(Z|\theta)|X) \\ &= Var(\theta|X) + E(\theta|X) \\ &= \frac{\alpha + n\bar{x}}{(\beta + n)^2} + \frac{\alpha + n\bar{x}}{\beta + n} = \frac{(\alpha + n\bar{x})(\beta + n + 1)}{(\beta + n)^2}. \end{aligned} \tag{5}$$

- (d) The data in the table below are the number of fatal accidents on scheduled airline flights between 1976 and 1985.

1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
24	25	31	31	22	21	26	20	16	22

Let  $Z$  be the number of fatal accidents in 1986 and  $X_i$  the number of fatal accidents in 1975+ $i$ . Adopting the model given above, with  $\alpha = 101$  and  $\beta = 5$  and assuming that  $Z|x$  may be approximated by a  $N(E(Z|X), Var(Z|X))$ , find an approximate 95% prediction interval for the number of accidents in 1986, that is an interval  $(z_L, z_U)$  such that  $P(z_L < Z < z_U | x) = 0.95$ .

Taking  $\alpha = 101$  and  $\beta = 5$  we observe  $\sum_{i=1}^{10} x_i = 238$  with  $n = 10$ . Hence,  $\alpha + n\bar{x} = 101 + 238 = 339$  and  $\beta + n = 5 + 10 = 15$  so that  $\theta|x \sim Gamma(339, 15)$ . From (4) and (5) we have  $E(Z|X) = \frac{339}{15} = \frac{113}{5}$  and  $Var(Z|X) = \frac{339 \times 16}{15^2} = \frac{1808}{75}$  so we use  $Z|x \sim N(\frac{113}{5}, \frac{1808}{75})$  approximately. Recalling that if  $Y \sim N(0, 1)$  then  $P(-1.96 < Y < 1.96) = 0.95$  the (approximate) 95% prediction interval is

$$\frac{113}{5} \pm 1.96 \sqrt{\frac{1808}{75}} = (12.98, 32.22).$$

2. Let  $X_1, \dots, X_n$  be a finitely exchangeable sequence of random quantities and consider any  $i \neq j \in \{1, \dots, n\}$ .

- (a) Explain why the marginal distribution of any  $X_i$  does not depend upon  $i$  so that  $E(X_i)$  and  $Var(X_i)$  do not depend upon  $i$ .

If  $X_1, \dots, X_n$  are finitely exchangeable then their joint density satisfies

$$f(x_1, \dots, x_n) = f(x_{\pi(1)}, \dots, x_{\pi(n)})$$

for all permutations  $\pi$  defined on the set  $\{1, \dots, n\}$ . For a given permutation there exists  $j \in \{1, \dots, n\}$  such that  $x_{\pi(j)} = x_i$ . Let  $x_{-i}$  denote the set  $\{x_1, \dots, x_n\} \setminus x_i$ . The marginal distribution of  $X_i$  is given by

$$\begin{aligned} f_{X_i}(x_i) &= \int_{x_{-i}} f(x_1, \dots, x_n) dx_{-i} \\ &= \int_{x_{-i}} f(x_{\pi(1)}, \dots, x_{\pi(j-1)}, x_{\pi(j)}, x_{\pi(j+1)}, \dots, x_{\pi(n)}) dx_{-i} \\ &= \int_{x_{-i}} f(x_{\pi(1)}, \dots, x_{\pi(j-1)}, x_i, x_{\pi(j+1)}, \dots, x_{\pi(n)}) dx_{-i} \\ &= f_{X_j}(x_i). \end{aligned}$$

As this holds for all permutations then it holds for all  $j \in \{1, \dots, n\}$ . Thus, the marginal distribution of  $X_i$  does not depend upon  $i$ . As  $E(X_i)$  and  $Var(X_i)$  are constructed over the marginal of  $X_i$  then it follows immediately that these also do not depend upon  $i$ .

- (b) **Explain why the joint distribution of any  $X_i, X_j$  does not depend upon either  $i$  or  $j$  so that  $Cov(X_i, X_j)$  does not depend on either  $i$  or  $j$ .**

For  $i \neq j$ , let  $x_{-i,-j}$  denote the set  $\{x_1, \dots, x_n\} \setminus \{x_i, x_j\}$ . For a given permutation there exists  $k \neq l \in \{1, \dots, n\}$  such that  $x_{\pi(k)} = x_i$  and  $x_{\pi(l)} = x_j$ . Suppose without loss of generality that  $k < l$ . The joint distribution of  $X_i$  and  $X_j$  is given by

$$\begin{aligned} f_{X_i, X_j}(x_i, x_j) &= \int_{x_{-i,-j}} f(x_1, \dots, x_n) dx_{-i,-j} \\ &= \int_{x_{-i,-j}} f(x_{\pi(1)}, \dots, x_{\pi(k-1)}, x_{\pi(k)}, x_{\pi(k+1)}, \dots, x_{\pi(l-1)}, x_{\pi(l)}, x_{\pi(l+1)}, \dots, x_{\pi(n)}) dx_{-i,-j} \\ &= \int_{x_{-i,-j}} f(x_{\pi(1)}, \dots, x_{\pi(k-1)}, x_i, x_{\pi(k+1)}, \dots, x_{\pi(l-1)}, x_j, x_{\pi(l+1)}, \dots, x_{\pi(n)}) dx_{-i,-j} \\ &= f_{X_k, X_l}(x_i, x_j). \end{aligned}$$

As this holds for all permutations then it holds for all  $k \neq l \in \{1, \dots, n\}$ . Thus, the joint distribution of  $X_i$  and  $X_j$  does not depend upon  $i \neq j$ . As  $Cov(X_i, X_j)$  is constructed over the joint distribution of  $X_i$  and  $X_j$  then it follows immediately that  $Cov(X_i, X_j)$  does not depend on either  $i$  or  $j$ .

- (c) **Let  $Y = \sum_{i=1}^n X_i$ . By considering  $Var(Y) \geq 0$ , or otherwise, show that**

$$Corr(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}} \geq \frac{-1}{n-1}.$$

Taking the variance of  $Y = \sum_{i=1}^n X_i$  we have

$$\begin{aligned} Var(Y) &= Var\left(\sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n Var(X_i) + \sum_{i=1}^n \sum_{j \neq i}^n Cov(X_i, X_j). \end{aligned}$$

Now, from part (a),  $Var(X_i)$  does not depend upon  $i$  and, from part (b),  $Cov(X_i, X_j)$  does not depend upon either  $i$  or  $j$ . Thus,

$$Var(Y) = nVar(X_i) + n(n-1)Cov(X_i, X_j).$$

As  $Var(Y) \geq 0$  it follows that  $nVar(X_i) + n(n-1)Cov(X_i, X_j) \geq 0$ . Rearranging this inequality gives

$$\frac{Cov(X_i, X_j)}{Var(X_i)} \geq \frac{-1}{n-1}.$$

As  $Var(X_i)$  does not depend upon  $i$  then  $Var(X_i) = \sqrt{Var(X_i)Var(X_j)}$  so that

$$Corr(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}} \geq \frac{-1}{n-1}.$$

Notice that an immediate consequence of this result is that if  $X_1, X_2, \dots$  are infinitely exchangeable then  $Corr(X_i, X_j) \geq 0$  as we require  $Corr(X_i, X_j) \geq \frac{-1}{n-1}$  for all  $n$ .

3. **Suppose  $X | \mu \sim N(\mu, \sigma^2)$  and  $Y | \mu, \delta \sim N(\mu + \delta, \sigma^2)$ , where  $\sigma^2$  is known and  $X$  and  $Y$  are conditionally independent given  $\mu$  and  $\delta$ .**

- (a) **Find the joint distribution of  $X$  and  $Y$  given  $\mu$  and  $\delta$ .**

As  $X$  and  $Y$  are conditionally independent given  $\mu$  and  $\delta$  then

$$\begin{aligned} f(x, y | \mu, \delta) &= f(x | \mu, \delta) f(y | \mu, \delta) \\ &= f(x | \mu) f(y | \mu, \delta) \\ &= \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} [(x - \mu)^2 + (y - \mu - \delta)^2] \right\}. \end{aligned}$$

- (b) **Consider the improper noninformative joint prior distribution**

$$f(\mu, \delta) \propto 1.$$

**Find, up to a constant of proportionality, the joint posterior distribution of  $\mu$  and  $\delta$  given  $x$  and  $y$ . Are  $\mu | x, y$  and  $\delta | x, y$  independent?**

As  $f(\mu, \delta) \propto 1$  then

$$\begin{aligned} f(\mu, \delta | x, y) &\propto f(x, y | \mu, \delta) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} [(x - \mu)^2 + (y - \mu - \delta)^2] \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} [\mu^2 - 2x\mu + (\mu + \delta)^2 - 2y(\mu + \delta)] \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x + y)\mu + 2\mu\delta + \delta^2 - 2y\delta] \right\}. \quad (6) \end{aligned}$$

Thus, as a consequence of the  $\mu\delta$  term in (6), we cannot write  $f(\mu, \delta | x, y) \propto g(\mu | x, y)h(\delta | x, y)$  so that  $\mu | x, y$  and  $\delta | x, y$  are not independent.

Note that  $f(\mu, \delta | x, y)$  is quadratic in  $\mu$  and  $\delta$  so that

$$\begin{aligned} f(\mu, \delta | x, y) &\propto \exp \left\{ -\frac{1}{2\sigma^2} (\mu - x \ \delta - (y - x)) \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \mu - x \\ \delta - (y - x) \end{pmatrix} \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} (\mu - x \ \delta - (y - x)) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} \mu - x \\ \delta - (y - x) \end{pmatrix} \right\} \end{aligned}$$

which is the kernel of a bivariate Normal distribution. We have

$$\mu, \delta | x, y \sim N_2 \left( \begin{pmatrix} x \\ y - x \end{pmatrix}, \begin{pmatrix} \sigma^2 & -\sigma^2 \\ -\sigma^2 & 2\sigma^2 \end{pmatrix} \right).$$

Using the properties of multivariate Normal distributions, we can read off the marginal distributions so  $\mu | x, y \sim N(x, \sigma^2)$  and  $\delta | x, y \sim N(y - x, 2\sigma^2)$ . We shall derive these marginal distributions directly in parts (c) and (d) as these give insights into the types of techniques used when the marginals are not so conventional to obtain.

(c) **Find the marginal posterior distribution  $f(\delta | x, y)$ .**

$$\begin{aligned} f(\delta | x, y) &= \int_{-\infty}^{\infty} f(\mu, \delta | x, y) d\mu \\ &\propto \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x + y)\mu + 2\mu\delta + \delta^2 - 2y\delta] \right\} d\mu \\ &= \exp \left\{ -\frac{1}{2\sigma^2} [\delta^2 - 2y\delta] \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x + y - \delta)\mu] \right\} d\mu \\ &= \exp \left\{ -\frac{1}{2\sigma^2} [\delta^2 - 2y\delta] \right\} \times \\ &\quad \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{\sigma^2} \left[ \left( \mu - \frac{(x + y - \delta)}{2} \right)^2 - \frac{(x + y - \delta)^2}{4} \right] \right\} d\mu \\ &= \exp \left\{ -\frac{1}{4\sigma^2} [2\delta^2 - 4y\delta - (x + y - \delta)^2] \right\} \times \\ &\quad \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{\sigma^2} \left( \mu - \frac{(x + y - \delta)}{2} \right)^2 \right\} d\mu \\ &\propto \exp \left\{ -\frac{1}{4\sigma^2} [2\delta^2 - 4y\delta - (x + y - \delta)^2] \right\} \\ &\propto \exp \left\{ -\frac{1}{4\sigma^2} [\delta^2 - 2(2y - (x + y))\delta] \right\} \end{aligned}$$

which is a kernel of a  $N(y - x, 2\sigma^2)$  density. Hence,  $\delta | x, y \sim N(y - x, 2\sigma^2)$ .

(d) **Find the marginal posterior distribution  $f(\mu | x, y)$ .**

$$\begin{aligned} f(\mu | x, y) &= \int_{-\infty}^{\infty} f(\mu, \delta | x, y) d\delta \\ &\propto \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x + y)\mu + 2\mu\delta + \delta^2 - 2y\delta] \right\} d\delta \end{aligned}$$

$$\begin{aligned}
&= \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x+y)\mu] \right\} \times \\
&\quad \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [\delta^2 - 2(y-\mu)\delta] \right\} d\delta \\
&= \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x+y)\mu] \right\} \times \\
&\quad \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [\delta - (y-\mu)]^2 + \frac{1}{2\sigma^2} (y-\mu)^2 \right\} d\delta \\
&= \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x+y)\mu - (y-\mu)^2] \right\} \times \\
&\quad \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [\delta - (y-\mu)]^2 \right\} d\delta \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x+y)\mu - (y-\mu)^2] \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} [\mu^2 - 2(x+y-y)\mu] \right\}
\end{aligned}$$

which is a kernel of a  $N(x, \sigma^2)$  density. Hence,  $\mu | x, y \sim N(x, \sigma^2)$ .

- (e) **Consider a future observation  $Z$  where  $Z | \mu, \delta \sim N(\mu - \delta, \sigma^2)$  and  $Z$  is conditionally independent of  $X$  and  $Y$  given  $\mu$  and  $\delta$ . Find the predictive distribution  $f(z | x, y)$ .**

Notice that

$$f(z | \mu, \delta) \propto \exp \left\{ -\frac{1}{2\sigma^2} [z^2 - 2(\mu - \delta)z + (\mu - \delta)^2] \right\}$$

so that, as  $Z$  is conditionally independent of  $X$  and  $Y$  given  $\mu$  and  $\delta$  and using (6),

$$\begin{aligned}
f(z | x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z | \mu, \delta) f(\mu, \delta | x, y) d\mu d\delta \\
&\propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [z^2 - 2(\mu - \delta)z + (\mu - \delta)^2] \right\} \times \\
&\quad \exp \left\{ -\frac{1}{2\sigma^2} [2\mu^2 - 2(x+y)\mu + 2\mu\delta + \delta^2 - 2y\delta] \right\} d\mu d\delta \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [z^2 + 3\mu^2 - 2(x+y+z)\mu + 2\delta^2 - 2(y-z)\delta] \right\} d\mu d\delta \\
&= \exp \left\{ -\frac{1}{2\sigma^2} z^2 \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [3\mu^2 - 2(x+y+z)\mu] \right\} d\mu \\
&\quad \times \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma^2} [2\delta^2 - 2(y-z)\delta] \right\} d\delta \\
&= \exp \left\{ -\frac{z^2}{2\sigma^2} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{3}{2\sigma^2} \left[ \left( \mu - \frac{(x+y+z)}{3} \right)^2 - \frac{(x+y+z)^2}{9} \right] \right\} d\mu \\
&\quad \times \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{\sigma^2} \left[ \left( \delta - \frac{(y-z)}{2} \right)^2 - \frac{(y-z)^2}{4} \right] \right\} d\delta
\end{aligned}$$

$$\begin{aligned}
&\propto \exp\left\{-\frac{1}{2\sigma^2}z^2\right\} \exp\left\{\frac{3}{2\sigma^2}\frac{(x+y+z)^2}{9}\right\} \exp\left\{\frac{1}{\sigma^2}\frac{(y-z)^2}{4}\right\} \\
&\propto \exp\left\{-\frac{1}{12\sigma^2}[z^2 - 4(x+y)z + 6yz]\right\} \\
&= \exp\left\{-\frac{1}{12\sigma^2}[z^2 - 2(2x-y)z]\right\}
\end{aligned}$$

which is a kernel of a  $N(2x-y, 6\sigma^2)$  density. Hence,  $Z|x, y \sim N(2x-y, 6\sigma^2)$ .

4. Let  $X_1, \dots, X_n$  be exchangeable so that the  $X_i$  are conditionally independent given a parameter  $\theta$ . Suppose that each  $X_i|\theta \sim U(0, \theta)$ .

- (a) Show that  $M = \max_i(X_i)$  is a sufficient statistic for  $X = (X_1, \dots, X_n)$ .

As  $X_i|\theta \sim U(0, \theta)$  then

$$f(x_i|\theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x_i \leq \theta; \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\mathbb{I}_{(0,\theta)}(x)$  denote the indicator function for the event  $0 \leq x \leq \theta$ , so  $\mathbb{I}_{(0,\theta)}(x) = 1$  if  $0 \leq x \leq \theta$  and 0 otherwise. Then we can write  $f(x_i|\theta) = \frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x_i)$  so that

$$\begin{aligned}
f(x|\theta) &= \prod_{i=1}^n \frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x_i) \\
&= \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{I}_{(0,\theta)}(x_i).
\end{aligned}$$

Let  $\mathbb{I}_{\{a \geq b\}}$  denote the indicator function for the event  $a \geq b$  so  $\mathbb{I}_{\{a \geq b\}} = 1$  if  $a \geq b$  and 0 otherwise. Then

$$\prod_{i=1}^n \mathbb{I}_{(0,\theta)}(x_i) = \mathbb{I}_{\{\min_i(x_i) \geq 0\}} \mathbb{I}_{\{\theta \geq \max_i(x_i)\}}$$

so that

$$\begin{aligned}
f(x|\theta) &= \frac{1}{\theta^n} \mathbb{I}_{\{\min_i(x_i) \geq 0\}} \mathbb{I}_{\{\theta \geq \max_i(x_i)\}} \\
&= \frac{1}{\theta^n} \mathbb{I}_{\{\theta \geq m\}} \mathbb{I}_{\{\min_i(x_i) \geq 0\}}
\end{aligned}$$

where  $m = \max_i(x_i)$ . Thus,  $f(x|\theta) = g(\theta, m)h(x)$  where  $g(\theta, m) = \frac{1}{\theta^n} \mathbb{I}_{\{\theta \geq m\}}$  and  $h(x) = \mathbb{I}_{\{\min_i(x_i) \geq 0\}}$  so that  $M = \max_i(X_i)$  is sufficient for  $X = (X_1, \dots, X_n)$  for learning about  $\theta$ .

- (b) Show that the Pareto distribution,  $\theta \sim \text{Pareto}(a, b)$ ,

$$f(\theta) = \frac{ab^a}{\theta^{a+1}}, \quad \theta \geq b$$

is a conjugate prior distribution.

Utilising an indicator function, the prior may be expressed as

$$f(\theta) = \frac{ab^a}{\theta^{a+1}} \mathbb{I}_{\{\theta \geq b\}}.$$

The posterior is

$$\begin{aligned} f(\theta | x) &\propto \frac{1}{\theta^n} \mathbb{I}_{\{\theta \geq m\}} \times \frac{ab^a}{\theta^{a+1}} \mathbb{I}_{\{\theta \geq b\}} \\ &\propto \frac{1}{\theta^{a+n+1}} \mathbb{I}_{\{\theta \geq m\}} \mathbb{I}_{\{\theta \geq b\}} \\ &= \frac{1}{\theta^{a+n+1}} \mathbb{I}_{\{\theta \geq \max(b, m)\}} \end{aligned}$$

which is a kernel of a  $Pareto(a + n, \max(b, m))$  density so that  $\theta | x \sim Pareto(a + n, \max(b, m))$ . Thus, the prior and posterior are in the same family and so, relative to the  $U(0, \theta)$  likelihood, the Pareto distribution is a conjugate family.