M3S3/M4S3 : SOLUTIONS 4

1. Equate terms on the two sides of the equation yields the constants, that is, compare coefficients of terms in x^2, x and 1 as follows:

$$x^2 \ : \ A + B = C$$

$$x \ : \ 2Aa + 2Bb = 2Cc \quad \therefore \quad c = \frac{Aa + Bb}{C} = \frac{Aa + Bb}{A + B}$$

$$1 \ : \ Aa^2 + Bb^2 = Cc^2 + d \quad \therefore \quad d = Cc^2 - (Aa^2 + Bb^2) = (A + B)\left(\frac{Aa + Bb}{A + B}\right)^2 - (Aa^2 + Bb^2)$$

and as

$$(A + B)\left(\frac{Aa + Bb}{A + B}\right)^2 - (Aa^2 + Bb^2) = \frac{AB}{A + B}(a - b)^2$$

it follows that

$$A(x - a)^2 + B(x - b)^2 = (A + B)\left(x - \frac{Aa + Bb}{A + B}\right)^2 + \frac{AB}{A + B}(a - b)^2. \qquad (1)$$

For the Bayesian calculation,

$$\begin{aligned} L_n(\mu) &= \prod_{i=1}^{n}\left(\frac{1}{2\pi}\right)^{1/2}\exp\left\{-\frac{1}{2}(x_i - \mu)^2\right\} = \left(\frac{1}{2\pi}\right)^{n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} \\ &= \left(\frac{1}{2\pi}\right)^{n/2}\exp\left\{-\frac{1}{2}\left[n(\overline{x} - \mu)^2 + \sum_{i=1}^{n}(x_i - \overline{x})^2\right]\right\} \\ &\propto \exp\left\{-\frac{n}{2}(\overline{x} - \mu)^2\right\} \\ p_\mu(\mu) &= \left(\frac{1}{2\pi\tau^2}\right)^{1/2}\exp\left\{-\frac{1}{2\tau^2}(\mu - \theta)^2\right\} \propto \exp\left\{-\frac{1}{2\tau^2}(\mu - \theta)^2\right\} \end{aligned}$$

and therefore,

$$p_{\mu|\boldsymbol{X}}(\mu|\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\left[n(\overline{x} - \mu)^2 + \frac{1}{\tau^2}(\mu - \theta)^2\right]\right\} \propto \exp\left\{-\frac{1}{2}\left[\left(n + \frac{1}{\tau^2}\right)\left(\mu - \frac{n\overline{x} + \theta/\tau^2}{n + 1/\tau^2}\right)\right]\right\}$$

after completing the square using the earlier formula in equation (1), with

$$A = n \quad a = \overline{x} \quad B = 1/\tau^2 \quad b = \theta.$$

In the vector case, completing the square is again straightforward, except that we must remember to adhere to the rules of matrix manipulation. For example, the quadratic term in

$$(\boldsymbol{x} - \boldsymbol{c})^\mathsf{T} C(\boldsymbol{x} - \boldsymbol{c}) + \boldsymbol{d} = (\boldsymbol{x} - \boldsymbol{a})^\mathsf{T} A(\boldsymbol{x} - \boldsymbol{a}) + (\boldsymbol{x} - \boldsymbol{b})^\mathsf{T} B(\boldsymbol{x} - \boldsymbol{b})$$

is

$$\boldsymbol{x}^\mathsf{T} A\boldsymbol{x} + \boldsymbol{x}^\mathsf{T} B\boldsymbol{x} = \boldsymbol{x}^\mathsf{T}(A + B)\boldsymbol{x}.$$

so $C = A + B$. The term in $\boldsymbol{x}$ is given by

$$\boldsymbol{c}^\mathsf{T} C\boldsymbol{x} = \boldsymbol{a}^\mathsf{T} A\boldsymbol{x} + \boldsymbol{b}^\mathsf{T} B\boldsymbol{x}$$

therefore

$$C\boldsymbol{c} = A\boldsymbol{a} + B\boldsymbol{b} \qquad \therefore \qquad \boldsymbol{c} = (A+B)^{-1}(A\boldsymbol{a} + B\boldsymbol{b}).$$

Finally

$$\boldsymbol{c}^\top C\boldsymbol{c} + \boldsymbol{d} = \boldsymbol{a}^\top A\boldsymbol{a} + \boldsymbol{b}^\top B\boldsymbol{b}$$

so that

$$\boldsymbol{d} = \boldsymbol{a}^\top A\boldsymbol{a} + \boldsymbol{b}^\top B\boldsymbol{b} - \boldsymbol{c}^\top C\boldsymbol{c} = \boldsymbol{a}^\top A\boldsymbol{a} + \boldsymbol{b}^\top B\boldsymbol{b} - (A\boldsymbol{a} + B\boldsymbol{b})^\top (A+B)^{-1}(A\boldsymbol{a} + B\boldsymbol{b}).$$

2. By definition

$$p_{\boldsymbol{\theta}|\underset{\sim}{X}}(\boldsymbol{\theta}|\underset{\sim}{x}) = \frac{L_n(\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\displaystyle\int L_n(\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \tag{2}$$

where

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^n f_{X|\boldsymbol{\theta}}(x_i|\boldsymbol{\theta}) = \left\{\prod_{i=1}^{n_1} f_{X|\boldsymbol{\theta}}(x_i|\boldsymbol{\theta})\right\}\left\{\prod_{i=n_1+1}^{n_1+n_2} f_{X|\boldsymbol{\theta}}(x_i|\boldsymbol{\theta})\right\} = L_{n_1}(\boldsymbol{\theta})L_{n_2}(\boldsymbol{\theta})$$

say. The numerator in equation (2) is therefore

$$L_n(\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = L_{n_1}(\boldsymbol{\theta})L_{n_2}(\boldsymbol{\theta})p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \propto L_{n_2}(\boldsymbol{\theta})p_{\boldsymbol{\theta}|\underset{\sim}{X}_1}(\boldsymbol{\theta}|\underset{\sim}{x}_1)$$

and the result follows.

3. (i) For the Poisson likelihood/Gamma prior

$$\begin{aligned} L_n(\lambda) &\propto \lambda^s \exp\{-n\lambda\} \\ p_\lambda(\lambda) &\propto \lambda^{\alpha-1}\exp\{-\beta\lambda\} \end{aligned}$$

where $s = \displaystyle\sum_{i=1}^n x_i$, so

$$p_{\lambda|\underset{\sim}{X}}(\lambda|\underset{\sim}{x}) \propto \lambda^{s+\alpha-1}\exp\{-(n+\beta)\lambda\} \equiv Gamma(s+\alpha, n+\beta).$$

Therefore for future variables $\underset{\sim}{X}^\star$ taking values in the Cartesian product

$$\mathbb{X}^{(n^\star)} \equiv \mathbb{X} \times \ldots \times \mathbb{X}$$

where $\mathbb{X} \equiv \{0, 1, 2, \ldots\}$, we have

$$\begin{aligned} f_{\underset{\sim}{X}^\star|\underset{\sim}{X}}(\underset{\sim}{x}^\star|\underset{\sim}{x}) &= \int_0^\infty \left\{\prod_{i=1}^{n^\star} f_{X|\lambda}(x_i^\star|\lambda)\right\} p_{\lambda|\underset{\sim}{X}}(\lambda|\underset{\sim}{x})\, d\lambda \\ &= \int_0^\infty \left\{\frac{\lambda^{s^\star}\exp\{-n^\star\lambda\}}{c(\underset{\sim}{x}^\star)}\right\} \frac{(n+\beta)^{s+\alpha}}{\Gamma(s+\alpha)}\lambda^{s+\alpha-1}\exp\{-(n+\beta)\lambda\}\, d\lambda \\ &= \int_0^\infty \left\{\frac{(n+\beta)^{s+\alpha}}{\Gamma(s+\alpha)c(\underset{\sim}{x}^\star)}\right\}\lambda^{s+s^\star+\alpha-1}\exp\{-(n+n^\star+\beta)\lambda\}\, d\lambda \end{aligned}$$

where $s^\star = \displaystyle\sum_{i=1}^{n^\star} x_i^\star$ and

$$c(\underset{\sim}{x}^\star) = \prod_{i=1}^{n^\star} x_i^\star!.$$

Integrating out λ from this equation yields

$$\begin{aligned} f_{\underset{\sim}{X}^\star|\underset{\sim}{X}}(\underset{\sim}{x}^\star|\underset{\sim}{x}) &= \frac{(n+\beta)^{s+\alpha}}{\Gamma(s+\alpha)c(\underset{\sim}{x}^\star)} \int_0^\infty \lambda^{s+s^\star+\alpha-1} \exp\{-(n+n^\star+\beta)\lambda\}\, d\lambda \\ &= \frac{(n+\beta)^{s+\alpha}}{\Gamma(s+\alpha)c(\underset{\sim}{x}^\star)} \frac{\Gamma(s+s^\star+\alpha)}{(n+n^\star+\beta)^{s+s^\star+\alpha}}. \end{aligned}$$

as the integrand is proportional to a Gamma pdf.

(ii) For the Binomial likelihood/Beta prior

$$\begin{aligned} L_n(\theta) &\propto \theta^s(1-\theta)^{N-s} \\ p_\theta(\theta) &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \end{aligned}$$

where $s = \sum_{i=1}^{n} x_i$ and $N = nK$. Then

$$p_{\theta|\underset{\sim}{X}}(\theta|\underset{\sim}{x}) \propto \theta^{s+\alpha-1}(1-\theta)^{N-s+\beta-1} \equiv Beta(s+\alpha, N-s+\beta).$$

Therefore for future variables $\underset{\sim}{X}^\star$ taking values in the Cartesian product

$$\mathbb{X}^{(n^\star)} \equiv \mathbb{X} \times \ldots \times \mathbb{X}$$

where $\mathbb{X} \equiv \{0, 1, 2, \ldots, K\}$, we have

$$\begin{aligned} f_{\underset{\sim}{X}^\star|\underset{\sim}{X}}(\underset{\sim}{x}^\star|\underset{\sim}{x}) &= \int_0^1 \left\{ \prod_{i=1}^{n^\star} f_{X|\theta}(x_i^\star|\theta) \right\} p_{\theta|\underset{\sim}{X}}(\theta|\underset{\sim}{x})\, d\theta \\ &= \int_0^1 \left\{ c(\underset{\sim}{x}^\star)\theta^{s^\star}(1-\theta)^{N^\star-s^\star} \right\} \frac{\Gamma(N+\alpha+\beta)}{\Gamma(s+\alpha)\Gamma(N-s+\beta)} \theta^{s+\alpha-1}(1-\theta)^{N-s+\beta-1}\, d\theta \\ &= \int_0^1 \frac{c(\underset{\sim}{x}^\star)\Gamma(N+\alpha+\beta)}{\Gamma(s+\alpha)\Gamma(N-s+\beta)} \theta^{s+s^\star+\alpha-1}(1-\theta)^{N+N^\star-s-s^\star+\beta-1}\, d\theta \end{aligned}$$

where $s^\star = \sum_{i=1}^{n^\star} x_i^\star$, $N^\star = n^\star K$ and

$$c(\underset{\sim}{x}^\star) = \prod_{i=1}^{n^\star} \binom{K}{x_i^\star}.$$

Integrating out θ from this equation yields

$$\begin{aligned} f_{\underset{\sim}{X}^\star|\underset{\sim}{X}}(\underset{\sim}{x}^\star|\underset{\sim}{x}) &= \frac{c(\underset{\sim}{x}^\star)\Gamma(N+\alpha+\beta)}{\Gamma(s+\alpha)\Gamma(N-s+\beta)} \int_0^1 \theta^{s+s^\star+\alpha-1}(1-\theta)^{N+N^\star-s-s^\star+\beta-1}\, d\theta \\ &= \frac{c(\underset{\sim}{x}^\star)\Gamma(N+\alpha+\beta)}{\Gamma(s+\alpha)\Gamma(N-s+\beta)} \frac{\Gamma(s+s^\star+\alpha)\Gamma(N+N^\star-s-s^\star+\beta)}{\Gamma(N+N^\star+\alpha+\beta)} \end{aligned}$$

as the integrand is proportional to a Beta pdf.

4. Within the exponential family the density

$$f_{\boldsymbol{X}|\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left\{ \boldsymbol{t}(\boldsymbol{x})^\mathsf{T} \boldsymbol{a}(\boldsymbol{\theta}) + c(\boldsymbol{\theta}) + d(\boldsymbol{x}) \right\}$$

yields a likelihood for iid data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of the form

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{\boldsymbol{X}|\boldsymbol{\theta}}(\boldsymbol{x}_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} \exp\left\{\boldsymbol{t}(\boldsymbol{x}_i)^\top \boldsymbol{a}(\boldsymbol{\theta}) + c(\boldsymbol{\theta}) + d(\boldsymbol{x}_i)\right\}$$
$$= \exp\left\{\boldsymbol{T}^\top \boldsymbol{a}(\boldsymbol{\theta}) + nc(\boldsymbol{\theta}) + \boldsymbol{D}\right\}$$

where

$$\boldsymbol{T} = \sum_{i=1}^{n} \boldsymbol{x}_i \qquad \boldsymbol{D} = \sum_{i=1}^{n} d(\boldsymbol{x}_i).$$

Now, if the prior takes the form

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \propto \exp\left\{\boldsymbol{\alpha}^\top \boldsymbol{a}(\boldsymbol{\theta}) + \beta c(\boldsymbol{\theta})\right\}$$

then

$$p_{\boldsymbol{\theta}|\underset{\sim}{X}}(\boldsymbol{\theta}|\underset{\sim}{x}) \propto L_n(\boldsymbol{\theta}) p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \propto \exp\left\{\{\boldsymbol{T} + \boldsymbol{\alpha}\}^\top \boldsymbol{a}(\boldsymbol{\theta}) + (n + \boldsymbol{\beta})c(\boldsymbol{\theta})\right\} = \left\{\boldsymbol{\alpha}_{new}^\top \boldsymbol{a}(\boldsymbol{\theta}) + \beta_{new} c(\boldsymbol{\theta})\right\}$$

which is of the same functional form as the prior with hyperparameters updated as follows

$$\boldsymbol{\alpha} \longrightarrow \boldsymbol{\alpha}_{new} = \boldsymbol{T} + \boldsymbol{\alpha} \qquad \beta \longrightarrow \beta_{new} = n + \beta$$

5. The likelihood for the model is given by

$$L_n(\mu, \phi) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\phi}\right)^{1/2} \exp\left\{-\frac{1}{2\phi}(x_i - \mu)^2\right\} = \left(\frac{1}{2\pi\phi}\right)^{n/2} \exp\left\{-\frac{1}{2\phi} \sum_{i=1}^{n} (x_i - \mu)^2\right\}$$
$$= \left(\frac{1}{2\pi\phi}\right)^{n/2} \exp\left\{-\frac{1}{2\phi}\left[n(\overline{x} - \mu)^2 + \sum_{i=1}^{n} (x_i - \overline{x})^2\right]\right\}.$$

Therefore a conjugate prior that combines with this likelihood takes the form

$$p_{\mu,\phi}(\mu, \phi) = p_\phi(\phi) p_{\mu|\phi}(\mu|\phi) \propto \left(\frac{1}{\phi}\right)^{\alpha/2+1} \exp\left\{-\frac{\beta}{2\phi}\right\} \left(\frac{1}{\phi}\right)^{1/2} \exp\left\{-\frac{1}{2\gamma\phi}(\mu - \theta)^2\right\}.$$

for hyperparameters $\alpha, \beta, \gamma > 0$ and $\theta \in \mathbb{R}$, where

$$p_\phi(\phi) \propto \left(\frac{1}{\phi}\right)^{\alpha/2+1} \exp\left\{-\frac{\beta}{2\phi}\right\} \equiv IGamma(\alpha/2, \beta/2)$$
$$p_{\mu|\phi}(\mu|\phi) \propto \left(\frac{1}{\phi}\right)^{1/2} \exp\left\{-\frac{1}{2\gamma\phi}(\mu - \theta)^2\right\} \equiv N(\theta, \gamma\phi)$$

and the $IGamma(a, b)$ distribution is the distribution of $1/V$ if $V \sim Gamma(a, b)$ with density function

$$\frac{b^a}{\Gamma(a)} \left(\frac{1}{v}\right)^{a+1} \exp\left\{-\frac{b}{v}\right\}$$

- in the conjugate prior $a = \alpha/2$ and $b = \beta/2$. Therefore, the posterior distribution is

$$\begin{aligned}
p_{\mu,\phi|\underset{\sim}{X}}(\mu,\phi|\underset{\sim}{x}) &\propto L_n(\mu,\phi)p_{\mu,\phi}(\mu,\phi) \\
&\propto \left(\frac{1}{\phi}\right)^{n/2} \exp\left\{-\frac{1}{2\phi}\left[n(\overline{x}-\mu)^2 + S\right]\right\} \\
&\qquad \times \left(\frac{1}{\phi}\right)^{\alpha/2+1} \exp\left\{-\frac{\beta}{2\phi}\right\} \left(\frac{1}{\phi}\right)^{1/2} \exp\left\{-\frac{1}{2\gamma\phi}(\mu-\theta)^2\right\} \\
&= \left(\frac{1}{\phi}\right)^{(n+\alpha+1)/2+1} \exp\left\{-\frac{1}{2\phi}\left[n(\overline{x}-\mu)^2 + S + \beta + \frac{1}{\gamma}(\mu-\theta)^2\right]\right\}
\end{aligned}$$

where

$$S = \sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Now, using the identity from Q1, completing the square in μ, we have

$$n(\overline{x}-\mu)^2 + \frac{1}{\gamma}(\mu-\theta)^2 = M(\mu-m)^2 + d$$

where

$$M = n + 1/\gamma \qquad m = \frac{n\overline{x} + \theta/\gamma}{n + 1/\gamma} \qquad d = \frac{n/\gamma}{n+1/\gamma}(\overline{x}-\theta)^2.$$

Therefore the posterior distribution takes the form

$$\begin{aligned}
p_{\mu,\phi|\underset{\sim}{X}}(\mu,\phi|\underset{\sim}{x}) &\propto \left(\frac{1}{\phi}\right)^{(n+\alpha+1)/2+1} \exp\left\{-\frac{1}{2\phi}\left[M(\mu-m)^2 + d + S + \beta\right]\right\} \\
&\propto p_{\phi|\underset{\sim}{X}}(\phi|\underset{\sim}{x})p_{\mu|\underset{\sim}{X},\phi}(\mu|\underset{\sim}{x},\phi)
\end{aligned}$$

where, by inspection

$$\begin{aligned}
p_{\phi|\underset{\sim}{X}}(\phi|\underset{\sim}{x}) &\propto \left(\frac{1}{\phi}\right)^{(n+\alpha)/2+1} \exp\left\{-\frac{(d+S+\beta)}{2\phi}\right\} \equiv IGamma((n+\alpha)/2, (d+S+\beta)/2) \\
p_{\mu|\underset{\sim}{X},\phi}(\mu|\underset{\sim}{x},\phi) &\propto \left(\frac{1}{\phi}\right)^{1/2} \exp\left\{-\frac{M}{2\phi}(\mu-m)^2\right\} \equiv N(m,\phi/M).
\end{aligned}$$

Therefore, in the conjugate analysis,

$$\begin{aligned}
p_{\phi|\underset{\sim}{X}}(\phi|\underset{\sim}{x}) &\equiv IGamma(\alpha_1/2, \beta_1/2) \\
p_{\mu|\underset{\sim}{X},\phi}(\mu|\underset{\sim}{x},\phi) &\equiv N(\theta_1, \gamma_1\phi)
\end{aligned}$$

where

$$\alpha_1 = n + \alpha \qquad \beta_1 = d + S + \beta \qquad \theta_1 = \frac{n\overline{x} + \theta/\gamma}{n+1/\gamma} \qquad \gamma_1 = 1/M$$

so that

$$\begin{aligned}
p_{\mu,\phi|\underset{\sim}{X}}(\mu,\phi|\underset{\sim}{x}) &= \frac{(\beta_1/2)^{\alpha_1/2}}{\Gamma(\alpha_1/2)}\left(\frac{1}{\phi}\right)^{\alpha_1/2+1} \exp\left\{-\frac{\beta_1}{2\phi}\right\}\left(\frac{1}{2\pi\gamma_1\phi}\right)^{1/2} \exp\left\{-\frac{1}{2\gamma_1\phi}(\mu-\theta_1)^2\right\} \\
&= \frac{(\beta_1/2)^{\alpha_1/2}}{\Gamma(\alpha_1/2)}\left(\frac{1}{2\pi\gamma_1}\right)^{1/2}\left(\frac{1}{\phi}\right)^{(\alpha_1+1)/2+1} \exp\left\{-\frac{1}{2\phi}\left[\beta_1 + \frac{1}{\gamma_1}(\mu-\theta_1)^2\right]\right\}.
\end{aligned}$$

Hence

$$\begin{aligned}
p_{\mu|\underset{\sim}{X}}(\mu|\underset{\sim}{x}) &= \int_0^\infty p_{\mu,\phi|\underset{\sim}{X}}(\mu,\phi|\underset{\sim}{x})\, d\phi \\
&= \int_0^\infty \frac{(\beta_1/2)^{\alpha_1/2}}{\Gamma(\alpha_1/2)} \left(\frac{1}{2\pi\gamma_1}\right)^{1/2} \left(\frac{1}{\phi}\right)^{(\alpha_1+1)/2+1} \exp\left\{-\frac{1}{2\phi}\left[\beta_1 + \frac{1}{\gamma_1}(\mu-\theta_1)^2\right]\right\} d\phi \\
&= \frac{(\beta_1/2)^{\alpha_1/2}}{\Gamma(\alpha_1/2)} \left(\frac{1}{2\pi\gamma_1}\right)^{1/2} \int_0^\infty \left(\frac{1}{\phi}\right)^{(\alpha_1+1)/2+1} \exp\left\{-\frac{1}{2\phi}\left[\beta_1 + \frac{1}{\gamma_1}(\mu-\theta_1)^2\right]\right\} d\phi \\
&= \frac{(\beta_1/2)^{\alpha_1/2}}{\Gamma(\alpha_1/2)} \left(\frac{1}{2\pi\gamma_1}\right)^{1/2} \frac{\Gamma((\alpha_1+1)/2)}{\left(\dfrac{\beta_1}{2} + \dfrac{(\mu-\theta_1)^2}{2\gamma_1}\right)^{(\alpha_1+1)/2}} \\
&= \frac{\Gamma((\alpha_1+1)/2)}{\Gamma(\alpha_1/2)\Gamma(1/2)} \gamma_1^{\alpha_1} \left(\beta_1\gamma_1 + (\mu-\theta_1)^2\right)^{-(\alpha_1+1)/2}
\end{aligned}$$

as $\Gamma(1/2) = \pi^{1/2}$, and the integrand is proportional to a Gamma density. Thus the posterior marginal distribution for μ is in fact a generalized Student-t distribution with parameters $(\alpha_1, \theta_1, \beta_1, \gamma_1)$.

6. (i) If

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2}$$

where here (and in the rest of the question) $|A| = |\det A|$, the absolute value of the determinant of A, and $\boldsymbol{\phi} = \boldsymbol{g}(\boldsymbol{\theta})$ so that $\boldsymbol{\theta} = \boldsymbol{h}(\boldsymbol{\phi})$, say, then by the multivariate transformation theorem,

$$p_{\boldsymbol{\phi}}(\boldsymbol{\phi}) \propto |I(\boldsymbol{h}(\boldsymbol{\phi}))|^{1/2}\, |J_{\boldsymbol{\phi}}| \tag{3}$$

where $J_{\boldsymbol{\phi}}$ is the Jacobian for the transformation between $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, that is

$$[J_{\boldsymbol{\phi}}]_{jk} = \frac{\partial\theta_j}{\partial\phi_k} = \frac{\partial h_j(\boldsymbol{\phi})}{\partial\phi_k}$$

We know that the reparameterization is a 1-1 (and hence invertible) transformation

Now, the Fisher information for $\boldsymbol{\phi}$ can also be found from first principles. As usual, it is formed as the expectation of the matrix of second derivatives of the log-likelihood with $(j,k)^{th}$ element

$$\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k} = \frac{\partial}{\partial\theta_k}\left\{\frac{\partial l_n(\boldsymbol{\theta})}{\partial\theta_j}\right\}.$$

Now,

$$\begin{aligned}
\frac{\partial^2 l_n(\boldsymbol{\phi})}{\partial\phi_j\partial\phi_k} = \frac{\partial}{\partial\phi_k}\left\{\frac{\partial l_n(\boldsymbol{\phi})}{\partial\phi_j}\right\} &= \frac{\partial}{\partial\phi_k}\left\{\frac{\partial l_n(\boldsymbol{\theta})}{\partial\theta_j}\frac{\partial\theta_j}{\partial\phi_j}\right\} \\
&= \frac{\partial}{\partial\phi_k}\left\{\frac{\partial l_n(\boldsymbol{\theta})}{\partial\theta_j}\right\}\frac{\partial\theta_j}{\partial\phi_j} + \frac{\partial l_n(\boldsymbol{\theta})}{\partial\theta_j}\frac{\partial^2\theta_j}{\partial\phi_j\partial\phi_k} \\
&= \left\{\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right\}\frac{\partial\theta_j}{\partial\phi_j}\frac{\partial\theta_k}{\partial\phi_k} + \frac{\partial l_n(\boldsymbol{\theta})}{\partial\theta_j}\frac{\partial^2\theta_j}{\partial\phi_j\partial\phi_k}
\end{aligned}$$

and thus, taking expectations to compute the Fisher information, as

$$E_{f_{X|\boldsymbol{\theta}}}\left[\frac{\partial l_n(\boldsymbol{\theta})}{\partial\theta_j}\right] = 0 \qquad \forall\, j$$

by properties of the score function, and thus

$$E_{f_{X|\boldsymbol{\theta}}}\left[\frac{\partial^2 l_n(\boldsymbol{\phi})}{\partial\phi_j\partial\phi_k}\right] = E_{f_{X|\boldsymbol{\theta}}}\left[\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right]\frac{\partial\theta_j}{\partial\phi_j}\frac{\partial\theta_k}{\partial\phi_k} = \frac{\partial\theta_j}{\partial\phi_j}\, E_{f_{X|\boldsymbol{\theta}}}\left[\frac{\partial^2 l_n(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right]\frac{\partial\theta_k}{\partial\phi_k}$$

and hence

$$I(\boldsymbol{\phi}) = D(\boldsymbol{\theta})^{\mathsf{T}} I(\boldsymbol{\theta}) D(\boldsymbol{\theta})$$

where D is the square matrix with $(j,k)^{th}$ element

$$\frac{\partial\theta_j}{\partial\phi_k}.$$

Hence

$$|I(\boldsymbol{\phi})| = \left|D(\boldsymbol{\theta})^{\mathsf{T}} I(\boldsymbol{\theta}) D(\boldsymbol{\theta})\right| = \left|D(\boldsymbol{\theta})^{\mathsf{T}}\right| |I(\boldsymbol{\theta})|\,|D(\boldsymbol{\theta})| = |I(\boldsymbol{\theta})|\,|D(\boldsymbol{\theta})|^2$$

and

$$|I(\boldsymbol{\phi})|^{1/2} = |I(\boldsymbol{\theta})|^{1/2}\,|D(\boldsymbol{\theta})|\,.$$

However, by inspection of equation (3), we note that as

$$|D(\boldsymbol{\theta})| = |J_{\boldsymbol{\phi}}|$$

so in fact

$$p_{\boldsymbol{\phi}}(\boldsymbol{\phi}) \propto |I(\boldsymbol{\phi})|^{1/2}$$

and we see that Jeffrey's prior is **invariant** to reparameterization.

(ii) For the normal model, using parameters $\boldsymbol{\theta} = (\mu, \sigma^2)^{\mathsf{T}}$

$$f_{X|\boldsymbol{\theta}}(x|\boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

and hence

$$\begin{aligned}
l(\mu,\sigma^2) &= \text{const} - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2 \\
\frac{\partial l}{\partial\mu} &= \frac{x-\mu}{\sigma^2} \\
\frac{\partial^2 l}{\partial\mu^2} &= -\frac{1}{\sigma^2} \\
\frac{\partial l}{\partial\sigma^2} &= -\frac{1}{2\sigma^2} - \frac{1}{2\sigma^4}(x-\mu)^2 \\
\frac{\partial^2 l}{\partial(\sigma^2)^2} &= \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-\mu)^2 \\
\frac{\partial^2 l}{\partial\mu\partial\sigma^2} &= -\frac{x-\mu}{\sigma^4}
\end{aligned}$$

Hence

$$I(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{1}{\sigma^2} & 0 \\ 0 & \dfrac{1}{2\sigma^4} \end{bmatrix}$$

and Jeffrey's prior is

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2} \propto \frac{1}{\sigma^3} \tag{4}$$

Instead, using parameters $\boldsymbol{\phi} = (\mu, \sigma)^{\mathsf{T}}$

$$\begin{aligned} l(\mu, \sigma) &= \text{const} - \log \sigma - \frac{1}{2\sigma^2}(x-\mu)^2 \\ \frac{\partial l}{\partial \sigma} &= -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2 \\ \frac{\partial^2 l}{\partial \sigma^2} &= \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(x-\mu)^2 \end{aligned}$$

Hence

$$I(\boldsymbol{\phi}) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$$

and Jeffrey's prior is

$$p_{\boldsymbol{\phi}}(\boldsymbol{\phi}) \propto |I(\boldsymbol{\phi})|^{1/2} \propto \frac{1}{\sigma^2}. \tag{5}$$

This demonstrates that the earlier result relating to reparamterization holds; the prior is preserved. To see this, let $\lambda = \sigma^2$; the original prior in equation (4) says that

$$p_{\boldsymbol{\theta}}(\mu, \lambda) = \frac{c_1}{\lambda^{3/2}}.$$

Under the reparameterization $\lambda \longrightarrow \sigma = \sqrt{\lambda}$, the Jacobian of the transformation is 2σ

$$p_{\boldsymbol{\theta}}(\mu, \sigma) = \frac{c}{\sigma^3} \times 2\sigma = \frac{c_2}{\sigma^2}$$

which coincides with equation (5).