Jeffrey’s prior derivations

Contents

1 Proof of invariance under transformation 2

2 Binomial probability 3

3 Binomial odds 4

4 Poisson rate 5

5 Normal mean 6

6 Normal variance 7

7 Normal standard deviation 8

8 Normal mean and variance 9

9 Linear regression with unknown variance 10

10 Marginal distribution of a normal mean 11

11 Marginal posterior of the regression coefficients 13

1 Proof of invariance under transformation

Let $\gamma = g(\theta)$ be a transformation of the parameter θ. Below we show that directly putting a Jeffreys prior (JP) on γ is equivalent to placing a JP prior on θ and then transforming to γ. This is the one-dimensional case, with multiple parameters the same steps apply but with Jacobian matrices.

Defining $l(Y|\theta) = \log[f(Y|\theta)]$, the JP for θ is

$$\pi_1(\theta) \propto \sqrt{\mathcal{I}_1(\theta)} \propto \sqrt{-\mathrm{E}_{Y|\theta}\left[\frac{d^2 l(Y|\theta)}{d\theta^2}\right]}.$$

Similarly, the JP for γ is

$$\pi_2(\gamma) \propto \sqrt{\mathcal{I}_2(\gamma)} \propto \sqrt{-\mathrm{E}_{Y|\gamma}\left[\frac{d^2 l(Y|\gamma)}{d\gamma^2}\right]}.$$

To connect the two priors, write $\mathcal{I}_1(\theta)$ in terms of γ. The second-order chain rule gives

$$\frac{d^2 l(Y|\theta)}{d\theta^2} = \left(\frac{d^2 l(Y|\gamma)}{d\gamma^2}\right)\left(\frac{d\gamma}{d\theta}\right)^2 + \left(\frac{dl(Y|\gamma)}{d\gamma}\right)\left(\frac{d^2\gamma}{d\theta^2}\right).$$

The expected value with respect to $f(Y|\gamma)$ of second term is zero, since

$$\mathrm{E}_{Y|\gamma}\left[\frac{dl(Y|\gamma)}{d\gamma}\right] = \mathrm{E}_{Y|\gamma}\left[\frac{df(Y|\gamma)/d\gamma}{f(Y|\gamma)}\right] = \int\left[\frac{df(Y|\gamma)/d\gamma}{f(Y|\gamma)}\right] f(Y|\gamma)dY = \int \frac{df(Y|\gamma)}{d\gamma}dY$$

and if we exchange integration and differentiation,

$$\frac{d\int f(Y|\gamma)dY}{d\gamma} = 0$$

since $\int f(Y|\gamma)dY = 1$. Returning to the information,

$$\mathcal{I}_1(\theta) = -\mathrm{E}\left[\left(\frac{d^2 l(Y|\gamma)}{d\gamma^2}\right)\left(\frac{d\gamma}{d\theta}\right)^2\right] = \mathcal{I}_2(\gamma)\left(\frac{d\gamma}{d\theta}\right)^2.$$

Therefore, if we start with a JP π_1 on θ and perform a change of variables to γ, we get prior

$$\pi_3(\gamma) \propto \sqrt{\mathcal{I}_1(\gamma)}\frac{d\theta}{d\gamma} = \sqrt{\mathcal{I}_2(\gamma)} \propto \pi_2(\gamma).$$

Thus shows that a JP prior on θ and transforming to γ is equivalent to placing a JP directly on γ.

2 Binomial probability

The model is $Y|\theta \sim \text{Binomial}(n, \theta)$. This gives log-likelihood

$$l(Y|\theta) = c + Y \log(\theta) + (n - Y) \log(1 - \theta) \tag{1}$$

for constant c that does not depend on θ. The first derivative is

$$l'(Y|\theta) = \frac{Y}{\theta} - \frac{n - Y}{1 - \theta} \tag{2}$$

and the second derivative is

$$l''(Y|\theta) = -\frac{Y}{\theta^2} - \frac{n - Y}{(1 - \theta)^2}. \tag{3}$$

This gives expected information (recalling $\text{E}(Y|\theta) = n\theta$)

$$\mathcal{I}(\theta) = -\text{E}\left[l''(Y|\theta)\right] = \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1 - \theta)^2} = \frac{n}{\theta} + \frac{n}{1 - \theta} = \frac{n}{\theta(1 - \theta)}. \tag{4}$$

The JP is thus

$$\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)} \propto \theta^{-1/2}(1 - \theta)^{-1/2} \propto \theta^{1/2-1}(1 - \theta)^{1/2-1} \tag{5}$$

and so $\theta \sim \text{Beta}(1/2, 1/2)$.

3 Binomial odds

Say $Y|\theta \sim$ Binomial(n, θ) but our primary interest is in the odds, $\gamma = \theta/(1-\theta) > 0$. Solving for θ gives $\theta = \gamma/(1+\gamma)$. The model written in terms of γ is $Y|\gamma \sim$ Binomial$(n, \gamma/(1+\gamma))$. The JP for γ can be derived two ways.

(1) Using a change of variables: Since we know the JP for θ is a proper PDF and JPs are invariant to transformation, we could simply use the JP for θ and the univariate change of variables formula to arrive at the JP for γ. Using the change of variables formula (i.e., wikipedia), if $\theta \sim$ Beta(a, b), then $\gamma = \theta/(1-\theta)$ follows a BetaPrime(a, b) distribution with PDF

$$\pi(\gamma) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\gamma^{a-1}(1+\gamma)^{-a-b}.$$

Using $a = b = 1/2$ gives JP

$$\pi(\gamma) \propto \gamma^{-1/2}(1+\gamma)^{-1}.$$

(2) Derivation from scratch: The log-likelihood is

$$\begin{aligned} l(Y|\gamma) &= c + Y\log\{\gamma/(1+\gamma)\} + (n-Y)\log\{1-\gamma/(1+\gamma)\} \\ &= c + Y\log(\gamma) - Y\log(1+\gamma) - (n-Y)\log(1+\gamma) \\ &= c + Y\log(\gamma) - n\log(1+\gamma) \end{aligned}$$

for constant c that does not depend on γ. The derivatives are

$$l'(Y|\gamma) = \frac{Y}{\gamma} - \frac{n}{1+\gamma} \quad \text{and} \quad l''(Y|\gamma) = -\frac{Y}{\gamma^2} + \frac{n}{(1+\gamma)^2}. \tag{6}$$

This gives expected information (recalling $\mathrm{E}(Y|\gamma) = n\gamma/(1+\gamma)$)

$$\mathcal{I}(\gamma) = -\mathrm{E}\left[l''(Y|\gamma)\right] = \frac{n\gamma}{1+\gamma}\frac{1}{\gamma^2} - \frac{n}{(1+\gamma)^2} = n\frac{1+\gamma-\gamma}{\gamma(1-\gamma)^2} = n\gamma^{-1}(1+\gamma)^{-2}. \tag{7}$$

The Jeffreys' prior is thus

$$\pi(\gamma) \propto \sqrt{\mathcal{I}(\gamma)} \propto \gamma^{-1/2}(1+\gamma)^{-1} \tag{8}$$

and so, as using the change of variables formula, $\gamma \sim$ BetaPrime$(1/2, 1/2)$.

4 Poisson rate

The model is $Y|\theta \sim \text{Poisson}(\theta)$. This gives log-likelihood

$$l(Y|\theta) = c + Y\log(\theta) - \theta \tag{9}$$

for constant c that does not depend on θ. The first derivative is

$$l'(Y|\theta) = \frac{Y}{\theta} - 1 \tag{10}$$

and the second derivative is

$$l''(Y|\theta) = -\frac{Y}{\theta^2}. \tag{11}$$

This gives expected information (recalling $\text{E}(Y|\theta) = \theta$)

$$\mathcal{I}(\theta) = -\text{E}\left[l''(Y|\theta)\right] = \frac{\theta}{\theta^2} = \frac{1}{\theta}. \tag{12}$$

The Jeffreys' prior is thus

$$\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)} \propto \theta^{-1/2}. \tag{13}$$

This is an improper prior. It can be seen as the limiting distribution of the prior $\theta \sim \text{Gamma}(1/2, b)$ for b tending to zero.

5 Normal mean

The model is $Y_i|\mu \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$ with σ known. This gives log-likelihood

$$l(Y|\mu) = c - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2 \tag{14}$$

for constant c that does not depend on μ. The first derivative is

$$l'(Y|\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \mu) \tag{15}$$

and the second derivative is

$$l''(Y|\mu) = -n/\sigma^2. \tag{16}$$

This gives expected information $\mathcal{I}(\mu) = -\text{E}\left[l''(Y|\mu)\right] = n/\sigma^2$. The Jeffreys' prior is thus

$$\pi(\mu) \propto \sqrt{\mathcal{I}(\mu)} \propto \frac{n^{1/2}}{\sigma} \propto 1 \tag{17}$$

and so $\pi(\mu) \propto 1$ for all μ.

6 Normal variance

The model is $Y_i|v \overset{iid}{\sim} \text{Normal}(\mu, v)$ for known μ. This gives log-likelihood

$$l(Y|v) = c - \frac{n}{2}\log(v) - \frac{1}{2v}\sum_{i=1}^{n}(Y_i - \mu)^2 \tag{18}$$

for constant c that does not depend on v. The first derivative is

$$l'(Y|v) = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(Y_i - \mu)^2 \tag{19}$$

and the second derivative is

$$l''(Y|v) = \frac{n}{2v^2} - \frac{1}{v^3}\sum_{i=1}^{n}(Y_i - \mu)^2. \tag{20}$$

This gives expected information (recalling $\mathrm{E}\{(Y-\mu)^2|v) = v\}$)

$$\mathcal{I}(v) = -\mathrm{E}\left[l''(Y|v)\right] = -\frac{n}{2v^2} + \frac{1}{v^3}(nv) = \frac{n}{2v^2}. \tag{21}$$

The Jeffreys' prior is thus

$$\pi(v) \propto \sqrt{\mathcal{I}(v)} \propto \frac{1}{v}. \tag{22}$$

Typically we write $v = \sigma^2$ in which case $\pi(\sigma^2) \propto 1/\sigma^2$.

7 Normal standard deviation

The model is $Y_i|\sigma \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$ for known μ. This gives log-likelihood

$$l(Y|\sigma) = c - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2 \tag{23}$$

for constant c that does not depend on σ. The first derivative is

$$l'(Y|\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(Y_i - \mu)^2 \tag{24}$$

and the second derivative is

$$l''(Y|\sigma) = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(Y_i - \mu)^2. \tag{25}$$

This gives expected information (recalling $\text{E}\{(Y-\mu)^2|\sigma) = \sigma^2\}$)

$$\mathcal{I}(\sigma) = -\text{E}\left[l''(Y|\sigma)\right] = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4}(n\sigma^2) = \frac{2n}{\sigma^2}. \tag{26}$$

The Jeffreys' prior is thus

$$\pi(\sigma) \propto \sqrt{\mathcal{I}(\sigma)} \propto \frac{1}{\sigma}. \tag{27}$$

We now have JPs for the variance and standard deviation. Since the JP is invariant to transformation these should be equivalent. To see this, start with $\pi(\sigma)$ above and transform to $v = \sigma^2$. The prior for v is

$$\pi_v(v) = \pi_\sigma(\sigma)\left|\frac{d\sigma}{dv}\right| \propto \frac{1}{\sigma}\left|\frac{d\sqrt{v}}{dv}\right| \propto \frac{1}{\sqrt{v}}\left|\frac{1}{\sqrt{v}}\right| \propto \frac{1}{v},$$

which is the JP for the variance, v.

8 Normal mean and variance

The model is $Y_i|\mu, v \overset{iid}{\sim}$ Normal(μ, v) (usually we write $v = \sigma^2$). This gives log-likelihood

$$l(Y|\mu,\sigma) = c - \frac{n}{2}\log(v) - \frac{1}{2v}\sum_{i=1}^{n}(Y_i-\mu)^2 \tag{28}$$

for constant c that does not depend on μ or v. The first derivatives are

$$\frac{\partial l(Y|\mu,v)}{\partial\mu} = \frac{1}{v}\sum_{i=1}^{n}(Y_i-\mu) \quad \text{and} \quad \frac{\partial l(Y|\mu,v)}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(Y_i-\mu)^2. \tag{29}$$

The second-order derivatives are

$$\frac{\partial^2 l(Y|\mu,v)}{\partial\mu^2} = -n/v \tag{30}$$

$$\frac{\partial^2 l(Y|\mu,v)}{\partial v^2} = \frac{n}{2v^2} - \frac{1}{v^3}\sum_{i=1}^{n}(Y_i-\mu)^2 \tag{31}$$

$$\frac{\partial^2 l(Y|\mu,v)}{\partial\mu\partial v} = -\frac{1}{v^2}\sum_{i=1}^{n}(Y_i-\mu). \tag{32}$$

Thus the expected information has elements

$$-\mathrm{E}\left[\frac{\partial^2 l(Y|\mu,v)}{\partial\mu^2}\right] = n/v \tag{33}$$

$$-\mathrm{E}\left[\frac{\partial^2 l(Y|\mu,v)}{\partial v^2}\right] = -\frac{n}{2v^2} + \frac{1}{v^3}nv = \frac{n}{2v^2} \tag{34}$$

$$-\mathrm{E}\left[\frac{\partial^2 l(Y|\mu,v)}{\partial\mu\partial v}\right] = 0. \tag{35}$$

Therefore, $\mathcal{I}(\mu,v)$ is diagonal with diagonal elements n/v and $n/(2v^2)$, so its determinent is $|\mathcal{I}(\mu,v)| = n^2/(2v^3)$ and the JP is

$$\pi(\mu,v) \propto \sqrt{\mathcal{I}(\mu,v)} \propto v^{-3/2}. \tag{36}$$

9 Linear regression with unknown variance

The model is $Y_i|\boldsymbol{\beta}, v \overset{iid}{\sim}$ Normal$(\mathbf{X}_i\boldsymbol{\beta}, v)$. This gives log-likelihood

$$l(\mathbf{Y}|\boldsymbol{\beta}, v) = c - \frac{n}{2}\log(v) - \frac{1}{2v}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2 \tag{37}$$

for constant c that does not depend on $\boldsymbol{\beta}$ or v. The first derivatives are

$$\frac{\partial l(\mathbf{Y}|\boldsymbol{\beta}, v)}{\partial \beta_j} = \frac{1}{v}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\boldsymbol{\beta})X_{ij} \quad \text{and} \quad \frac{\partial l(Y|\mu, v)}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2. \tag{38}$$

The second-order derivatives are

$$\frac{\partial^2 l(Y|\boldsymbol{\beta}, v)}{\partial \beta_j \partial \beta_k} = -\sum_{i=1}^{n} X_{ij}X_{ik}/v \tag{39}$$

$$\frac{\partial^2 l(Y|\boldsymbol{\beta}, v)}{\partial v^2} = \frac{n}{2v^2} - \frac{1}{v^3}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2 \tag{40}$$

$$\frac{\partial^2 l(Y|\beta, v)}{\partial \beta_j \partial v} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \mathbf{X}_i\boldsymbol{\beta})X_{ij}. \tag{41}$$

This gives expected information has elements

$$-\mathrm{E}\left[\frac{\partial^2 l(Y|\boldsymbol{\beta}, v)}{\partial \beta_j \partial \beta_k}\right] = \sum_{i=1}^{n} X_{ij}X_{ik}/v \tag{42}$$

$$-\mathrm{E}\left[\frac{\partial^2 l(Y|\boldsymbol{\beta}, v)}{\partial v^2}\right] = -\frac{n}{2v^2} + \frac{1}{v^3}nv = \frac{n}{2v^2} \tag{43}$$

$$-\mathrm{E}\left[\frac{\partial^2 l(Y|\boldsymbol{\beta}, v)}{\partial \beta_j \partial v}\right] = 0. \tag{44}$$

Therefore, the $(p+1) \times (p+1)$ information matrix is

$$\mathcal{I}(\mu, v) = \begin{pmatrix} \sum_{i=1}^{n} \mathbf{X}_i^T\mathbf{X}_i/v & 0 \\ 0 & 2n/v^2 \end{pmatrix}$$

and its determinant is proportional to $v^{-(p+2)}$, giving JP

$$\pi(\boldsymbol{\beta}, v) \propto \sqrt{\mathcal{I}(\boldsymbol{\beta}, v)} \propto v^{-(p+2)/2}. \tag{45}$$

10 Marginal distribution of a normal mean

Assume $Y_i \overset{iid}{\sim}$ Normal(μ, σ^2) and Jeffreys' prior $\pi(\mu, \sigma^2) \propto (\sigma^2)^{-3/2}$. Letting $\bar{Y} = \sum_{i=1}^n Y_i/n$ and $\hat{\sigma}^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2/n$, we show that

$$\mu|\mathbf{Y} \sim t_n\left(\bar{Y}, \hat{\sigma}/\sqrt{n}\right),$$

i.e., a Student t distribution with location $\bar{\mathbf{Y}}$, scale $\hat{\sigma}\sqrt{n}$ and n degrees of freedom.

Denoting $\tau = \sigma^2$, the joint posterior is

$$\begin{aligned}
p(\mu, \tau|\mathbf{Y}) &\propto \left\{\tau^{-n/2}\exp\left[-\frac{\sum_{i=1}^n (Y_i - \mu)^2}{2\tau}\right]\right\}\left\{\tau^{-3/2}\right\} \\
&\propto \tau^{-(n+1)/2-1}\exp\left[-\frac{\sum_{i=1}^n (Y_i - \mu)^2}{2\tau}\right] \\
&\propto \tau^{-A-1}\exp\left[-\frac{B}{\tau}\right],
\end{aligned}$$

where $A = (n+1)/2$ and $B = \sum_{i=1}^n (Y_i - \mu)^2/2$. As a function of τ, the joint distribution resembles an InvGamma(A, B) PDF. Integrating over τ gives

$$\begin{aligned}
p(\mu|\mathbf{Y}) &\propto \int p(\mu, \tau|bY)d\tau \\
&\propto \int \tau^{-A-1}\exp(-B/\tau)d\tau \\
&\propto \frac{\Gamma(A)}{B^A}\int \frac{B^A}{\Gamma(A)}\tau^{-A-1}\exp(-B/\tau)d\tau \\
&\propto \frac{\Gamma(A)}{B^A} \\
&\propto \left[\sum_{i=1}^n (Y_i - \mu)^2\right]^{-(n+1)/2}.
\end{aligned}$$

The marginal PDF is a quadratic function of μ raised to the power $-(n+1)/2$, suggesting that the posterior is a t distribution with n degrees of freedom. Completing the square gives

$$\begin{aligned}
\sum_{i=1}^{n}(Y_i-\mu)^2 &= \sum_{i=1}^{n} Y_i^2 - 2\sum_{i=1}^{n} Y_i\mu + n\mu^2 \\
&= n\left[\sum_{i=1}^{n} Y_i^2/n - 2\bar{Y}\mu + \mu^2\right] \\
&= n\left[\sum_{i=1}^{n} Y_i^2/n - \bar{Y}^2 + \bar{Y}^2 - 2\bar{Y}\mu + \mu^2\right] \\
&= n\left[\sum_{i=1}^{n} Y_i^2/n - \bar{Y}^2 + (\mu-\bar{Y})^2\right] \\
&= n\left[\hat{\sigma}^2 + (\mu-\bar{Y})^2\right],
\end{aligned}$$

since $\hat{\sigma}^2 = \sum_{i=1}^{n}(Y_i-\bar{Y})^2/n = \sum_{i=1}^{n} Y_i^2/n - \bar{Y}^2$. Inserting this expression back into the marginal posterior gives

$$\begin{aligned}
p(\mu|\mathbf{Y}) &\propto \left[\sum_{i=1}^{n}(Y_i-\mu)^2\right]^{-(n+1)/2} \\
&\propto \left[\hat{\sigma}^2 + (\mu-\bar{Y})^2\right]^{-(n+1)/2} \\
&\propto \left[1 + \frac{1}{n}\left(\frac{\mu-\bar{Y}}{\hat{\sigma}/\sqrt{n}}\right)^2\right]^{-(n+1)/2}.
\end{aligned}$$

This is Student's t distribution with location parameter $\bar{Y}$, scale parameter $\hat{\sigma}/\sqrt{n}$, and n degrees of freedom.

11 Marginal posterior of the regression coefficients

Assume $\mathbf{Y}|\boldsymbol{\beta}, \sigma^2 \sim \text{Normal}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$ and Jeffreys' prior $\pi(\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-p/2-1}$. Letting $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ and $\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/n$, we show that

$$\boldsymbol{\beta}|\mathbf{Y} \sim t_n\left\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}\right\},$$

i.e., p-dimensional t-distribution with location vector $\hat{\boldsymbol{\beta}}$, scale matrix $\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}$ and n degrees of freedom.

Denoting $\tau = \sigma^2$, the joint posterior is

$$\begin{aligned} p(\boldsymbol{\beta}, \tau|\mathbf{Y}) &\propto \left\{\tau^{-n/2}\exp\left[-\frac{1}{2\tau}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right]\right\}\tau^{-p/2-1} \\ &\propto \tau^{-A-1}\exp\left[-\frac{B}{\tau}\right], \end{aligned}$$

where $A = (n+p)/2$ and $B = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})/2$. Marginalizing over σ^2 gives

$$\begin{aligned} p(\boldsymbol{\beta}|\mathbf{Y}) &= \int p(\boldsymbol{\beta}, \tau|\mathbf{Y})d\tau \\ &\propto \frac{\Gamma(A)}{B^A}\int \frac{B^A}{\Gamma(A)}\tau^{-A-1}\exp\left[-\frac{B}{\tau}\right]d\tau \\ &\propto B^{-A} \\ &\propto \left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right]^{-(n+p)/2}. \end{aligned}$$

The quadratic form is factored as

$$\begin{aligned} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) &= \mathbf{Y}^T\mathbf{Y} - 2\mathbf{Y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{W}\boldsymbol{\beta} \\ &= \mathbf{Y}^T\mathbf{Y} - 2\hat{\boldsymbol{\beta}}^T\mathbf{W}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{W}\boldsymbol{\beta} \\ &= \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\beta}}^T\mathbf{W}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}^T\mathbf{W}\hat{\boldsymbol{\beta}} - 2\hat{\boldsymbol{\beta}}^T\mathbf{W}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{W}\boldsymbol{\beta} \\ &= n\hat{\sigma}^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{W}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \end{aligned}$$

where $\mathbf{W} = \mathbf{X}^T\mathbf{X}$ and $n\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\beta}}^T\mathbf{W}\hat{\boldsymbol{\beta}}$. Therefore,

$$\begin{aligned} p(\boldsymbol{\beta}|\mathbf{Y}) &\propto \left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right]^{-(n+p)/2} \\ &\propto \left[n\hat{\sigma}^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{W}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]^{-(n+p)/2} \\ &\propto \left[1 + \frac{1}{n\hat{\sigma}^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\mathbf{W}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]^{-(n+p)/2}. \end{aligned}$$

The marginal posterior of $\boldsymbol{\beta}$ is thus the p-dimensional t-distribution with location vector $\hat{\boldsymbol{\beta}}$, scale matrix $\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}$, and n degrees of freedom.