Linear Regression

The classical linear model is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y}$ is $n \times 1$, $\boldsymbol{X}$ is $n \times p$, $\boldsymbol{\beta}$ is $p \times 1$ and

$$\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

Let $\boldsymbol{M} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ and $\tau = 1/\sigma^2$. Assume that

- $\pi(\boldsymbol{\beta}) \propto 1$,
- τ is known, and
- $\boldsymbol{X}$ is of full rank p.

Then the posterior distribution of $\boldsymbol{\beta}$ (given $\boldsymbol{y}$) is $N(\widehat{\boldsymbol{\beta}}, \tau^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1})$, where

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

We use the familiar technique of completing the square to prove the last statement.

$$
\begin{aligned}
\pi(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \exp\left[-\frac{\tau}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\right] \\
&\propto \exp\left[-\frac{\tau}{2}\left(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}-2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{y}\right)\right] \\
&\propto \exp\left[-\frac{\tau}{2}\left(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}-2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)\right] \\
&\propto \exp\left[-\frac{\tau}{2}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^T\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\right]
\end{aligned}
$$

It follows immediately that the posterior of $\boldsymbol{\beta}$ is $N(\widehat{\boldsymbol{\beta}},\tau^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1})$.

Jeffreys prior for τ known

The log-likelihood is

$$
\begin{aligned}
\log f(\boldsymbol{y}|\boldsymbol{\beta}) &= -\frac{n}{2}\log(2\pi)+\frac{n}{2}\log\tau \\
&\quad -\frac{\tau}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}).
\end{aligned}
$$

$$\frac{\partial}{\partial \beta_j} \log f(\boldsymbol{y}|\boldsymbol{\beta}) = \tau \sum_{i=1}^{n} y_i x_{ij} - \tau \sum_{i=1}^{p} \beta_i \sum_{k=1}^{n} x_{kj} x_{ki}$$

$$\frac{\partial^2}{\partial \beta_i \beta_j} \log f(\boldsymbol{y}|\boldsymbol{\beta}) = -\tau \sum_{k=1}^{n} x_{ki} x_{kj}$$

So, the information matrix is

$$\boldsymbol{I}(\boldsymbol{\beta}) = \tau(\boldsymbol{X}^T \boldsymbol{X}),$$

and the Jeffreys prior is

$$\pi(\boldsymbol{\beta}) \propto |\boldsymbol{X}^T \boldsymbol{X}|^{1/2} \propto 1.$$

Verify that when $\boldsymbol{\beta}$ and τ are both unknown the Jeffreys prior is

$$\pi(\boldsymbol{\beta}, \tau) \propto \tau^{p/2-1}.$$

Let $t_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the p-variate t-distribution with k degrees of freedom, mean $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$ (p. 577 GCSR).

When using the noninformative prior $\pi(\boldsymbol{\beta}, \tau) \propto \tau^{-1}$, we have

$$\boldsymbol{\beta}|\boldsymbol{y} \sim t_{n-p}(\widehat{\boldsymbol{\beta}}, s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}),$$

where $s^2 = \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{M})\boldsymbol{Y}/(n-p)$, and

$$\tau|\boldsymbol{y} \sim \text{gamma}\left(\frac{n-p}{2}, \frac{(n-p)s^2}{2}\right).$$

The joint posterior is

$$\begin{aligned} \pi(\boldsymbol{\beta}, \tau|\boldsymbol{y}) \quad \propto \quad & \tau^{n/2-1} \\ & \times \exp\left[-\frac{\tau}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right]. \end{aligned}$$

Simple algebra shows that

$$(\boldsymbol{y}-\boldsymbol{X}\beta)^T(\boldsymbol{y}-\boldsymbol{X}\beta)=$$

$$\boldsymbol{y}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}+(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^T\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})=Q(\boldsymbol{\beta}).$$

So,

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \int_0^\infty \tau^{n/2-1}\,\mathrm{exp}\left[-\frac{\tau}{2}Q(\boldsymbol{\beta})\right]d\tau \\
&\propto Q(\boldsymbol{\beta})^{-n/2} \\
&\propto \Big[1+\frac{1}{s^2(n-p)} \\
&\quad \times(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^T\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\Big]^{-(n-p+p)/2},
\end{aligned}$$

and hence $\boldsymbol{\beta}|\boldsymbol{y}\sim t_{n-p}(\widehat{\boldsymbol{\beta}},s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$.

We have also

$$\begin{aligned}\pi(\tau|\boldsymbol{y}) &\propto \tau^{n/2-1}\exp\left[-\frac{\tau}{2}(n-p)s^2\right] \\ &\quad \times \int_{\Re^p}\exp\left[-\frac{\tau}{2}Q_1(\boldsymbol{\beta})\right]\,d\boldsymbol{\beta},\end{aligned}$$

where

$$Q_1(\boldsymbol{\beta}) = (\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^T\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}).$$

Now,

$$\int_{\Re^p}\exp\left[-\frac{\tau}{2}Q_1(\boldsymbol{\beta})\right]\,d\boldsymbol{\beta} =$$

$$(2\pi)^{p/2}|(\boldsymbol{X}^T\boldsymbol{X})^{-1}/\tau|^{1/2} =$$

$$(2\pi)^{p/2}|\boldsymbol{X}^T\boldsymbol{X}|^{-1/2}\tau^{-p/2},$$

and therefore

$$\pi(\tau|\boldsymbol{y}) \propto \tau^{(n-p)/2-1}\exp\left[-\frac{\tau}{2}(n-p)s^2\right].$$

Prediction

Prediction is often an important problem in regression analysis. The Bayesian paradigm provides an ideal mechanism for constructing predictions and prediction regions.

Suppose we have $\boldsymbol{X}$ and observations $\boldsymbol{y}$ from the linear model on p. 166N.

We wish to predict the $q \times 1$ vector $\boldsymbol{Z}$ of future observations corresponding to the $q \times p$ matrix of covariates $\boldsymbol{X}_f$. We have

$$\boldsymbol{Z} = \boldsymbol{X}_f \boldsymbol{\beta} + \boldsymbol{\epsilon}_f,$$

where $\boldsymbol{\epsilon}_f$ is independent of $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_f \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$. We assume that $\boldsymbol{X}_f$ is of full column rank.

Again using the noninformative prior $\pi(\boldsymbol{\beta}, \tau) \propto \tau^{-1}$, we will show that $\boldsymbol{Z}|\boldsymbol{y}$ has the q-variate t distribution

$$t_{n-p}\left(\boldsymbol{X}_f \widehat{\boldsymbol{\beta}}, s^2 \left(\boldsymbol{I} + \boldsymbol{X}_f (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}_f^T\right)\right).$$

The proof of this result is actually somewhat tedious, and so not every step will be shown.

To begin with, the conditional density of $\boldsymbol{Z}$ given $\boldsymbol{y}$ is

$$m(\boldsymbol{z}|\boldsymbol{y}) = \int_0^\infty \int_{\Re^p} f(\boldsymbol{z}|\boldsymbol{\beta}, \tau)\pi(\boldsymbol{\beta}, \tau|\boldsymbol{y})\, d\boldsymbol{\beta} d\tau.$$

The integrand is proportional to

$$\tau^{(q+n)/2-1} \exp\left[-\tfrac{\tau}{2}(\boldsymbol{z} - \boldsymbol{X}_f\boldsymbol{\beta})^T(\boldsymbol{z} - \boldsymbol{X}_f\boldsymbol{\beta})\right] \\ \times \exp\left\{-\tfrac{\tau}{2}\left[(n-p)s^2 + Q_1(\boldsymbol{\beta})\right]\right\}.$$

The quadratic form in the first exp term may be written

$$(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}})^T \boldsymbol{X}_f^T \boldsymbol{X}_f (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}) + \boldsymbol{z}^T(I - \boldsymbol{M}_f)\boldsymbol{z} = \\ Q_2(\boldsymbol{\beta}) + \boldsymbol{z}^T(I - \boldsymbol{M}_f)\boldsymbol{z},$$

using notation that parallels that on p. 166N.

"Completing the square" for matrices

Consider $\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} - 2\boldsymbol{u}^T\boldsymbol{\alpha}$. Want to "complete the square."

$$\begin{aligned}
\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} - 2\boldsymbol{u}^T\boldsymbol{\alpha} &= \boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} - 2\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{\alpha} \\
&= \boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} - 2\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{\alpha} \\
&\quad +\boldsymbol{\alpha}^T\boldsymbol{A}^{-1}\boldsymbol{\alpha} - \boldsymbol{\alpha}^T\boldsymbol{A}^{-1}\boldsymbol{\alpha} \\
&= (\boldsymbol{u} - \boldsymbol{A}^{-1}\boldsymbol{\alpha})^T\boldsymbol{A}(\boldsymbol{u} - \boldsymbol{A}^{-1}\boldsymbol{\alpha}) \\
&\quad -\boldsymbol{\alpha}^T\boldsymbol{A}^{-1}\boldsymbol{\alpha}
\end{aligned}$$

Now apply this result to

$$Q_1(\boldsymbol{\beta}) + Q_2(\boldsymbol{\beta}) =$$

$$\boldsymbol{\beta}^T\boldsymbol{X}_f^T\boldsymbol{X}_f\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\boldsymbol{X}_f^T\boldsymbol{X}_f\widehat{\boldsymbol{\beta}}_{\boldsymbol{z}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}^T\boldsymbol{X}_f^T\boldsymbol{X}_f\widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}$$

$$\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}^T\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}}.$$

Using the notation on the first half of p. 174N, write

$$\boldsymbol{u} = \boldsymbol{\beta}, \quad \boldsymbol{A} = \boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{X}^T \boldsymbol{X}$$

and

$$\boldsymbol{\alpha} = \boldsymbol{X}_f^T \boldsymbol{X}_f \widehat{\boldsymbol{\beta}}_z + \boldsymbol{X}^T \boldsymbol{X} \widehat{\boldsymbol{\beta}}.$$

Now we have

$$Q_1(\boldsymbol{\beta}) + Q_2(\boldsymbol{\beta}) =$$

$$\begin{gathered}(\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{\alpha})^T \boldsymbol{A} (\boldsymbol{\beta} - \boldsymbol{A}^{-1}\boldsymbol{\alpha}) - \boldsymbol{\alpha}^T \boldsymbol{A}^{-1} \boldsymbol{\alpha} \\ + \widehat{\boldsymbol{\beta}}^T \boldsymbol{X}^T \boldsymbol{X} \widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}^T \boldsymbol{X}_f^T \boldsymbol{X}_f \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}.\end{gathered}$$

Deft matrix algebra shows that

$$-\boldsymbol{\alpha}^T \boldsymbol{A}^{-1}\boldsymbol{\alpha} + \widehat{\boldsymbol{\beta}}^T \boldsymbol{X}^T \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}^T \boldsymbol{X}_f^T \boldsymbol{X}_f \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}} =$$

$$(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}})^T \boldsymbol{\Lambda}_f \boldsymbol{X}^T \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}),$$

where

$$\boldsymbol{\Lambda}_f = \boldsymbol{I} - \boldsymbol{X}^T \boldsymbol{X}(\boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{X}^T \boldsymbol{X})^{-1}.$$

A key in the "deft algebra" is recognizing the fact that

$$\boldsymbol{X}_f^T \boldsymbol{X}_f \left[\boldsymbol{I} - (\boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}_f^T \boldsymbol{X}_f\right] =$$

$$\boldsymbol{\Lambda}_f \boldsymbol{X}^T \boldsymbol{X}.$$

At first, this seems like voodoo, but it's just the matrix analog of the identity

$$a\left(1 - \frac{a}{a+b}\right) = b\left(1 - \frac{b}{a+b}\right).$$

So, now we see that the integrand of the integral on p. 173N is proportional to

$$\tau^{(q+n)/2-1} \exp\left[-\frac{\tau}{2}(n-p)s^2\right]$$
$$\times \exp\left[-\frac{\tau}{2}(\boldsymbol{\beta}-\boldsymbol{A}^{-1}\boldsymbol{\alpha})^T\boldsymbol{A}(\boldsymbol{\beta}-\boldsymbol{A}^{-1}\boldsymbol{\alpha})\right]$$
$$\times \exp\left[-\frac{\tau}{2}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{z}})^T\boldsymbol{\Lambda}_f\boldsymbol{X}^T\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{z}})\right]$$
$$\times \exp\left[-\frac{\tau}{2}\boldsymbol{z}^T(\boldsymbol{I}-\boldsymbol{M}_f)\boldsymbol{z}\right].$$

More matrix algebra shows that

$$(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{z}})^T\boldsymbol{\Lambda}_f\boldsymbol{X}^T\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{z}})+\boldsymbol{z}^T(\boldsymbol{I}-\boldsymbol{M}_f)\boldsymbol{z}=$$
$$\widehat{\boldsymbol{\beta}}^T\boldsymbol{\Lambda}_f\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}}$$
$$-2\boldsymbol{z}^T\boldsymbol{X}_f(\boldsymbol{X}_f^T\boldsymbol{X}_f+\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}}$$
$$+\boldsymbol{z}^T(\boldsymbol{I}-\boldsymbol{X}_f(\boldsymbol{X}_f^T\boldsymbol{X}_f+\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T)\boldsymbol{z}.$$

We need a couple of identities to finish the proof.

1. $$\left[\boldsymbol{I} - \boldsymbol{X}_f(\boldsymbol{X}_f^T\boldsymbol{X}_f + \boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T\right]^{-1} = \boldsymbol{I} + \boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T$$

2. $$\boldsymbol{X}_f = \left(\boldsymbol{I} + \boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T\right)\boldsymbol{X}_f \times(\boldsymbol{X}_f^T\boldsymbol{X}_f + \boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}.$$

Proof of 1.

The equality is true iff

$$\left[\boldsymbol{I} - \boldsymbol{X}_f\boldsymbol{A}^{-1}\boldsymbol{X}_f^T\right]\left[\boldsymbol{I} + \boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T\right] = \boldsymbol{I}.$$

The lhs immediately above is

$$\boldsymbol{I} - \boldsymbol{X}_f\boldsymbol{A}^{-1}\boldsymbol{X}_f^T + \boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T$$

$$-\boldsymbol{X}_f\boldsymbol{A}^{-1}\boldsymbol{X}_f^T\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T =$$

$$\boldsymbol{I} + \boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T$$

$$-\boldsymbol{X}_f\left[\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{X}_f^T\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right]\boldsymbol{X}_f^T.$$

It suffices to show that

$$\boldsymbol{A}^{-1}\left[\boldsymbol{I}+\boldsymbol{X}_f^T\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right]=(\boldsymbol{X}^T\boldsymbol{X})^{-1}.$$

But the last equality is true iff

$$\begin{aligned}\boldsymbol{I}+\boldsymbol{X}_f^T\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1} &= \boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\\ &= \boldsymbol{I}+\boldsymbol{X}_f^T\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}.\end{aligned}$$

Q.E.D.

Proof of 2.

This equality is true iff

$$\left(\boldsymbol{I}+\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T\right)\boldsymbol{X}_f=\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}.$$

Now,

$$\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}=\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T\boldsymbol{X}_f+\boldsymbol{X}_f.$$

But

$$\left(\boldsymbol{I}+\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T\right)\boldsymbol{X}_f=\\ \boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T\boldsymbol{X}_f+\boldsymbol{X}_f,$$

and hence Q.E.D.

Now, continuing from the last expression on p. 177N, and using yet again our "complete the square" trick,

$$\widehat{\boldsymbol{\beta}}^T \boldsymbol{\Lambda}_f \boldsymbol{X}^T \boldsymbol{X} \widehat{\boldsymbol{\beta}} - 2\boldsymbol{z}^T \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}^T \boldsymbol{X} \widehat{\boldsymbol{\beta}}$$
$$+\boldsymbol{z}^T (\boldsymbol{I} - \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}_f^T) \boldsymbol{z} =$$

$$\widehat{\boldsymbol{\beta}}^T \boldsymbol{\Lambda}_f \boldsymbol{X}^T \boldsymbol{X} \widehat{\boldsymbol{\beta}}$$
$$+(\boldsymbol{z} - \mu_{\boldsymbol{z}})^T (\boldsymbol{I} - \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}_f^T)(\boldsymbol{z} - \mu_{\boldsymbol{z}})$$
$$-\mu_{\boldsymbol{z}}^T (\boldsymbol{I} - \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}_f^T) \mu_{\boldsymbol{z}}, \qquad (*)$$

where

$$\begin{aligned} \mu_{\boldsymbol{z}} &= (\boldsymbol{I} - \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}_f^T)^{-1} \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}^T \boldsymbol{X} \widehat{\boldsymbol{\beta}} \\ &= (\boldsymbol{I} + \boldsymbol{X}_f (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}_f^T) \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}^T \boldsymbol{X} \widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{X}_f \widehat{\boldsymbol{\beta}}. \end{aligned}$$

We can write $(*)$ as

$$\widehat{\boldsymbol{\beta}}^T \left[\boldsymbol{\Lambda}_f \boldsymbol{X}^T \boldsymbol{X} - \boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{\Lambda}_f^T \right] \widehat{\boldsymbol{\beta}}$$
$$+(\boldsymbol{z} - \boldsymbol{X}_f \widehat{\boldsymbol{\beta}})^T (\boldsymbol{I} - \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}_f^T)(\boldsymbol{z} - \boldsymbol{X}_f \widehat{\boldsymbol{\beta}}).$$

Now, since $\boldsymbol{\Lambda}_f = \boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{A}^{-1}$,

$$\boldsymbol{\Lambda}_f \boldsymbol{X}^T \boldsymbol{X} - \boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{\Lambda}_f^T =$$

$$\begin{gathered} \boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}^T \boldsymbol{X} - \boldsymbol{X}_f^T \boldsymbol{X}_f \\ + \boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}_f^T \boldsymbol{X}_f. \end{gathered}$$

Claim:

$$\boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{X}_f^T \boldsymbol{X}_f \boldsymbol{A}^{-1} \boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{X}_f^T \boldsymbol{X}_f.$$

The claim is true iff

$$\boldsymbol{A}^{-1} \boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{A}^{-1} \boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{I},$$

which is true iff

$$\boldsymbol{X}_f^T \boldsymbol{X}_f + \boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{A},$$

which is true by definition.

Putting everything together, $m(\boldsymbol{z}|\boldsymbol{y})$ is proportional to

$$\int_0^\infty \int_{\Re^p} \tau^{(q+n)/2-1} \exp\left[-\frac{\tau}{2}(n-p)s^2\right]$$
$$\times \exp\left[-\frac{\tau}{2}(\boldsymbol{\beta}-\boldsymbol{A}^{-1}\boldsymbol{\alpha})^T \boldsymbol{A}(\boldsymbol{\beta}-\boldsymbol{A}^{-1}\boldsymbol{\alpha})\right]$$
$$\times \exp[-\frac{\tau}{2}(\boldsymbol{z}-\boldsymbol{X}_f\widehat{\boldsymbol{\beta}})^T(\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{A}^{-1}\boldsymbol{X}_f^T)$$
$$\times(\boldsymbol{z}-\boldsymbol{X}_f\widehat{\boldsymbol{\beta}})]\, d\boldsymbol{\beta} d\tau,$$

which is proportional to

$$\int_0^\infty \tau^{(q-p+n)/2-1} \exp\left[-\frac{\tau}{2}(n-p)s^2\right]$$
$$\times \exp[-\frac{\tau}{2}(\boldsymbol{z}-\boldsymbol{X}_f\widehat{\boldsymbol{\beta}})^T(\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{A}^{-1}\boldsymbol{X}_f^T)$$
$$\times(\boldsymbol{z}-\boldsymbol{X}_f\widehat{\boldsymbol{\beta}})]\, d\tau.$$

The last integral is proportional to the q-variate t-distribution

$$t_{n-p}(\boldsymbol{X}_f\widehat{\boldsymbol{\beta}}, s^2(\boldsymbol{I}+\boldsymbol{X}_f(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}_f^T)).$$