The Normal Regression Model as a LUCK-model

Allowing for the Construction of Imprecise Conjugate Priors

Gero Walter
Department of Statistics, Ludwig-Maximilians-University
Munich, Germany
gero.walter@campus.lmu.de

18th May 2007

Abstract

This note is one of two supplements to the paper "Linear Regression Analysis under Sets of Conjugate Priors", by G. Walter, T. Augustin, and A. Peters, in revision for ISIPTA 07, in which Bayesian inference in linear regression models is extended by considering imprecise conjugated priors.

A central step towards that general framework was formulated in Theorem 2 there, where it was shown that the "normal regression model" constitutes a LUCK-model. The name LUCK-model was introduced to denote situations for which the update step in Bayesian learning is especially elegant and simple (described in detail in [5] by Definition 1), since such models can be extended to imprecise priors, by applying a variation and an extension of a method for inference in i.i.d. exponential families presented at ISIPTA'05 by Quaeghebeur and de Cooman.

In the following, the proof of that Theorem 2 will be given in detail. Before that, as a preparation, the derivation of the "normal regression model" and some other parts needed from [5] will be repeated.

1 Introduction

1.1 Bayesian Analysis of Regression Models

The regression model is noted as follows:

$$z = \mathbf{X}\beta + \varepsilon\,, \quad \mathbf{X} \in \mathbb{R}^{k\times p}\,,\ \beta \in \mathbb{R}^{p}\,,\ z \in \mathbb{R}^{k}\,,\ \varepsilon \in \mathbb{R}^{k}\,,$$

where z is the response, $\mathbf{X}$ the so-called design matrix with the p regressors collected column by column, and β is the p-dimensional vector of regression coefficients.

Each error term ε_i is assumed to be normally distributed with mean value 0 and variance σ^2; all error terms are taken to be uncorrelated:

$$\varepsilon_i \overset{i.i.d}{\sim} \mathrm{N}(0,\sigma^2) \quad \Longrightarrow \quad \varepsilon \sim \mathrm{N}_k(\mathbf{0},\sigma^2\mathbf{I})) \quad (\sigma^2 \text{ known})$$

Because $\mathbf{X}$ is considered fixed and non-stochastic, we get the following likelihood of z:

$$z\,|\,\beta \sim \mathrm{N}_k(\mathbf{X}\beta,\sigma^2\mathbf{I})$$

In principle several conjugate priors to this likelihood exist. The standard choice (see, e.g., [1, p. 244ff]), on which we also focus in this paper[1], is a multivariate normal distribution on β with parameters $\beta^{(0)}$ and $\mathbf{\Sigma}^{(0)}$. (As in [5], we will denote the prior parameters with the upper index $^{(0)}$, whereas the posterior parameters will be denoted with upper index $^{(1)}$.) The model of Bayesian regression analysis based on the standard prior will be called *normal regression model*.

$$\beta \ \sim \ \mathrm{N}_p\left(\beta^{(0)},\sigma^2\mathbf{\Sigma}^{(0)}\right) \qquad \text{with } \beta^{(0)} \in \mathbb{R}^p\,,\ \mathbf{\Sigma}^{(0)} \in \mathbb{R}^{p\times p} \text{ positive definite (p.d.)}\,,$$

i.e.

$$p(\beta) \ = \ \frac{1}{|\mathbf{\Sigma}^{(0)}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\left(\beta-\beta^{(0)}\right)^{\mathsf{T}}\mathbf{\Sigma}^{(0)^{-1}}\left(\beta-\beta^{(0)}\right)\right\}. \tag{1}$$

The posterior, calculated by applying Bayes's rule $p(\beta\,|\,z) \propto p(z\,|\,\beta)\,p(\beta)$, is then the multivariate normal distribution $\mathrm{N}_p\left(\beta^{(1)},\sigma^2\mathbf{\Sigma}^{(1)}\right)$ with posterior parameters

$$\begin{aligned}
\beta^{(1)} &= \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}+\mathbf{\Sigma}^{(0)^{-1}}\right)^{-1}\left(\mathbf{X}^{\mathsf{T}}z+\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)}\right) && (2)\\
&= \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}+\mathbf{\Sigma}^{(0)^{-1}}\right)^{-1}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\hat\beta+\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)}\right)\\
&= (\mathbf{I}-\mathbf{A})\beta^{(0)}+\mathbf{A}\hat\beta_{LS} \qquad \text{with } \mathbf{A}=\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}+\mathbf{\Sigma}^{(0)^{-1}}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{X}\\
\mathbf{\Sigma}^{(1)} &= \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}+\mathbf{\Sigma}^{(0)^{-1}}\right)^{-1}. && (3)
\end{aligned}$$

[1] In the second supplement [4], an alternative approach is considered more in detail.

1.2 Classical Bayesian Inference and LUCK-models

For the application presented in [5], it is quite convenient to distinguish certain standard situations (called *models with 'Linearly Updated Conjugate prior Knowledge'* (LUCK) here) of Bayesian updating on a parameter ϑ based on a sample w with likelihood $f(w \mid \vartheta)$ by

$$p(\vartheta \mid w) \propto f(w \mid \vartheta) \cdot p(\vartheta)\,, \tag{4}$$

where the prior $p(\vartheta)$ and the posterior $p(\vartheta \mid w)$ fit nicely together in the sense that

i) they belong to the same class of parametric distributions, a case where they are called *conjugate*, and, in addition,

ii) the updating of one parameter ($y^{(0)}$ below) of the prior is linear.

More precisely, we introduced the following definition:

Definition 1 *Consider classical Bayesian inference on a parameter ϑ based on a sample w as described in (4), and let the prior $p(\vartheta)$ be characterized by the (vectorial) parameter $\vartheta^{(0)}$. The pair $\big(p(\vartheta), p(\vartheta \mid w)\big)$ is said to constitute a* LUCK-model of size q in the natural parameter ψ with prior parameters $n^{(0)} \in \mathbb{R}^+$ and $y^{(0)}$ and sample statistic $\tau(w)$, *iff there exist $q \in \mathbb{N}$ as well as transformations of ϑ into ψ and $\mathbf{b}(\psi)$ and of $\vartheta^{(0)}$ into $n^{(0)}$ and $y^{(0)}$, such that $p(\vartheta)$ and $p(\vartheta \mid w)$ can be rewritten in the following way:*[2]

$$p(\vartheta) \propto \exp\left\{n^{(0)}\left[\langle \psi, y^{(0)}\rangle - \mathbf{b}(\psi)\right]\right\} \tag{5}$$

and

$$p(\vartheta \mid w) \propto \exp\left\{n^{(1)}\left[\langle \psi, y^{(1)}\rangle - \mathbf{b}(\psi)\right]\right\} \tag{6}$$

with

$$n^{(1)} = n^{(0)} + q \qquad \textit{and} \qquad y^{(1)} = \frac{n^{(0)} y^{(0)} + \tau(w)}{n^{(0)} + q}\,. \tag{7}$$

Theorem 2, relating LUCK-models with the normal regression model described in Section 1.1, was formulated in the following way:

Theorem 2 *Consider the normal regression model described by the prior $p(\beta)$ from (1) with prior parameters $\beta^{(0)}$ and $\mathbf{\Sigma}^{(0)}$, and the multivariate normal posterior defined by (2) and (3).*
Fixing a value $n^{(0)}$, $\big(p(\beta), p(\beta \mid z)\big)$ constitutes a LUCK*-model of size* 1 *with prior parameters*

$$y^{(0)} = \frac{1}{n^{(0)}} \begin{pmatrix} \mathbf{\Lambda}^{(0)} \\ \mathbf{\Lambda}^{(0)}\beta^{(0)} \end{pmatrix} =: \begin{pmatrix} y_a^{(0)} \\ y_b^{(0)} \end{pmatrix} \tag{8}$$

and $n^{(0)}$ and sample statistic

$$\tau(z) = \tau(\mathbf{X}, z) = \begin{pmatrix} \mathbf{X}^{\mathsf{T}}\mathbf{X} \\ \mathbf{X}^{\mathsf{T}} z \end{pmatrix} =: \begin{pmatrix} \tau_a(\mathbf{X}, z) \\ \tau_b(\mathbf{X}, z) \end{pmatrix}\,. \tag{9}$$

[2] $\langle a, b\rangle$ denotes the scalar product of a and b.

2 Proof of Theorem 2 from [5]

For proving Theorem 2, two steps have to be performed: First, showing that the prior of (1) fits the form of the prior in Definition 1; and second, showing that the update step described by Equations (2) and (3) is equivalent to an update step in a LUCK-model.

2.1 Proving Step 1: Equivalence of the Prior

For a more direct link to the model introduced by [2], the factor $\mathbf{c}\left(n^{(0)}, y^{(0)}\right)$ that was omitted in Definition 1, is kept here.

So, this is the 'target' form we want to achieve:

$$p(\psi) \;=\; \mathbf{c}\left(n^{(0)}, y^{(0)}\right) \exp\left\{n^{(0)}\left[\left\langle \psi \,,\, y^{(0)} \right\rangle - \mathbf{b}(\psi)\right]\right\}, \tag{10}$$

where ψ is some transformation of β, and $n^{(0)}$ and $y^{(0)}$ are transformations of the prior parameters $\beta^{(0)}$ and $\mathbf{\Sigma}^{(0)}$.

Starting with the prior (1), by collecting all parts that depend on the prior parameters only, we get

$$\begin{aligned}
p(\beta) &= \frac{1}{|\mathbf{\Sigma}^{(0)}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\left(\beta-\beta^{(0)}\right)^{\mathsf{T}} \mathbf{\Sigma}^{(0)^{-1}} \left(\beta-\beta^{(0)}\right)\right\} \\
&= \frac{1}{|\mathbf{\Sigma}^{(0)}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}} \\
&\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\left[\beta^{\mathsf{T}}\mathbf{\Sigma}^{(0)^{-1}}\beta - \beta^{\mathsf{T}}\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)} - \beta^{(0)^{\mathsf{T}}}\mathbf{\Sigma}^{(0)^{-1}}\beta + \beta^{(0)^{\mathsf{T}}}\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)}\right]\right\} \\
&= \underbrace{\frac{1}{|\mathbf{\Sigma}^{(0)}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}(\sigma^2)^{\frac{p}{2}}} \cdot \exp\left(-\frac{1}{2\sigma^2}\beta^{(0)^{\mathsf{T}}}\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)}\right)}_{=:\,\mathbf{c}(\beta^{(0)},\,\mathbf{\Sigma}^{(0)})} \\
&\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\beta^{\mathsf{T}}\mathbf{\Sigma}^{(0)^{-1}}\beta + \frac{1}{2\sigma^2}\beta^{\mathsf{T}}\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)} + \frac{1}{2\sigma^2}\beta^{(0)^{\mathsf{T}}}\mathbf{\Sigma}^{(0)^{-1}}\beta\right\}.
\end{aligned}$$

As $\beta^{\mathsf{T}}\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)}$ and $\beta^{(0)^{\mathsf{T}}}\mathbf{\Sigma}^{(0)^{-1}}\beta$ are scalar and it holds that $\left(\beta^{\mathsf{T}}\mathbf{\Sigma}^{(0)^{-1}}\beta^{(0)}\right)^{\mathsf{T}} = \beta^{(0)^{\mathsf{T}}}\mathbf{\Sigma}^{(0)^{-1}}\beta$, we can simply add these two terms.

For simplification of notation, we introduce the matrix $\mathbf{\Lambda}^{(0)} := \mathbf{\Sigma}^{(0)^{-1}}$ with its elements $\left(\mathbf{\Lambda}^{(0)}\right)_{ij} =: \lambda_{ij}^{(0)}$.

Furthermore, let us fix the index letters as follows:

h, i, j	index for $1, \ldots, p = \dim(\beta)$
l, m	index for $1, \ldots, k = \dim(z)$

With these denotations, we get

$$\begin{aligned}
p(\beta) &= \mathbf{c}(\beta^{(0)}, \mathbf{\Sigma}^{(0)}) \cdot \exp\Big\{ -\frac{1}{2\sigma^2}\beta^{\mathsf{T}}\mathbf{\Lambda}^{(0)}\beta + \frac{1}{\sigma^2}\beta^{\mathsf{T}}\mathbf{\Lambda}^{(0)}\beta^{(0)} \Big\} \\
&= \mathbf{c}(\beta^{(0)}, \mathbf{\Sigma}^{(0)}) \cdot \exp\Big\{ -\sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\beta_i\beta_j}{2\sigma^2} \cdot \lambda_{ij}^{(0)} + \sum_{i=1}^{p} \frac{\beta_i}{\sigma^2} \cdot \Big(\sum_{j=1}^{p} \lambda_{ij}^{(0)} \beta_j^{(0)} \Big) \Big\} \\
&= \mathbf{c}(\beta^{(0)}, \mathbf{\Sigma}^{(0)}) \cdot \exp\Big\{ -\sum_{i=1}^{p} \frac{\beta_i^2}{2\sigma^2} \cdot \lambda_{ii}^{(0)} - \sum_{i=1}^{p-1}\sum_{j=i+1}^{p} \frac{\beta_i\beta_j}{\sigma^2} \cdot \lambda_{ij}^{(0)} \\
&\qquad\qquad + \sum_{i=1}^{p} \frac{\beta_i}{\sigma^2} \cdot \Big(\sum_{j=1}^{p} \lambda_{ij}^{(0)} \beta_j^{(0)} \Big) \Big\}
\end{aligned}$$

Therefore, we have

$$\mathbf{c}\left(\beta^{(0)}, \mathbf{\Sigma}^{(0)}\right) = \frac{1}{|\mathbf{\Sigma}^{(0)}|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}} (\sigma^2)^{\frac{p}{2}}} \cdot \exp\left(-\frac{1}{2\sigma^2} \beta^{(0)\mathsf{T}} \mathbf{\Sigma}^{(0)-1} \beta^{(0)} \right)$$

and $$\mathbf{b}(\psi) = 0\,,$$

as there is no summand in the exponent that depends on components of β only.

Now $n^{(0)}$ has to be introduced 'artificially' to make the last step in the alignment to the form of (10). To this end, the factor $\frac{1}{n^{(0)}}$ is multiplied to each summand in the exponent to form the term $\langle \psi\,,\, y^{(0)} \rangle$. Because ψ, as a transformation of β, must remain unchanged during an update step, $\frac{1}{n^{(0)}}$ is associated with the parts depending on $\beta^{(0)}$ and $\mathbf{\Sigma}^{(0)}$, thus forming $y^{(0)}$. Therefore, the above noted transformations of β and the prior parameters $\beta^{(0)}$ and $\mathbf{\Sigma}^{(0)}$ are

$$\psi = \begin{bmatrix} -\frac{\beta_1^2}{2\sigma^2} \\ \vdots \\ -\frac{\beta_p^2}{2\sigma^2} \\ -\frac{\beta_1\beta_2}{\sigma^2} \\ \vdots \\ -\frac{\beta_{p-1}\beta_p}{\sigma^2} \\ -\frac{\beta_1}{\sigma^2} \\ \vdots \\ -\frac{\beta_p}{\sigma^2} \end{bmatrix}, \quad y^{(0)} = \begin{bmatrix} \frac{\lambda_{11}^{(0)}}{n^{(0)}} \\ \vdots \\ \frac{\lambda_{pp}^{(0)}}{n^{(0)}} \\ \frac{\lambda_{12}^{(0)}}{n^{(0)}} \\ \vdots \\ \frac{\lambda_{p-1,p}^{(0)}}{n^{(0)}} \\ \frac{1}{n^{(0)}} \sum_{j=1}^{p} \lambda_{1j}^{(0)} \beta_j^{(0)} \\ \vdots \\ \frac{1}{n^{(0)}} \sum_{j=1}^{p} \lambda_{pj}^{(0)} \beta_j^{(0)} \end{bmatrix}, \quad \begin{matrix} \left.\right\} p \text{ summands} \\ \left.\right\} \frac{p(p-1)}{2} \\ \left.\right\} p \end{matrix}$$

and so equivalence with (10) is reached.

2.2 Proving Step 2: Equivalence of the Update Step

The update step for a LUCK-model of size q leads to the following posterior:

$$
\begin{aligned}
p(\psi \,|\, w) \;&=\; \mathbf{c}\big((n^{(0)}+q), y^{(1)}\big) \exp\Big\{(n^{(0)}+1)\big[\,\langle \psi\,,\, y^{(1)}\rangle - \mathbf{b}(\psi)\big]\Big\}\,, \\
&\text{with} \quad y^{(1)} = \frac{n^{(0)}y^{(0)} + \tau(w)}{n^{(0)}+q}\,.
\end{aligned} \tag{11}
$$

To verify that the normal regression model constitutes a LUCK-model, we have to check whether the normal prior (1), after being updated along (2) and (3), has the form of (11). Because $n^{(0)}$ was introduced 'artificially', the normal regression model can be shown to constitute a LUCK-model of any size q; the choice of q does not affect the posterior inferences in terms of $\beta^{(1)}$ and $\mathbf{\Sigma}^{(1)}$, because in the "retranslation" step 4 in Section 3.2 of [5], $n^{(1)} = n^{(0)} + q$ will be canceled out: $\mathbf{\Sigma}^{(1)}$ can be derived from $\mathbf{\Lambda}^{(1)} = n^{(1)} \cdot y_a^{(1)}$; the same applies for the derivation of $\beta^{(1)}$ from $y_b^{(1)}$. Therefore, we will set $q =: 1$ w.l.o.g.

Step 2 of the proof can be done for each of the summands of the exponent separately. So, in the following for each summand of the exponent the equivalence of the update step in the normal regression model with the update step in a LUCK-model of size 1 is shown. As a by-product, the form of $\tau(w)$ can be determined. $\tau(w)$ is here, not as intended by Quaeghebeur and de Cooman, already known from the likelihood, but so far undetermined, because considerations were started from the conjugate prior directly. Although only z is considered as the stochastic part of the sample and $\mathbf{X}$ is considered fix, both will explicitly appear as the arguments of the function $\tau(\cdot)$ to clarify the dependency on $\mathbf{X}$.

Rewriting (2) and (3) with $\mathbf{\Lambda}^{(0)}$ instead of ${\mathbf{\Sigma}^{(0)}}^{-1}$, we get

$$
\begin{aligned}
\beta^{(1)} \;&=\; \left(\mathbf{X}^\mathsf{T}\mathbf{X} + \mathbf{\Lambda}^{(0)}\right)^{-1} \left(\mathbf{X}^\mathsf{T} z + \mathbf{\Lambda}^{(0)}\beta^{(0)}\right) \\
\Lambda^{(1)} \;&=\; \mathbf{X}^\mathsf{T}\mathbf{X} + \mathbf{\Lambda}^{(0)}\,.
\end{aligned}
$$

Again some simplification of notation is applied: The elements of the design matrix $\mathbf{X}$ are denoted with lower case x: $(\mathbf{X})_{l,i} =: x_{l,i}$; furthermore the following equivalence is of use for checking the equivalence in the third group of summands:

$$
\begin{aligned}
\beta_j^{(1)} \;&=\; \left({\mathbf{\Lambda}^{(1)}}^{-1}\left(\mathbf{X}^\mathsf{T} z + \mathbf{\Lambda}^{(0)}\beta^{(0)}\right)\right)_j \\
&=\; \sum_{i=1}^{p} \left({\mathbf{\Lambda}^{(1)}}^{-1}\right)_{ji} \left(\mathbf{X}^\mathsf{T} z + \mathbf{\Lambda}^{(0)}\beta^{(0)}\right)_i \\
&=\; \sum_{i=1}^{p} \left({\mathbf{\Lambda}^{(1)}}^{-1}\right)_{ji} \left(\sum_{l=1}^{k} x_{li} z_l + \sum_{h=1}^{p} \lambda_{ih}^{(0)} \beta_h^{(0)}\right).
\end{aligned}
$$

On the following two pages, we will have the summands in the exponent of the prior on the left side, which are updated by the 'observation $(\mathbf{X}, z)$' according to the LUCK-model by directly replacing the prior parameters $n^{(0)}$ and $y^{(0)}$ with the posterior parameters $n^{(1)}$ and $y^{(1)}$. Then, the updated parameters are 'decoded' using Equations (2) and (3), and the resulting terms are transformed to reveal that they can actually be obtained by performing the update step of a LUCK-model of size 1.

$$
\begin{array}{lcll}
\underbrace{-\frac{\beta_1^2}{2\sigma^2}}_{\psi_1}\cdot\underbrace{\frac{1}{n^{(0)}}\lambda_{11}^{(0)}}_{y_1^{(0)}} & \xrightarrow{\text{observation}} & -\frac{\beta_1^2}{2\sigma^2}\cdot\underbrace{\frac{1}{n^{(1)}}\lambda_{11}^{(1)}}_{y_1^{(1)}} & = \underbrace{-\frac{\beta_1^2}{2\sigma^2}}_{\psi_1}\cdot\frac{1}{n^{(0)}+1}\Big(n^{(0)}\cdot\underbrace{\frac{1}{n^{(0)}}\,\lambda_{11}^{(0)}}_{y_1^{(0)}}+\underbrace{\sum_{l=1}^{k}x_{l1}^2}_{\tau_1(\mathbf{X},z)}\Big) \\
\vdots \quad \vdots & \vdots & & \quad \vdots \qquad\qquad \vdots \qquad \vdots \\
\underbrace{-\frac{\beta_p^2}{2\sigma^2}}_{\psi_p}\cdot\underbrace{\frac{1}{n^{(0)}}\lambda_{pp}^{(0)}}_{y_p^{(0)}} & \longrightarrow & -\frac{\beta_p^2}{2\sigma^2}\cdot\underbrace{\frac{1}{n^{(1)}}\lambda_{pp}^{(1)}}_{y_p^{(1)}} & = \underbrace{-\frac{\beta_p^2}{2\sigma^2}}_{\psi_p}\cdot\frac{1}{n^{(0)}+1}\Big(n^{(0)}\cdot\underbrace{\frac{1}{n^{(0)}}\,\lambda_{pp}^{(0)}}_{y_p^{(0)}}+\underbrace{\sum_{l=1}^{k}x_{lp}^2}_{\tau_p(\mathbf{X},z)}\Big) \\
\underbrace{-\frac{\beta_1\beta_2}{\sigma^2}}_{\psi_{p+1}}\cdot\underbrace{\frac{1}{n^{(0)}}\lambda_{12}^{(0)}}_{y_{p+1}^{(0)}} & \longrightarrow & -\frac{\beta_1\beta_2}{\sigma^2}\cdot\underbrace{\frac{1}{n^{(1)}}\lambda_{12}^{(1)}}_{y_{p+1}^{(1)}} & = \underbrace{-\frac{\beta_1\beta_2}{\sigma^2}}_{\psi_{p+1}}\cdot\frac{1}{n^{(0)}+1}\Big(n^{(0)}\cdot\underbrace{\frac{1}{n^{(0)}}\,\lambda_{12}^{(0)}}_{y_{p+1}^{(0)}}+\underbrace{\sum_{l=1}^{k}x_{l1}\,x_{l2}}_{\tau_{p+1}(\mathbf{X},z)}\Big) \\
\vdots \quad \vdots & \vdots & & \quad \ddots \qquad\qquad \ddots \qquad \ddots \\
\underbrace{-\frac{\beta_{p-1}\beta_p}{\sigma^2}}_{\psi_{\frac{p(p+1)}{2}}}\cdot\underbrace{\frac{1}{n^{(0)}}\lambda_{p-1,p}^{(0)}}_{y_{\frac{p(p+1)}{2}}^{(0)}} & \longrightarrow & -\frac{\beta_{p-1}\beta_p}{\sigma^2}\cdot\underbrace{\frac{1}{n^{(1)}}\lambda_{p-1,p}^{(1)}}_{y_{\frac{p(p+1)}{2}}^{(1)}} & = \underbrace{-\frac{\beta_{p-1}\beta_p}{\sigma^2}}_{\psi_{\frac{p(p+1)}{2}}}\cdot\frac{1}{n^{(0)}+1}\Big(n^{(0)}\cdot\underbrace{\frac{1}{n^{(0)}}\,\lambda_{p-1,p}^{(0)}}_{y_{\frac{p(p+1)}{2}}^{(0)}}+\underbrace{\sum_{l=1}^{k}x_{l,p-1}\,x_{l,p}}_{\tau_{\frac{p(p+1)}{2}}(\mathbf{X},z)}\Big)
\end{array}
$$

$$
\underbrace{-\frac{\beta_1}{\sigma^2}}_{\psi_{\frac{p(p+1)}{2}+1}} \cdot \underbrace{\frac{1}{n^{(0)}} \sum_{j=1}^{p} \lambda_{1j}^{(0)} \beta_j^{(0)}}_{y^{(0)}_{\frac{p(p+1)}{2}+1}} \xrightarrow{\text{observation}} -\frac{\beta_1}{\sigma^2} \cdot \frac{1}{n^{(1)}} \sum_{j=1}^{p} \lambda_{1j}^{(1)} \beta_j^{(1)}
$$

$$
\begin{aligned}
&= -\frac{\beta_1}{\sigma^2} \cdot \frac{1}{n^{(0)}+1} \sum_{j=1}^{p} \lambda_{1j}^{(1)} \sum_{i=1}^{p} \left(\mathbf{\Lambda}^{(1)^{-1}}\right)_{ji} \cdot \Big(\sum_{l=1}^{k} x_{li} z_l + \sum_{h=1}^{p} \lambda_{ih}^{(0)} \beta_h^{(0)} \Big) \\
&= -\frac{\beta_1}{\sigma^2} \cdot \frac{1}{n^{(0)}+1} \sum_{i=1}^{p} \underbrace{\sum_{j=1}^{p} \left(\mathbf{\Lambda}^{(1)}\right)_{1j} \left(\mathbf{\Lambda}^{(1)^{-1}}\right)_{ji}} \cdot \Big(\sum_{l=1}^{k} x_{li} z_l + \sum_{h=1}^{p} \lambda_{ih}^{(0)} \beta_h^{(0)} \Big) \\
&\qquad \left[\begin{array}{ll} i = 1 : & \sum_{j=1}^{p} \left(\mathbf{\Lambda}^{(1)}\right)_{1j} \left(\mathbf{\Lambda}^{(1)^{-1}}\right)_{ji} = 1 \\ i \neq 1 : & \sum_{j=1}^{p} \left(\mathbf{\Lambda}^{(1)}\right)_{1j} \left(\mathbf{\Lambda}^{(1)^{-1}}\right)_{ji} = 0 \end{array} \right] \\
&= -\frac{\beta_1}{\sigma^2} \cdot \frac{1}{n^{(0)}+1} \Big(\sum_{l=1}^{k} x_{l1} z_l + \sum_{h=1}^{p} \lambda_{1h}^{(0)} \beta_h^{(0)} \Big) \\
&= \underbrace{-\frac{\beta_1}{\sigma^2}}_{\psi_{\frac{p(p+1)}{2}+1}} \cdot \frac{1}{n^{(0)}+1} \Big(n^{(0)} \cdot \underbrace{\frac{1}{n^{(0)}} \sum_{j=1}^{p} \lambda_{1j}^{(0)} \beta_j^{(0)}}_{y^{(0)}_{\frac{p(p+1)}{2}+1}} + \underbrace{\sum_{l=1}^{k} x_{l1} z_l}_{\tau_{\frac{p(p+1)}{2}+1}(\mathbf{X}, z)} \Big)
\end{aligned}
$$

$$
\vdots \qquad \vdots \qquad \vdots \qquad \ddots \qquad \ddots \qquad \ddots
$$

$$
\underbrace{-\frac{\beta_p}{\sigma^2}}_{\psi_{\frac{p(p+3)}{2}}} \cdot \underbrace{\frac{1}{n^{(0)}} \sum_{j=1}^{p} \lambda_{pj}^{(0)} \beta_j^{(0)}}_{y^{(0)}_{\frac{p(p+3)}{2}}} \longrightarrow -\frac{\beta_p}{\sigma^2} \cdot \underbrace{\frac{1}{n^{(1)}} \sum_{j=1}^{p} \lambda_{pj}^{(1)} \beta_j^{(1)}}_{y^{(1)}_{\frac{p(p+3)}{2}}} = \underbrace{-\frac{\beta_p}{\sigma^2}}_{\psi_{\frac{p(p+3)}{2}}} \cdot \frac{1}{n^{(0)}+1} \Big(n^{(0)} \cdot \underbrace{\frac{1}{n^{(0)}} \sum_{j=1}^{p} \lambda_{pj}^{(0)} \beta_j^{(0)}}_{y^{(0)}_{\frac{p(p+3)}{2}}} + \underbrace{\sum_{l=1}^{k} x_{lp} z_l}_{\tau_{\frac{p(p+3)}{2}}(\mathbf{X}, z)} \Big)
$$

So, also equivalence of the update step is obtained.

For an overview, the $\frac{p(p+3)}{2}$-dimensional transformations ψ, $y^{(0)}$ and $\tau(\mathbf{X}, z)$ are presented again in the following:

$$\psi = \begin{bmatrix} -\frac{\beta_1^2}{2\sigma^2} \\ \vdots \\ -\frac{\beta_p^2}{2\sigma^2} \\ -\frac{\beta_1\beta_2}{\sigma^2} \\ \vdots \\ -\frac{\beta_{p-1}\beta_p}{\sigma^2} \\ -\frac{\beta_1}{\sigma^2} \\ \vdots \\ -\frac{\beta_p}{\sigma^2} \end{bmatrix}, \quad y^{(0)} = \begin{bmatrix} \frac{\lambda_{11}^{(0)}}{n^{(0)}} \\ \vdots \\ \frac{\lambda_{pp}^{(0)}}{n^{(0)}} \\ \frac{\lambda_{12}^{(0)}}{n^{(0)}} \\ \vdots \\ \frac{\lambda_{p-1,p}^{(0)}}{n^{(0)}} \\ \frac{1}{n^{(0)}} \sum_{j=1}^p \lambda_{1j}^{(0)} \beta_j^{(0)} \\ \vdots \\ \frac{1}{n^{(0)}} \sum_{j=1}^p \lambda_{pj}^{(0)} \beta_j^{(0)} \end{bmatrix}, \quad \tau(\mathbf{X}, z) = \begin{bmatrix} \sum_{l=1}^k x_{l1}^2 \\ \vdots \\ \sum_{l=1}^k x_{lp}^2 \\ \sum_{l=1}^k x_{l1} x_{l2} \\ \vdots \\ \sum_{l=1}^k x_{l,p-1} x_{l,p} \\ \sum_{l=1}^k x_{l1} z_l \\ \vdots \\ \sum_{l=1}^k x_{lp} z_l \end{bmatrix}$$

3 A Quick Detour on the Impossibility to Handle Unknown σ^2 in the Same Way

The whole argumentation only was possible because of $\mathbf{b}(\psi) = 0$. In the LUCK-model, the respective summand in the exponent is $n^{(0)} \cdot \mathbf{b}(\psi)$ for the prior and $n^{(1)} \cdot \mathbf{b}(\psi)$ for the posterior, whereas in an ordinary Bayesian update step, all terms depending on β only would remain unchanged. This is the crucial point why the so-called normal inverse gamma regression model, where σ^2 is considered unknown, too (and where the prior on σ^2 is modeled through an inverse gamma distribution), cannot be shown to be a LUCK-model as well. (See [3, Appendix] for a detailed description of this finding.)

References

[1] A. O'Hagan. *Bayesian Inference*. Kendall´s Advanced Theory of Statistics Vol. 2B, Arnold, 1994.

[2] E. Quaeghebeur and G. de Cooman. Imprecise Probability Models for Inference in Exponential Families. In: F. G. Cozman, R. Nau, and T. Seidenfeld (eds.) *Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications (ISIPTA'05)*, 2005.

[3] G. Walter. Robuste Bayes-Regression mit Mengen von Prioris — Ein Beitrag zur Statistik unter komplexer Unsicherheit. *Diploma thesis, Department of Statistics, Ludwig-Maximilians-University Munich*, 2006. `http://www.statistik.lmu.de/~thomas/team/diplomathesis_GeroWalter.pdf`

[4] G. Walter. Sketch of an Alternative Approach to Linear Regression Analysis under Sets of Conjugate Priors. *Discussion Paper*, 2007. `http://www.statistik.lmu.de/~thomas/team/isipta07_conjugate_prior.pdf`

[5] G. Walter, T. Augustin, and A. Peters. Linear Regression Analysis under Sets of Conjugate Priors. Under Revision for: *Proceedings of the Fifth International Symposium on Imprecise Probabilities and their Applications (ISIPTA'07)*, 2007.