Sketch of an Alternative Approach to Linear Regression Analysis under Sets of 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.

In that paper, two different conjugate priors on the regression parameter β were considered and generalized:

i) the standard choice, advocated, e.g., by [2], and the generalization of which was presented in detail in [6], and

ii) a prior constructed along the lines of [3, 1], which is to be presented and generalized in more detail in this supplement: Doing this, it will turn out that a similar powerful framework like for i) can also be built on this choice: at least for two regressors, the prior can be shown to be normal with linearly updated parameters, and so the extension to imprecise priors can be performed in a similar way.

Derivation of a Conjugate Prior to the Likelihood Arising from Linear Regression

The regression model is noted as follows:

$$z_i = x_i^\mathsf{T}\beta + \varepsilon_i\,, \quad x_i \in \mathbb{R}^p\,,\ \beta \in \mathbb{R}^p\,,\ \varepsilon_i \sim \mathrm{N}(0,\sigma^2)\,,$$

where z_i is the response, x_i the vector of the p covariates for observation i, and β is the p-dimensional vector of adjacent regression coefficients.

In this derivation as in [6], σ^2 is assumed to be known.

As x_i is considered to be non-stochastic, it holds that $z_i \sim \mathrm{N}(x_i^\mathsf{T}\beta, \sigma^2)$, and so the likelihood for k i.i.d. samples of this kind is the same as (9) in [6]:

$$\begin{aligned}
&f(z\,|\,\beta)\\
&= \prod_{i=1}^{k} f(z_i\,|\,\beta)\\
&= \frac{1}{(2\pi)^{\frac{k}{2}}\sigma^k}\exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{k}(z_i - x_i^\mathsf{T}\beta)^2\right\}\\
&= \frac{1}{(2\pi)^{\frac{k}{2}}\sigma^k}\exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{k} z_i^2 - \frac{1}{\sigma^2}\sum_{i=1}^{k} z_i\cdot x_i^\mathsf{T}\beta + \frac{1}{2\sigma^2}\sum_{i=1}^{k}(x_i^\mathsf{T}\beta)^2\right\}\\
&= \underbrace{\frac{1}{(2\pi)^{\frac{k}{2}}\sigma^k}\exp\left\{\frac{\sum_{i=1}^{k} z_i^2}{2\sigma^2}\right\}}_{\prod_{i=1}^{k}\mathbf{a}(z_i)}\cdot\exp\left\{\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{k} z_i\cdot\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)}_{\langle\psi,\,\tau^k(z)\rangle} - \underbrace{\frac{1}{2\sigma^2}\sum_{i=1}^{k}\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)^2}_{k\cdot\mathbf{b}(\psi)}\right\},
\end{aligned}$$

which corresponds to the form of the likelihood as requested by [3],

$$f(z\,|\,\psi) = \prod_{i=1}^{k} a(z_i)\cdot\exp\{\langle\psi,\tau^k(z)\rangle - k\cdot\mathbf{b}(\psi)\}\,, \tag{1}$$

where $\psi = \psi(\beta)$ is a function of β.

Although σ^2 is known, it is not possible to attach it to either β or some parts of $\mathbf{X}$ (the so-called design matrix of dimension $(k\times p)$, where row i consists of the observation x_i, $i = 1,\ldots,k$), because it would show up quadratic in $\langle\psi,\tau^k(z)\rangle$, but linear in $\mathbf{b}(\psi)$.

Analyzing the term $\langle \psi, \tau^k(z) \rangle$ more closely, we get

$$\begin{aligned}
\langle \psi, \tau^k(z) \rangle &= \frac{1}{\sigma^2} \sum_{i=1}^{k} z_i \cdot \left(\sum_{j=1}^{p} x_{ij} \beta_j \right) \\
&= \frac{1}{\sigma^2} \sum_{j=1}^{p} \sum_{i=1}^{k} z_i x_{ij} \beta_j \\
&= \sum_{j=1}^{p} \beta_j \cdot \left(\sum_{i=1}^{k} \frac{z_i \cdot x_{ij}}{\sigma^2} \right),
\end{aligned}$$

leading to

$$\begin{aligned}
\psi_j &= \beta_j \,, \quad j = 1, \ldots, p \,, \\
\tau^k(z)_j &= \frac{1}{\sigma^2} \left(\mathbf{X}^\mathsf{T} z \right)_j \,, \quad j = 1, \ldots, p \,, \\
\mathbf{b}(\psi) &= \frac{1}{2k\sigma^2} \sum_{i=1}^{k} \left(\sum_{j=1}^{p} x_{ij} \psi_j \right)^2 .
\end{aligned}$$

According to [3], a conjugate prior (and posterior) can be obtained by the following equation:

$$p(\psi) = \mathbf{c}(n, y) \cdot \exp\left\{ n \cdot \left[\langle \psi, y \rangle - \mathbf{b}(\psi) \right] \right\} .$$

Here, n and y are the parameters that define the prior distribution on ψ, and $\mathbf{b}(\psi)$ is to be taken from the likelihood. As in our case it holds that $\psi = \beta$, we have already the prior on β, writing

$$p(\beta) = \mathbf{c}(n, y) \cdot \exp\left\{ n \cdot \left[\langle \beta, y \rangle - \mathbf{b}(\beta) \right] \right\} . \tag{2}$$

The parameter space $\mathcal{Y}$ for y is taken as $\mathrm{co}(\mathcal{T})$, the convex hull of the space of $\tau(z)$ without the boundary (see [3] for more details), resulting in $\mathcal{Y} = \{ y : y \in \mathbb{R}^p \}$, because $\tau(z)_j = \frac{1}{\sigma^2} \left(x^\mathsf{T} z \right)_j \in \mathbb{R}$.

(2) can, at least for the case of two regressors ($p = 2$), be shown to be a multivariate normal distribution on β. Starting with arbitrary p, the term in the exponent of a normal distribution on β can have one of the two following forms:

$$\frac{1}{g(n, y)} \sum_{j=1}^{p} \left(\beta_j - h_j(n, y) \right)^2 \tag{3}$$

$$\text{or} \qquad \left(\beta - f(n, y) \right)^\mathsf{T} \mathbf{S}(n, y) \left(\beta - f(n, y) \right) , \tag{4}$$

where f and h are p-dimensional functions of n and / or y forming the expected value of β. $g(n, y)$ is a onedimensional function playing the role of the variance in the case of the components of β being uncorrelated, and $\mathbf{S}(n, y)$ is forming the inverse of the covariance matrix for a multivariate normal with correlations between the components of β.

Considering the term in the exponent in (2):

$$n \cdot [\langle \beta, y \rangle - \mathbf{b}(\beta)] = n \cdot \sum_{j=1}^{p} \beta_j \cdot y_j - \frac{n}{2k\sigma^2} \sum_{i=1}^{k} \left(\sum_{j=1}^{p} x_{ij} \beta_j \right)^2 , \tag{5}$$

we see that there must be summands without, with linear and with squared components of β in (5) if the square is expanded. As this can be only provided by the second form of multivariate normal exponents, we will focus on (4) and expand it in order to compare coefficients with an expanded version of (5).[1] As this is tricky for the general case of arbitrary p (being the number of covariates / regressors), we focus attention on the case of two regressors as we did in the detailed study presented in [6], Sections 3.3, 4–6.

Denoting the elements of $\mathbf{S}(n, y)$ with s_{ij}, $i, j = 1, 2$, we get for (4)

$$\begin{aligned}
&\big(\beta - f(n,y)\big)^{\mathsf{T}} \mathbf{S}(n,y) \big(\beta - f(n,y)\big) \\
&\quad = \big(\beta_1 - f_1(n,y)\big)^2 s_{11} + 2\big(\beta_1 - f_1(n,y)\big)\big(\beta_2 - f_2(n,y)\big) s_{12} + \big(\beta_2 - f_2(n,y)\big)^2 s_{22} \\
&\quad = \beta_1^2 s_{11} - 2\beta_1 f_1(n,y) s_{11} + f_1(n,y)^2 s_{11} \\
&\qquad\quad + 2\beta_1\beta_2 s_{12} - 2\beta_1 f_2(n,y) s_{12} - 2\beta_2 f_1(n,y) s_{12} + 2 f_1(n,y) f_2(n,y) s_{12} \\
&\qquad\quad + \beta_2^2 s_{22} - 2\beta_2 f_2(n,y) s_{22} + f_2(n,y)^2 s_{22} \\
&\quad = \beta_1^2 s_{11} + \beta_2^2 s_{22} + 2\beta_1\beta_2 s_{12} \\
&\qquad\quad - 2\beta_1\big(f_1(n,y) s_{11} + f_2(n,y) s_{12}\big) - 2\beta_2\big(f_1(n,y) s_{12} + f_2(n,y) s_{22}\big) \\
&\qquad\quad + f_1(n,y)^2 s_{11} + f_2(n,y)^2 s_{22} + 2 f_1(n,y) f_2(n,y) s_{12} \,,
\end{aligned}$$

whereas for (5), we get

$$\begin{aligned}
&n \cdot [\langle \beta, y \rangle - \mathbf{b}(\beta)] \\
&\quad = n \cdot (\beta_1 y_1 + \beta_2 y_2) - \frac{n}{2k\sigma^2} \sum_{i=1}^{k} (x_{i1}\beta_1 + x_{i2}\beta_2)^2 \\
&\quad = n\beta_1 y_1 + n\beta_2 y_2 - \frac{n}{2k\sigma^2} \sum_{i=1}^{k} x_{i1}^2\beta_1^2 + 2x_{i1}x_{i2}\beta_1\beta_2 + x_{i2}^2\beta_2^2 \\
&\quad = n\beta_1 y_1 + n\beta_2 y_2 - \frac{n}{2k\sigma^2}\beta_1^2 \sum_{i=1}^{k} x_{i1}^2 - \frac{n}{k\sigma^2}\beta_1\beta_2 \sum_{i=1}^{k} x_{i1}x_{i2} - \frac{n}{2k\sigma^2}\beta_2^2 \sum_{i=1}^{k} x_{i2}^2 \,.
\end{aligned}$$

[1] (5) must be expanded because we must arrive at a sum on the components of β, and not on the components of z as we have in the second summand in (5). In order to obtain a sum running on j, we will have to expand the squared summands in the sum over i and reorder them.

Comparing coefficients for squared components of β, we obtain

$$\begin{aligned} s_{11} &= -\frac{n}{2k\sigma^2}\sum_{i=1}^{k} x_{i1}^2 \quad &&= -\frac{n}{2k\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X})_{11} &&= -\frac{n}{2k\sigma^2}\mathbf{x}_{11} \\ s_{22} &= -\frac{n}{2k\sigma^2}\sum_{i=1}^{k} x_{i2}^2 \quad &&= -\frac{n}{2k\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X})_{22} &&= -\frac{n}{2k\sigma^2}\mathbf{x}_{22} \\ s_{12} &= -\frac{n}{2k\sigma^2}\sum_{i=1}^{k} x_{i1}x_{i2} &&= -\frac{n}{2k\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X})_{12} &&= -\frac{n}{2k\sigma^2}\mathbf{x}_{12} \end{aligned}$$

$$\text{or} \quad \mathbf{S}(n,y) = -\frac{n}{2k\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X})\,,$$

where $\mathbf{x}_{ij}$, $i,j = 1,2$ denotes the components of $\mathbf{X}^\mathsf{T}\mathbf{X}$.

Using the same notation as in [5] or [6] for the multivariate normal distribution, where the covariance matrix is denoted by $\sigma^2\mathbf{\Sigma}$, with $\mathbf{\Sigma}$ defining the covariance structure, we have

$$\mathbf{\Sigma} = \left(\frac{n}{k}\cdot\mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1} = \frac{k}{n}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\,.$$

Comparing coefficients for the terms linear in β_1 and β_2, respectively, it follows that

$$\begin{aligned} ny_1 &= \frac{n}{k\sigma^2}\mathbf{x}_{11}f_1(n,y) + \frac{n}{k\sigma^2}\mathbf{x}_{12}f_2(n,y) \\ ny_2 &= \frac{n}{k\sigma^2}\mathbf{x}_{12}f_1(n,y) + \frac{n}{k\sigma^2}\mathbf{x}_{22}f_2(n,y) \end{aligned}$$

leading to

$$\begin{aligned} f_1(y) = \quad & k\sigma^2\frac{\mathbf{x}_{22}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2}\,y_1 - k\sigma^2\frac{\mathbf{x}_{12}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2}\,y_2 \\ f_2(y) = -& k\sigma^2\frac{\mathbf{x}_{12}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2}\,y_1 + k\sigma^2\frac{\mathbf{x}_{11}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2}\,y_2\,, \end{aligned}$$

that is, in matrix notation,

$$f(y) = A\,y \qquad \text{and} \qquad y = A^{-1}f(y)\,,$$

with

$$A = \begin{pmatrix} \dfrac{k\sigma^2\mathbf{x}_{22}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2} & -\dfrac{k\sigma^2\mathbf{x}_{12}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2} \\ -\dfrac{k\sigma^2\mathbf{x}_{12}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2} & \dfrac{k\sigma^2\mathbf{x}_{11}}{\mathbf{x}_{11}\mathbf{x}_{22}-\mathbf{x}_{12}^2} \end{pmatrix} = k\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\,.$$

The system $y \mapsto f(y)$ is solvable if A is invertible, that is, if $\mathbf{X}^\mathsf{T}\mathbf{X}$ is invertible, which is same the condition as for the calculation of the classical least squares estimator $\hat{\beta}_{LS} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}z$.

Therefore, under this regularity condition, it holds that

$$f(y) = k\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\, y \qquad \text{and} \qquad y = \frac{1}{k\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X}) f(y)\,.$$

As the result, we have

$$\beta \sim N_p\left(A\, y\,,\ \frac{k\sigma^2}{n}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\right)$$

as the conjugate prior, or, with y as the expectation value,

$$A^{-1}\beta \sim N_p\Big(y\,,\ \underbrace{A^{-1}\frac{k\sigma^2}{n}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}A^{-1}}_{=\frac{1}{nk\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X})}\Big)\,.$$

Now, if a conjugate prior of the form (2) is updated with the likelihood (1) according to Bayes's rule $p(\beta \,|\, z) \propto f(z \,|\, \beta) \cdot p(\beta)$, then the posterior has the same form as the prior, but with the parameters n and y updated from their prior values $n^{(0)}$ and $y^{(0)}$ to the posterior values $n^{(1)}$ and $y^{(1)}$ as follows:[2]

$$\begin{aligned} n^{(1)} = n^{(0)} + k \qquad \text{and} \qquad y^{(1)} &= \frac{n^{(0)}y^{(0)} + \tau^k(z)}{n^{(0)} + k} \\ &= \frac{n^{(0)}}{n^{(0)} + k}\, y^{(0)} + \frac{k}{n^{(0)} + k} \cdot \frac{1}{k\sigma^2}(\mathbf{X}^\mathsf{T} z)\,. \end{aligned} \tag{6}$$

Just as an example for the similarity (not equivalence!) between the results obtained here and the model produced by choice i), the posterior expected value for β_1 is presented here:

$$\begin{aligned} &f_1(y^{(1)}) \\ &= \frac{(k\sigma^2\mathbf{x}_{22})\left[\frac{n^{(0)}}{n^{(0)}+k}\, y_1^{(0)} + \frac{1}{n^{(0)}+k} \cdot \frac{1}{\sigma^2}(\mathbf{X}^\mathsf{T} z)_1\right] - (k\sigma^2\mathbf{x}_{12})\left[\frac{n^{(0)}}{n^{(0)}+k}\, y_2^{(0)} + \frac{1}{n^{(0)}+k} \cdot \frac{1}{\sigma^2}(\mathbf{X}^\mathsf{T} z)_2\right]}{\mathbf{x}_{11}\mathbf{x}_{22} - \mathbf{x}_{12}^2} \end{aligned}$$

This can be compared to the form of $\beta_1^{(1)}$ to be found below Remark 3 in [6], or, making the similarity more obvious, with (4.22) in [4, p. 81], where b_1 plays the role of $y_1^{(0)}$ and b_2 the role of $y_2^{(0)}$.

As is described in detail in [3], the model of Bayesian updating obtained with such *linearly* updated prior parameters can be generalized in a straighforward way to an imprecise probability calculus by using sets of priors (instead of a single prior as in classical Bayesian learning): When sets of priors are defined via sets of parameters, and these sets of parameters are defined by lower and upper bounds, the lower and upper

[2] As in [6], we will denote the prior parameters with the upper index $^{(0)}$, whereas the posterior parameters will be denoted with upper index $^{(1)}$.

bounds of the sets of posterior parameters can be obtained directly from (6). The posterior parameter $y^{(1)}$ is then also the posterior expected value of the posterior distribution.[3]

As seen above, we have $A\,y$ as the expected value of the normal distribution on β. To get interpretable results, we must therefore consider sets of $A\,y^{(0)}$ and 'translate' them into sets of $y^{(0)}$, which are linearly updated to $y^{(1)}$ and finally 'retranslated' to sets of $A\,y^{(1)}$ again, corresponding to the posterior expected value of β. Because the 'translations' are linear, we can get from the lower and upper bounds on $A\,y^{(0)}$ directly to the lower and upper bounds on $A\,y^{(1)}$.[4] Noted as a step-by-step procedure, we must,

1. fix the lower and upper bounds for $f(y^{(0)})$ based on the prior knowledege on β; $n^{(0)}$ must be chosen fix[5] and determines the prior variance matrix for β, as it holds that $\mathbb{V}(\beta) = \frac{k\sigma^2}{n^{(0)}}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$;
2. 'translate' the bounds for $f(y^{(0)})$ into bounds for $y^{(0)}$ by $y^{(0)} = \frac{1}{k\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X})f(y^{(0)})$;
3. perform the linear update step on $n^{(0)}$ and the bounds for $y^{(0)}$ to obtain $n^{(1)}$ and bounds for $y^{(1)}$;
4. 'retranslate' the bounds for $y^{(1)}$ into interpretable bounds for $f(y^{(1)})$.

Performing these four steps, we get for the posterior expected value of β

$$\begin{aligned}
\mathbb{E}[\beta \mid z] &= f(y^{(1)}) \\
&= k\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\, y^{(1)} \\
&= k\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\left(\frac{n^{(0)}}{n^{(0)}+k}\, y^{(0)} + \frac{k}{n^{(0)}+k}\cdot\frac{1}{k\sigma^2}(\mathbf{X}^\mathsf{T} z)\right) \\
&= k\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\frac{n^{(0)}}{n^{(0)}+k}\cdot\frac{1}{k\sigma^2}(\mathbf{X}^\mathsf{T}\mathbf{X})f(y^{(0)}) + k\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\frac{k}{n^{(0)}+k}\cdot\frac{1}{k\sigma^2}(\mathbf{X}^\mathsf{T} z) \\
&= \frac{n^{(0)}}{n^{(0)}+k}\cdot f(y^{(0)}) + \frac{k}{n^{(0)}+k}\cdot\underbrace{(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T} z}_{\hat{\beta}_{LS}},
\end{aligned}$$

and therefore, the posterior expected value is the weighted mean of some chosen prior expected value $f(y^{(0)})$ and the common least squares estimator for β with the weights $n^{(0)}$ and k, respectively.

[3] See [3] why this is the case also for the general case of distributions belonging to a so-called exponential family.

[4] It is only necessary to take into account that either $\mathbf{x}_{12}$ or $-\mathbf{x}_{12}$ are negative.

[5] We commented on this in [6, Section 6]

As there are no 'translations' performed on $n^{(0)}$, the posterior variance matrix of β is simply updated to

$$\begin{aligned} \mathbb{V}(\beta \,|\, z) &= \frac{k\sigma^2}{n^{(1)}}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1} \\ &= \frac{k\sigma^2}{n^{(0)}+k}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\,. \end{aligned} \tag{7}$$

Therefore, the lower and upper bounds for the posterior expectation value of β can be derived easily from the lower and upper bounds for the prior expectation value $f(y^{(0)})$, whereas the prior variance matrix does not vary in a set and is simply updated to the posterior variance matrix through $n^{(1)} = n^{(0)} + k$. Hence in this model, imprecise calculus is only possible for the expectation value, but quite simple to perform.

As some exemplary results, the decrease of imprecision for $f(y^{(0)})$ obtained by the update step is quantified by

$$\overline{f(y^{(1)})} - \underline{f(y^{(1)})} = \frac{n^{(0)}}{n^{(0)}+k}\left(\overline{f(y^{(0)})} - \underline{f(y^{(0)})}\right)\,,$$

where, e.g., $\overline{f(y^{(1)})}$ represents the upper bound for $f(y^{(1)})$. So, for $n^{(0)} = k$, imprecision is reduced to its half after the update step.

The prior variance of some regression coefficient β_j is, for the same choice of $n^{(0)}$, also reduced to its half after the update step, as can be seen in (7).

References

[1] J. Bernardo and A. Smith. *Bayesian Theory.* Wiley & Sons, 1993.

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

[3] 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.

[4] 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`

[5] G. Walter. The Normal Regression Model as a LUCK-model. *Discussion Paper*, 2007.
`http://www.statistik.lmu.de/~thomas/team/isipta07_proof.pdf`

[6] 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.