Linear Regression Analysis under Sets of Conjugate Priors

Gero Walter
Department of Statistics
Ludwig-Maximilians-University
Munich, Germany
gero.walter@stat.uni-muenchen.de

Thomas Augustin
Department of Statistics
Ludwig-Maximilians-University
Munich, Germany
thomas@stat.uni-muenchen.de

Annette Peters
GSF – National Research Center for Environment and Health
Neuherberg, Germany
peters@gsf.de

Abstract

REGRESSION is *the* central concept in applied statistics for analyzing multivariate, heterogeneous data: The influence of a group of variables on one other variable is quantified by the regression parameter β. Here we extend standard Bayesian inference on β in linear regression models by considering imprecise conjugate priors. Inspired by 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, we develop a general framework for handling linear regression models including analysis of variance models, and discuss obstacles in direct implementation of the method. Then properties of the interval-valued point estimates for a two-regressor model are derived and illustrated with simulated data. As a practical example we take a small data set from the AIRGENE study and consider the influence of age and body mass index on the concentration of an inflammation marker.

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.

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

$\mathbf{X}$ is considered fixed and non-stochastic, and so

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

In principle several conjugate priors to this likelihood exist. The standard choice (see, e.g., [2, p. 244ff]), on which we focused in our work, is

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

i.e. $p(\beta) =$

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

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

$$\beta^{(1)} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X} + \Lambda^{(0)}\right)^{-1}\left(\mathbf{X}^{\mathsf{T}}z + \Lambda^{(0)}\beta^{(0)}\right) \quad (2)$$

$$\Sigma^{(1)} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X} + \Lambda^{(0)}\right)^{-1}, \quad \text{where } \Lambda^{(0)} = \Sigma^{(0)^{-1}}. \quad (3)$$

The model of Bayesian regression analysis based on the standard prior will be called *normal regression model*.
(A different conjugate prior is derived in [6].)

2. Classical Bayesian Inference and LUCK-models

WE 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 $f(w\,|\,\vartheta)$ by

$$p(\vartheta\,|\,w) \propto f(w\,|\,\vartheta)\cdot p(\vartheta)\,, \quad (4)$$

where the prior $p(\vartheta)$ and the posterior $p(\vartheta\,|\,w)$ 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.

Definition 1 *Consider classical Bayesian inference on ϑ based on a sample w as described in (4), and let the prior $p(\vartheta)$ be characterized by the (vectorial) parameter $\vartheta^{(0)}$. Call $(p(\vartheta), p(\vartheta\,|\,w))$ 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 $q \in \mathbb{N}$, transformations $\vartheta \mapsto \psi$, $\vartheta \mapsto \mathbf{b}(\psi)$ and $\vartheta^{(0)} \mapsto (n^{(0)}, y^{(0)})$ such that*

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

$$\textit{and} \quad p(\vartheta\,|\,w) \propto \exp\left\{n^{(1)}\left[\langle\psi, y^{(1)}\rangle - \mathbf{b}(\psi)\right]\right\}, \quad \textit{where} \quad (6)$$

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

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

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

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

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

Proof: The proof is given in [5].

3. Imprecise Priors for Inference in LUCK-models

TO create sets of priors, we rely on the work of Quaeghebeur and de Cooman [3], who consider certain LUCK-models for Bayesian inference based on i.i.d. observations from *regular, linear canonical exponential families* [1, p. 202 and p. 272f]. The central idea of [3] is that the parameterization in terms of $y^{(0)}$ and $n^{(0)}$ is perfectly suitable to be generalized to credal sets of priors. As the crucial point is that these parameters are updated *linearly*, [3]'s technique can be applied to any LUCK-model, leading to the same imprecise calculus as in the IDM (which is contained as a special case, $y^{(0)} \leftrightarrow t$, $n^{(0)} \leftrightarrow s$): Let $y^{(0)}$ vary in some set $\mathcal{Y}^{(0)} \subset \mathcal{Y}$ and take as the imprecise prior the credal set consisting of all convex mixtures of all $p(\vartheta)$ from (5) created by varying $y^{(0)}$ in $\mathcal{Y}^{(0)}$. The posterior credal set is then characterized by its extreme points, the set of posteriors $p(\vartheta\,|\,w)$ arising from (6) by varying $y^{(1)}$ in $\mathcal{Y}^{(1)}$, where

$$\mathcal{Y}^{(1)} = \left\{\frac{n^{(0)}y^{(0)} + \tau(w)}{n^{(0)} + n} \,\middle|\, y^{(0)} \in \mathcal{Y}^{(0)}\right\} \subset \mathcal{Y}. \quad (10)$$

$\mathcal{Y}^{(1)}$ can be seen as a shifted and rescaled version of $\mathcal{Y}^{(0)}$:

$$\mathcal{Y}^{(1)} = \frac{n^{(0)}}{n^{(0)}+n}\cdot\mathcal{Y}^{(0)} + \frac{n}{n^{(0)}+n}\cdot\frac{1}{n}\sum_{i=1}^{n}\tau(w_i), \quad (11)$$

which suggests a vivid interpretation of $n^{(0)}$ as "prior strength" or "pseudocounts", as it plays the same role for the prior as n for the sample.

The set $\mathcal{Y}^{(0)}$ should reflect the prior information on the parameters, but must be bounded to avoid vacuous posterior inference, as choosing $\overline{y}_j^{(0)} = \infty$ would lead to $\overline{y}_j^{(1)} = \infty$.

4. The Imprecise Normal Regression Model

FOR multidimensional $\mathcal{Y}$, [3] suggest to bound $\mathcal{Y}^{(0)}$ by some global constraints. Their suggestion for the multivariate normal distribution is adopted here, leading to

$$\frac{1}{n^{(0)}}\Lambda^{(0)} \quad \text{positive definite (p.d.)}, \quad \text{and} \quad (12)$$

$$\frac{1}{n^{(0)}}\left(\Lambda^{(0)} - \frac{1}{n^{(0)}}\Lambda^{(0)}\beta^{(0)}\beta^{(0)^{\mathsf{T}}}\Lambda^{(0)}\right) \quad \text{p.d.} \quad (13)$$

To apply the normal regression model as an imprecise probability model, we have to proceed as follows:

1. Prior knowledge on β must be expressed as a set of values of $\beta^{(0)}$ and $\Lambda^{(0)}$.
2. This set must be "translated" into a set of values of $y^{(0)}$ such that the resulting set $\mathcal{Y}^{(0)}$ satisfies (12) and (13).
3. Then each $y^{(0)}$ in $\mathcal{Y}^{(0)}$ is linearly updated by (7) to $y^{(1)}$.
4. The obtained set $\mathcal{Y}^{(1)}$ must be "retranslated" into an interpretable set of values of $\beta^{(1)}$ and $\Lambda^{(1)}$.

Defining the set by element-wise lower and upper bounds, e.g.,

$$\beta_j^{(0)} \in \left[\underline{\beta}_j^{(0)}, \overline{\beta}_j^{(0)}\right] \qquad j = 1, \ldots, p,$$

the "translation" step 2. turns out to be quite difficult, as it holds that

$$\underline{y}_{bi}^{(0)} = \min_{\beta^{(0)},\Lambda^{(0)}} \frac{1}{n^{(0)}}\sum_{j=1}^{p}\lambda_{ij}^{(0)}\beta_j^{(0)}$$

$$\overline{y}_{bi}^{(0)} = \max_{\beta^{(0)},\Lambda^{(0)}} \frac{1}{n^{(0)}}\sum_{j=1}^{p}\lambda_{ij}^{(0)}\beta_j^{(0)},$$

and maximization and minimization must be executed only on combinations of values between the bounds on $\beta^{(0)}$ and $\Lambda^{(0)}$ that are admissible according to (12) and (13), who form polynomial constraints of degree p when checking wether all eigenvalues are positive. Thus, analytical results for arbitrary p are not within reach.

5. Application

TO obtain interpretable analytical expressions, we focused on the case of two regressors, the results of which can be found in detail in [4]. The resulting model features reasonable properties, was tested in a short simulation study and was applied to a small data set from the AIRGENE study, assessing the influence of age and body mass index on the concentration of an inflammation marker.

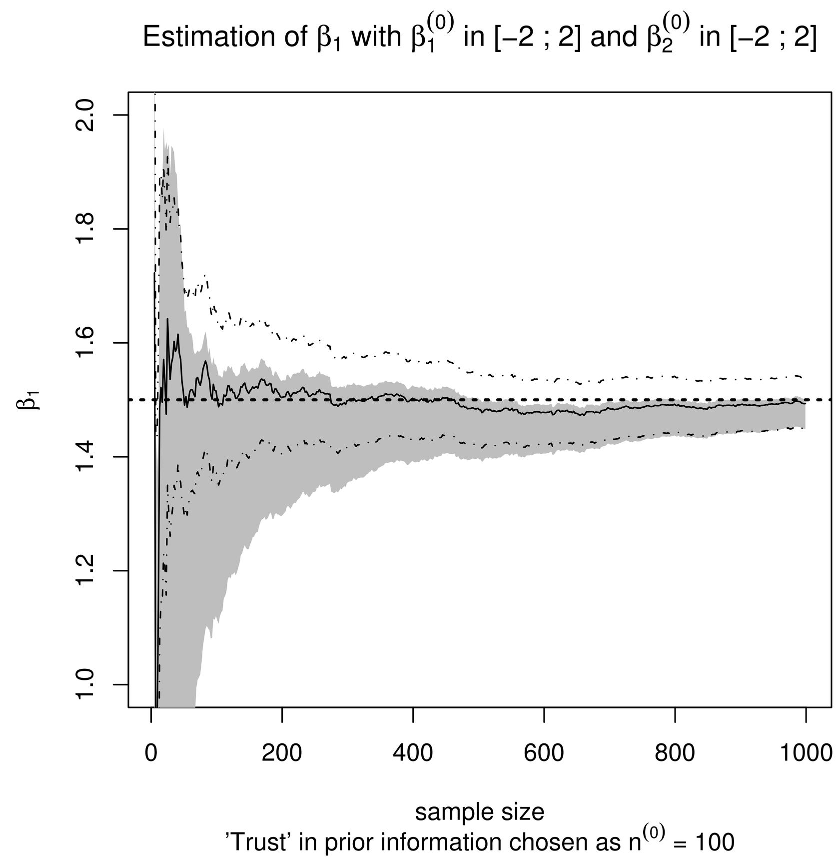

Figure 1: *Illustration of asymptotic behavior of interval-valued regression parameter estimates.*

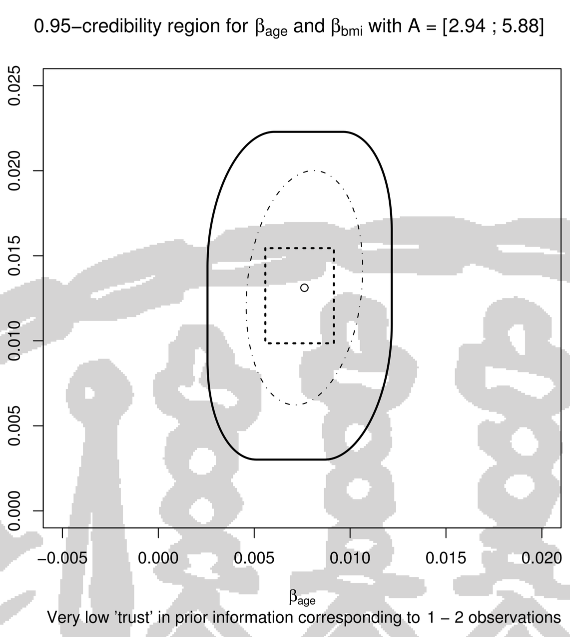

Figure 2: *Exemplary results for the* AIRGENE *data.*

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.) *Proc. of the 4th Intern. Symp. 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, LMU 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. 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`