

Chapter 1

Appendix

The error term in the by [Gelman *et al.* \(2017\)](#) proposed definition of the R^2 is defined as $\text{Var}(\sum_{i=1}^n e_n^s)$. I think we could also use $\sum (y - \hat{y}^s)^2 / (n - 1)$ as an estimate for the error. For the maximum likelihood estimate $\text{Var}(y_i - \hat{y}_i) = \sum (y_i - \hat{y}_i)^2 / (n - 1)$. This is because the mean of the residuals is 0. When samples of the posterior parameters are used, the mean of the residuals is not exactly zero. $\text{Var}(y_i - \hat{y}_i) = \sum (y_i - \hat{y}_i)^2 / (n - 1)$ is than a little bit bigger than $\text{Var}(y_i - \hat{y}_i)$. In practice the values should only differ by a very small amount. We do not expect the errors to have a systematic bias. However, the residuals are just a sample of the error. The mean of the residuals must not be exactly 0 when the samples of the posteriors are used for the regression coefficients.

1.1 Abstract

Quantifying the importance of predictors in regression has been an active area of research for a long time ([Grömping, 2015](#)). The variance decomposition metric *LMG* provides useful information about possible associations between the variables. The LMG metric is implemented in the R packages `hier.part` and `relaimpo`. Bayesian methods gained high popularity in many applied research areas in recent years.

This master thesis shows how the LMG metric can be applied to a linear regression model that is fitted in the Bayesian framework. The LMG metric requires calculation of R^2 for all the possible submodels. The conditional variance formula can be applied to calculate the R^2 values of the submodels from the posterior samples of the model containing all predictors. The mutual interdependence of the submodels is then respected for each posterior sample.

The implementation of the LMG metric in the Bayesian framework is presented on simulated and on empirical data. Using weakly informative priors resulted in very similar LMG values as the values obtained by using bootstrap in `relaimpo`.

Some possible extension of the LMG metric to repeated measures studies are additionally presented. There are certain difficulties involved in quantifying the R^2 in longitudinal data. However, the extension seems to be reasonably possible for the simple random intercept model and marginal models.

Bibliography

Gelman, A., Goodrich, B., Gabry, J., and Ali, I. (2017). R-squared for Bayesian regression models *. Technical report. [1](#)

Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, **7**, 137–152. [1](#)