Chapter 1

Discussion and Conclusion

1.1 Other variable importance metrics in the Bayesian framework

Different variable importance metrics exist (Grömping, 2015). The focus of this master thesis was on the LMG formula. A lot of variable importance metrics are based on the R^2 of the full model compared to the submodels. Instead of the LMG formula, another variable importance metric could as well be used on the R^2 of all submodels for each posterior sample. Commonality analysis (Nimon *et al.*, 2008) and dominance analysis (Grömping, 2015) seem to be interesting variance decomposition metrics. Both provide some extensions to the LMG framework. Different options exist further for the description of the LMG distribution. The `relaimpo` package provides for example some more bootstrap options, like pairwise differences, that could easily be transferred to the Bayesian framework.

1.2 Conclusion

The Bayesian frameworks provides the option to include the uncertainty about parameters. Using the conditional variance formula allows to calculate the R^2 of all the submodels from the posterior parameter distributions of the full model. Instead of fitting $2^p - 1$ models, only the full model needs to be fitted. The interdependence of the submodels to each other is then automatically respected. The R^2 of the submodels does not decrease when adding predictors. The important property of non-negativity shares is then respected in the Bayesian framework.

A disadvantage about calculating the R^2 of all the submodels with the conditional variance formula seems to be the restriction to the linear model. Although this may be a topic of further research. A further disadvantage of the Bayesian framework compared to the classical LMG implementation, are the higher computational costs. For non-stochastic predictors, the calculations are possible in a reasonable amount of time. Parallel computing could be used to speed up the calculations. For stochastic predictors, the computation times will be much higher than in the classical framework.

Assuming non-stochastic or stochastic predictors can have a big impact in small samples on the the uncertainty of the explained variance and the LMG values. Although the posterior regression parameter distributions is the same in both cases (under some assumptions described in chapter 2), the explained variance of a model is directly dependent on the covariance matrix.

Inference about the covariance of the predictors $\mathbf{X}$ is therefore an important part when stochastic predictors are assumed. However, this does in general not seem to be an easy problem. Non-parametric bootstrap provides a practical solution in the Bayesian framework.

When the sample size is large enough the classical and the Bayesian framework should lead to very similar values. The Bayesian framework allows to include prior information, that may especially be relevant for small sample sizes. The credible interval of the Bayesian framework, compared to the confidence intervals, may further be easier to interpret in the mathematically correct way.

A lot of studies are concerned with within-subject changes. The extension of the LMG formula to those kind of problems is not straightforward. It is depending on the complexity of the data. However, for the simple random intercept model the extension seems to be easily possible when the focus is on the fixed effects. In the Bayesian and maximum likelihood framework it seems reasonable to compare only the explained variance of the fixed effects of the submodels against the total variance, containing $\sigma_f^2 + \sigma_\alpha^2 + \sigma_\epsilon^2$. Otherwise, there may be problems with the non-negativity property of the shares.

Bibliography

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

Nimon, K., Lewis, M., Kane, R., and Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, **40**, 457–466. 1