**Susanne Konrath und Fabian Scheipl**

# Bayesian Regularization

## Abstract:

In regression models with a larger number of covariates, some kind of regularisation or variable selection strategy is typically desired, both to avoid overfitting and to select only "important" covariates. Several popular possibilities that try to combine regularisation and variable selection have been developed throughout recent years in the context of penalised likelihood approaches. The general idea is to add a penalty term to the likelihood that penalises the model complexity and therefore avoids overfitting. Two well-known examples include the quadratic penalty (i.e. the L-2-norm of the regression coefficients) and the LASSO (corresponding to the L1-norm penalty). While the former has the advantage of allowing for a closed form solution, the so-called ridge estimator, the latter allows combining regularisation and variable selection. Due to the shape of the contours of the penalty function, "small" covariates can be estimated to be exactly zero when maximising the likelihood subject to the LASSO penalty. While efficient algorithms, such as least angle regression for Gaussian regression models and modifications for exponential family regression and continuous survival time Coxmodels have been developed that allow for the routine estimation of LASSO-penalised regression models, the extension to more flexible models is typically difficult. Such extensions are for example required if some covariate effects should be modelled in a nonparametric fashion. This leads to semiparametric additive models that combine linear effects of covariates that are to be regularised with nonparametric effects of further covariates. Another example are geoadditive models that add spatial effects to the semiparametric predictor. The situation becomes even more complicated if semiparametric hazard regression models are to be considered. Such models extend the Cox model by the possibility to jointly estimate the baseline hazard rate and the covariate effects (modelling the baseline hazard rate as a nonparametric function, e.g. based on penalised splines), and also possibly contain additional semiparametric or spatial effects. To allow for the inclusion of such non-standard effects in the regularisation framework, it is convenient to reformulate the problem in a Bayesian setting. In this case, the penalty terms augmented to the likelihood correspond to log-prior terms that express the prior knowledge about the regression coefficients. Adding a large penalty then simply means that -- a priori -- we expect a lot of regression coefficients to be zero. The most popular priors correspond to a Gaussian prior (ridge regression) and a Laplace or Double Exponential prior (LASSO), respectively. The major advantage of the Bayesian formulation is not the Bayesian interpretation but that Markov chain Monte Carlo (MCMC) simulation techniques are available as a general and versatile tool for estimation. Due to the modular structure of MCMC algorithms, it is easy to extend the model at some place without having to re-implement the rest of the estimation algorithm. The major drawback of the sampling-based approach is that the sharp variable selection property of the usual LASSO approach is lost. This is due to the fact that MCMC does not maximize the posterior (or the penalised likelihood) but estimates the posterior expectation or median of the regression coefficients. However, regularisation of the regression model still takes place and coefficients corresponding to covariates with minor effect are typically shrunken close to zero.

Based on previous work on semiparametric regression, our aim is consider regression models that combine structured additive predictors with the possibility to include regularisation priors for some of the fixed or nonparametric effects. As a first step to reach this goal we present results with some priors of the exponential power family and priors based on scale mixtures of the normal distribution for the linear model.