Bayesian Effect Fusion for Categorial and Ordinal Predictors

Daniela Pauger[1], Helga Wagner[1]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] Department of Applied Statistics, Johannes Kepler University, Linz, Austria
`<daniela.pauger,helga.wagner>@jku.at`

Abstract

We extend Bayesian variable selection methods to categorial and ordinal predictor variables. We use a hierarchical spike and slab prior distribution on the regression effects to reduce the possibly high-dimensional vector of coefficients by fusion of similar effects and removing of variables with no non-zero effects. The demonstrate the effectiveness of our methods we model the income in Austria based on SILC (= survey on income and living conditions) data.
Keywords: spike and slab prior distribution; sparse modelling; dummy coding; EU-SILC

1 Introduction

In regression type models often many of the collected variables are categorial, measured on an ordinal or nominal scale. The usual modelling strategy using dummy variables for each category can easily lead to a high-dimensional vector of regression effects. A sparse representation of the model can be achieved by fusing category levels with essentially the same effect into one category and by removing variables where none of the levels has a non-zero effect.

2 Model

Let y denote the normal response in a standard linear regression model with $j = 1, \ldots, p$ categorial covariates c_j. We assume that the j-th covariate has $K_j + 1$ categories $0, \ldots, K_j$ and the first category 0 defines the reference category. We specify the linear regression model as

$$y = \mu + \sum_{j=1}^{p} \sum_{k=1}^{K_j} X_{jk} \theta_{j,k0} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

with regressors X_{jk} defined as in [2] using split coding for ordinal covariates, i.e.

$$X_{jk} = \begin{cases} 1, & \text{for } c_j \geq k \\ 0, & \text{otherwise,} \end{cases} \qquad k = 1, \ldots K_j$$

and usual dummy coding for nominal covariates.

For the nominal covariates the regression effects corresponding to the K_j dummy variables can be interpreted as the effect contrast of category k and the reference category. To allow for fusion of level effects we define for all $k > l$ by $\theta_{j,kl}$ the effect contrast of categories k and l of covariate c_j, which leads to the restriction

$$\theta_{j,k0} - \theta_{j,l0} - \theta_{j,kl} = 0, \qquad \text{for all } 0 < l < k \leq K_j.$$

We subsume all parameters $\theta_{j,kl}$ with $0 \leq l < k \leq K_j$ in the vector $\boldsymbol{\theta}_j$.

3 Priors and Inference

To encourage sparsity in the coefficient vectors we specify for each element of $\boldsymbol{\theta}_j$ a spike and slab prior distribution [1] hierarchically as

$$p(\theta_{j,kl}|\delta_{j,kl}, \tau^2) \sim \delta_{j,kl}\mathcal{N}(0, \tau^2) + (1 - \delta_{j,kl})\mathcal{N}(0, r\tau^2),$$

where r is a small value and $\delta_{j,kl}$ is an indicator for the slab component with prior distribution $p(\delta_{j,kl} = 1) = w_j$. w_j corresponds to the weight of variable j and is either fixed or assigned a hyperprior, such as a Beta hyperprior $w_j \sim \mathcal{B}(a_{0j}, b_{0j})$. For the variance parameter τ^2 we investigate various prior choices. Effect fusion is accomplished by the spike component: If $\delta_{j,kl} = 0$, the effect $\theta_{j,kl}$ is assigned to the spike component and hence shrunk to zero. Thus the corresponding level effects are fused.

Inference is accomplished by sampling from the posterior distribution using MCMC methods based on a Gibbs sampling scheme. To guarantee the restriction on the parameters $\boldsymbol{\theta}_j$ we use either the kriging algorithm described in [4] or soft restrictions based on an augmented model described in [3].

4 Application

We use data from EU-SILC in 2010 to model personal income of full-time employees in Austria. The data set provides a wide range of variables on financial and living aspects of households as well as demographic characteristics of individuals. We restrict our analysis to full-time employeesand model the logarithm of the annual income using sex, age (grouped), Austrian federal state of residence, citizenship and highest education achieved as potential regressors.

References

[1] George, I. and McCulloch, R.E. (1993). "Variable Selection Via Gibbs Sampling." *Journal of the American Statistical Association*, **88**(423), 881–889.

[2] Gertheiss, J. and Tutz, G. (2009). "Penalized regression with ordinal predictors." *International Statistical Review*, 345–365.

[3] Gertheiss, J. and Tutz, G. (2010). "Sparse Modeling of Categorial Explanatory Variables." *The Annals of Applied Statistics*, Vol.4, No.4, 2150–2180.

[4] [Rue, H. and Held, L.] (2005). *Gaussian Markov Random Fields. Theorie and Applications.* Chapman and Hall/CRC. New York.