

# SELECTIVE INFERENCE AFTER VARIABLE SELECTION BY THE RANDOMIZED GROUP LASSO METHOD

Gerda Claeskens<sup>1</sup>, Sarah Pirenne<sup>1</sup>, Snigdha Panigrahi<sup>2</sup>, and Yiling Huang<sup>2</sup>

<sup>1</sup> ORStat and Leuven Statistics Research Center, KU Leuven, (e-mail: gerda.claeskens@kuleuven.be; sarah.pirenne@kuleuven.be)

<sup>2</sup> Department of Statistics, University of Michigan, (e-mail: psnigdha@umich.edu; yilingh@umich.edu)

**ABSTRACT:** The randomized group lasso method performs a selection of groups of variables in a model and returns estimates of the coefficients in the selected model. The lasso estimator is a special case when all groups have size one. Typically, one is interested in inference for only those model coefficients that appear in the selected model and non-selected coefficients are further ignored. In selective inference one obtains valid confidence intervals and P-values for the model coefficients after selection, when conditioning on the event of the selection. We consider this problem in the framework of a general class of loss functions and distributions, including the generalized linear models, but also quasi-likelihood models that can deal with overdispersed data, for example. Our method allows the models to contain both categorical or grouped covariates as well as continuous covariates. We use an additional randomization during the group lasso estimation stage, which allows us to define a post-selection likelihood. We show that this likelihood function can be used for selective inference when conditioning on the event of the selection. An additional bonus is the selective point estimator obtained from this likelihood, which accounts by construction for the selection of the variables by the group lasso method. The confidence intervals for the regression coefficients in the selected model can be constructed in the familiar Wald-type way and we show that they have bounded lengths. We illustrate the selective inference method for grouped lasso on data from the national health and nutrition examination survey.

**KEYWORDS:** group lasso estimation, likelihood estimation, post-selection inference, selective inference