BIAS OF DIMENSION ESTIMATION USING PENALIZED LIKELIHOOD METHODS

P.Salzman[†1], A.Almudevar[1]

[1] *Department of Biostatistics and Computational Biology, University of Rochester,Rochester, NY, USA*

[†] E-mail: *psalzman@bst.rochester.edu*

It has been noted that model selection using penalized likelihood tends to select models that are larger than the true model. Zhang (1992) studied penalized likelihood method (maximize log-likelihood of a model minus λ times the dimension) for nested linear regression and established that as the number of parameters increases the AIC ($\lambda = 2$) overestimates the dimension of the model by 0.7117. In general, the dimension estimate will converge for any fixed λ. In this talk we consider all subset model selection in linear regression. We will show that in this case the average bias is at least linear in the number of possible parameters. Moreover, in the case when the regressors are orthogonal we will express the bias explicitly and propose an unbiased estimator of model dimension based on penalized likelihood.