3e cycle romand de statistique et probabilités appliquées
Villars-sur-Ollon, 28.2 – 2.3.2005

Edward I. George
Department of Statistics, University of Pennsylvania

Model Uncertainty, Selection and Combination

Advances in computing technology over the past few decades have allowed for the consideration of an increasingly wider variety of statistical models to consider for data. One wonders whether this proliferation of potential models is a blessing or a curse. A blessing because it offers a more powerful arsenal to reduce model uncertainty. But also a curse because there are simply more opportunities to be led astray. With so many models to consider, the stakes are high.

These lectures will begin by describing a number of settings where large model uncertainty problems arise naturally. These will include canonical regression settings where there is uncertainty about which regressors to include and about the form of the conditional distributions. We will then proceed to consider and compare a wide variety of selection criteria including AIC, BIC, cross validation and their many cousins that have evolved for selecting a "best" model. Distinct in their motivations, we will consider common evaluation frameworks within which their performance properties can be better understood.

We will then turn to Bayesian approaches for model uncertainty. In partcular, we will focus on hierarchical Bayes mixture priors that lead to posterior distribution characterizations of post data model uncertainty. Utility considerations will be seen to motivate model selection and model averaging within this framework. Various approaches will be described for selecting prior distributions on parameter and model spaces when the goal is to allow the posterior to concentrate on the more promising regions of the model space.

Finally, we will consider various calculation strategies for implementing the various methods in large problems. Here, we will discuss the inevitable role for heuristics, such as greedy algorithms and MCMC stochastic search, when exhaustive calculation ceases to be feasible.

References

George, E.I. (2000). The Variable Selection Problem. *Journal of the American Statistical Association* 95 1304-1307.

George, E.I and Foster, D.P. (2000). Calibration and Empirical Bayes Variable Selection. *Biometrika* 87 731-748.

Chipman, H., George, E.I. and McCulloch, R.E. (2001). The Practical Implementation of Bayesian Model Selection (with discussion). In *Model Selection* (P. Lahiri, ed.) IMS Lecture Notes -- Monograph Series, Volume 38, 65-134.

Clyde, M. and George, E.I. (2004). Model Uncertainty, *Statistical Science,* 19 1 81-94.