

STANFORD UNIVERSITY  
DEPARTMENT OF STATISTICS  
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, March 7, 2000  
Sequoia Hall Rm. 200  
(Cookies at 3:45 p.m. in 1st Floor Lounge)

*Amir Najmi*  
*Electrical Engineering*  
*Stanford University*

**Model Selection, Statistical Inference and Data Compression**

Abstract:

The problems of model selection and subsequent estimation of model parameters are often treated separately. A better-motivated approach might address these two problems within a single framework. One such framework is based on the maximum likelihood principle. Pioneered by Herbert Robbins for the compound estimation problem, the Empirical Bayes approach provides us with a different framework. This talk derives an Empirical Bayes criterion for "compound model selection", the selection of a single model for multiple data sets. Under suitable approximations, the model selection criterion can be specialized to derive a penalty on the maximum log likelihood of any model. The same approximations also yield reasonable shrinkage estimators.

The particular articulation of the model selection procedure presented in this talk bears a deep connection to Shannon's theory of lossy data compression (rate-distortion theory). It can be shown formally that in a precise statistical sense, the best model is that for which a compression scheme exists with the smallest distortion plus rate, the distortion measure being minus the log likelihood of the model density. This result implies that the theoretical apparatus and algorithms of data compression may be used to approach problems of statistical inference and model selection. This work is in keeping with Donoho's "Shannon Estimator" (presented in a previous Statistics Seminar).