

“Shrinkage” in Modern Statistical Theory and Practice

Lawrence D. Brown

Abstract and (approximate) plan of lectures

Abstract:

Nearly 50 years ago Charles Stein discovered a startling mathematical fact relating to the most common of all statistical procedures. For simultaneously estimating several means based on sampled data the intuitively appealing and universally recommended statistical estimator was to use each sample mean as the estimate of its corresponding population mean. Stein discovered a way to do better. This estimator, and the variety of modifications and generalizations of it are often referred to as statistical shrinkage procedures because of the way they act to improve on the usual procedures. In the decades since his discovery this has led to new understanding of the nature of statistical procedures, to better understanding of the properties of high dimensional spaces and of probability distributions on those spaces, and to the creation of entirely new classes of statistical procedures for a variety of problems.

This series of lectures is designed to explore the mathematical bases for shrinkage estimation and the variety of statistical contexts in which this idea plays a key role in contemporary statistical methodology. The concept of shrinkage began as a minimax solution to the ordinary problem of the estimation of several Normal means. These lectures thus begin with an exploration of this problem and then continue to study the various manifestations of this idea in various statistical directions. These include Empirical Bayes and Hierarchical Bayes procedures, Gaussian and non-Gaussian random effects and latent variables models, and nonparametric function estimation including recently developed adaptive procedures involving wavelet or other orthogonal basis systems.

The emphasis in the lectures will be on intuition and heuristic connections among the various manifestations of shrinkage, but attention will also be paid to a careful overall description of the mathematical results supporting the general theory. Some proofs will be presented in detail, but more technical ones will only be sketched or described in general outline. Several extended examples involving data applications will be included. These include treatments related to baseball batting averages, analysis of housing prices, call center data, stock market volumes and the classical regression data gathered by F. Galton.

Prerequisites: Stat 550 (Intro to mathematical statistics) and Stat 551 (Intro to theory of Linear Models) contain the required background material for these lectures.

Plan for the semester: Most of the material for the lectures is contained in preliminary form in a set of notes written over the past year. These (or up-dated versions of them) will be made available on the web as background reading. For much of the semester I will answer questions, expand on material in these notes, and supplement them with additional material, as needed. The earlier portions of the notes contain some “exercises” that may help to better focus reading and understanding of the material. Later in the semester students will make presentations drawn from the extensive literature on aspects

of the general topic. Focus of these presentations may be on theory, methodology or applications.

These lectures will cover a selection of the following topics, as time permits.

Topical Coverage of the Lectures:

Topic 1: The normal-means estimation problem

- Statistical decision theory; invariance and minimaxity
- Stein's unbiased estimate of risk and the James-Stein estimator
- Geometric and statistical intuition for the estimator
- Confidence sets (introduction)

Topic 2: Bayes and generalized Bayes procedures

- Complete class theorem
- Admissible estimators
- Harmonic and superharmonic priors
- Consequences for estimation and prediction

Topic 3: Empirical Bayes and Hierarchical Bayes procedures

- Hierarchical Bayes estimation (Strawderman + more recent developments)
- Empirical Bayes interpretations (Efron-Morris, Lindley-Smith and others)
- Penalized likelihood, and its connection to Bayesian methods
- Conflict between minimaxity and Bayesian goals

Topic 4: Gaussian Linear Models

- Canonical-form models
- Shrinkage for estimation and prediction
- Other loss functions and inferential objectives
- Ancillarity paradox

Topic 5: Random effects models

- One-way random effects model
- BLUP^s as empirical Bayes shrinkage estimators
- Higher order and Mixed Models
- Latent-variables interpretations
- Neyman-Scott paradox; Ordinary regression and BLUP shrinkage

Topic 6: Nonparametric function estimation I – Smoothing splines

- Non-parametric estimation; formulation
- Smoothing splines as penalized likelihood estimators
- Choice of the penalization parameter (cross-validation)
- Smoothing splines as empirical Bayes shrinkage estimators
- Smoothing splines as a Kalman filter estimator

The following topics are a logical sequel to the preceding material. But there will probably not be time to cover this material in any detail. (Also, some of it may be included – albeit from a different perspective – in T. Cai’s Special Topics course.)

Topic 7: Nonparametric function estimation II – Asymptotic minimax theory
Function space formulation (via Lipschitz, Sobolev and Besov spaces)
Theory for estimation at a point (Donoho-Liu)
Theory for estimation of the function (Efromovich-Pinsker & Donoho-Liu-MacGibbon)
Asymptotic ratewise-minimaxity of spline estimators

Topic 8: Nonparametric function estimation III – Adaptive estimation
Adaptive estimation, description and formulation
Shrinkage and adaptation; wavelets and other blockwise shrinkage procedures (Donoho-Johnstone, Cai-Low-Zhao)
Adaptivity through penalized likelihood (Lin-Brown)