



2007

A Unified View of Regression, Shrinkage, Empirical Bayes, Hierarchical Bayes, and Random Effects

Lawrence D. Brown
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/statistics_papers

 Part of the [Statistics and Probability Commons](#)

Recommended Citation

Brown, L. D. (2007). A Unified View of Regression, Shrinkage, Empirical Bayes, Hierarchical Bayes, and Random Effects. *Oberwolfach Reports*, 50 12-14. Retrieved from http://repository.upenn.edu/statistics_papers/171

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/statistics_papers/171
For more information, please contact repository@pobox.upenn.edu.

A Unified View of Regression, Shrinkage, Empirical Bayes, Hierarchical Bayes, and Random Effects

Abstract

A wide range of statistical problems involve estimation of means or conditional means of multidimensional normal distributions. There are many commonly employed classes of statistical models and related approaches to such problems. This talk surveys the interrelations among some of these approaches, and proposes some issues for further investigation.

Disciplines

Statistics and Probability

A unified view of regression, shrinkage, empirical bayes, hierarchical bayes, and random effects

LAWRENCE D. BROWN

A wide range of statistical problems involve estimation of means or conditional means of multidimensional normal distributions. There are many commonly employed classes of statistical models and related approaches to such problems. This talk surveys the interrelations among some of these approaches, and proposes some issues for further investigation.

The survey begins with a review of the background of shrinkage estimation. Stein (1956) surprised the statistical world with his discovery that the ordinary least squares estimator of a multivariate normal mean is not admissible in the usual setting. James and Stein (1961) then produced their classic estimator which often provides significant improvement over the ordinary estimator. 'Shrinkage' is a core feature of the estimator. An empirical Bayes interpretation of shrinkage was first proposed by Stein (1962) and Lindley (1962). The interpretation was effectively exploited by Efron and Morris (1972) and subsequently by many others. The empirical Bayes interpretation and its hierarchical fully Bayes first cousin, as first developed for this problem by Strawderman (1972), provide an important link to the manifestations of shrinkage in the various contemporary methodologies. The Bayesian viewpoint is also completely consistent with a random-effects view of the situation. These perspectives in turn allow for a shrinkage motivation of familiar ordinary linear regression.

Some analytic theory and data analyses illustrate the main points. The first of the data-based illustrations uses Galton's original data on adult heights. (See Hanley (2004) for the data.) The goal is to use heights of daughters within a family to predict the heights of the sons within that family. The second illustration sketches an analysis of US baseball batting averages, with the goal being to use each batters first half-season batting records in order to predict their second half-season performance. (See Brown (2007) for a thorough analysis of this data.) After preliminary manipulations both these examples involve estimation of means, and out-of-sample predictions, based on heteroscedastic Gaussian data. The data is moderately high dimensional (151 families and 567 batters, respectively).

It is (now) well-known that the observed sample means are themselves not desirable estimators in such contexts. For homoscedastic data shrinkage estimation ala James and Stein provides canonical frequently motivated estimators that dominate the sample means. Shrinkage is intimately related to three other approaches to estimation (and other inference) for such data, which we termed the "three siblings". These are Empirical Bayes, Hierarchical Bayes, and Random Effects. The close connection among these three and their close relation to minimax shrinkage provides increased motivation for them. However, this does not provide much basis for choosing any one version from one among them as the version of choice. Indeed, in the canonical homoscedastic setting there is little practical difference in performance among them. There are, however, significant practical differences in heteroscedastic settings.

In the homoscedastic setting ordinary regression can also be viewed as a shrinkage estimator. The view here is the converse of that in Stigler (1990) in which shrinkage is interpreted as a version of ordinary regression. The interpretation of regression as a version of shrinkage augments the understanding of (any one of) the three siblings in heteroscedastic settings, and also further motivates their use. This shrinkage idea is encapsulated in rough form in the regression paradox that dates back to Galton's original treatments of his data. In heteroscedastic settings (as in the examples treated in our presentation) the general shrinkage idea behind regression seems appropriate, but its insistence on fitting a linear estimation/prediction form is not desirable.

For heteroscedastic problems, such as those considered here, there are significant numerical differences among different implementations of the different procedures. The most pronounced difference is that between the classical proposals for minimax shrinkage (as, eg in Berger (1985, Theorem 5.20)) and the various formulas for the three siblings. This difference has been noted by many researchers. See, eg Casella (1980). Roughly, the classical minimax proposals shrink the most on the dimensions where the variance is smallest. This type of behavior contrasts with all the other proposals here which shrink the least on those dimensions, and the classical minimax procedure is neither intuitively appealing nor numerically efficient in the examples.

To remedy this, a different type of risk function is proposed as a criterion for minimaxity (and admissibility) in problems such as ours that involve estimation of several means of qualitatively exchangeable importance, a-priori. Suppose it is desired to estimate the coordinate values $\{\theta_i : i = 1, \dots, p\}$ of the vector θ . The ordinary squared-error risk function for a procedure δ , is $R(\theta, \delta) = E_\theta \left(\|\delta(X) - \theta\|^2 \right)$. We propose instead to judge a procedure by its ensemble risk. There are alternate versions of ensemble risk that can be motivated from different perspectives, and may lead to somewhat different results.

One version of this risk is

$$\overline{\overline{R}}(\gamma^2, \delta) = \int R(\theta, \delta) \phi_p(\theta; 0, \gamma^2) d\theta$$

where $\phi_p(\theta; 0, \gamma^2)$ denotes the p -dimensional normal density with iid coordinates having mean 0 and variance γ^2 . (In this version the ensemble risk is a function of only one hyper-parameter, γ^2 .)

Another version of ensemble risk can be defined as follows. Let $\theta_{(\bullet)}$ denote the p -dimensional vector whose coordinates are the increasingly ordered coordinate values of θ . Then define this version of ensemble risk as a function of the values of $\theta_{(\bullet)}$ by

$$\overline{R}(\theta_{(\bullet)}, \delta) = \frac{1}{p!} \sum_{\psi: \psi_{(\bullet)} = \theta_{(\bullet)}} R(\psi, \delta).$$

We conjecture that many of the standard shrinkage type estimators are minimax and nearly admissible for both \overline{R} and $\overline{\overline{R}}$. (An appropriately chosen hierarchical Bayes estimator should be minimax and admissible.)

In the baseball batting example it is possible to provide an interesting comparison of the out-of-sample performance of several versions of empirical Bayes, hierarchical Bayes and ordinary shrinkage estimators. It turns out that a non-parametric empirical Bayes estimator suggested in Brown and Greenshtein (2007) performs best, with the ordinary shrinkage estimator and a method-of-moments parametric empirical Bayes estimator not far behind. Other versions of empirical Bayes and hierarchical Bayes perform less well, although - as anticipated - all of the methods dominate the ordinary, naive estimator. (Other numerical investigations we have performed suggest that the explanation for the weaker performance of some of the methods may be a robustness issue related to structural features of the baseball context that are not reflected in the motivation for these methods.)

Finally, it is noted that the general perspectives here extend considerably beyond the specific data structures of the examples. These perspectives apply to a much wider variety of settings in which shrinkage is also appropriate. These settings include multiple regression, longitudinal and panel data models, spatial models (especially those appropriate for “Kriging”), penalized likelihood methods (“regularization”) involving quadratic penalty functions (especially smoothing splines), and various nonparametric regression and density estimation problems. Other settings involving varieties of shrinkage should be considered as being also related.

REFERENCES

- [1] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, NY 1985.
- [2] L.D. Brown, *In-season prediction of batting averages: a field test of empirical Bayes and Bayes methodologies*, Ann. Appl. Statist., (2007), to appear.
- [3] L.D. Brown and E. Greenshtein, *Nonparametric empirical Bayes and compound decision approaches to estimation of normal means*, manuscript (2007).
- [4] G. Casella, *Minimax ridge regression estimation*, Ann. Statist. **8** (1980), 1036–1056
- [5] B. Efron and C. Morris, *Stein’s estimation rule and its competitors-an empirical Bayes approach*, Jour. Amer. Stat. Assoc., **68** (1973), 117–130.
- [6] J.A. Hanley, *Transmuting women into men: Galton’s family data on human stature*, Amer. Statist., **58** (2004), 237–243.
- [7] W. James and C. Stein, *Estimation with quadratic loss*, Proc. Fourth Berk. Symp. Math. Statist. Prob., University of Calif. Press, **1** (1961), 311–319.
- [8] D. Lindley, *Discussion on Professor Stein’s paper*, Jour. Royal Statist. Soc, B, **24** (1962), 285–287.
- [9] C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proc. Third Berk. Symp. Math. Statist. Prob., University of Calif. Press, **1** (1956), 197–206
- [10] C. Stein, *Confidence sets for the mean of a multivariate normal distribution (with discussion)*, Jour. Royal Statist. Soc, B, **24** (1962), 265–296
- [11] S.M. Stigler, *The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators*, Statist. Science, **5** (1990), 147–155
- [12] W.E. Strawderman, *Proper Bayes estimators of the multivariate normal mean*, Ann. Math. Statist., **42** (1971), 385–388.