BAYESIAN STATISTICS 9,
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

Shrink globally, act locally: a discussion

Christian P. Robert & Julyan Arbel
CREST, Paris and Université Paris-Dauphine, CEREMADE
xian@ceremade.dauphine.fr, arbel@ensae.fr

Summary

In this discussion of Polson and Scott, we emphasize the links with the classical shrinkage literature.

It is quite pleasant to witness the links made by Polson and Scott between the current sparse modeling strategies and the more classical (or James-Stein) shrinkage literature of the 70's and 80's that was instrumental in the first author's (CPR) personal Bayesian epiphany! Nevertheless, we have some reservation about this unification process in that (a) MAP estimators do not fit a decision-theoretic framework and (b) the classical shrinkage approach is some adverse to sparsity. Indeed, as shown in Judge and Bock (1978), the so-called *pre-test* estimators that took the value zero with positive probability are inadmissible and dominated by smooth shrinkage estimators under the classical losses. While the efficiency of priors (respective to others) is not clearly defined in Polson and Scott's paper, the use of a mean sum of squared errors in Table 1 seems to indicate the authors favour the quadratic loss (Berger, 1985) at the core of the James-Stein literature. It would be of considerable interest to connect sparseness and minimaxity, if at all possible.

As detailed in, e.g., Robert (2001, Chapters 8 and 10), differential expressions linking $\mathbb{E}[\beta|y]$ and the marginal density abound in the shrinkage literature, as in e.g. Brown and Hwang (1982), Berger (1985), George (1986a,b), Bock (1988), in connection with the superharmonicity minimaxity condition (Haff and Johnstone, 1986, Berger and Robert, 1990). Connections between tail [robustness] behaviour and admissibility are introduced in Brown (1971) and developed in Hwang (1982), while boundary conditions appear in Karlin (1958) (see also Berger, 1982). In particular, Berger and Robert (1990) link the minimaxity of the Bayes estimator of a normal mean under conjugate priors, $\beta \sim \mathcal{N}(\mu, \sigma^2\Lambda)$, with the fact that the hyperprior density $\pi(\sigma^2|\mu)$ is increasing. As mentioned by the authors in Polson and Scott (2009), a related set of sufficient conditions for minimaxity (including an assumption of monotonicity on the prior density of the sampling variance σ^2) is given by Fourdrinier et al. (2008).

C.P. Robert is supported by the 2007–2010 grant ANR-07-BLAN-0237-01 "SP Bayes".

We quite agree with Polson and Scott about the dangers of using plug-in (*a.k.a.* empirical Bayes) procedures, given that the shrinkage literature has persistently shown the inefficiency and suboptimality of such procedures. We do wonder however about the connection of the double expectation formula

$$\mathbb{E}_{\tau|y}[\hat{\beta}(\tau^2)] = \mathbb{E}_{\tau|y}[\mathbb{E}_{\Lambda|\tau,y}\{\beta|y,\tau\}]$$

with the Rao–Blackwell theorem made in Section 2.4 of the paper, since this classical hierarchical decomposition of the Bayes estimator can be found for instance in Lindley and Smith (1972) as well as in Berger (1985).

Finally, Theorems 3 and 4 provide new possibilities for penalty functions based on Lévy processes, and seem to open very exciting connections with the mathematical finance literature.

REFERENCES

Berger, J. 1982. Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. In *Statistical Decision Theory and Related Topics*, eds. S. Gupta and J. Berger, vol. 3, 109–142. New York: Academic Press.

—. 1985. *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. Springer-Verlag, New York.

Berger, J. and C. Robert. 1990. Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.* 18: 617–651.

Bock, M. 1988. Shrinkage estimators: pseudo-Bayes rules for normal vectors. In *Statistical Decision Theory and Related Topics*, eds. S. Gupta and J. Berger, vol. 4, 281–297. New York: Springer-Verlag.

Brown, L. 1971. Admissible estimators, recurrent diffusions, and insoluble boundary-value problems. *Ann. Math. Statist.* 42: 855–903.

Brown, L. and J. Hwang. 1982. A unified admissibility proof. In *Statistical Decision Theory and Related Topics*, eds. S. Gupta and J. Berger, vol. 3, 205–230. New York: Academic Press.

Fourdrinier, D., O. Kortbi, and W. Strawderman. 2008. Bayes minimax estimators of the mean of a scale mixture of multivariate normal distributions. *Journal of Multivariate Analysis* 99: 74–93.

George, E. 1986a. Combining minimax shrinkage estimators. *J. American Statist. Assoc.* 81: 437–445.

—. 1986b. Minimax multiple shrinkage estimators. *Ann. Statist.* 14: 188–205.

Haff, L. and R. Johnstone. 1986. The superharmonic condition for simultaneous estimation of means in exponential families. *Canadian J. Statist.* 14: 43–54.

Hwang, J. 1982. Semi-tail upper bounds on the class of admissible estimators in discrete exponential families, with applications to Poisson and negative binomial distributions. *Ann. Statist.* 10: 1137–1147.

Judge, G. and M. Bock. 1978. *Implications of Pre-Test and Stein Rule Estimators in Econometrics*. Amsterdam: North-Holland.

Karlin, S. 1958. Admissibility for estimation with quadratic loss. *Ann. Statist.* 29: 406–436.

Lindley, D. and A. Smith. 1972. Bayes estimates for the linear model. *J. Royal Statist. Society Series B* 34: 1–41.

Polson, N. and J. Scott. 2009. Alternative global–local shrinkage priors using hypergeometric Beta mixtures. Tech. Rep. 2009-14, Duke University, Department of Statistical Science.

Robert, C. 2001. *The Bayesian Choice*. 2nd ed. Springer-Verlag, New York.