

Discussion of “Estimating Random Effects via Adjustment for Density Maximization” by C. Morris and R. Tang

Claudio Fuentes and George Casella

We congratulate Morris and Tang for an interesting addition to empirical Bayes methods, and for tackling a difficult and nagging problem in variance estimation. The ADM adjustment appears to bring on interesting properties, not just in variance estimation but also in estimation of the means. In this discussion we want to focus on the latter topic, and see how the ADM-derived estimators of a normal mean perform in a decision-theoretic way. To facilitate this we will stay with the simple model

$$(1) \quad y_i | \theta_i \sim N(\theta_i, V), \quad \theta_i \sim N(0, A).$$

1. THE JAMES–STEIN ESTIMATOR AS GENERALIZED BAYES (NOT!)

We first address the comment of Morris and Tang in Section 2.5, that the prior $A \sim \text{Unif}(0, \infty)$ is strongly suggested because the James–Stein estimator is the posterior mean if we take $A \sim \text{Unif}(-V, \infty)$. Professor Morris has noted this before, and in the interest of understanding, we want to show this calculation and comment on its relevance.

Writing $\mathbf{y} = (y_1, \dots, y_k)$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, the posterior expected loss from model (1), with the $A \sim \text{Unif}(-V, \infty)$ prior, is

$$(2) \quad \int_{-V}^{\infty} \int_{\mathbb{R}^p} |\boldsymbol{\theta} - \delta(\mathbf{y})|^2 \cdot \frac{e^{-|\mathbf{y}-\boldsymbol{\theta}|^2/(2V)}}{(2\pi V)^{k/2}} \frac{e^{-|\boldsymbol{\theta}|^2/(2A)}}{(2\pi A)^{k/2}} d\boldsymbol{\theta} dA,$$

and factoring the exponent in (2) and writing $B = V/(V + A)$ shows that

$$\boldsymbol{\theta} | \mathbf{y}, A \sim N((1 - B)\mathbf{y}, V(1 - B)),$$

Claudio Fuentes is Ph.D. candidate, Department of Statistics, University of Florida, Gainesville, Florida 32611, USA (e-mail: cfuentes@stat.ufl.edu). George Casella is Distinguished Professor, Department of Statistics, University of Florida, Gainesville, Florida 32611, USA (e-mail: casella@stat.ufl.edu).

$$A | \mathbf{y} \sim \left(\frac{1}{V + A} \right)^{k/2} e^{-(1/(2(V+A)))|\mathbf{y}|^2}.$$

The Bayes rule is the posterior mean, which we can calculate as

$$\begin{aligned} E(\boldsymbol{\theta} | \mathbf{y}) &= E[E(\boldsymbol{\theta} | \mathbf{y}, A)] \\ &= E[(1 - B)\mathbf{y} | \mathbf{y}] = [1 - E(B | \mathbf{y})]\mathbf{y}. \end{aligned}$$

We now, very carefully, calculate $E(B | \mathbf{y})$, yielding

$$\begin{aligned} E(B | \mathbf{y}) &\propto \int_{-V}^{\infty} \left(\frac{V}{V + A} \right) \left(\frac{1}{V + A} \right)^{k/2} \\ &\quad \cdot e^{-(1/(2(V+A)))|\mathbf{y}|^2} dA \\ &= V \int_{1/V}^{\infty} t^{k/2-1} e^{-t|\mathbf{y}|^2/2} dt \\ &\quad + V \int_0^{1/V} t^{k/2-1} e^{-t|\mathbf{y}|^2/2} dt, \end{aligned}$$

where we make the transformation $t = 1/(V + A)$, with the first integral coming from $A \in (-V, 0)$. Noting that the integrand is the kernel of a chi-squared density, we finally have

$$(3) \quad \begin{aligned} E(B | \mathbf{y}) &\propto \frac{V\Gamma(k/2)2^{k/2}}{(|\mathbf{y}|^2)^{k/2}} [P(\chi_k^2 \geq |\mathbf{y}|^2/V) \\ &\quad + P(\chi_k^2 \leq |\mathbf{y}|^2/V)], \end{aligned}$$

where χ_k^2 is a chi-squared random variable with k degrees of freedom. Since the chi-squared probabilities sum to 1, normalizing this expectation (dividing by $\frac{\Gamma(k/2-1)2^{k/2-1}}{(|\mathbf{y}|^2)^{k/2-1}}$) results in $E(B | \mathbf{y}) = V(k - 2)/|\mathbf{y}|^2$, yielding the James–Stein estimator. There are a number of things to note:

1. If this were a valid calculation, it would contradict such important papers as Brown (1971) and Strawderman and Cohen (1971), which provided complete characterizations of admissible generalized Bayes estimators.

2. In fact, Strawderman and Cohen [(1971), Section 4.5], explicitly tell us that the James–Stein estimator cannot be generalized Bayes.

3. In fact, the calculation leading to $E(B|\mathbf{y}) = V(k-2)/|\mathbf{y}|^2$ is invalid. To see this note that, starting from (1), with $A \sim U(-V, \infty)$, the prior on θ is

$$\int_{-V}^{\infty} \frac{e^{-|\theta|^2/(2A)}}{(2\pi A)^{k/2}} dA,$$

and, even if we take k to be even to avoid complex integration, it is straightforward to verify that the integral over $(-V, 0)$ is infinite.

What does this tell us about the James–Stein estimator? The “bad” part of the integral, which leads to the piece in (3) corresponding to $P(\chi_k^2 \geq |\mathbf{y}|^2/V)$, is to be avoided. We can informally interpret this as pointing to the region where $|\mathbf{y}|^2/V$ is small, resulting in shrinkage factors that could be greater than 1 (in absolute value), and result in the James–Stein estimator both changing the sign and expanding \mathbf{y} . When we lop off this part, we are led to estimators such as the positive-part James–Stein estimator, or admissible estimators, like those based on (32) in Morris and Tang.

2. MINIMAXITY OF ADM

The lesson from the previous section is to avoid estimators that do not control the shrinker to be between 0 and 1. So we turn to ADM and ask if it can do this. We find, interestingly, that the ADM approach will, almost automatically, give us a minimax estimator and, moreover, it controls the shrinker.

Typically, minimax estimators have been constructed using empirical Bayes arguments and a bit of customizing, or using formal Bayes derivations with priors like $A \sim \text{Unif}(0, \infty)$. The derivation of Morris and Tang in Section 2.7 is a straightforward differentiation, and we can apply the following theorem. [This is Theorem 5.5, Chapter 5, Lehmann and Casella (1998), and can be traced back to Baranchik (1970).]

THEOREM 1. *Under model (1), the estimator*

$$\delta(\mathbf{y}) = \left(1 - \frac{Vg(|\mathbf{y}|)}{|\mathbf{y}|^2}\right)\mathbf{y}$$

is minimax under the loss $|\boldsymbol{\theta} - \delta(\mathbf{y})|^2$ if

1. *the function $g(|\mathbf{y}|)$ is nondecreasing,*
2. $0 \leq g(|\mathbf{y}|) \leq 2(k-2)$.

In the notation of Morris and Tang, we are considering the case $r = 0$, and writing $T = |\mathbf{y}|^2/(2V)$, the ADM shrinkage factor is

$$(4) \quad \hat{B} = \frac{1}{T} \left[\frac{2(m-c+1)T}{T+m+1+\sqrt{(T-m-1)^2+4cT}} \right].$$

Morris and Tang note that \hat{B} is monotone decreasing in T , but for minimaxity we need the function in square brackets, which corresponds to $g(\cdot)$ of the theorem, to be nondecreasing. As $T \rightarrow \infty$, the function converges to $m-c+1$. If this is the maximum, and is less than $2(k-2)$, then the estimator will be minimax. In fact, it is straightforward (but tedious) to show that the derivative of the function in square brackets is always nonnegative, so the function is nondecreasing and the estimator is minimax. For $c = 1$ the bound can be satisfied by taking $m = (k-2)/2$, for $k \geq 3$.

Unfortunately, this is as far as we can go. The estimator based on (4), which is reminiscent of a ridge regression estimator, cannot be admissible. Again we can trace this back to Strawderman and Cohen (1971), and also Berger and Srinivasan (1978). The problem is that (being a bit informal here) admissible estimators must be analytic in the complex plane which is not the case with those based on (4).

Lastly, we wanted to see the risk performance of ADM. Morris and Tang set $m = (k-2)/2$, but there is actually a range of values of m for which the estimator is minimax. To clarify, denote $(k-2)/2 = m^*$, the Morris and Tang choice, and consider the m in (4) to be a variable. Then, for $c = 1$ the estimator is minimax for all $m \leq 2(k-2)$. In Figure 1 we see, in the left panel, the risk of five ADM estimators, along with the risk of the James–Stein estimator for comparison. There we see that the choice of m completely orders the ADM risk, with $m^* = (k-2)/2$ being the best choice, resulting in an estimator with risk similar to that of James–Stein. In the right panel we compare the ADM estimator, with $m^* = (k-2)/2$, to the James–Stein estimator, its positive-part version, and the admissible estimator with B given in (32). There we see that ADM compares favorably with the James–Stein estimator, is uniformly dominated in risk by the admissible estimator, but not by the positive-part estimator, whose risk crosses that of ADM for large $|\theta|$.

3. IS ADM AUTOMATIC?

The automatic appearance of the ADM minimax estimator gives support to the claim of Morris and Tang that “ADM maintains the spirit of MLE while making

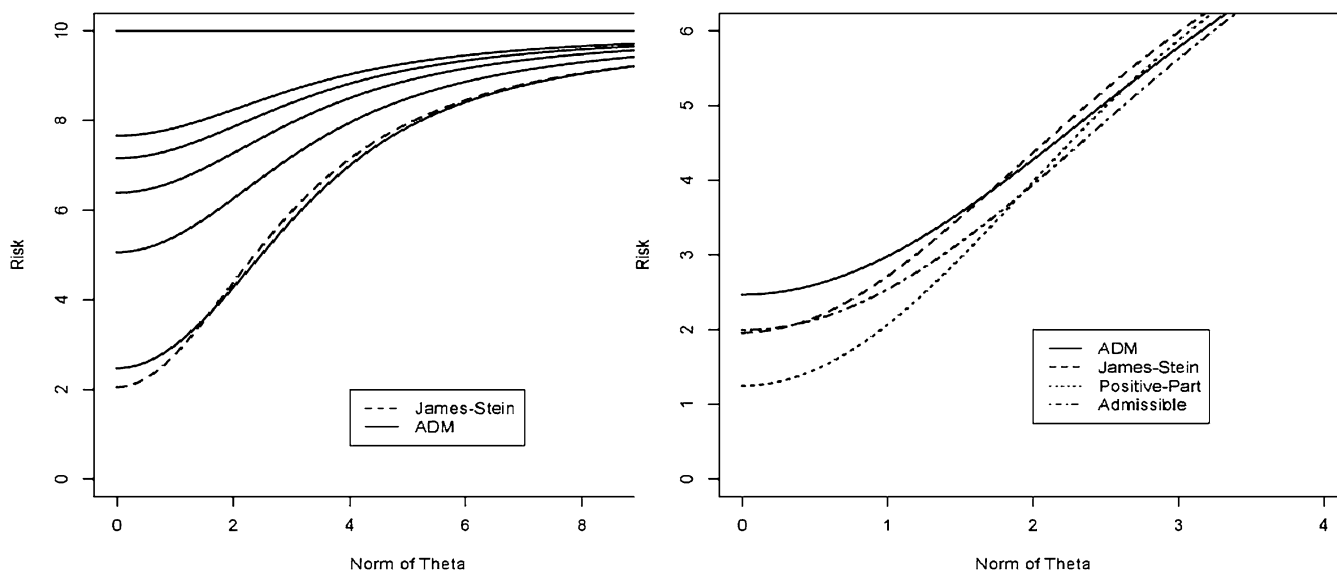


FIG. 1. For dimension $k = 10$, the left panel shows the risk of the James–Stein estimator (dashed line) and five ADM estimators (solid lines), for $m = (k - 2)/2, (k - 2)/4, (k - 2)/6, (k - 2)/8, (k - 2)/10$. The risk function increases uniformly as the denominator increases, so $m^* = (k - 2)/2$ gives the smallest risk. The right panel shows the risk of the ADM estimator with $m^* = (k - 2)/2$ (solid), the James–Stein estimator (dashed), the positive-part estimator (dotted), and the admissible estimator with B of (32) (dash-dot).

small sample improvements.” In fact, examination of the ADM shrinker B , and its risk functions, shows that “automatic” ADM produces an estimator that does not shrink as strongly as either the admissible estimator or the positive-part and, hence, can have smaller risk for larger values of the norm of θ . It is not clear to us that such small sample properties as minimaxity will continue to hold for other models, for example for the estimation of a Poisson mean, where many similar minimaxity results hold. However, the results of Morris and Tang are encouraging and certainly deserve further investigation.

ACKNOWLEDGMENT

This work is supported by NSF Grant MMS 1028329.

REFERENCES

- BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645. [MR0253461](#)
- BERGER, J. O. and SRINIVASAN, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* **6** 783–801. [MR0478426](#)
- BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903. [MR0286209](#)
- LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York. [MR1639875](#)
- STRAWDERMAN, W. E. and COHEN, A. (1971). Admissibility of estimators of the mean vector of a multivariate normal distribution with quadratic loss. *Ann. Math. Statist.* **42** 270–296. [MR0281293](#)