Journal of Statistical Research *ISSN 0256 - 422 X*
2010, Vol. 44, No. 1, pp. 129-134
Bangladesh

ADMISSIBLE ESTIMATION OF A MULTIVARIATE NORMAL MEAN

BARRY C. ARNOLD

Department of Statistics, University of California, Riverside 92506, USA
Email: barry.arnold@ucr.edu

SUMMARY

The James-Stein (1961) estimate of a multivariate normal mean has spawned a spectrum of related admissibility and inadmissibility results. Nevertheless it has not had much impact at the grass-roots level of applied statistics and continues to be almost totally ignored (after almost a half century) in introductory textbooks. Possible reasons for its low profile are discussed.

Keywords and phrases: Loss function, complexity, Bayesian, unbiased, invariant.

AMS Classification: Place Classification here. Leave as is, if there is no classification

1 Introduction

In a celebrated paper by James and Stein (1961) (see also the earlier paper by Stein (1956)), the following remarkable fact was brought to the attention of the statistical community. If $\underline{X}^{(1)}, \underline{X}^{(2)}, \ldots, \underline{X}^{(n)}$ are independent and identically distributed with common k-dimensional ($k > 2$) normal distribution with mean vector $\underline{\mu}$ and variance covariance matrix I, then the vector of sample means, $\overline{\underline{X}} = \frac{1}{n}\sum_{j=1}^{n} \underline{X}^{(j)}$, is inadmissible as an estimate of $\underline{\mu}$. This unexpected observation was greeted initially with a degree of disbelief, but was relatively quickly given a Bayesian interpretation which made it more palatable, and eventually it became accepted by many as a natural phenomenon. It became accepted by many theoreticians but perhaps it was not so readily accepted by practitioners.

Brad Efron once asked "Why isn't everyone a Bayesian?" A variety of explanations were proffered (Efron, 1986). Here we might ask, "Why doesn't everyone use some version of the James-Stein shrinkage estimate?" It indeed appears to be far from being universally accepted procedure. And it certainly hasn't filtered down to be routinely included in statistical "cook-books" which offer statistical advice to scientists in all manner of different areas of research.

This is the case despite Efron and Morris'(1973a,b) masterful selling of the product with an example dealing with batting average prediction for major league baseball players.

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

Exchangeability, in some form or other, seems to be lurking in the background of many of the arguments in favor of the James-Stein approach. A small nagging hint of doubt even applies to the baseball players. If exchangeability is even remotely approximately true, then how do we justify the enormous salary variability encountered among ballplayers. Alex Rodriguez's agent must surely be successful in his arguments against exchangeability and in favor of a gigantic difference between the salary of his client and those of more exchangeable utility players.

People seem to continue using $\overline{\underline{X}}$, despite its demonstrated inadmissibility. In general then, it appears to be reasonable to ask "Why do people persist in using inadmissible estimates?" This general question will be considered in Section 2. In Section 3, the particular case of the James-Stein estimate will be considered. Some further comments are included in Section 4.

2 $T(\underline{X})$ is inadmissible, why do you insist on using it?

Before discussing the multivariate normal example, it will be instructive to review another case in which demonstrably inadmissible estimates continue to be used and recommended. Perhaps embarrassingly, the example is found near the beginning of most elementary statistics textbooks. And Charles Stein is hovering in the background here also. We will briefly review this well known scenario.

Suppose that $X_1, X_2, \ldots, X_n$ is a sample of size n from a $Normal(\mu, \sigma^2)$ population, where both μ and σ^2 are unknown. Estimation of μ appears to be non-controversial. Pretty much any estimation strategy in use (unless Tom Bayes enters the discussion) will lead to the recommendation of $\overline{X}$ as an estimate of μ. Estimation of σ^2 is more problematic. Different estimation strategies suggest different estimates. Maximum likelihood estimation produces $T_1(\underline{X}) = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2$. Unbiasedness aficionados will "correct" this and instead will use $T_2(\underline{X}) = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$. Others consider estimates of the form $cT_1(\underline{X})$ and chose c to minimize the mean squared error of this as an estimate of σ^2. They then decide upon use of $T_3(\underline{X}) = \frac{1}{n+1}\sum_{j=1}^{n}(X_j - \overline{X})^2$. Mean squared error considerations indicate that $T_3(\underline{X})$ is preferred to $T_1(\underline{X})$ which is preferred to $T_2(\underline{X})$. Moreover, none of these three estimates are admissible under squared error loss since (Charles Stein once more, see Stein (1964)) any estimate of the form

$$T^{(a)}(\underline{X}) = \min\{\frac{1}{n+2}\sum_{j=1}^{n}(X_j - a)^2, \frac{1}{n+1}\sum_{j=1}^{n}(X_j - \overline{X})^2\}$$

for $a \in \mathcal{R}$ will be preferable. And, worse still, none of these $T^{(a)}$'s is admissible either.

In the face of all this evidence, the unbiased estimate $T_2(\underline{X})$ continues to be frequently the estimate of choice. People persist in using this inadmissible estimate. What is going on here?

There are several possible explanations and perhaps more than one of them is true or partially true. People may wish to continue using a demonstrably bad estimate because

that is the way it has always been done. They may use a bad estimate because they are stupid and/or ill-informed. But, perhaps the estimate doesn't look bad to them. Put another way, perhaps their continued use of an estimate that is inadmissible under a certain loss structure actually indicates not that they are irrational but instead indicates that the loss structure assumed in proving inadmissibility does not reflect their evaluation of the "costs" of errors from their viewpoint. As a case in point, an unbiasedness enthusiast can be viewed as one for whom the cost of using any biased estimate is infinite and, within the class of unbiased estimates, he/she will seek one with smallest variance if such exists. Minimum variance unbiased estimates are the only admissible estimates for such a person. It is not so easy to justify the use of $T_1(\underline{X})$ (the m.l.e.) instead of $T_3(\underline{X})$ (the best invariant estimate). Some loss function distinct from squared error loss is needed to justify this otherwise seemingly irrational choice. And what about those Stein-type estimates $T^{(a)}(\underline{X})$? There seem to be just too many estimates that are preferred to $T_1(\underline{X})$,$T_2(\underline{X})$ and $T_3(\underline{X})$, but are still themselves inadmissible. How should one choose among them? In despair, perhaps, people return to $T_1(\underline{X})$,$T_2(\underline{X})$ or $T_3(\underline{X})$. Perhaps their loss function involves some kind of complexity cost. Simple estimates are judged to be, in some sense, preferable to complex ones. (See Meeden and Arnold (1979) for some discussion of estimation with a loss function incorporating complexity cost in a regression context).

In the variance estimation example, we have identified two possible justifications for using an inadmissible estimate. One possibility is that the loss function used to conclude inadmissibility is not an appropriate one and does not reflect the albeit inchoate feelings about losses on the part of the user of the estimate. The second possibility is that a plethora of better estimates are available, but they suffer from two faults. First, they are often unattractively complicated and second, there is little or no available guidance on which one to select.

With this background in mind, let us turn to consider the famous multivariate inadmissibility example.

3 On not using the James-Stein estimate or its relatives.

We return to the James-Stein scenario where we have $\underline{X}^{(1)}, \underline{X}^{(2)}, \ldots, \underline{X}^{(n)}$ i.i.d. k-dimensional random variables with $\underline{X}^{(j)} \sim Normal^{(k)}(\underline{\mu}, I), \quad j = 1, 2, \ldots, n$ and we wish to estimate $\underline{\mu}$. The maximum likelihood estimate of $\underline{\mu}$ is $\underline{\hat{\mu}} = \overline{\underline{X}}$, the vector of sample means (i.e.,$\overline{\underline{X}} = \frac{1}{n}\sum_{j=1}^{n} \underline{X}^{(j)}$. James and Stein (1961) argue that this natural estimate is inadmissible when $k > 2$, being dominated by the estimate

$$\underline{\tilde{\mu}}_{JS} = \left(1 - \frac{k-2}{n\overline{\underline{X}}^T\overline{\underline{X}}}\right)\overline{\underline{X}}. \tag{3.1}$$

Based on our above observations in the normal variance example, we might expect that some users might not be impressed by this result and might continue to use $\underline{\hat{\mu}} =$

$\overline{X}$ as their estimate of $\underline{\mu}$. Their reasons might be based on an abhorrence of complexity and/or dissatisfaction (explicitly stated or vaguely sensed) with the loss function used in the inadmissibility argument. Complexity considerations seem to be of small consequence here. The estimate $\underline{\tilde{\mu}}_{JS}$ is a little more complicated but it is also a little more sophisticated in appearance and that might even be a plus in its favor. But, what about the loss function?

The mathematical result that the estimate (3.1) is to be preferred to $\underline{\hat{\mu}}$, under the assumed loss structure, is unassailable. It argues in favor of "borrowing strength" in estimating each coordinate of $\underline{\mu}$ by utilizing information from all the coordinates of the n data points. This often makes sense, as for example in the baseball setting (superstars notwithstanding). But it does not always seem reasonable. It might be interpreted as suggesting pooling unrelated experiments to "gain strength". The prospect of using data coming from a survival study of tractors in Uzbekistan to "improve" our estimates for a study of scholastic achievement in Palo Alto would seem strange and unacceptable to most.The most obvious fault associated with indiscriminate use of the James-Stein estimate is the questionable nature of the loss function, $\sum_{i=1}^{k}(\tilde{\mu}_i - \mu_i)^2$, in such situations. It treats all the μ_i's equally and inherently assumes some kind of exchangeability in the setting. It is indeed a mathematically tractable loss function. It is the natural extension of unidimensional squared error loss to accommodate higher dimensional parameter estimation. But, is it the right loss function? The paucity of adoptions of the James-Stein estimates might well suggest that, for many people in many situations, it is just not the right loss function. Surely it is not the right one in settings such as the one alluded to above involving data from Uzbekistan which is presumably of no interest to us. We would like to have good estimates for the Palo Alto parameters. We care not about parameters associated with the independent Uzbekistan data. A strong case can often be made against use of the mathematically attractive loss function $\sum_{i=1}^{k}(\tilde{\mu}_i - \mu_i)^2$.

If it is not the right loss function, what would be a better choice? Needless to say, there is no globally acceptable recommendation possible here. Loss functions, if they are to be truly reflective of the costs associated with various decisions, are necessarily subjective. They will vary from individual to individual. Since many, if not most, researchers use the maximum likelihood estimate $\underline{\hat{\mu}} = \overline{X}$ in preference to the James-Stein estimate, it is appealing to introspect on what kind of a loss function will render $\underline{\hat{\mu}}$ admissible. There are presumably many possibilities. Brown (1980) identified a class of loss functions for which $\underline{\hat{\mu}}$ is admissible. They include ones of the form

$$L(\underline{\tilde{\mu}}, \underline{\mu}) = \sum_{i=1}^{k} \frac{(1+\mu_i^2)^{\tau}}{\Sigma(1+\mu_j^2)^{\tau}}(\tilde{\mu}_i - \mu_i)^2,$$

where $\tau \geq 1/2$. Lehmann and Casella (1998, pp. 353-354) is a convenient source for more detailed discussion of loss and risk structures for which $\underline{\hat{\mu}}$ is admissible. It seems reasonable to state that none of the loss functions discussed in these sources would compel general acceptance.

At issue here is not whether we can precisely identify the loss function appropriate for

a particular user in a particular situation. What is relevant is the observation that if the user persists in using $\overline{X}$, this should not be taken as an indication of faulty reasoning on the part of the user, but instead should be recognized as an indication that he/she "hears a different drummer" than James and Stein, in that his/her loss function differs perhaps markedly from that used by James and Stein to justify the claimed inadmissibility of $\overline{X}$.

4 Is there a general message here?

Perhaps not. Perhaps the hope of identifying the class of admissible estimates for a given individual with an imperfectly articulated loss function is hopeless. But when we label a particular estimate as inadmissible, we must be very careful to explain the loss structure involved in arriving at the decision and we must be willing to permit users to reject "admissible" estimates and prefer certain "inadmissible" estimates if they feel that the corresponding loss structure is inappropriate for them. I think that even the most doctrinaire James-Stein enthusiast would admit that there are scenarios in which the $\sum_{i=1}^{k}(\tilde{\mu}_i - \mu_i)^2$, loss function is questionable and would, in such situations, not insist on discarding $\overline{X}$ as an estimate. Use of $\overline{X}$ is roughly equivalent to rejection of the James-Stein loss structure and the frequency with which this occurs and has occurred over the last half-century may be taken as an indicator that this particular loss structure is often judged to be inappropriate. It is a beautiful result, but at best, it is of limited applicability.

Acknowledgments

I am grateful to Glen Meeden and Joe Eaton for comments on a first draft of this note.

References

[1] Brown, L. D. (1980) Examples of Berger's phenomenon in the estimation of independent normal means. *Ann. Statist.*, **8**, 572–585.

[2] Efron, B. (1986) Why isn't everyone a Bayesian? With discussion and a reply by the author. *Amer. Statist.,* **40***, 1–11.*

[3] Efron, B. and C. Morris (1972) Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika,* **59***, 335–347.*

[4] Efron, B. and C. Morris (1973a) Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68***, 117–130.*

[5] Efron, B. and C. Morris (1973b) Combining possibly related estimation problems. With discussion by D. V. Lindley, J. B. Copas, James M. Dickey, M. Stone, A. P. Dawid, A. F. M. Smith, A. Birnbaum, M. S. Bartlett, G. N. Wilkinson, J. A. Nelder, C. Stein, T. Leonard, G. A. Barnard and R. L. Plackett. *J. Roy. Statist. Soc.* **B.35***, 379–421.*

[6] James, W. and C. Stein (1961) Estimation with quadratic loss. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, Univ. California Press, Berkeley, Calif., 361–379.

[7] Lehmann, E.L. and G. Casella (1998) *Theory of Point Estimation, 2nd Edn.* Springer, New York.

[8] Meeden, G, and B.C. Arnold (1979) The admissibility of a preliminary test estimator when the loss incorporates a complexity cost. *J. Amer. Statist. Assoc,* **74**, 872–874.

[9] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, University of California Press, Berkeley, Calif., 197-206.

[10] Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16**, 155-160.