Comment: How Much Can the Improvements Be Realized?

Jiunn T. Hwang

In reading this elegant and enjoyable article of Professors Maatta and Casella, I agree with most of what is depicted. My presumption, in agreement with that of the authors, is that the size of improvement of the alternative variance estimators over the standard one could be substantial when the number of the unknown means is large. However, calculations below seem to indicate the opposite.

I was also particularly interested in the authors' comments describing the parallel development of the estimation of variance problem and of the multivariate normal mean problem. I will follow the same line by comparing the relative improvements in these two problems under a linear model. It suffices to consider its canonical form

$$S^2/\sigma^2 = X_1' X_1/\sigma^2 \sim x_\nu^2$$

and

$$X_2 \sim N(\mu, \sigma^2),$$

where X_2 and hence μ is p-dimensional and is independent of S^2. (I am using the authors' notation depicted in the paragraph containing (5.1).) Both μ and σ^2 are unknown parameters.

Variance estimator. Stein's estimator for σ^2 is denoted as

$$\hat{\sigma}^2 = \phi_S(Z)S^2 = \mathrm{Min}\left(\frac{S^2}{\nu+2}, \frac{S^2+Y^2}{\nu+p+2}\right),$$

where $Y^2 = X_2' X_2$. When $p = 1$, its improvement over $\delta_1(X) = S^2/(\nu + 2)$ is small, only 4% as demonstrated by Rukhin (1987a). However, this estimator only "borrows strength" from one sample mean. If it borrowed from a large number of means, would the improvement be substantial?

Rukhin and Ananda (1989) offered an answer to this question by letting $p \to \infty$. They showed that in the most favorable situation, $\mu = 0$, the asymptotic improvement is 50%.

However, in practice, one typically confronts a p and ν which are comparable. Perhaps a more realistic asymptote would be $p \to \infty$ with $\nu/p = r$, where r is a fixed positive constant usually greater than one. Using the fact that $R(\delta_1, \sigma^2) = 2\sigma^2/(\nu + 2)$ and the standard asymptotic theory, one can establish the following theorem.

THEOREM 1. *Assume that* $p \to \infty$ *and* $\nu = rp$ *when* $r > 0$. *Suppose that* $|\mu|^2/\sqrt{p} \to \eta$. *Then*

$$\frac{R(\hat{\sigma}^2, \sigma^2)}{R(\delta_1, \sigma^2)} \to E\,\mathrm{Min}^2\left(Z_1, Z_1\frac{r}{r+1} + Z_2\frac{\sqrt{r}}{r+1} + \frac{\eta}{\sigma^2}\right),$$

where Z_1 *and* Z_2 *are iid standard normal random variables.*

James–Stein estimator. The James–Stein estimator for μ in this case is

$$\hat{\mu} = \left(1 - \frac{(p-2)S^2/(\nu+2)}{X_2' X_2}\right)X_2.$$

Conditioning on S^2 and using integration by parts (Stein, 1981) and then integrating out S^2, one can show that

$$E|\hat{\mu} - \mu|^2 = E|X_2 - \mu|^2 - E\frac{\nu\sigma^2/(\nu+2)}{|X|^2/p}.$$

Similar to Casella and Hwang (1982), one can use the identity to establish Theorem 2.

THEOREM 2. *Let* $p \to \infty$ *and* $\nu \to \infty$. *Assume also* $|\mu|^2/p \to c$. *Then*

$$\frac{R(\hat{\mu}, \mu)}{R(X, \mu)} \to \frac{c}{c + \sigma^2}.$$

Comparison of two improvements. First about the variance estimation. As shown in Rukhin and Ananda (1989), the maximum improvement occurs at $\mu = 0$. Using Theorem 1, and simulation based on 160 thousand pairs of (Z_1, Z_2), we get Table 1. These figures are substantial when r is near zero. (Theorem 1 actually does not apply to $r = 0$. However, when r is close to zero, one expects the improvement to be close to 50% due to Table 1.)

As I commented earlier, r is usually greater than one, and hence we expect the maximum improvement to be less than 25% in such a situation.

Jiunn T. Hwang is Professor, Department of Mathematics, Cornell University, Ithaca, New York 14853-7901.

TABLE 1
Maximum asymptotic percentage of improvement of $\hat{\sigma}^2$ over δ_1

	r							
	0	0.5	1	1.5	2	3	4	5
% improvement	50	33	25	21	17	13	10	7.7

Returning to the James–Stein estimator, one notes that the maximum asymptotic improvement will be 100% by Theorem 2.

However, the maximum reduction is usually not realizable in both cases as explained below. To achieve the asymptotic maximum improvement, μ must satisfy for the variance estimation case

$$\frac{|\mu|^2}{\sqrt{p}} \to 0,$$

and for the mean estimation case

$$\frac{|\mu|^2}{p} \to 0.$$

Conditions (1) and (2) are related to the distance between μ and the origin. Of course, there is nothing special about the origin. If we have prior information indicating that μ is close to $g = (g_1, \cdots, g_p)$, we should incorporate the guesses g_i's in the design of improved estimators $\hat{\sigma}_g^2$ and $\hat{\mu}_g$ (which are similar to $\hat{\sigma}^2$ and $\hat{\mu}$, respectively) by translation. The estimators $\hat{\sigma}_g$ and $\hat{\mu}_g$ will achieve the maximum improvement asymptotically if respectively conditions (1) and (2) with μ replaced by $\mu - g$ are met. These two conditions basically imply that

$$\mu_i - g_i \to 0,$$

at least when $|\mu_i - g_i|$ is monotonic. Hence the guesses g_i will be arbitrarily close to μ_i as i increases.

Such accurate prior information is, however, quite rare. Perhaps a more realistic assumption is that $\mu_i - g_i$ are iid with the mean zero and the standard deviation $\tau > 0$. (Below we will assume $g_i = 0$ without loss of generality.) Consequently, the constant η in Theorem 1 is infinite and hence the asymptotic improvement of $\hat{\sigma}^2$ is zero for the variance estimation problem.

The exact (nonasymptotic) improvement is also small in agreement with the asymptotic assertion. Based on simulating 160 thousand pairs of (S^2, Y^2),

TABLE 2
Percentage improvement of Stein's variance estimator $\hat{\sigma}^2$ over the standard estimator, when $\nu = p$ and $|\mu|^2 = p\sigma^2$

	p							
	2	3	5	10	20	30	40	50
% improvement	2.9	4.2	5.3	4.6	2.4	1.3	1.1	0.0

Table 2 reports the improvement of the variance estimator for the special case that $|\mu|^2/p = \sigma^2$.

The conclusion that the realizable improvement is small is not expected to differ for the Brewster–Zidek estimator. Nor will it be different for the confidence interval problems. In fact, the best improved confidence intervals among those numerically examined by Goutis (1989) offer only 5.3% maximum length reduction for moderate p and ν (or n) (see Maatta and Casella's Table 1). Analytic calculation of risk reduction taking into consideration the guessing error described above is difficult to do. However, a smaller reduction is expected.

For the same sequence of μ_i's (namely, iid with zero mean and standard deviation τ), the constant c in Theorem 2 is $|\mu|^2/p \to c = \tau^2$. The James–Stein estimator has improvement

$$1 - \frac{R(\hat{\mu}, \mu)}{R(X, \mu)} \to \frac{\sigma^2}{\tau^2 + \sigma^2}.$$

Hence if $\tau^2 = \sigma^2$, the realizable improvement is 50%!

In summary, the James–Stein estimator seems to offer sizable improvements that are attainable in practice whereas Stein's variance estimator does not. Given the evidence reported here, it would not be surprising if the James–Stein estimator receives much more attention in application.

ACKNOWLEDGMENT

This research was supported by NSF Grant DMS-88-09016.

ADDITIONAL REFERENCES

CASELLA, G. and HWANG, J. (1982). Limit expressions for the risk of James–Stein estimators. *Canad. J. Statist.* **10** 305–309.

RUKHIN, A. L. and ANANDA, M. M. A. (1989). Risk behavior of variance estimators in multivariate normal distributions. Technical Report, Univ. Maryland, Baltimore County.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.