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.