## Inadmissibility of the sample variance

For any observations $X_1, ..., X_n$ with $n \geq 2$, assumed to be i.i.d. with finite variance $\sigma^2$, let $s_X^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X})^2$, the usual unbiased estimator of $\sigma^2$. Y. G. Yatracos (2005) proved that $s_X^2$ is inadmissible as an estimator of $\sigma^2$ by providing an estimator which has smaller mean-squared error in all relevant cases. More precisely, he proved the following:

**Theorem 1** (Yatracos). *There is a constant $c_n$ depending on $n$, namely $c_n = \frac{(n+2)(n-1)}{n(n+1)}$, such that for any $n \geq 2$ and for all $X_1, ..., X_n$ i.i.d. with $E(X_1^4) < +\infty$ and variance $\sigma^2$ with $\sigma > 0$, the mean-square error of $c_n s_X^2$ as an estimator of $\sigma^2$ is less than that of $s_X^2$.*

*Proof.* First let's find the mean-square error as a function of an unspecified $c_n$. We can assume without loss of generality that $\mu := EX_1 = 0$ because replacing all $X_j$ by $X_j - \mu$ does not change either $\sigma^2$ or $s_X^2$. Then let $\tau := \tau_X := E(X_1^4)/\sigma^4$. We have

$$E\left((c_n s_X^2 - \sigma^2)^2\right) = c_n^2 E(s_X^4) - 2c_n\sigma^4 + \sigma^4, \tag{1}$$

and

$$E\left(s_X^4\right) := E\left(\left(s_X^2\right)^2\right) = \frac{1}{(n-1)^2}\left[n(n-1)T_{12} + nT_4\right] \tag{2}$$

where

$$T_{12} := E\left\{(X_1 - \overline{X})^2(X_2 - \overline{X})^2\right\}$$

and

$$T_4 := E\left\{(X_1 - \overline{X})^4\right\}.$$

Next,

$$\begin{aligned}
T_{12} &= E[(X_1^2 - 2X_1\overline{X} + \overline{X}^2)(X_2^2 - 2X_2\overline{X} + \overline{X}^2)] \\
&= \sigma^4 - 2E(X_1X_2^2\overline{X}) - 2E(X_2X_1^2\overline{X}) + 4E(X_1X_2\overline{X}^2) \\
&\quad -2E(X_1\overline{X}^3) - 2E(X_2\overline{X}^3) + E(X_1^2\overline{X}^2) + E(X_2^2\overline{X}^2) + E(\overline{X})^4 \\
&= \sigma^4 - 4E(X_1X_2^2\overline{X}) + 4E(X_1X_2\overline{X}^2) - 4E(X_1\overline{X}^3) + 2E(X_1^2\overline{X}^2) + E(\overline{X}^4) \\
&= \sigma^4 - 4\frac{\sigma^4}{n} + 8\frac{\sigma^4}{n^2} - 4\left[\frac{\tau\sigma^4 + 3(n-1)\sigma^4}{n^3}\right] \\
&\quad +2\left[\frac{\tau\sigma^4 + (n-1)\sigma^4}{n^2}\right] + \frac{n\tau\sigma^4 + 3n(n-1)\sigma^4}{n^4},
\end{aligned}$$

where $3n(n-1) = \binom{n}{2}\binom{4}{2}$. Thus

$$T_{12} = \sigma^4 \Bigg\{ \left[1 - \frac{4}{n} + \frac{8}{n^2} - 12\frac{(n-1)}{n^3} + \frac{2(n-1)}{n^2} + \frac{3(n-1)}{n^3}\right] + \tau\left[-\frac{4}{n^3} + \frac{2}{n^2} + \frac{1}{n^3}\right] \Bigg\}$$

$$= \sigma^4 \left\{1 - \frac{2}{n} - \frac{3}{n^2} + \frac{9}{n^3} + \tau\left[\frac{2}{n^2} - \frac{3}{n^3}\right]\right\}. \tag{3}$$

That finishes the representation of $T_{12}$. For another term of (2) we have

$$T_4 = E(X_1^4) - 4E(X_1^3\overline{X}) + 6E(X_1^2\overline{X}^2) - 4E(X_1\overline{X}^3) + E(\overline{X}^4)$$

$$= \sigma^4 \left(\tau - 4\frac{\tau}{n} + 6\left\{\frac{\tau + n - 1}{n^2}\right\} - 4\left\{\frac{\tau + 3(n-1)}{n^3}\right\} + \frac{\tau + 3(n-1)}{n^3}\right)$$

$$= \sigma^4 \left[\frac{6(n-1)}{n^2} - \frac{9(n-1)}{n^3} + \tau\left\{1 - \frac{4}{n} + \frac{6}{n^2} - \frac{4}{n^3} + \frac{1}{n^3}\right\}\right]$$

$$= \sigma^4 \left[\tau\left\{1 - \frac{4}{n} + \frac{6}{n^2} - \frac{3}{n^3}\right\} + \frac{6}{n} - \frac{15}{n^2} + \frac{9}{n^3}\right]. \tag{4}$$

Inserting the results of (3) and (4) into (2) gives

$$E\left[(c_n s_X^2 - \sigma^2)^2\right] = \sigma^4\left[c_n^2 A_n - 2c_n + 1\right]$$

where

$$A_n = \frac{n}{n-1}\left[1 - \frac{2}{n} - \frac{3}{n^2} + \frac{9}{n^3} + \tau\left(\frac{2}{n^2} - \frac{3}{n^3}\right)\right] + \frac{n}{(n-1)^2}\left[\frac{6}{n} - \frac{15}{n^2} + \frac{9}{n^3} + \tau\left(1 - \frac{4}{n} + \frac{6}{n^2} - \frac{3}{n^3}\right)\right].$$

This implies

$$E\left[(c_n s_X^2 - \sigma^2)^2\right] = \sigma^4\left[\frac{c_n^2}{(n-1)^2}(B_n + \tau C_n) - 2c_n + 1\right] \tag{5}$$

where

$$B_n = n(n-1) - 2(n-1) - \frac{3(n-1)}{n} + \frac{9(n-1)}{n^2} + 6 - \frac{15}{n} + \frac{9}{n^2}$$

and

$$C_n = \frac{2(n-1)}{n} - \frac{3(n-1)}{n^2} + n - 4 + \frac{6}{n} - \frac{3}{n^2}.$$

Then

$$B_n = n^2 - 3n + 2 - 3 + 6 + \frac{1}{n}(3 + 9 - 15) + \frac{1}{n^2}(-9 + 9)$$

$$= n^2 - 3n + 5 - \frac{3}{n}$$

and

$$C_n = n - 4 + 2 + \frac{1}{n}(6 - 2 - 3) + \frac{1}{n^2}(3 + 11) = n - 2 + \frac{1}{n} + \frac{14}{n^2}.$$

Let

$$D_n := n^2 - 3n + 5 - \frac{3}{n} + \tau\left(n - 2 + \frac{1}{n} + \frac{14}{n^2}\right).$$

Thus the mean-square error (MSE) on the left in (1) or (5) equals

$$\sigma^4\left[\frac{c_n^2}{(n-1)^2}D_n - 2c_n + 1\right].$$

So, in order to get a smaller MSE for a given $c_n < 1$ than for $c_n = 1$, we ask whether

$$\frac{c_n^2}{(n-1)^2}D_n - 2c_n + 1 < \frac{D_n}{(n-1)^2} - 1,$$

the right side being the result if $c_n$ on the left is replaced by 1. in other words whether

$$\frac{c_n^2 - 1}{(n-1)^2}D_n < 2c_n - 2,$$

or, reversing signs and dividing by $1 - c_n$, whether

$$\frac{1 + c_n}{(n-1)^2}D_n > 2.$$

We can also write

$$D_n = n^2 - 2n + 3 - \frac{2}{n} + \frac{14}{n^2} + (\tau - 1)\left(n - 2 + \frac{1}{n} + \frac{14}{n^2}\right).$$

Now $n - 2 + \frac{1}{n} + \frac{14}{n^2} > 0$ for all $n \geq 2$ because $n^3 - 2n^2 + n + 14 = n^2(n-2) + n + 14 > 0$. Thus

$$D_n \geq E_n := n^2 - 2n + 3 - \frac{2}{n} + \frac{14}{n^2}$$

since $\tau \geq 1$. Now plugging in Yatracos's value $c_n = \frac{(n+2)(n-1)}{n(n+1)}$ we're asking whether

$$\left[1 + \frac{(n+2)(n-1)}{n(n+1)}\right]E_n > 2(n-1)^2,$$

or equivalently whether

$$[n(n+1) + (n+2)(n-1)]E_n > 2(n-1)^2 n(n+1),$$

or whether

$$(2n^2 + 2n - 2)E_n > 2(n^4 - n^3 - n^2 + n),$$

or whether

$$n^4 - n^3 + n(-2+3+2) + 14 - 2 - 3 + \frac{1}{n}(14+2) - \frac{14}{n^2} > n^4 - n^3 - n^2 + n.$$

This reduces to

$$n^2 + 2n + 9 + \frac{16}{n} - \frac{14}{n^2} > 0$$

which is clearly true for all all $n \geq 2$, proving the theorem. $\square$

**Remarks**. It do not know at this writing whether with Yatracos's $c_n$ the resulting estimator is admissible.

The mean-square error of $c_n s_X^2$ depends on the distribution of $X_1$ only through the number $\tau$. For each $\tau \geq 1$, it seems possible to determine a $c_n = c_n(\tau)$ which minimizes the mean-square error.

**Notes**. I appreciate that Y. Yatracos sent me an advance copy, in 2004 or 2005, of his work. In 18.466, Fall 2005, Songzi Du wrote a term paper in which he gave an exposition of Yatracos's paper, with proofs, which helped my understanding. Both Yatracos's paper, and Songzi Du's exposition, use a method of augmented samples. Whereas, I looked for a direct proof of Yatracos's theorem without augmented samples.

## REFERENCES

Du, Songzi (2005). Mean-squared error reduction for unbiased estimators. Term paper, 18.466, Fall 2005.

Yatracos, Yannis G. (2005). Artificially augmented samples, and mean squared error reduction. *J. Amer. Statistic. Assoc.* **100**, pp. 1168–1175.