Variance of Sample Variance

Eungchun Cho* Moon Jung Cho†

Abstract
The variance of variance of finite samples taken from a finite population with replacement is expressed in terms of the sample size and the second and fourth order moments of population.

Key Words: Sample Variance, Sample with replacement, Randomization Variance, Moments, Finite Population

1. Introduction

We give a formula of the variance of with-replacement sample variance in terms of the sample size and the second and fourth moments of the population about the mean. The derivation of the formula does not require working with the more elaborate "polykay" approach of Tucky [3] [4] [5] [6]. Formula for the variance of the variance of *without-replacement* samples from a finite population given in Cho et al. [1] is quoted for comparison at the end of this paper.

2. Main Theorem

Let A be a finite set $\{a_1, \ldots, a_N\}$ and s a sample of n elements $\{x_1, \ldots, x_n\}$ taken from A *with replacement.* n is not bounded by the population size N, though often in practice $n << N$. The sample s is viewed as a realization of independent identically distributed random variables $X_1, \ldots, X_n$ on A. Following notation will be used.

$$\begin{aligned}
\overline{X} &= \frac{\sum_{i=1}^n X_i}{n}, \quad S^2 = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}{n-1}, \quad \mu = \frac{\sum_{i=1}^N a_i}{N} \\
\mu_k &= \frac{\sum_{i=1}^N (a_i - \mu)^k}{N}, \quad \mu'_k = \frac{\sum_{i=1}^N {a_i}^k}{N}
\end{aligned}$$

Theorem 1 *Let S^2 be the variance of of* with-replacement *samples of size n from a set A of real numbers $a_1, a_2, \ldots, a_N$. The variance of S^2*

$$\begin{aligned}
Var\left(S^2\right) &= \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}{\mu_2}^2\right) &(1)\\
&= \frac{1}{n}\left(\mu_4 - {\mu_2}^2\right) + \mathbf{O}\left(n^{-2}\right) &(2)
\end{aligned}$$

Proof. Let $Z_i = X_i - \mu$ for $i = 1, 2, \ldots n$ so that $E\left(Z_i\right) = 0$. Since $Var\left(S^2\right) = E\left(S^4\right) - {\mu_2}^2$, we derive an expression of $E(S^4)$ in terms of n and the moments. We can write

$$S^2 = \frac{n\sum_{i=1}^n {Z_i}^2 - \left(\sum_{i=1}^n Z_i\right)^2}{n\,(n-1)}$$

and by squaring

$$\begin{aligned}
S^4 &= \frac{n^2\left(\sum_{i=1}^n {Z_i}^2\right)^2 - 2\,n\left(\sum_{i=1}^n {Z_i}^2\right)\left(\sum_{i=1}^n Z_i\right)^2 + \left(\sum_{i=1}^n Z_i\right)^4}{n^2\,(n-1)^2} \\
E(S^4) &= \frac{n^2\,E\left(\sum_{i=1}^n {Z_i}^2\right)^2 - 2\,n\,E\left(\left(\sum_{i=1}^n {Z_i}^2\right)\left(\sum_{i=1}^n Z_i\right)^2\right) + E\left(\sum_{i=1}^n Z_i\right)^4}{n^2\,(n-1)^2}
\end{aligned}$$

*Kentucky State University, Frankfort, KY, 40601, USA, Current address: eccho@math.snu.ac.kr, Seoul National University, Seoul, Korea, Eungchun Cho's work at Seoul National University was supported by The Korea Research Foundation and The Korean Federation of Science and Technology Socities Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund).
†U.S. Bureau of Labor Statistics 2 Massachusetts Avenue, NE Washington, DC 20212 USA

Since $Z_1, \ldots, Z_n$ are independent, we have

$$\begin{aligned} E\left(Z_i\, Z_j\right) &= 0, \qquad E\left({Z_i}^3\, Z_j\right) = 0, \qquad E\left({Z_i}^2\, Z_j\, Z_k\right) = 0 \\ E\left({Z_i}^2\, {Z_j}^2\right) &= {\mu_2}^2, \qquad E\left({Z_i}^4\right) = \mu_4, \quad \text{for distinct} \quad i, j, k. \end{aligned}$$

Routine algebraic simplification with the expected values given above yields

$$E\left(\sum_{i=1}^{n} {Z_i}^2\right)^2 = n\,\mu_4 + n\,(n-1)\,{\mu_2}^2 \tag{3}$$

$$E\left(\left(\sum_{i=1}^{n} {Z_i}^2\right)\left(\sum_{i=1}^{n} Z_i\right)^2\right) = n\,\mu_4 + n\,(n-1)\,{\mu_2}^2 \tag{4}$$

$$E\left(\sum_{i=1}^{n} Z_i\right)^4 = n\,\mu_4 + 3\,n\,(n-1)\,{\mu_2}^2 \tag{5}$$

Substitution of (3), (4) and (5) into the expansion of $E\left(S^4\right)$ and simplification give

$$E\left(S^4\right) = \frac{(n-1)\,\mu_4 + \left(n^2 - 2n + 3\right){\mu_2}^2}{n\,(n-1)} \tag{6}$$

and

$$\begin{aligned} Var\left(S^2\right) &= E\left(S^4\right) - {\mu_2}^2 \\ &= \frac{(n-1)\,\mu_4 + \left(n^2 - 2n + 3\right){\mu_2}^2}{n\,(n-1)} - {\mu_2}^2 \\ &= \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}{\mu_2}^2\right) \\ &= \frac{1}{n}\left(\mu_4 - {\mu_2}^2\right) + \frac{2}{n(n-1)}{\mu_2}^2 \qquad \| \end{aligned}$$

To obtain an expression of the formula of $Var(S^2)$ in terms of μ and the moments μ_2', μ_3' and μ_4' about zero, we substitute

$$\begin{aligned} \mu_2 &= \mu_2' - \mu^2 \\ \mu_4 &= \mu_4' - 4\mu\mu_3' + 6\mu^2\mu_2' - 3\mu^4 \end{aligned}$$

into (1) and get

$$\begin{aligned} Var\left(S^2\right) = \; & \frac{1}{n}\,\mu_4' - \frac{4}{n}\,\mu\mu_3' - \frac{n-3}{n\,(n-1)}\,{\mu_2'}^2 + \\ & + \frac{4\,(2\,n-3)}{n\,(n-1)}\,\mu^2\mu_2' - \frac{2\,(2\,n-3)}{n\,(n-1)}\,\mu^4 \end{aligned} \tag{7}$$

3. Comparison with Without-replacement Samples

Here we compare (1) with the variance of variance of *without-replacement* samples given in [1]. Let $Var_{wo}\left(S^2\right)$ denote the variance of variance of *without-replacement* samples of size n from A. The following is a simplified (improved) version from [1].

$$Var_{wo}\left(S^2\right) = c_1\,\mu_4 + c_3\,{\mu_2}^2 \tag{8}$$

where

$$\begin{aligned} c_1 &= \frac{N\,(N-n)\,(Nn - N - n - 1)}{n\,(n-1)\,(N-3)\,(N-2)\,(N-1)} \\ c_3 &= -\frac{N\,(N-n)\left(N^2 n - 3\,n - 3\,N^2 + 6\,N - 3\right)}{n\,(n-1)\,(N-1)^2\,(N-2)\,(N-3)} \end{aligned} \tag{9}$$

We note

$$\begin{aligned} \lim_{N\to\infty} Var_{wo}\left(S^2\right) &= \frac{1}{n}\,\mu_4 - \frac{n-3}{n\,(n-1)}\,{\mu_2}^2 \\ &= Var\left(S^2\right) \end{aligned}$$

as expected. The difference of $Var_{wo}\left(S^2\right)$ and $Var\left(S^2\right)$ is of order $1/N$, that is, $|Var_{wo}\left(S^2\right)-Var\left(S^2\right)|$ is $\mathbf{O}(N^{-1})$. In most practical situations where $n = cN^{\alpha}$ for some $c > 0$ and $0 < \alpha < 1$, $|Var_{wo}\left(S^2\right) - Var\left(S^2\right)|$ is $\mathbf{O}(n^{-\frac{1}{\alpha}})$. For example, if $n = \sqrt{N}$, then the difference of $Var_{wo}\left(S^2\right)$ and $Var\left(S^2\right)$ is $\mathbf{O}(n^{-2})$. As we did for $Var(S^2)$, we represent $Var_{wo}\left(S^2\right)$ in terms of the moments μ_2' and μ_4' about zero by substitution of (7) into (8).

$$Var_{wo}\left(S^2\right) = c_1\,\mu_4' + c_2\,\mu\mu_3' + c_3\,{\mu_2'}^2 + c_4,\mu^2\mu_2' \; + c_5\,\mu^4 \tag{10}$$

where c_1 and c_3 are as before (9) and

$$\begin{aligned} c_2 &= -4\,\frac{N\,(N-n)\,(Nn-N-n-1)}{n\,(n-1)\,(N-3)\,(N-2)\,(N-1)} \\ c_4 &= 4\,\frac{N^2\,(N-n)\,(2\,Nn-3\,N-3\,n+3)}{n\,(n-1)\,(N-1)^2\,(N-2)\,(N-3)} \\ c_5 &= -2\,\frac{N^2\,(N-n)\,(2\,Nn-3\,N-3\,n+3)}{n\,(n-1)\,(N-1)^2\,(N-2)\,(N-3)} \end{aligned}$$

Here again, each c_i converges to the corresponding coefficient in (7).

$$\begin{aligned} \lim_{N\to\infty} c_2 &= -\frac{4}{n} \qquad \lim_{N\to\infty} c_4 = \frac{4(2n-3)}{n(n-1)} \\ \lim_{N\to\infty} c_5 &= -\frac{2n-3}{n(n-1)} \end{aligned}$$

4. Acknowledgment

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics. The authors thank John Eltinge for many helpful suggestions that improved the paper.

References

[1] Eungchun Cho, Moon Jung Cho and John Eltinge, The variance of sample variance from a finite population, Proceedings of Joint American Statistical Association and International Statistical Institute Conference, Toronto, Canada, August, 2004

[2] W. G. Cochran, *Sampling Techniques (3rd ed.)*, John Wiley, 1977.

[3] J. W. Tukey, Some sampling simplified, *Journal of the American Statistical Association*, 45 (1950), 501-519.

[4] J. W. Tukey, Variances of variance components: I. Balanced designs. *The Annals of Mathematical Statistics*, 27 (1956), 722-736.

[5] J. W. Tukey, Variances of variance components: II. Unbalanced single classifications. *The Annals of Mathematical Statistics*, 28 (1957), 43-56.

[6] J. W. Tukey, Variance components: III. The third moment in a balanced single classification. *The Annals of Mathematical Statistics*, 28 (1957), 378-384.