Remark On Variance Bounds

R. Sharma and R. Bhandari
Department of Mathematics & Statistics
Himachal Pradesh University
Shimla -5,
India - 171005
email: rajesh_hpu_math@yahoo.co.in

Abstract. It is shown that the formula for the variance of combined series yields surprisingly simple proofs of some well known variance bounds.

AMS classification 60E15

Key words and phrases : Mean, Variance, Samuelson's inequality.

1 Introduction

It is well known that if we have two sets of data X and Y containing n_1 and n_2 observations with means $\overline{X}$ and $\overline{Y}$, and variances $S_{n_1}^2$ and $S_{n_2}^2$, respectively, then the combined variance of $n_1 + n_2$ observations is given by

$$S_{n_1+n_2}^2 = \frac{n_1}{n_1+n_2} S_{n_1}^2 + \frac{n_2}{n_1+n_2} S_{n_2}^2 + \frac{n_1 n_2}{(n_1+n_2)^2} \left(\overline{X} - \overline{Y}\right)^2 . \tag{1.1}$$

Let $X = \{x_j\}$ be a sample of size one and let Y be the sample of size $n-1$ drawn from the population $\{x_1, x_2, ..., x_n\}$ such that $X \cap Y = \phi$. Then $n_1 = 1, n_2 = n-1, \overline{X} = x_j, \overline{Y} = \frac{1}{n} \sum\limits_{i \neq j} x_j, S_1^2 = 0$, and it follows from (1.1) that

$$S_n^2 = \frac{n-1}{n} S_{n-1}^2 + \frac{1}{n-1} \left(x_j - \overline{x}\right)^2 , \tag{1.2}$$

where $\overline{x} = \frac{1}{n} \sum\limits_{i=1}^{n} x_i$ and $S_n^2 = \frac{1}{n} \sum\limits_{i=1}^{n} \left(x_i - \overline{x}\right)^2$are respectively the mean and variance of the data $\{x_1, x_2, ..., x_n\}$.
Each summand in (1.2) is non-negative, so

$$S_n^2 \geq \frac{1}{n-1} \left(x_j - \overline{x}\right)^2 , \tag{1.3}$$

for all $j = 1, 2, ..., n$.
The inequality (1.3) is known as Samuelson's inequality (1968) in statistical literature. The inequality (1.3) was also established in mathematical literature by Laguerre (1880) in some different context and notations. Several alternative proofs of this inequality were given in literature, see Arnold and Balakrishna (1989), and Rassias and Srivastava (1999).
It may be noted here that the identity (1.2) also implies that

$$S_n^2 \geq \frac{n-1}{n} S_{n-1}^2 .$$

Thus, if S_m^2 is the variance of a sample of size m drawn from a population of size n, then

$$S_n^2 \geq \frac{m}{n} S_m^2 .$$

Let $X = \{x_j, x_k\}$ be a sample of size 2 and let Y be the sample of size $n-2$ drawn from the population $\{x_1, x_2, ..., x_n\}$ such that $X \cap Y = \phi$. Then $n_1 = 2$, $n_2 = n-2$, $\overline{X} = \frac{x_j + x_k}{2}$,

$\overline{Y} = \frac{1}{n-2} \sum\limits_{i \neq j,k} x_i$, $S_2^2 = \frac{(x_j - x_k)^2}{4}$, and it follows from (1.1) that

$$S_n^2 = \frac{n-2}{2} S_{n-2}^2 + \frac{1}{2n} (x_j - x_k)^2 + \frac{2}{n-2} \left(\overline{x} - \frac{x_j + x_k}{2} \right)^2, \tag{1.4}$$

for all $j = 1, 2, ..., n$ and $n \geq 3$.

Each summand in (1.4) is non-negative, so

$$S_n^2 \geq \frac{1}{2n} (x_j - x_k)^2, \tag{1.5}$$

for all $j = 1, 2, ..., n$. From (1.5), for $m \leq x_i \leq M, i = 1, 2, ..., n$, we have

$$S_n^2 \geq \frac{1}{2n} (M - m)^2. \tag{1.6}$$

The inequality (1.6) is due to Nagy (1918). See also Nair (1948) and Thompson (1935). Likewise, from (1.4), we have for $n \geq 3$,

$$S_n^2 \geq \frac{1}{2n} (M - m)^2 + \frac{2}{n-2} \left(\overline{x} - \frac{m + M}{2} \right)^2. \tag{1.7}$$

The inequality (1.7) provides a refinement of (1.6), see Sharma et al. (2008).

Mallows and Richter (1969) proves an extension of the Samuelson inequality (1.3). This says that if γ_r is the mean of any subset of r numbers chosen from the set $\{x_1, x_2, ..., x_n\}$, then

$$S_n^2 \geq \frac{r}{n-r} (\gamma_r - \overline{x})^2, \tag{1.8}$$

for $r = 1, 2, ..., n-1$.

From (1.1), we have

$$S_{n_1+n_2}^2 \geq \frac{n_1 n_2}{(n_1 + n_2)^2} \left(\overline{X} - \overline{Y} \right)^2. \tag{1.9}$$

Let X be a sample of size r and let Y be the sample of size $n-r$ drawn from the population $\{x_1, x_2, ..., x_n\}$ such that $X \cap Y = \phi$. Then $n_1 = r$, $n_2 = n-r$, $\overline{X} - \overline{Y} = \frac{n}{n-r} (\gamma_r - \overline{x})$,and so (1.8) follows from (1.9).

Likewise, we can deduce Boyd-Hawkins inequalities (1971) from (1.9). This says that if $x_1 \leq x_2 \leq ... \leq x_n$, then

$$\overline{x} - \sqrt{\frac{n-k}{k}} S_n \leq x_k \leq \overline{x} + \sqrt{\frac{k-1}{n-k+1}} S_n$$

for $k = 2.3, ..., n-1$.

The variance bounds have various extensions and applications in statistics, polynomials and matrix theory. We see that formula (1.1) provides further insight, and is very useful in the study of these inequalities. In this way we can study various further refinements, generalisations and extensions of the variance bounds.

References

[1] Arnold, B.C., Balakrishnan, N., Bounds and Approximations for Order Statistics, Lecture Notes in Statistics, 53, Springer-Verlag, New York, (1989).

[2] Boyd, A.V., Bounds for order statistics, Publikacije Elektrotehnickog Fakulteta Univerziteta U Beogradu, Seriya Matematika I Fizika (Belgrade), 365, 31-32, (1971).

[3] Hawkins, D.M., On the bounds of the range of order statistics, J. Amer. Statist. Assoc., 66, 644-645, (1971).

[4] Laguerre, E N, Surune methode pour Obtenir par approximation les racines d'une equation algebrique qui a toutes ses raciness reelles [in French], Nouv. Ann. de Math., 19, 161-171 & 193-202, (1880).

[5] Mallows, C.L., Richter, D., Inequalities of Chebyshev type involving conditional expectations, The Annals of Mathematical Statistics, 40, 1922-1932, (1969).

[6] Nagy, J. V. S., Uber algebraische Gleichungen mit lauter reelen Wurzeln [in German], Jahresbericht der Deutschen Mathematiker - Vereinigung, 27, 37-43, (1918).

[7] Nair, K.R., The distribution of the extreme deviate from the sample mean and its studentized form, Biometrika, 35, 118-144, (1948).

[8] Rassias, T.M., Srivastava, H.M., Analytic and Geometric Inequalities and Applications, Kluwer Academic Publishers, Dordrecht, (1999).

[9] Samuelson, P.A., How deviant can you be?, J. Amer. Statist. Assoc., 63, 1522-1525, (1968).

[10] Sharma, R., Shandil, R.G., Devi, S., Ram, S., Kapoor, G., Barnett, N.S., Some bounds on the sample Variance in terms of the Mean & Extreme values, Advances in Inequalities from Probability Theory & Statistics, Edited by N.S. Barnett & Dragomir, 187-193, (2008).

[11] Thompson, W.R., On a criteria for the rejection of the observations and the distribution of the ratio of deviation to sample standard deviation, Ann. Math. Satist., 6, 214-219, (1935).