CU-52-61-Nonr-266(33)MS

A NOTE ON THE EXACT VARIANCE OF PRODUCTS[1]

by

Leo A. Goodman

Columbia University

Mathematical Statistics

April 27, 1961

This research was sponsored in part by the Office of Naval Research under Contract Number Nonr-266(33), Project Number NR 042-034. Reproduction in whole or part is permitted for any purpose of the United States Government.

[1]This note was written while the author was a Visiting Professor of Mathematical Statistics and Sociology at Columbia University on leave of absence from the University of Chicago. For their very helpful comments, the author is indebted to Gerald J. Glasser, Joseph E. Keilin, and Hilary L. Seal.

A number of readers of [2] have written the author inquiring about the possibility of generalizing the results presented there. It therefore seemed worthwhile to prepare the present brief note indicating how some of the results in [2] can be generalized.

Let $x_1, x_2, \ldots, x_K$ be K random variables. Let us denote the expected value of x_i by $E(x_i) = X_i$, the variance of x_i by V_i , and the square of the coefficient of variation of x_i by $V_i/X_i = G_i$. (For the sake of simplicity, we assume that $X_i \neq 0$, although some of the results presented do not require this assumption.) We shall make use of the simple identity

$$(1) \qquad \prod_{i=1}^{K} x_i = \prod_{i=1}^{K} X_i \prod_{i=1}^{K} (\delta_i + 1) = \prod_{i=1}^{K} (\Delta_i + X_i) ,$$

where $\delta_i = (x_i - X_i)/X_i$ and $\Delta_i = (x_i - X_i)$. If the x_i are mutually independent, we find using identity (1) that the variance of $\prod_{i=1}^{K} x_i$ will be equal to

$$(2) \qquad V\left(\prod_{i=1}^{K} x_i\right) = E\left\{\prod_{i=1}^{K} x_i^2\right\} - \prod_{i=1}^{K} X_i^2 = \prod_{i=1}^{K} X_i^2 \left[\prod_{i=1}^{K} (G_i + 1) - 1\right] ,$$

which can also be written as

$$V\left(\prod_{i=1}^{K} x_i\right) = \prod_{i=1}^{K} (V_i + X_i^2) - \prod_{i=1}^{K} X_i^2$$

$$(3) \qquad = \sum_i V_i \prod_{j \neq i} X_j^2 + \sum_{i_1, i_2} V_{i_1} V_{i_2} \prod_{j \neq i_1, i_2} X_j^2 + \sum_{i_1, i_2, i_3} V_{i_1} V_{i_2} V_{i_3} \prod_{j \neq i_1, i_2, i_3} X_j^2 + \ldots + V_1 V_2 \ldots V_K$$

$$= \prod_{i=1}^{K} X_i^2 \left[\sum_i G_i + \sum_{i_1, i_2} G_{i_1} G_{i_2} + \sum_{i_1, i_2, i_3} G_{i_1} G_{i_2} G_{i_3} + \ldots + G_1 G_2 \ldots G_K\right]$$

where the summation, $\sum_{i_1, i_2, \ldots, i_S}$, is over all values of

$i_1 \neq i_2 \neq i_3 \ldots \neq i_S$ ranging over $1, 2, \ldots, K$, and where

$\prod_{j \neq i_1, i_2, \ldots, i_S}$ is the product over the K-S values different

from the S values $i_1, i_2, \ldots, i_S$. Equation (3) here is a generalization of equations (2) and (15) in [2] and equation (a) in [7]; equation (2) here appeared earlier in [3] where it was used to study the case where the distribution of $\prod_{i=1}^{K} x_i$ was (approximately) logarithmic-normal.

We now present an unbiased estimator of $V(\prod_{i=1}^{K} x_i)$ based on unbiased estimators, $\bar{x}_i$ and v_i , of X_i and V_i , respectively, where $\bar{x}_i$ is the sample mean and v_i is the sample variance in a sample of n_i observations each having mean X_i and variance $V_i (i=1,2,\ldots,K)$. When the K samples $(i=1,2,\ldots,K)$ are mutually independent, we find that

$$(4)\quad v(\prod_{i=1}^{K} x_i) = \prod_{i=1}^{K} (v_i + z_i) - \prod_{i=1}^{K} z_i$$

$$= \prod_{i=1}^{K} \left[\bar{x}_i^2 + v_i (n_i - 1)/n_i\right] - \prod_{i=1}^{K} \left[\bar{x}_i^2 - v_i/n_i\right]$$

is an unbiased estimator of $V(\prod_{i=1}^{K} x_i)$, where $z_i = \bar{x}_i^2 - v_i/n_i$.

This follows from the fact that $E(\bar{x}_i^2)-X_i^2 = V_i/n_i$. Equation (4) here is a generalization of equation (5) in [2] .

The case where the x_i are not mutually independent is more complicated. From identity (1) we see that the variance of $\prod_{i=1}^{K} x_i$ is

$$(5) \qquad V\left(\prod_{i=1}^{K} x_i\right) = \prod_{i=1}^{K} X_i^2 \left[E\left\{\prod_{i=1}^{K} (\delta_i + 1)^2\right\} - B^2\right]$$

$$= E\left\{\prod_{i=1}^{K} (\Delta_i + X_i)^2\right\} - M^2 ,$$

where $M = E\left\{\prod_{i=1}^{K} x_i\right\}$ and $B=M/\prod_{i=1}^{K} X_i$. The special case of (5) where $K = 2$ was studied in [2] . We now consider the case where $K = 3$. By straightforward calculation, we find that, when $K = 3$, equation (5) can be rewritten as

$$(6) \quad V\left(\prod_{i=1}^{3} x_i\right)=\prod_{i=1}^{3} X_i^2\left[\sum_{i=1}^{3} G_i+(B-1)(3-B)+\sum_{j,k,\ell} E\left\{\delta_1^j \delta_2^k \delta_3^\ell\right\} h(j,k,\ell)\right] ,$$

where the indices j, k, ℓ range over the values 0,1,2, and where $h(j,k,\ell)$ is a symmetric function of j,k,ℓ having the following values:

$$h(j,k,\ell) = \begin{cases} 0 \text{ for } (j,k,\ell) = (0,0,0),(0,0,1),(0,1,1) \\ 1 \text{ for } (j,k,\ell) = (0,2,2),(2,2,2) \\ 2 \text{ for } (j,k,\ell) = (0,1,2),(1,2,2) \\ 4 \text{ for } (j,k,\ell) = (1,1,1),(1,1,2) \end{cases} .$$

Equation (6) here is a generalization of equation (18) in [2], the formula given there for the variance of the product of two random variables (not necessarily independent). In the same way that equation (18) was used in [2] to derive other variance formulas for various product estimators (e.g., equations (20) and (21) in [2]), equation (6) here can also be used to derive other variance formulas for product estimators where, for example, three estimators (rather than two) are multiplied together. We shall not go into these details in this brief note.

References

[1] Barnett, H.A.R., "The variance of the product of two independent variables and its application to an investigation based on sample data," Journal of the Institute of Actuaries, 81 (1955), p. 190.

[2] Goodman, Leo A., "On the exact variance of products," Journal of the American Statistical Association, 55(1960), pp. 708-713.

[3] Shellard, G.D., "Estimating the product of several random variables," Journal of the American Statistical Association, 47(1952), pp. 216-221.