I. SĂCUIU

Prof. Dr. R. THEODORESCU

The Variance of a Random Variables Product and the Efficiency of Certain Estimations

A criterion of choice taking into account the speed with which the estimation is obtained can obviate the drawbacks—possible in certain circumstances—in using the variance when appreciating the quality of an estimator.

The formula of the variance of a random variables product, known in case the variables are assumed to be independent, is still in force if the random variables are assumed to be uncorrelated in a certain way.

The economic applications occur when we have to do with estimators in which random variables products arise, and the eventual use of the formula in a certain non-corelation hypothesis enlarges considerably the field of applicability. The number of measurements and the number of elementary operations required for a certain estimator is often essential in practical applications, and this may be useful for the choice of the estimator from a given set of estimators.

Consider a probability space $(\Omega, K, P,)$ and two independent random variables ξ, η, defined on this space. It is known that

$$D^2(\xi\eta) = E^2(\xi)D^2(\eta) + D^2(\xi)E^2(\eta) + D^2(\xi)D^2(\eta)$$

This relation is true under far more general conditions. Thus, we have:

Lemma 1. If ξ and η are two non-correlated random variables, both themselves and their squares, then

$$D^2(\xi\eta) = E^2(\xi)\,D^2(\eta) + D^2(\xi)\,E^2(\eta) + D^2(\xi)\,D^2(\eta)$$

Proof. Since the random variables ξ, η are non-correlated, we can write

$$E(\xi\eta) = E(\xi)\,E(\eta)$$

and, since ξ^2, η^2 are non-correlated, it also follows that

$$E(\xi^2\eta^2) = E(\xi^2)\,E(\eta)^2$$

However,

$$E(\xi^2) = D^2(\xi) + E^2(\xi)$$
$$E(\eta^2) = D^2(\eta) + E^2(\eta)$$

and then

$$E(\xi^2)\,E(\eta^2) = (D^2(\xi) + E^2(\xi))\,(D^2(\eta) + E^2(\eta)) =$$
$$= D^2(\xi)D^2(\eta) + E^2(\xi)D^2(\eta) + D^2(\xi)E^2(\eta) + E^2(\xi)E^2(\eta)$$

According to the hypothesis made above we have, however,

$$E^2(\xi)\,E^2(\eta) = (E(\xi)\,E(\eta))^2 = E^2(\xi\eta)$$

and we can write in this case the equality

$$E(\xi^2\eta^2) - E^2(\xi\Phi) = D^2(\xi\eta) = E^2(\xi)\,D^2(\eta) + E^2(\eta)\,D^2(\xi) +$$
$$+ D^2(\xi)\,D^2(\eta)$$

This equality can be generalized as follows:

Lemma 2. If ξ_i, $1 \leqslant i \leqslant n$ are random variables for which the couples $\xi_1^k\xi_2^k \ldots \xi_{j-1}^k$ and ξ_j^k, $2 \leqslant j \leqslant n$, $k = 1, 2$ are non-correlated, then

$$D^2\left(\prod_{i=1}^{n}\xi_i\right) = \sum_{i=1}^{n} E^2(\xi_1)\ldots E^2(\xi_{i-1})\,D^2(\xi_i)\,E^2(\xi_{i+1})\ldots E^2(\xi_n) +$$

$$+ \sum_{\substack{i_1,\,i_2=1\\ i_1<i_2}}^{n} E^2(\xi_1)\ldots E^2(\xi_{i_1-1})\,D^2(\xi_{i_1})E^2(\xi_{i_1+1})\ldots E^2(\xi_{i_2-1})\,D^2(\xi_{i_2})E^2(\xi_{i_2+1})\ldots$$

$$\ldots E^2(\xi_n) + \ldots + \prod_{i=1}^{n} D^2(\xi_i)$$

Proof. Taking into account the hypothesis made and Lemma 1, we can write the succession of qualities

$$D^2\left(\prod_{i=1}^{n}\xi_i\right) = D^2\left(\prod_{i=1}^{n-1}\xi_i\cdot\xi_n\right) = E^2\left(\prod_{i=1}^{n-1}\xi_i\right)D^2(\xi_n) + D^2\left(\prod_{i=1}^{n-1}\xi_i\right)E^2(\xi_n) +$$

$$+ D^2\left(\prod_{i=1}^{n-1}\xi_i\right)D^2(\xi_n) = E^2\left(\prod_{i=1}^{n-2}\xi_i\right)E^2(\xi_{n-1})D^2(\xi_n) + \left[E^2\left(\prod_{i=1}^{n-2}\xi_i\right)D^2(\xi_{n-1}) +\right.$$

$$\left. + D^2\left(\prod_{i-1}^{n-2}\xi_i\right)E^2(\xi_{n-1}) + D^2\left(\prod_{i=1}^{n-2}\xi_i\right)D^2(\xi_{n-1})\right]\,E^2(\xi_n) +$$

$$+\left[E^2\left(\prod_{i=1}^{n-2}\xi_i\right)D^2(\xi_{n-1})+D^2\left(\prod_{i=1}^{n-2}\xi_i\right)E^2(\xi_{n-1})+D^2\left(\prod_{i=1}^{n-2}\xi_i\right)D^2(\xi_{n-1})\right]D^2(\xi_n)=$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$=\sum_{i=1}^{n}E^2(\xi_1)\ldots E^2(\xi_{i-1})\,D^2(\xi_i)\,E^2(\xi_{i-1})\ldots E^2(\xi_n)+$$

$$+\sum_{\substack{i_1,i_2=1\\ i_1<i_2}}^{n}E^2(\xi_1)\ldots E^2(\xi_{i_1-1})\,D^2(\xi_{i_1})\,E^2(\xi_{i_1+1})\ldots E^2(\xi_{i_2-1})\,D^2(\xi_{i_2})\,E^2(\xi_{i_2+1})\ldots E^2(\xi_n)+$$

$$+\ldots+\prod_{i=1}^{n}D^2(\xi)$$

2. Let us assume now that the random variables ξ_i. $1\leqslant i\leqslant n$ satisfy the conditions of Lemma 2 and let us write $m_i=E(\xi_i)$ $1\leqslant i\leqslant n$. Then, according to Chebyshev's inequality, it follows that we can write

$$P\left(\left|\prod_{i=1}^{n}\xi_i-\prod_{i=1}^{n}m_i\right|\geqslant\varepsilon\right)<\frac{D^2\left(\prod_{i=1}^{n}\xi_i\right)}{\varepsilon^2}$$

Assuming, for instance, the random variables ξ_i, $1\leqslant i\leqslant n$ to be symmetrical and $\max\limits_{1\leqslant i\leqslant n}D^2(\xi_i)<k<1$, $n\in N^*$ it follows that

$$\prod_{i=1}^{n}\xi_i\xrightarrow{P}0$$

Consider the independent random variables $\xi_j\in N(m_j;\ \sigma_j)$ $1\leqslant j\leqslant k$ and a sample ξ_{j_1}, ξ_{j_2}, ..., ξ_{jn_j} of the size n_j drawn from the population, characterized by the random variable ξ_j $1\leqslant j\leqslant k$.

If, on the basis of the above samples, we make up the sample means

$$\overline{X}_j=\frac{1}{n_j}\sum_{l=1}^{n_j}\xi_{jl},\qquad 1\leqslant j\leqslant k$$

it is clear that these sample means are independent random variables and that $\overline{X}_j\in N\left(m_j;\ \frac{\sigma_j}{\sqrt{n_j}}\right)$; $1\leqslant j\leqslant k$. From the same Chebyshev's inequality

$$P\left(\left|\prod_{j=1}^{k}\overline{X}_j-\prod_{j=1}^{k}m_j\right|\geqslant\varepsilon\right)<\frac{D^2\left(\prod_{j=1}^{k}\overline{X}_j\right)}{\varepsilon^2}$$

Assuming now $\inf_{1\leqslant j\leqslant k} n_j \to \infty$ and, taking into account relation

$$D^2\left(\prod_{j=1}^{k} \overline{X}_j\right) = \sum_{j=1}^{k} m_1^2 \dots m_{j-1}^2 \frac{\sigma_j^2}{n_j} m_{j+1}^2 \dots m_k^2 +$$

$$+ \sum_{\substack{j_1, j_2 = 1 \\ j_1 < j_2}}^{n} m_1^2 \dots m_{j_1-1}^2 \frac{\sigma_{j_1}^2}{n_{j_1}} m_{j_1+1}^2 \dots m_{j_2}^2 \frac{\sigma_{j_2}^2}{n_{j_2}^2} m_{j+1}^2 \dots m_{j_2}^2 + \dots + \prod_{j=1}^{k} \frac{\sigma_j^2}{n_j}$$

it follows that

$$\prod_{j=1}^{k} \overline{X}_j \xrightarrow{P} \prod_{j=1}^{k} m_j$$

3. In a previous article [5] the following example was considered: We assume to have in view the estimation of area A of a rectangle whose real sizes are ξ and η.

(The example considered is chosen so that it can facilitate a best understanding of the possibility of application. The applications proper to economy occur whenever products of measurements arise. This occurs when applying, for instance, the geometric mean.)

By means of a sample (ξ_i, η_i), $1 \leqslant i \leqslant n$ of the size n drawn from the population characterized by vector (ξ, η) we obtain the empirical areas

$$A_1^* = \frac{1}{n} \sum_{i=1}^{n} A_i = \frac{1}{n} \sum_{i=1}^{n} \xi_i \eta_i$$

and

$$A_2^* = \overline{\xi}\,\overline{\eta} = \frac{1}{n^2} \left(\sum_{i=1}^{n} \xi_i\right) \left(\sum_{i=1}^{n} \eta_i\right)$$

It is obvious that A_1^* and A_2^* are unbiased and consistent estimators for area A and, besides,

$$D^2(A_2^*) < D^2(A_1^*)$$

Considering the estimation of the volume of a parallelepipedon whose real sizes are ξ, η, ζ, then, by means of a sample (ξ_i, η_i, ζ_i), $1 \leqslant i \leqslant n$, drawn from the population characterized by vector (ξ, η, ζ), we can find the empirical volumes

$$V_1^* = \frac{1}{n} \sum_{i=1}^{n} V_i = \frac{1}{n} \sum_{i=1}^{n} \xi_i \eta_i \zeta_i$$

$$V_2^* = A_2^* \overline{\zeta} = \frac{1}{n^2} \left(\sum_{i=1}^{n} \xi_i \eta_i\right) \left(\sum_{i=1}^{n} \zeta_i\right)$$

$$V_3^* = \overline{\xi}\,\overline{\eta}\,\overline{\zeta} = \frac{1}{n^3} \left(\sum_{i=1}^{n} \xi_i\right) \left(\sum_{i=1}^{n} \eta_i\right) \left(\sum_{i=1}^{n} \zeta_i\right)$$

(It goes without saying that V_2^* can be obtained by carrying out also the other two possible combinations, but they will be equivalent from the standpoint of our reasoning.)

Obviously, the same as in the case of the area estimation, the empirical volumes V_1^*, V_2^*, V_3^* are unbiased and consistent estimators of volume V and, besides,

$$D^2(V_3^*) < D^2(V_2^*) < D^2(V_1^*)$$

The inequalities obtained lead us to assert that, from the precision standpoint, it is advisable to use the least variance estimator (A_2^* with the area and V_3^* with the volume).

The problem of the estimator choice may be set also from a different point of view, namely, to choose the estimator requiring the shortest computation time on an electronic computer. To this end we shall reason in agreement with the estimators of an area and of a volume.

When estimating an area, the number of operations required for the computation of the estimator is:

$2n+1$ operations, out of which n additions, m multiplications and one division for estimator A_1^*;

$2n+3$ operations, out of which $2n$ additions, two multiplications and one division for estimator A_2^*.

REMARK 1. Estimator A^*_2 can be written under the form

$$A_2^* = \left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\xi_i\right)\left(\sum_{i=1}^{n}\eta_i\right)$$

Hence, it follows that the number of operations to be carried out $2n + 3$ may be considered as made up either of $2n$ additions, one multiplication and two divisions, or of $2n$ additions, two multiplications and one division, as considered before.

The choice made is justified both by the fact that, in general, the time required for carrying out a multiplication is shorter than the time required for carrying out a division, and that the division may incur greater computation errors.

For the estimators of a volume, the number of operations to be performed is:

$3n + 1$ operations, out of which n additions, $2n$ multiplications and one division for estimator V_1^*;

$3n + 3$ operations, out of which $2n$ additions, $n + 2$ multiplications and one division for estimator V_2^*;

$3n + 5$ operations, out of which $3n$ additions, four multiplications and one division for estimator V_3^*;

REMARK 2. For estimators V_2^* and V_3^* we maintain the assertion made in Remark 1 as to the possibility of a choice between the operations of multiplication and division.

The time required for carrying out an operation depends on the nature of the operation and on the computer.

If we denote by a, b, c,* the time required for carrying out an addition a multiplication or a division, respectively, then

$$a < b < c$$

Let us denote by t_λ the time required for computing the λ estimator. We can see immediately that

$$t_\lambda = \varphi_\lambda (a,\ b,\ c;\ n)$$

where $\varphi_\lambda(a,\ b,\ c;\ n)$ is an increasing function with respect to each of the variables a, b, c, n.

If we denote by Λ the set of estimatos of the value considered, then we shall choose the estimator $\lambda_0 \in \Lambda$ so that

$$\varphi_{\lambda_0}(a,\ b,\ c;\ n) = \inf_{\lambda \in \Lambda} \varphi_\lambda(a,\ b,\ c;\ n)$$

for a, b, c, and n defined.

Of course, between a, b, c there may often exist various relations, which simplifies to an important extent the problem considered. The additional relations between the parameters a, b, c may depend both on the type of computer and the computation program.

Coming back to the estimation of an area A and of a volume V, the set s, in these cases, respectively

$$\Lambda = \{A_1^*,\ A_2^*\} \text{ and } \Lambda = \{V_1^*,\ V_{,2}^*\ V_3^*\}$$

In such cases we have

$$t_{A_1^*} = \varphi_{A_1^*}(a,\ b,\ c;\ n) = n\ (a + b) + c$$

$$t_{A_2^*} = \varphi_{A_2^*}(a,\ b,\ c;\ n) = 2\ n\ a + 2\ b + c$$

and

$$t_{V_1^*} = \varphi_{V_1^*}(a,\ b,\ c;\ n) = n(a + 2\ b) + c$$

$$t_{V_2^*} = \varphi_{V_2^*}(a,\ b,\ c;\ n) = n(2a + b)\ + 2b + c$$

$$t_{V_3^*} = \varphi_{V_3^*}(a,\ b,\ c;\ n) = 3na + 4b + c$$

Making up the difference

$$t_{A_1^*} - t_{A_2} = n(b - a) - 2b$$

it follows that it is advantageous to use the estimator A_1^* if $t_{A_1^*} - t_{A_2^*} \leqslant 0$ that is if

$$n \leqslant \frac{2b}{b - a}$$

and it is advantageous to use the estimator A_2^* if

$$t_{A_1^*} - t_{A_2^*} > 0$$

---

* Only these three operations occur in our reasoning here

that is if

$$n > \frac{2b}{b-a}$$

In a similar manner we obtain

$$t_{V_1^*} - t_{V_2^*} = n(b-a) - 2b$$

and

$$t_{V_2^*} - t_{V_3^*} = n(b-a) - 2b$$

So, if $n \leqslant \frac{2b}{b-a}$ it is preferable to use the estimator V_1^*, while if $n > \frac{2b}{b-a}$ it is preferable to use the estimator V_3^*.

Certainly, here we have considered φ_λ as a function of a, b, c and n. However, if other operations besides the mentioned ones occur when computing the estimator $\lambda \in \Lambda$, then φ_λ will depend on more variables.

As already pointed out, in order to choose the estimator out of a given set, either the dispersions of the estimators $D^2(\lambda)$, $\lambda \in \Lambda$ or the functions $t_\lambda = \varphi_\lambda\ (a, b, c; n)\ \lambda \in \Lambda$ have been considered separately.

It is more reasonable, yet, to use both $D^2(\lambda)$ and $\varphi_\lambda(a, b, c; n)$ simultaneously.

In order to take this into account, we bring in the product $D^2(\lambda)\, t_\lambda = \psi_\lambda(a, b, c; n)\ \lambda \in \Lambda$ and choose the estimator $\lambda_0 \in \Lambda$ for which

$$\psi_{\lambda_0}(a, b, c; n) = \inf_{\lambda \in \Lambda} \psi_\lambda(a, b, c; n)$$

and, consequently, all the reasonings made so far hold true with some obvious modifications.

REFERENCES

1. CREANGĂ-SÂMBOAN A. : *Asupra dispersiei produsului de variabile aleatoare independente* "Studii şi cercetări matematice", Vol 18, No.9, 1966.
2. GOODMAN, A.L.: *On the exact variance of products.* "Journal of the American Statistical Association", Vol. 55, No. 292, Dec. 1960.
3. IOSIFESCU, M. , MIHOC, G. and THEODORESCU R. : *Teoria probabilităţilor şi statistica matematică.* Editura Tehnică, Bucureşti, 1966.
4. ONICESCU, O. and MIHOC, G.: *Lecţii de statistică matematică.* Editura Tehnică, Bucureşti, 1958.
5. SĂCUIU I.: *Dispersia unui produs oarecare de variabile aleatoare independente.* "Studii şi cercetări matematice", Vol. 21, No. 6, 1969.

LA VARIANCE D'UN PRODUIT DE VARIABLES ALEATOIRES ET L'EFFICACITE DE CERTAINES ESTIMATIONS

On démontre une formule génerale pour la variance d'un produit de variables aléatoires qui sont non-corrélées d'une certaine façon.

Elle est appliquée ensuite, pour prouver la convergence en probabilité d'un produit de variables aléatoires.

Enfin, on étudie l'efficacité calculatrice et on donne comme illustration l'estimation d'une aire et d'un volume.

ДИСПЕРСИЯ ПРОИЗВЕДЕНИЯ СЛУЧАЙНЫХ ПЕРЕМЕННЫХ И ЭФФЕКТИВНОСТЬ НЕКОТОРЫХ ОЦЕНОК

Демонстрируется общая формула дисперсии произведения случайных переменных, о которых предполагается, что они не независимы, но некоррелированны определенным образом.

Эта формула затем применяется для демонстрации конвергенции вероятности произведения со случайными переменными.

Исследуется также эффективность расчета и в качестве примера приводится оценка поверхности и объема.