The influence of ratios and combined ratios on the distribution of the product of two independent Gaussian random variables

Amílcar Oliveira[1,4], Teresa A. Oliveira[2], and José A. Seijas-Macias[3]
[1] Universidade Aberta and CEAUL, Portugal
[2] Universidade Aberta and CEAUL, Portugal
[3] University of A Corunha and UNED, Spain
[4] Corresponding author: Amílcar Oliveira, e-mail: aoliveira@uab.pt

Abstract

The distribution of the product of Gaussian random variables has been of interest to various authors in different research application areas. Among other results, it is known that the distribution of product of Gaussian random variables is good for small values of variation coefficients. By using simulation techniques, in this work our aim is to study which ratios have more influence on the presence of normality for the product of two independent Gaussian variables and to quantify this influence. We will consider the variation coefficient value, the individual ratios (means divided by standard deviation) and the combined ratio (product of the two means divided by variance) of two Gaussian variables considering the homocedastic and the heterocedastic cases.

Keywords: variation coefficient, simulation, normality, approximation methods

1. Introduction

This work is focused to study distribution of the product of two Gaussian uncorrelated variables. We are considering two normally distributed variables. The distribution of the product of two normal variables is not, in general, a normally distributed variable. However, under some conditions, is showed that the distribution of the product can be approximated by means of a Normal distribution. Craig (1936) was the first author to determine the algebraic form of the moment-generating function of the product, but he could not determine the distribution of the product. A conclusion standing out that the distribution of the product xy is a function of the coefficient of correlation of both variables and of two parameters that are proportional to the inverse of the coefficient of variation (δ) of each variable. Aroian (1947) advanced in the investigations of Craig and proved that when the inverse of the coefficients of variation are big, the function of density of $z = xy$ approximates to a Normal curve and, under certain conditions, the product approaches the standardized Pearson type III distribution.

These works are relatively old, but there are not at all well known among mathematicians. Until 2003, when the introduction of computer and numerical and symbolic calculus were extend, there were not new advances in this problem. In 2003, Ware and Lad published an article where the problem of the probability of the product of two normally distributed variables was approached.

Our work begins with a review of previous research in this area, and we found a group of theoretical and practical results of interest for solving this problem. Next section considers the degree of influence for the value of the coefficient of variation inverse, on the normality character of the product distribution and the combined ratio.

2. Theoretical Fundamentals

Let (X,Y) be a bivariate normal distribution with independent variables and parameters: $\mu_x, \mu_y, \sigma_x, \sigma_y$ for mean and standard deviation and where the covariance is null. Let $Z = XY$ denote the product of the two variables. Then it is possible to estimate the values of moments of the distribution Z, using the moment-generating function from Craig (1936) and Aroian et al. (1978).

$$M_Z(t) = \frac{\exp\left[\dfrac{t\mu_x\mu_y + \frac{1}{2}\left(\mu_y^2\sigma_x^2 + \mu_x^2\sigma_y^2\right)t^2}{1 - t^2\sigma_x^2\sigma_y^2}\right]}{\sqrt{1 - t^2\sigma_x^2\sigma_y^2}} \qquad (1)$$

If we define the variables: $\delta_x = \dfrac{\mu_x}{\sigma_x}$ and $\delta_y = \dfrac{\mu_y}{\sigma_y}$ then (1) could be rewritten as:

$$M_Z(t) = \frac{\exp\left[\dfrac{t\mu_x\mu_y + \left(t\delta_y^2\mu_x\mu_y + \delta_x^2(2\delta_y^2 + t\mu_x\mu_y)\right)}{2\delta_x^2\delta_y^2 - 2t^2\mu_x^2\mu_y^2}\right]}{\sqrt{1 - \dfrac{t^2\mu_x^2\mu_y^2}{\delta_x^2\delta_y^2}}} \qquad (2)$$

The associated moments of the product of variables (mean, variance and skewness) will be:

$$\begin{aligned} E[Z] &= \mu_x\mu_y, \\ Var[Z] &= \mu_y^2\sigma_x^2 + \left(\mu_x^2 + \sigma_x^2\right)\sigma_y^2 = (1 + \delta_x^2 + \delta_y^2)\sigma_x^2\sigma_y^2, \qquad (3) \\ \alpha_3[Z] &= \frac{6\delta_x\delta_y\sigma_x^3\sigma_y^3}{\left(\left(1 + \delta_x^2 + \delta_y^2\right)\sigma_x^2\sigma_y^2\right)^{3/2}} \end{aligned}$$

To calculate the effect of the inverse of variation coefficient δ over the normality of the product of two normally distributed variables, we have investigated the consequences of different values, basically, through the value of parameters (mean, variance and skewness) and we have observed the shape of distributions. Numerical integration methods were used on the calculation of the values of the product of normal variables.

Several examples were studied for the product of two Gaussian independent variables. We have considered three approaches in order to estimate the moments of the product and the shape of distribution:

a) Numerical Integration: We consider the method in Ware and Lad (2003) for estimation of the partial density function of the product of two variables. Solution to that integral requires using a numerical integration method. We have used Newton-Cotes 8 panel method, Chapra and Canale (2010). The number of points considered is variables with at least 6000 points and at more 100.000 points. Then, we calculate the mean, variance and skewness of the distribution.

b) Monte-Carlo Simulation: We generate a random sample of a million of elements using a Monte-Carlo method, for the two variables and consider the distribution of the product of these elements. We estimate directly the moments of distribution using the sample.

c) Moment Approach: We estimate the moment for the product using the moment-generating function (3) for the parameters of the distributions considered as part of the product.

We consider the normality of distribution as a result of the skewness of the product distribution. If skewness=0, the data are perfectly symmetrical. But for a skewness of not exactly zero, Bulmer (1979) suggests this rule of thumb:

- If skewness is less than -1 o greater than +1, the distribution is highly skewed.
- If skewness is between -1 and -1/2 or between +1/2 and +1, the distribution is moderately skewed.
- If skewness is between -1/2 and +1/2, the distribution is approximately symmetric.

This interpretation is valid when one has data for the whole population, that is, only for the moment approach case in our study. But when one has just a sample (Monte-Carlo simulation or numerical integration approaches), the sample skewness doesn't necessarily apply to the whole population. In that case, there is the need to divide the sample skewness α_3 by the standard error of skewness (SES) to get the test statistic, which measures how many standard errors separate the sample skewness from zero, Cramer (1997):

$$Z_{\alpha_3} = \frac{\alpha_3}{\sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}} \qquad (4)$$

The critical value is approximately 2:

- If $Z_{\alpha_3} < 2$, the population is very likely skewed negatively.
- If Z_{α_3} is between -2 and +2, no conclusion about skewness could be reached.
- If $Z_{\alpha_3} > 2$, the population is very likely skewed positively.

Value of standard error of skewness for Monte-Carlo Simulation, with a sample-size of 1000000 of points, is:

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \approx \sqrt{\frac{6}{n}} \approx 0.00245$$

Value of standard error of skewness for numerical integration, with a maximum sample-size of 100000 of points, is:

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \approx \sqrt{\frac{6}{n}} \approx 0.00774$$

When we use the criteria of SES in order to estimate the significance of skewness into a sample, we cannot deduce the presence or absence of normality of the total distribution. Then, considerations about normality of the product are based into the rule of thumb from Bulmer (1979).

3. Results.

Results for different analyzed cases:

a) Homocedastecity: $\sigma_x = \sigma_y = 1$.

Parameters	Type	Mean	Variance	Skewness
$\mu_x = 1, \mu_y = 100$	Numerical Integration Monte-Carlo Simulation Moment Approach	99.9781 100.057 100	9996.54 10006. 10002	-0.001383 -0.003683 0.000599
$\mu_x = 1, \mu_y = 50$	Numerical Integration Monte-Carlo Simulation Moment Approach	49.9949 49.9492 50	2501.98 2502.92 2502	0.002355 0.000326 0.002397
$\mu_x = 1, \mu_y = 5$	Numerical Integration Monte-Carlo Simulation Moment Approach	4.99342 4.99903 5	26.965 26.9908 27	0.208079 0.212197 0.213833

Normal approach is a good approximation for the two first cases, but when the value of mean of one variable is decreasing then skewness is increasing. In the third case, the coefficient is $\delta_x = 1, \delta_y = 5$, and following the test of skewness we could assume that distribution is positive skewed, although the normality could be assumed in the three cases, because the problem is only for sample simulations.

b) Same Mean: $\mu_x = \mu_y = 1$.

Parameters	Type	Mean	Variance	Skewness
$\sigma_x = 1, \sigma_y = 100$	Numerical Integration Monte-Carlo Simulation Moment Approach	0.75971 1.07526 1	13862.5 19923.6 20001	0.0064765 0.0299193 0.0212116
$\sigma_x = 1, \sigma_y = 50$	Numerical Integration Monte-Carlo Simulation Moment Approach	0.928407 1.11453 1	4506.74 5013.74 5001	0.026129 0.039337 0.0424137
$\sigma_x = 1, \sigma_y = 5$	Numerical Integration Monte-Carlo Simulation Moment Approach	0.991085 0.996119 1	50.6099 50.9828 51	0.38637 0.410971 0.411847

Normal approach is a good approximation for the two first cases, although numerical integral values are quite different, when value of parameter is high, caused by approximation techniques used. In this situation, we observe that the coefficient δ has a very low value for one of the variables: $\left\{\frac{1}{100} = 0.01, \frac{1}{50} = 0.02, \frac{1}{5} = 0.2\right\}$, but a relatively large value for variable x (that is, 1). Then, the effect of a large value for δ_x produces a normally distributed product. As the value of δ_y is increasing, skewness of the product distribution is decreasing to zero. Normality could be assumed for the three cases considered.

c) Different values for one variable: $\mu_x = \sigma_x = 1$.

Parameters	Type	Mean	Variance	Skewness
$\mu_y = 100, \sigma_y = 100$	Numerical Integration Monte-Carlo Simulation Moment Approach	67.7064 99.7819 100	15623.9 29936.4 30000	0.323073 1.14848 1.1547
$\mu_y = 10, \sigma_y = 10$	Numerical Integration Monte-Carlo Simulation Moment Approach	9.99492 10.0393 10	299.986 301.569 300	1.15425 1.15759 1.1547
$\mu_y = 0.1, \sigma_y = 0.1$	Numerical Integration Monte-Carlo Simulation Moment Approach	0.991085 0.996119 1	50.6099 50.9828 51	0.38637 0.41097 0.41184

For the three cases studied, variables present the same values for the inverse of variation coefficient $\delta = 1$. But variable x has the same value for mean and standard deviation in the three cases, and variable y is different for each case, in this situation, we have three values high (100), medium (10) and low (0.1). Only in the last case we can consider the presence of normality.

The presence of a mean and standard deviation high-valued produces an increment of skewness and the product distribution doesn't tend to normality.

We consider the influence over the Normal approximation for the product of two independent non-correlated Normal variables of the combined ratio. Let X and Y be two normal variables with mean μ_x and μ_y, respectively; and with the same variance $\sigma_x = \sigma_y = \sigma$. We define the combined ratio as:

$$\frac{\mu_x \mu_y}{\sigma^2} \qquad (5)$$

This ratio is used to study the joint influence of the two inverse coefficients of variation, that is, same magnitude for both variables or influence is independent for each one.

We have considered four cases: first one, combined ratio is 100, a very high value, and two δ coefficients are 1 and 100. The second one is a combined ratio of 1000 (very high value) with δ coefficients are 1 and 100 (like the first case). The third case is a low combined ratio of 0.1 with δ coefficients: 1 and 0.1. And the last one is a combined ratio of 10000 with δ coefficients: 1 and 1000.

a) Combined Ratio effect: $\mu_x = 1$.

Parameters	Type	Mean	Variance	Skewness
$\mu_y = 100, \sigma = 1$ combined ratio=100	Numerical Integration Monte-Carlo Simulation Moment Approach	99.9781 100.07 100	9996.54 9997.42 10002	-0.001383 -0.000896 0.000599
$\mu_y = 10, \sigma = 0.1$ combined ratio=100	Numerical Integration Monte-Carlo Simulation Moment Approach	9.995 10.00 10	1.0101 1.0084 1.0101	0.00591 0.00703 0.00591
$\mu_y = 0.1, \sigma = 1$ combined ratio=0.1	Numerical Integration Monte-Carlo Simulation Moment Approach	0.0949 0.0998 0.1	2.01 2.0121 2.01	0.21056 0.20437 0.21055
$\mu_y = 100, \sigma = 0.1$ combined ratio=10000	Numerical Integration Monte-Carlo Simulation Moment Approach	99.995 100.00 100	100.012 100.014 100.01	-0.00025 0.00142 0.000059

When combined ratio value is high, then the product presents a small skewness and small skewness is a guarantee for normality. On the other side, a small combined ratio value is associated to skewness higher for the product. In the four cases considered skewness is low, although normality could be assumed for all the four cases considered.

4. Considerations and conclusions

The approximation of the distribution of the two Gaussian variables is an ancient problem whose first resolutions trace back to the first-half of the 20th century. Previous works of determinate authors have showed that it is possible to try estimating the function of density of the product by means of diverse methods. In this work we have centered our approximation using numerical integration by means of Newton-Cotes, Monte-Carlo Simulation and Moment-generating function approaches, to calculate the parameters (mean, variance and skewness) of the product of two variables, in order to consider if it is normally distributed. We center our consideration into the value of skewness and considering the significance of it.

Our conclusions show that the product of two normally distributed variables is normally distributed when large a value in the inverse of the coefficients of variation for one of the variables is presented. Evolution of skewness is increasing for diminishing values of mean or standard deviation, when two variables have the same standard deviation or the same mean, respectively. When we considered two different mean and standard deviation for the two variables (one variable with mean and standard deviation unity), low skewness is associated to low values of the parameters of the other variable.

The presence of normality can be accepted for values of the coefficient of variation inverse of one of the variables being in the order of the unit, but not when the two variables presents high values for the inverse of the coefficient of variation, simultaneously; however, when the combined ratio is high and skewness is low, then the normality can be assumed.

Acknowledgements: Research partially supported by National Funds through FCT-Fundação para a Ciência e a Tecnologia, project PEst-OE /MAT /UI0006 /2011 (CEAUL)

References

Aroian, L.A. (1947) "The probability function of a product of two normal distributed variables", *Annals of Mathematical Statistics*, 18. 256-271

Aroian, L.A.; Taneja, V.S. y L.W. Cornwell (1978) "Mathematical forms of the distribution of the product of two normal variables", Communications in Statistics – Theorical Methods, A7 (2). 165-172.

Bulmer, M. G. (1979) *Principles of Statistics*. Dover. New York.

Chapra, S.C. and R.P. Canale (2010). *Numerical Methods for Engineers*, McGraw-Hill, 6 edition.

Craig, Cecil C. (1936) "On the Frequency Function of xy", *Annals of Mathematical Statistics*, 7. 1-15.

Cramer, Duncan (1997) *Basic Statistics for Social Research*. Routledge. London.

Ware, R. and F. Lad (2003) *Approximating the Distribution for Sums of Product of Normal Variables*, Research-Paper 2003-15. New Zealand: Department of Mathematics and Statistics, University of Canterbury, New Zealand.