THE ANNALS OF "DUNĂREA DE JOS" UNIVERSITY OF GALAŢI
FASCICLE V, TECHOLOLOGIES IN MACHINE BUILDING,
ISSN 1221-4566, 2013

STUDY OF ANALYTICAL FUNCTION OF CORRELATION FOR RANDOM PARAMETER PROCESS

Nicusor NISTOR[1], Constantin GHEORGHIEŞ[1]

[1]Physics, Chemistry and Environment Department, Faculty of Science and Environment, Dunarea de Jos University of Galati, Domneasca no. 111, cod 800201, Galati, Romania,
nistor.nicusor@ymail.ro

ABSTRACT

In this article, we are presenting one modeling method to obtain analytic functions, corresponding to the statistical correlation of parameters of an industrial process, where some parameters of the process are influenced by permanent variables and the entire process is also influenced by random events. The proposed method of this study, involves finding the polynomial functions for real distribution of probability, for the individual set of variables, according to the histogram of relative frequency. Those functions have the property that generated surfaces of the function from definition domain to be unitary - this will be the normalization condition. By eliminating the common parameter of functions, we obtain a correspondence relationship, between that functions, which we will call analytical correlation function of the two random variables. This method is a numerical solution of the problem given and applied successfully in statistical analysis of multi-parametric industrial processes, where he proved fully effective.

KEYWORDS: Mathematical model; statistical data; correlation index; modeling method; random value of parameters

1. INTRODUCTION

Physical parameters of one process are influenced by permanent variables (which are predictable), but in reality each real process involves variation of the parameters due to the random events. Experimental data obtained from a physical process are recorded for a period of time and these sets of data are - the starting point of our analysis.

The proposed method of this study, involves finding a polynomial functions for real distribution of probability, according with the histogram of relative frequency [2]. By eliminating the common parameter of that functions, we can obtain a correspondence relationship between the functions, which we call analytical correlation function of the two random variables (in the uniqueness of the studied phenomenon).

2. MODEL CONCEPT

In this work we address the problem of finding a correlation function between two string of random variables of a process, where physical phenomenon are influenced by external factors random, unpredictable.

In the study of these phenomena it is important to know the influence of the statistical event, expressed by a mathematical relationship. In such cases, using statistical analysis, we choose an index of correlation [3] between (-1...0...1), and we have to find a relationship which it is linking the statistical parameters involved in the process.

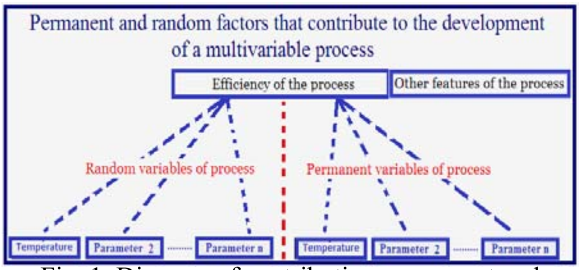

Fig. 1. Diagram of contributing permanent and random parameters to the development of a multivariable process.

It’s known that each industrial process, where efficiency of the process can be calculated by dedicated method is also influenced by random phenomena (Fig. 1).

In order to implement the calculation method, we can take for example two sets of statistical variables, the first one is the temperature of a random process determined by external factors and the second one is the efficiency of a process, that depends largely from the known parameters, but also it depends from the external random parameters.

The statistical data are structured into two matrices like in the form:

$$x = \begin{pmatrix} 29 & 30 & 31 & 32 & 33 & 34 \\ 0.05 & 0.25 & 0.31 & 0.21 & 0.15 & 0.03 \end{pmatrix}$$

$$y = \begin{pmatrix} 19 & 20 & 21 & 22 & 23 & 24 \\ 0.02 & 0.13 & 0.22 & 0.41 & 0.18 & 0.04 \end{pmatrix}$$

where (x) parameter is the temperature of the process, and (y) is the efficiency of the process.

The statistical average [3] of the random temperature parameter will be calculate with next formula:

$$x_m = \sum_{i=1}^{n} x_i p_{x_i} = \sum_{i=1}^{6} x_i p_{x_i} = 31.25\ {}^{o}C\,.$$

The statistichal average [3] of the efficiency parameter is calculate with next formula:

$$y_m = \sum_{i=1}^{n} y_i p_{x_i} = \sum_{i=1}^{6} y_i p_{y_i} = 21.78\ \%\,;$$

$$Cov(x,y) = \sum_{i=1}^{6} [(x_i - x_m)(y_i - y_m) p_{x_i} p_{y_i}]$$

The correlation [3] of these two strings of statistical data is calculated with formula:

$$Correl(x,y) = \sum_{i=1}^{6} \frac{Cov(x,y)}{\sigma_x \sigma_y} = 0.62334$$

In this way we can calculate the correlation of the two parameters, finding the index: 0.62334, which means: these two statistical parameters are direct correlated, but not strong.

However, in some cases, we are interested in finding analytic functions, linking the parameters of a random process, where it can not be writen a mathematical relationship, based on logical concepts.

Based on the property that for each random parameter, we can find the relative frequency polygon, as shown in Figure (2a, 2b), we intend to try to find the continuous distribution of the probability.

2. EXPERIMENTS AND RESULTS

We start to search the polynomial curve of distribution of probability, based on the relative frequency polygon, which can be drawn through logical reasoning – the probability of the natural phenomena do not occur in nature in steps, but with slow growth rate – and this is showing on the figure 2a, 2b, with yellow.

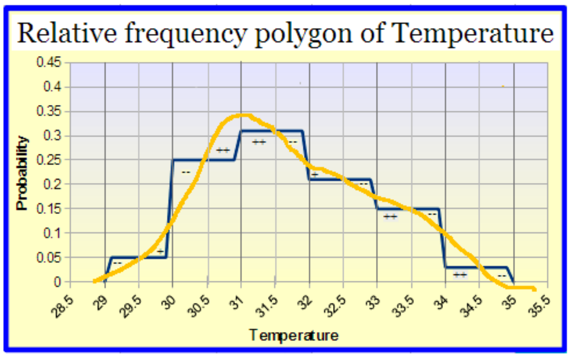

Fig. 2a. Relative frequency polygon for randomvariables (with blue); Approximation curve of the function of repartition for random variables (with yellow).

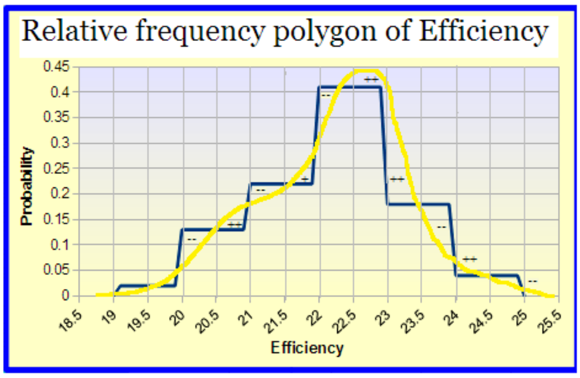

Fig. 3b. Relative frequency polygon for random variables (with blue); Approximation curve of the function of repartition for random variables (with yellow).

In this figure we can see the function of repartition (yellow), which can be approximate by polynomial interpolation [4], as following.

From the relative frequency polygon we can take the next representative points to find the continuous function of repartition:

<u>Temperature:</u>

x = [29, 30, 31, 32, 33, 34];

P(x) = [0.05, 0.25, 0.31, 0.21, 0.15, 0.03];

Through well known polynomial interpolation method [4], with these points we can build the function:

$$P(x) = -0.0043\ x^5 + 0.68\ x^4 - 42.74\ x^3 + 1340.29\ x^2 - 20990.77\ x^1 + 131349.17 \quad (1)$$

With the normalization condition following:

$$\int_{29}^{34} P(x)dx = \int_{29}^{34} (-0.0043\ x^5 + 0.68\ x^4 - 42.74\ x^3 + 1340.29\ x^2 - 20990.77\ x^1 + 131349.17\ )dx = \underline{1.0017;}$$

<u>Efficiency:</u>

y = [19, 20, 21, 22, 23, 24];

P(y) = [0.02, 0.13, 0.22, 0.41, 0.18, 0.04].

Through well known polynomial interpolation method [4], with these points we can build the function:

$$P(y) = 0.013\ y^5 - 1.48\ y^4 + 63.50\ y^3 - 1352.82\ y^2 + 14381.32\ y^1 - 61040.90 \quad (2)$$

With the normalization condition following:

$$\int_{19}^{24} P(y)dx = \int_{19}^{24} (0.013\ y^5 - 1.48\ y^4 + 63.50\ y^3 - 1352.82\ y^2 + 14381.32\ y^1 - 61040.90\ )dy = \underline{0.983}$$

Drawing the graph of these analytic functions, we obtain the two graphs (Figure 3a, 3b), which represents the approximate of the statistical distribution of these two functions.

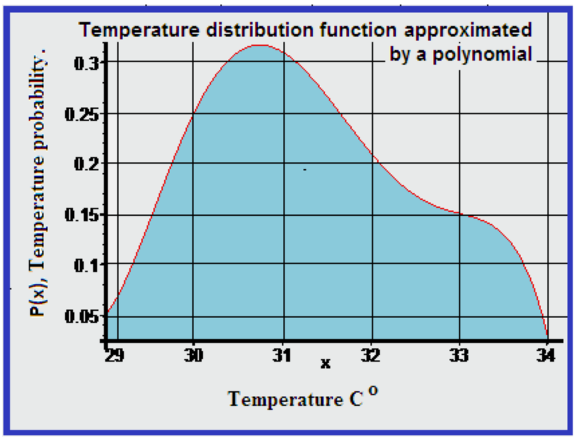

Fig. 3a. Graph of continuous distribution of probability for temperature parameter.

Finally by eliminating the common variable, between these two parameters we can obtain next equation of correlation which is connection from these two functions.

$$0.013\ y^5 - 1.48\ y^4 + 63.50\ y^3 - 1352.82\ y^2 + 14381.32\ y^1 - 61040.90 = -0.0043\ x^5 + 0.68\ x^4 - 42.74\ x^3 + 1340.29\ x^2 - 20990.77\ x^1 + 131349.17 \quad (3)$$

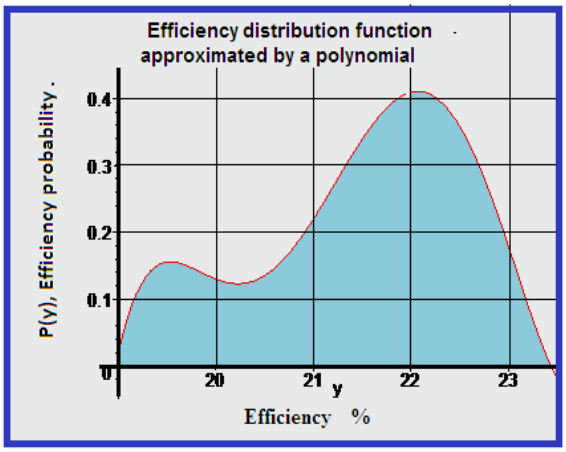

Fig. 3b.Graph of continuous distribution of probability for efficiency parameter.

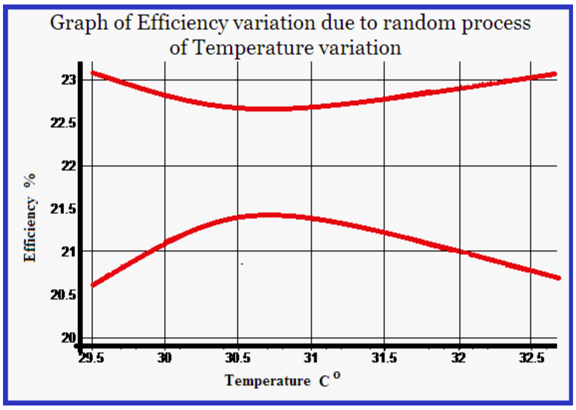

Fig. 4 Graph of function of correlation. Variation of the efficiency of the process, due to variation of the random temperatures.

Figure 4, shows the graph of new connection equation, where from the equality of the two forms of variation, we can get a new equation of connection from first random variable to the second random variable.

4. DISCUSSIONS

In the case, we need to find the value of the temperature or efficiency, when we know the proper probability, we can inverse the repartition polynomial as it follows:

Let us consider the next polynomial form:

$$y(x) = a_1\ x^{b_1} + a_2\ x^{b_2} + a_3\ x^{b_3} + \ldots + a_n\ x^{b_n}$$

In the Euclidian space, the vectorial form of the function [6] is as follow:

$$\bar{\Gamma}(x) = x\bar{e}_1 + y(x)\bar{e}_2$$

The tangent vector of the vectorial function will be:

$$d\bar{\Gamma}(y(x)) = [\bar{e}_1 + \frac{\delta(y(x))}{\delta x}\bar{e}_2]dx$$

From mathematical analisys [5], we can write:

$$\frac{\delta(x(y))}{\delta y} = \frac{1}{\frac{\delta(y(x))}{\delta x}}, \text{ thus}$$

$$d\bar{\Gamma}(x(y)) = [\bar{e}_1 + \frac{1}{\frac{\delta(y(x))}{\delta x}}\bar{e}_2]dy$$

thus we concluding that:

$$x(y) = \int_b^a \frac{1}{\frac{\delta(y(x))}{\delta x}} dy$$

In this way we can find the $x(y)$ form. This new function founded is helping us to calculate the variation of the argument when variation of the function is known.

5. CONCLUSION

In case, we need to evaluate - which part of one random variable is influenced from other parameters or features of the process variables, classically we can calculate the correlation factor, which is a number (for whole process).

With this method it is possible to find one linking between two strings of random data, in the hypothesis that two random sets are part of the same process, like different part of the process;

The obtained formula is a mathematical link between these two sets of variables, by which we can describe the variation at the one to the other one.

Because in the case of random natural processes without human intervention, a random relative frequency converges to the normal distribution, when the number of events tends to infinity, we can say that where the process are conducted by human intervention in the development process, the distribution function should have another form and this form, can be determined experimentally.

In the uniqueness of studied phenomenon (due to probability of both random parameters) the equation 1 is a functional link between the two random parameter and is showing us the influence of external disturbances parameters to the system process parameters and vice versa.

6. REFERENCES

[1] Gottschalk, L., Leblois, E., Skoien, J.O, *Correlation and covariance of runoff revisited*, Journal of Hydrology, vol. 398, no. 1-2, pp. 76–90, 2011.

[2] Rummel, C., Baier, G., Muler, M., *The influence of static correlations on multivariable correlation analysis of the EEG*, J Neurosci Meth 166 138–157: 21. doi:10.1016/j.jneumeth .2007.06.023.

[3] Frigioiu, C., Corneschi C., *Special Mathematics*, Publishing House of Dunarea de Jos University of Galati, 2006;

[4] Cretu, T., Falie, V., *Experimental data in physics*, Didactic and Pedagogic Publishing House, 1986.

[5] Guss, G., Stanasila, O., Stoica, T., *Mathematics. Mathematical analysis elements*, Didactic and Pedagogic Publishing House, 1986.

[6] Sabac, I. Gh., *Special Mathematics*, Didactic and Pedagogic Publishing House, 1965.