COMPOSITION OF RANDOM VARIABLE DISTRIBUTIONS IN QUALITY ENGINEERING

Prof. dr. eng. Liviu-Marius CÎRȚÎNĂ, *National University of Science and Technology POLITEHNIC, Bucharest, ROMANIA*
Prof. dr. eng. Oana Roxana CHIVU, *National University of Science and Technology POLITEHNICA Bucharest*
PhD Student Elena-Cosmina VÎNTURIȘ, *National University of Science and Technology POLITEHNIC, Bucharest, ROMANIA*
Phd Student Cristina FLEACĂ, *National University of Science and Technology POLITEHNIC, Bucharest, ROMANIA*

ABSTRACT: This paper presents the composition of random variable distributions as a fundamental tool in quality engineering and dimensional chain analysis. The study explores the characteristics of discrete and continuous random variables, including probability distribution functions, mean values, variance, and standard deviation. The convolution product method for composing multiple distributions is detailed, with practical examples including the classical "sum of two dice" problem. The research demonstrates that when composing more than five independent random distributions, the resulting distribution approaches a normal distribution according to the Central Limit Theorem, validated by Laplace and Liapunov. The methodology enables engineers to predict dimensional variation propagation through complex assembly chains, supporting robust design and quality assurance in industrial applications.

Keywords: random variables, probability distribution, convolution product, normal distribution, quality engineering

1. INTRODUCTION

In technological manufacturing processes, the resulting parts exhibit dimensional magnitudes that differ from one specimen to another. The dispersion of these dimensions constitutes a fundamental random phenomenon in quality engineering and manufacturing process control. Understanding and mathematical modeling of random variables becomes essential for predicting the behavior of finished products and optimizing production processes.

Processing precision is influenced by multiple sources of variation, including the rigidity of the technological system, tool wear, control precision, and measurement equipment. Each of these sources contributes to the final dispersion of the dimensional characteristics of the product.

The composition of random variable distributions represents a powerful mathematical tool for the analysis of dimensional chains and prediction of cumulative tolerances in complex assemblies. This methodology allows engineers to anticipate the variation of final product characteristics based on the variations of individual components.

2. CHARACTERISTICS OF RANDOM VARIABLES

2.1 Definition of Random Variable

A **random variable X** is defined as any magnitude that can take different values with certain probabilities of realization. A random variable can have discrete or continuous variation, depending on the nature of the phenomenon being studied.

In the case of an experiment, the random variable X can take the values $X_1, X_2, \dots, X_n$ with probabilities $P(X_1), P(X_2), \dots, P(X_n)$. This correspondence is expressed in the form of a distribution table:

Table 1: Distribution table of a discrete random variable.

X	x_1	x_2	$\cdots$	x_n
$P(X)$	$P(x_1)$	$P(x_2)$	$\cdots$	$P(x_n)$

The above expression is called a **probability distribution** and shows the correspondence between values and their respective probabilities. A fundamental property is that the sum of probabilities must equal unity:

$$\sum_{i=1}^{n} P(x_i) = 1$$

2.2 Distribution Function

The probability that the argument of the random variable takes values less than a certain value x, i.e., $X < x$, is denoted by $F(x) = P(X < x)$, where $F(x)$ is the distribution function of the random variable X.

For discrete variables, the distribution function is given by:

$$F(x) = \sum_{x_i < x} P(x_i)$$

For continuous variables, the distribution function is expressed through the integral of the probability density:

$$F(x) = \int_0^x f(t)dt$$

where $f(t)$ is the probability density. The function $F(x)$ has values between 0 and 1, being a monotonically increasing function.

2.3 Statistical Parameters for Discrete Variables

The important parameters that characterize the distribution of a discrete random variable are:

Mean value:

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

Median value M_e, according to STAS 2631-82:

$$M_e = x_{\frac{n+1}{2}} \text{ (for n odd)}$$

$$M_e = \frac{x_{n/2} + x_{(n/2)+1}}{2} \text{ (for n even)}$$

Central value x_c of the value series:

$$x_c = \frac{x_{\max} - x_{\min}}{2}$$

Variance $D(x)$ characterizes the tendency of values to scatter around the mean value:

$$D(x) = \frac{1}{N-1}\sum_{j=1}^{n} n_j (x_j - \bar{x})^2 = \frac{1}{N(N-1)}\left[N\sum x_j^2 - \left(\sum x_j\right)^2\right]$$

Standard deviation:

$$\sigma(x) = \sqrt{D(x)}$$

Relative standard deviation:

$$\lambda = \frac{\sigma}{x_{\max} - x_{\min}}$$

2.4 Parameters for Continuous Variables

For continuous random variables, the important parameters are:

Arithmetic mean value:

$$\mu = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$

Variance $D(x)$:

$$D(x) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx$$

Standard deviation**:**

$$\sigma = \sqrt{D(x)}.$$

3. COMPOSITION OF RANDOM VARIABLES

3.1 Fundamental Principles

The calculation of summing distributions of independent random variables consists of adding or composing two or more distributions. This operation is fundamental in the analysis of dimensional chains in technological processes.

Let X and Y be two independent random variables. The distribution of the variable $Z = X + Y$ is called the **composition** of X with Y.

For two simple random variables X and Y with distributions:

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

$$Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \\ q_1 & q_2 & \cdots & q_m \end{pmatrix}$$

The random variable $Z = X + Y$ has the distribution table:

$$Z = \begin{pmatrix} x_1 + y_1 & x_2 + y_1 & \cdots & x_n + y_m \\ p_{11} & p_{12} & \cdots & p_{nm} \end{pmatrix}$$

where $p_{nm} = p_n \cdot q_m$.

3.2 Calculation Formula

The probability of the composed distribution $P(z)$ from two independent random variable distributions is calculated with the relation:

$$P(z) = \sum_{i=1}^{n} P_1(i) \cdot P_2(z - i)$$

This fundamental formula allows the calculation of probability for any value of the sum Z.

3.3 Example of the Sum of Two Dice

The justification of the previous formula is illustrated by the classic example of the **"Sum of Two Dice"**. The probability that each number of the die appears on a throw is $P = 1/6$. The probability that the throw results in the sum $Z = 5$ is obtained as follows:

$$P(Z = 5) = P_1(1) \cdot P_2(4) + P_1(2) \cdot P_2(3)$$
$$+P_1(3) \cdot P_2(2) + P_1(4) \cdot P_2(1)$$
$$P(Z = 5) = \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{4}{36}$$

Table 2: Sum matrix for two dice - composition of two uniform distributions

Z	1	2	3	4	5	6
6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
1	2	3	4	5	6	7

The numbers on the die from 1 to 6 represent the scatter field of values. In this case, the scatter field is uniform, the distribution being rectangular in form. The product of composition is an **isosceles triangle** type distribution, with values from 2 to 12.

3.4 Convolution Product

If we consider two discrete distributions, their composition is a probability distribution of the superposition called a **convolution product**. Thus, if F_1 and F_2 are two discrete distributions, the convolution product is:

$$F(x) = \sum_{i=1}^{n} F_1(i) * F_2(x - i)$$

In the case of continuous distributions, the sum distribution becomes:

$$F(x) = \int_{-\infty}^{+\infty} F_1(y) * F_2(x - y)dy$$

For the composition of multiple distributions, an iterative process takes place:

$$F_{rez}(x) = F_1(x) * F_2(x) * F_3(x) * F_4(x) * \cdots$$

where the symbol $*$ designates the composition operation.

3.5 Composition of Continuous Variables

Let X and Y be two independent random variables with probability densities $f_x(x)$ and $f_y(y)$. The probability density of the sum $Z = X + Y$ is:

$$f_z(z) = \int_{-\infty}^{+\infty} f_x(x) \cdot f_y(z - x)dx$$

or equivalently:

$$f_z(z) = \int_{-\infty}^{+\infty} f_y(y) \cdot f_x(z - y)dy$$

For designating the composition operation, the symbolic notation is used:

$$f_z(z) = f_x(x) * f_y(y)$$

The distribution function of the probabilistic magnitude Z will be:

$$F(z) = \iint_{x+y \le z} f_x(x) * f_y(y)dxdy$$

4. PROPERTIES OF DISTRIBUTION COMPOSITION

4.1 Associativity and Commutativity

The composition operation fulfills the fundamental laws of associativity and commutativity:

Associativity:

$$[F_1(x) * F_2(x)] * F_3(x) = F_1(x) * [F_2(x) * F_3(x)]$$

Commutativity:

$$F_1(x) * F_2(x) = F_2(x) * F_1(x)$$

These properties significantly simplify calculations for complex dimensional chains.

4.2 Mean Value of the Convolution Product

The mean value of the convolution product is obtained by summing the mean values of the components:

$$\mu_{rez} = \sum_{i=1}^{n} \mu_i$$

Let $\sum_{k=1}^{n} C_k X_k$ where C_k are constants. The variance of this sum is:

$$D\left(\sum_{k=1}^{n} C_k X_k\right) = \sum_{k=1}^{n} C_k^2 D(X_k)$$

Particular case: If $C_1 = C_2 = \cdots = C_n = 1$, then:

$$D\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} D(X_k)$$

This means that **the variance of a sum of independent random variables equals the sum of the variances of the random variables**.

The resulting standard deviation is calculated with:

$$\sigma_{rez} = \sqrt{D(X_{rez})}$$

4.3 Error Propagation

For a general function $f(x_1, x_2, \dots, x_n)$, the standard deviation is calculated through the error propagation formula:

$$\sigma_f^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2$$

4.4 Central Limit Theorem

The calculation of the central value of the resulting distribution is done with the arithmetic mean formula. If from a set of events samples are taken, each having a central value, then the central value of the entire set is the arithmetic mean of the central values of the samples.

The law of distribution of the central value, stated by Laplace in 1812 and demonstrated by Liapunov in 1901, affirms that:

"A probabilistic magnitude is normally distributed in the vicinity if it represents a sum of a large number of variables independent of each other, each bringing to the sum a negligible quantity."

Thus, in the case of composing **more than five distributions**, each composed of independent random variables, the resulting distribution will be a **normal distribution** (Gaussian).

5. APPLICATIONS IN QUALITY ENGINEERING

5.1 Dimensional Chains

The fundamental result of the theory of distribution composition has major practical implications in quality engineering:

-For **dimensional chains with more than five primary elements**, the dimensional distribution of the resulting element will be a normal distribution, regardless of the type of distributions of individual components.

-For **chains with fewer than five elements** that present dimensional distributions different from the normal one, a detailed analysis through successive composition is necessary.

5.2 Tolerance Analysis

The distribution composition method allows:

1. Prediction of the dimensional distribution of the final product
2. Optimization of component tolerances for cost minimization
3. Evaluation of manufacturing process capability
4. Establishment of quality control strategies

5.3 Measurement Equipment Precision

The composition of measurement errors from multiple sources (calibration, repeatability, reproducibility, environmental conditions) allows estimation of the total uncertainty of the measurement system, essential for ensuring measurement quality.

5.4 Calculation Example

Consider the resulting radial clearance G in an assembly in which two machined components have approximately uniform variations, and adjustment is done with standardized spacer shims with discrete thicknesses, a case often encountered in practice in "shim packs" in steps of 0.05-0.20 mm.

The chain model is G = H - (A + B + S), where H is a housing dimension considered constant at assembly, A and B come from nearly "flat" processes in the tolerance interval, and S is a discrete variable that models the selection of adjustment shim from a finite set of thicknesses.

Data and Assumptions

- H = 15.20 mm (constant at assembly to isolate the effect of random composition under 5 components).
- A ~ U(9.95, 10.05), such that E[A] = 10.00 mm and the support is [9.95; 10.05], representing an approximately uniform distribution in tolerance.
- B ~ U(4.98, 5.02), such that E[B] = 5.00 mm and the support is [4.98; 5.02], also uniform in tolerance.
- S ∈ {0.10, 0.15, 0.20} mm with probabilities {0.2, 0.5, 0.3}, reflecting practical selection from standardized shim packages in standard steps.

Distributional Form

The sum of two independent uniform variables has triangular density on the support of the sum, so T = A + B has a triangular distribution on [9.95 + 4.98, 10.05 + 5.02] = [14.93, 15.07], with mode around 15.00.

The distribution of G = H - (T + S) becomes a mixture of three "shifted" triangulars corresponding to the discrete values of S, which produces a multimodal non-normal distribution for n = 3, typical for finite mixtures of components.

In this situation, the normal approximation is not adequate because n is small and the discrete component introduces distinct modes, so the evaluation is done through exact composition (convolution) or numerical simulation, not through normal z-scores.

Calculation Steps

- Means: E[T] = E[A] + E[B] = 10.00 + 5.00 = 15.00 mm, and E[S] = 0.2 × 0.10 + 0.5 × 0.15 + 0.3 × 0.20 = 0.165 mm, so E[G] = 15.20 - 15.00 - 0.165 = 0.035 mm, which indicates a low centering of the clearance.
- For S = 0.15, G ∈ [-0.02, 0.12] mm, also triangular but translated, contributing to a second mode.
- For S = 0.20, G ∈ [-0.07, 0.07] mm, the third triangular branch, so the global distribution is a multimodal mixture of three weighted triangulars {0.2, 0.5, 0.3}.

Conclusions from Results

- G is multimodal and cannot be adequately evaluated by normal approximation at n = 3; we recommend exact integration of triangular density over the acceptance interval, or a Monte Carlo simulation for conformity probabilities.
- Mean E[G] = 0.035; the global form is a mixture of triangulars; the sum of uniforms generates triangular densities (Irwin-Hall, case n = 2).
- The property "uniform + uniform = triangular" justifies the use of triangular densities in continuous composition for n = 2, as a first step before CLT applicability at higher n.
- Industrial shim sets have discrete standard thicknesses, which naturally leads to distribution mixtures in dimensional chains, with possible multiple modes in the functional resultant.

6. CONCLUSIONS

The composition of random variable distributions represents a fundamental mathematical tool in quality engineering and technological process analysis. The main results of this study are:

1. The **convolution product** offers a systematic method for calculating the distribution of the sum of independent random variables, applicable to both discrete and continuous variables.
2. The **Central Limit Theorem** guarantees that for chains with more than five primary elements, the resulting distribution converges to the normal distribution, significantly simplifying statistical analysis and facilitating the application of standard quality control methods.
3. The **associativity and commutativity properties** of the composition operation allow flexibility in complex calculations for multi-level dimensional chains.
4. The **variance propagation formula** ($\sigma_{rez}^2 = \sum \sigma_i^2$) constitutes the theoretical basis for statistical tolerance analysis in robust product design.
5. The methodology can be extended for the analysis of **any complex system** in which the final characteristic depends on the sum or linear combination of independent variables, including measurement uncertainty analysis and process capability evaluation.

REFERENCES

[1] Laplace, P.S. (1812). *Théorie Analytique des Probabilités*. Courcier, Paris.

[2] Liapunov, A.M. (1901). Nouvelle forme du théorème sur la limite de probabilité. *Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg*, Série 8, 12(5).

[3] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd Edition. John Wiley & Sons, New York.

[4] STAS 2631-82. Statistical Quality Control of Products. Terminology.

[5] Montgomery, D.C. (2019). *Introduction to Statistical Quality Control*, 8th Edition. John Wiley & Sons.

[6] Wadsworth, H.M., Stephens, K.S., Godfrey, A.B. (2002). *Modern Methods for Quality Control and Improvement*, 2nd Edition. John Wiley & Sons.

[7] ISO 3534-1:2006. Statistics - Vocabulary and symbols - Part 1: General statistical terms and terms used in probability.

[8] Papoulis, A., Pillai, S.U. (2002). *Probability, Random Variables and Stochastic Processes*, 4th Edition. McGraw-Hill.

[9] Ross, S.M. (2014). *Introduction to Probability Models*, 11th Edition. Academic Press.

[10] Evans, D.H. (1975). Statistical tolerancing: The state of the art. *Journal of Quality Technology*, 7(1), 1-12.