The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2024

Volume 30, Pages 9-16

ICBAST 2024: International Conference on Basic Sciences and Technology

Construction of an Integral Distribution Function of Random Variables Determining the Impact of Wind Power Plant on Birds in accordance with the Predictive Analysis by Experts

Vladimir Yeremeev
Melitopol State University

Andrey Naidysh
Melitopol State University

Oksana Stokan
Melitopol State University

Abstract: The team of scientists led by ornithologist V. D. Siokhin proposed a method for predicting the impact of planned wind power plants (WPP) on avifauna using expert assessment. The method for modeling the differential $p(x)$ and integral $F(x)$ distribution functions of the sum of several random variables with distribution densities $p_1(x_1)$, $p_2(x_2)$, $p_3(x_3)$…, determining the probability of bird interaction on the territory of a wind farm and adjacent buffer zones is developed in this article. The function p(x), which is a convolution of $p_1(x_1)$, $p_2(x_2)$, $p_3(x_3)$ ..., is represented as an improper integral with infinite limits. An algorithm for calculating $p(x)$ and $F(x)$ is proposed by replacing an improper integral with an integral having finite limits and subsequent numerical integration taking into account the specified accuracy. Testing of the method for calculating the functions $p(x)$ and $F(x)$ was carried out on two examples with known solutions. One of the tests was carried out on two differential functions of distribution subjecting the normal law with mathematical expectations a_1 and a_2, and variances σ_1^2 and σ_2^2. It is known that the distribution of such a sum obeys the normal law with mathematical expectation a equal to the sum of a_1 and a_2, and variance σ^2 equal to the sum of σ_1^2 and σ_2^2. The results of calculations of the functions $p(x)$ and $F(x)$ using numerical methods for the number of nodes of 100 and more coincide with known solutions with an error of 10^{-15}, which indicates the high accuracy of the proposed method. The found integral distribution function allows us to determine the probability characteristics of the impact of wind farms on birds.

Keywords: Integration, Mathematical statistics, Numerical method, Random variable distribution function.

Introduction

The team of scientists led by ornithologist V.D. Siokhin proposed a method for predicting the impact of planned wind power plants (WPP) on avifauna using expert assessment. The theory of distribution of a system of random variables given by a vector ($\alpha_1, \alpha_2, \ldots \alpha_n$) on Borel sets $B_1, B_2, \ldots Bn$ from $B(-\infty, +\infty)$ is described in many monographs and textbooks (Gnedenko & Kolmogorov, 1949; Kremer, 2004; Gmurman, 1999; Samarova, n.d.). A system of random variables defining a random vector ($\alpha_1, \alpha_2, \ldots, \alpha_n$) when the condition is met

$$P(\alpha_1 \in B_1, \alpha_2 \in B_2, \ldots \alpha_n \in B_n) = P(\alpha_1 \in B_1) \cdot P(\alpha_2 \in B_2) \cdot \ldots P(\alpha_n \in B_n), \quad (1)$$

it is called independent in aggregate.

The distribution function of a random vector ($\alpha_1, \alpha_2, \dots, \alpha_n$) in the space of n variables ($x_1, x_2, \dots, x_n$)$\in R^n$ is determined by equality

$$F_{\alpha_1,\alpha_2,..\alpha_n}(x_1, x_2, ...x_n) = P(\alpha_1 \le x_1, \alpha_2 \le x_2, \dots \alpha_n \le x_n) \quad (2)$$

For a continuous random vector, there is a non-negative function $p_{\alpha_1,\alpha_2,..\alpha_n}(x_1, x_2, ...x_n)$, which is called the distribution density, that for any Borel set $B\in(R^n)$ ensures the following equality

$$P\{(\alpha_1, \alpha_2, \dots \alpha_n) \in B\} = \int_B p_{\alpha_1,\alpha_2,..\alpha_n}(x_1, x_2, ...x_n) dx_1 dx_2, ...dx_n. \quad (3)$$

The distribution function for any random vector is determined by multidimensional integrals

$$F_{\alpha_1,\alpha_2,..\alpha_n}(x_1, x_2, ...x_n) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} p_{\alpha_1,\alpha_2,..\alpha_n}(x_1, x_2, ...x_n) dx_1 dx_2, ...dx_n. \quad (4)$$

The multidimensional integral of the distribution density over the entire area of its definition is equal to one:

$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} p_{\alpha_1,\alpha_2,..\alpha_n}(x_1, x_2, ...x_n) dx_1 dx_2, ...dx_n = 1. \quad (5)$$

Formulas (3)-(5) can be used to find the distribution function of the sum of random variables. In most cases, taking integrals in an analytical form is not possible and finding the vector (2) is reduced to numerical modeling of the true distribution law. Therefore, the analysis of calculation methods is an urgent task.

Method

Analysis of Recent Research and Publications

The integral $F_{ak}(x)$ and differential $p_{ak}(x)$ functions for the one-dimensional distribution of random variables, which are a special case of the multidimensional distribution (4), have the form

$$F_{\alpha_k}(x) = \int_{-\infty}^{x} p_{\alpha_k}(x) dx_k. \quad (6)$$

$$p_{\alpha_k}(x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} p_{\alpha_1,\alpha_2,..\alpha_n}(x_1, x_2, ...x_n) dx_1 ... dx_{k-1}, dx_{k+1}, ...dx_n \right) dx_k. \quad (7)$$

A large number of publications (Tregubova & Hartov, 2022; Rozovsky, 2022; Ganin & Polenin, 2015) have been devoted to the distribution of the sum of continuous random variables. It can be obtained from formulas (6), (7). Let us consider two independent random variables α_1, α_2 with distribution densities $p_{a1}(x_1)$, $p_{a2}(x_2)$. The differential distribution function of the sum $\alpha=\alpha_1+\alpha_2$ according to (7) is determined by the formula (Samarova, n.d)

$$p_\alpha(x) = \int_{-\infty}^{+\infty} p_{\alpha_1}(x_1) p_{\alpha_2}(x - x_1) dx_1, \quad (8)$$

which is actually a convolution of two functions $p_{a1}(x_1)$, $p_{a2}(x_2)$.

According to definition (6), the integral distribution function for the sum of two random variables α_1 and α_2 has the form

$$F_\alpha(x) = P(\alpha = \alpha_1 + \alpha_2 \le x) = \int_{-\infty}^{x} p_\alpha(x)dx. \qquad (9)$$

The differential distribution function (8) can be easily generalized to the case of the sum of three or more random variables.

Example. Let the random variable U be equal to the sum of three independent random variables $U=\alpha_1+\alpha_2+\alpha_3$ with known distribution densities $p_{a1}(x_1)$, $p_{a2}(x_2)$, $pa3(x3)$. Since the differential distribution function of the sum of two random variables $\alpha_1+\alpha_2$ is determined by formula (9), then the integral

$$\mathrm{p}_\alpha(u) = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\mathrm{p}_{\alpha_1}(x_1)\mathrm{p}_{\alpha_2}(t-x_1)dx_1\right)\mathrm{p}_{\alpha_3}(u-t)dt; u,t,x_1 \in [-\infty,+\infty]. \qquad (10)$$

determines the distribution density of the sum of three random variables $\alpha_1+\alpha_2+\alpha_3$.

Formulas (6)-(10) allow us to construct a numerical model for the integral and differential distribution functions of the sum of random variables in the space of n variables $(x_1, x_2, \ldots x_n) \in R^n$.

The Purpose of the Work and the Task Statement

The purpose of the research is to develop an algorithm for numerical modeling of distribution functions of the sum of random variables. To achieve the goal, it is necessary to solve the following tasks: to analyze numerical methods for calculating improper integrals; to select an algorithm for calculating integrals of type (8)-(10) and estimate its error in calculating the integral and differential distribution functions of the sum of random variables.

Results and Discussion

Let's consider an algorithm for calculating improper integrals of type (8)-(10) by numerical method. The first step involves converting them into certain integrals with a given accuracy.

As an example, we replace the integral (9) with its approximate value $\int_a^x p_a(x)dx; x \in [a,b]$ ensuring an error of no more than α for all x on the segment $[a,b]$:

$$|\int_a^x p_\alpha(x)dx - \int_{-\infty}^{x} p_\alpha(x)dx| < \alpha; x \in [a,b]. \qquad (11)$$

The exact value of the first definite integral in condition (11) can be represented as an integral sum [10]:

$$\int_a^x p_\alpha(x)dx = S = \sum_{i=1}^{n} p_\alpha(x_i^*)h_i, x \in [a,b], \qquad (12)$$

where $p_a(x_i^*)$ is the value of the integrand function at the point x_i^* belonging to the i interval h_i, $\sum h_i = b-a$.

Since the value of $F(x)$ belongs to the segment $[0,1]$, the value of the integral sum at $x=b$ must correspond the condition:

$$|1\text{-}S| < \alpha. \qquad (13)$$

The definition of an improper integral (12) is reduced to the calculation of the integral sum S. The accuracy of the calculation results depends on the correctness of replacing the infinite limits with finite values and the error in calculating the integral sum. One of the most common ways to calculate a certain integral is based on the use of Newton–Cotes quadrature formulas («Methods of numerical integration», n.d.).

The simplest quadrature formulas are implemented in the methods of trapezoids («Numerical methods of calculating the definite integral», n.d.), left, right and middle rectangles («Method of rectangles», n.d.) and parabolas («Numerical methods of calculating the definite integral», n.d.). These methods are widely used in the educational process and in solving many applied problems. In case of the method of average rectangles with a constant step $h_i=h$ the integral sum S in formula (12) is represented as

$$S \approx S^* = h\sum_{i=1}^{i=n} p(x_i + h/2). \tag{14}$$

The use of formula (14) provides an error of no more than (Fikhtenholtz, 1969):

$$\varepsilon = n\,|\,p''(x^*)\,|\,h^3/24 \tag{15}$$

where n – is the number of intervals, $nh=b\text{-}a$, $\varepsilon=p''(x^*)$ - is the maximum value of the second derivative function at the point $t^*\epsilon[a,b]$.

The actual error turns out to be less than the value determined by the integral sum (15). It was shown in (Yeremeev, 2023) that its value, found for functions in the case of an absolute value of the second derivative of the order of one with the number of intervals n =1000, is about 10-8, which is quite acceptable for numerical modeling of distribution functions of random variables. For large values of the second derivative, as well as when finding the distribution of the sum of several random variables, more accurate methods may be required, for example, Simpson or Gauss methods. Using a variable interval value can also significantly improve the accuracy. One of the simple ways to determine the h_i+1 interval relates its length to the length of the previous hi interval using the formula:

$$h_{i+1} = h_i\,|\,p'(t_i)/p'(t_{i+1})\,| \tag{16}$$

where $p'(t_i)$ and $p'(t_{i+1})$ are the first derivatives at the nodes t_i and t_{i+1}.

The algorithm for numerical modeling of distribution functions of the sum of random variables is presented in the form of the following steps:

- replacement of an improper integral with a definite integral with finite integration limits to ensure a given accuracy,
- choice of a numerical integration method,
- testing a numerical model using well-known accurate methods for calculating distribution functions.

The testing was performed using two examples. Let two random variables α_1, α_2 with distribution densities be given

$$p_1(x) = \begin{cases} 0, x < 0 \\ 0 \le x \le 1, \\ 0, x > 1 \end{cases} \tag{17}$$

$$p_2(x) = \begin{cases} 0, x < 0 \\ e^{-x}, 0 \le x < \infty. \end{cases} \tag{18}$$

Differential and integral distribution functions of the sum of two random variables defined by formulas (17), (18), (8), (9), have the form of:

$$p(x) = \begin{cases} 0, x < 0, \\ 1 - e^{-x}, x \in [0,1], \\ e^{-x}(e-1), x > 1. \end{cases} \quad (19)$$

$$F(x) = \begin{cases} 0, x < 0 \\ \int_0^x (1 - e^{-x})dx = t + e^{-x} - 1, 0 \le x \le 1 \\ e^{-1} + \int_1^x e^{-x}(e-1)dx = e^{-1} + (e-1)(e^{-1} - e^{-x}), 1 \le x < \infty. \end{cases} \quad (20)$$

The distribution function of the sum of two random variables (8), (9) was calculated in two ways: by the method of average rectangles, formula (14); using analytical formulas (19), (20).

The accuracy of presenting the results in the form of a numerical model for the distribution of the sum of random variables depends on the choice of the limits of integration of a certain integral [a, b] and the method of calculating the integral sum. The largest error in calculating the function $p(x)$ is expected at $x=b$. The value of the distribution density $p(x)$ depending on b for a=0 and the number of intervals n from 10 to 104 is given in Table 1.

Table 1. The results of calculations

b	b=5	b=10	b=15	b=20
$n=10^2$	0,0115776	0,0000780	0,0000005	$<10^{-15}$
$n=10^3$	0,0115776	0,0000780	0,0000005	$<10^{-15}$
$n=10^4$	0,0115777	0,0000780	0,0000005	$<10^{-15}$

According to Table 1, fairly good results are achieved at $b \ge 5$ and the number of partitions $n \ge 10^2$, where the error is provided at the level of $5 \cdot 10^{-8}$ or less. Close accuracy is obtained when calculating the integral function. The second test was carried out using the example of the distribution of the sum of two random variables with a normal distribution. Let there be two normal distributions with mathematical expectations a_1=0, a_2=3 and standard deviations σ_1=1 and σ_2=2:

$$p_1(x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}}, p_2(x_2) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x_2-3)^2}{8}}. \quad (21)$$

Figure 1 shows graphs of the distribution density for two functions $p_1(x_1)$, $p_2(x_2)$, corresponding to the normal law (21), and the distribution density $p(x)$ for the sum $x=x_1+x_2$, calculated using convolution (8) by the method of rectangles.

It is known (Lemons, 2003; «Sum of normally distributed random variables», n.d.) that the mathematical expectation a and the variance σ_2 of the sum of two random variables obeying the normal law are equal, respectively, to the sum of mathematical expectations and variances of its terms, i.e.: $a=a_1+a_2$, $\sigma_2=\sigma_{21}+\sigma_{22}$. Therefore, the distribution density of the sum of two random variables determined by formulas (21) has the form:

$$p(x) = \frac{1}{\sqrt{10\pi}} e^{-\frac{(x-3)^2}{10}}. \quad (22)$$

Calculations of the distribution density of the sum of two random variables $x=x_1+x_2$ in accordance with formula (22), and the same sum calculated using convolution (8) by the numerical method for the number of intervals n=10 differ only in the second significant digit. Increasing the number of intervals to n = 30 ensures an accuracy of at least 10^{-10} (Table 2).

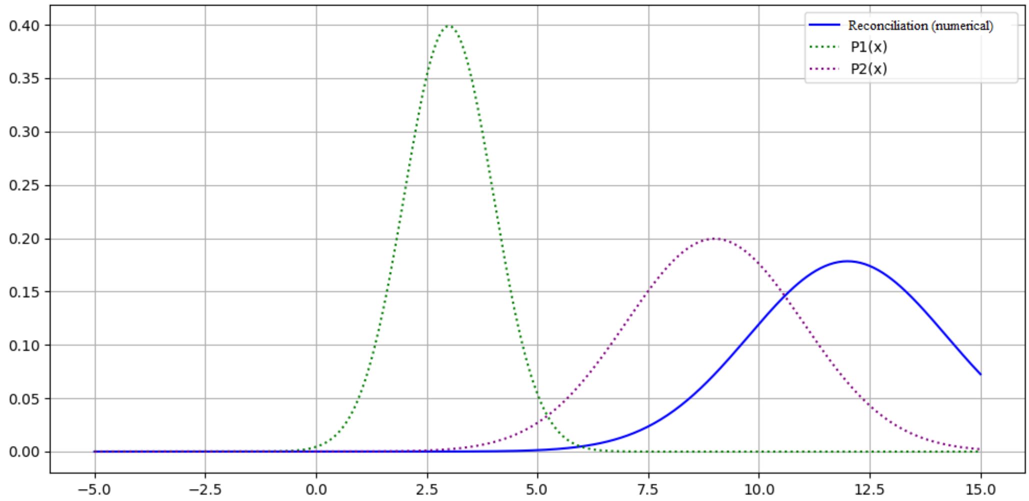

Figure 1. Distribution densities of the random numbers

Table 2. The results of calculating

x	3	6	9	12
p(x), formula (8)	0.0000541551	0.0048748912	0.072537074	0.17841241
p(x), n=30, convolution	0.0000541551	0.0048748912	0.072537074	0.17841241

With an increase in the number of intervals to 100, 1000, 10000, the accuracy of calculations increases to 10^{-15}-10^{-19}. The integral distribution function of the random variable *X* for the normal law (22) is determined by the formula:

$$F(x) = \frac{1}{2}(1 + \Phi(t)), \tag{23}$$

where $t=(x-3)/\sqrt{5}$, $\Phi(t)$- is the Laplace integral:

$$\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt.$$

Test calculations of the integral function for different values of n are given in Table 3.

Table 3. The results of calculating

x	3	6	9	12
Formula (23)	0,5000	0,9101	0,9963	1,0000
Formula (6), *n*=30	0,5053	0,9109	0,9965	0,9998
Formula (6), *n*=100	0,5016	0,9104	0,9964	0,9998
Formula (6), *n*=300	0,5005	0,9102	0,9964	0,9998
Formula (6), *n*=1000	0,5002	0,9102	0,9963	0,9998

According to the calculation results presented in Table 3, the error of the numerical method for determining the integral distribution function for the sum of two normally distributed random variables at $n = 30$ is about 10^{-3}. Increasing the number of intervals to $n = 1000$ reduces the error by an order to $2 \cdot 10^{-4}$. Based on the verification carried out (Tables 1-3, Figure 1), it can be argued that numerical modeling of the distribution function for the sum of several random variables provides a sufficiently high accuracy.

Conclusion

There it has been developed the method for determining the distribution functions of the sum of random variables by replacing an improper integral with an integral with finite integration limits and then choosing a numerical integration method. The method was tested using the examples with two distribution densities $p_{\alpha 1}(x_1)$, $p_{\alpha 2}(x_2)$, for which the law of distribution of the sum of random variables $\alpha_1+\alpha_2$ is known. In the first test, the distribution density $p_{\alpha 1}(x_1)$ corresponded to a uniform distribution on the interval [0,1], and $p_{\alpha 2}(x_2)$ to an exponential distribution on the half-interval [0, +∞). In the second test, we analyzed the sum of two differential distribution functions obeying the normal law. The accuracy of calculations of the distribution density of the sum of two random variables turned out to be about 10^{-15}. The accuracy of calculating the integral function was lower: in the first case, it was no worse than 10^{-7}, and in the second - about 10^{-4}. If necessary, the accuracy of calculations can be improved by optimizing the numerical algorithm, namely:

– the implementation of a more accurate method for calculating the integral sum,
– the use of a variable step, determined, for example, by the formula (16).

The found integral distribution function allows determining probabilistic characteristics of the impact of wind farms on birds that fly through them or adjacent buffer zones.

Recommendations

The obtained data can be used in applied mathematical statistics in the study of measurement error, the results of which depend on the action of several factors.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Basic Sciences and Technology (www.icbast.net) held in Antalya/Turkey on November 14-17, 2024.

* The work was carried out under the state order FRRS-2023-0035 «Assessment of the state of seasonal ornithological complexes and transcontinental migrations at monitoring sites of the Azov-Black Sea region and Sivash, development of software for modeling and managing natural complexes».

References

Fikhtengol'ts, G. M. (1969). Course of differential and integral calculus. *M.: Science*, *2*, 153.

Ganin, M.P., & Polenin, V.I. (2015). Distributions of sums of a random number of continuous random variables. *Scientific periodical, Ceteris Paribus. Phys.-Mat. Sciences, 3*.

Gmurman, V.E. (199). *Probability theory*. Moscow: Higher School.

Gnedenko, B.V., &Kolmogorov, A.N. (1949). *Limit distributions for sums of independent random variables*. Moscow: Fizmatgiz.

ITMO University. (n.d). *Methods of numerical integration*. Retrieved from http:// aco.ifmo.ru/el_books/numerical_methods/lectures/glava2.html.

Kremer, N.S. (2004). *Probability theory and mathematical statistics*. Moscow: Izd. UNITI-Dana.

Lemons, D. S., & Langevin, P. (2002). *An introduction to stochastic processes in physics*. JHU Press.

Mathprofi. (n.d.). *Method of rectangles*. Retrieved from http://mathprofi.ru/ metod_prjamougolnikov.html.

Rozovsky, L.V. (2022). Large evasions of the sum of independent random variables whose distributions have rapidly decreasing tails. *Probability Theory and Its Applications*, *67*(3), 456-470.

Samarova, S.S. (n.d). *Joint distribution of random variables*. Retrieved from https://old.mipt.ru/education/chair/mathematics/study/methods/%D0%A1%D0%A0%D0%A1%D0%92_%D0%A1%D0%B0%D0%BC%D0%BE%D1%80% D0%BE%D0%B2%D0%B0(2).pdf

Tomsk Polytechnic University. (n.d). *Numerical methods of calculating the definite integral*. Retrieved from https://portal.tpu.ru/SHARED/k/KRAYNOV/Study/Tab1/Tab/2nd_task.pdf.

Tregubova, K. A., & Khartov, A. A. (2022). Sums of independent random variables and the generalized Dickman laws. *Zapiski Nauchnykh Seminarov POMI*, *515,* 199-213.

Wikipedia. (n.d). *Sum of normally distributed random variables*. Retrieved from https://en.wikipedia.org/wiki/Sum_of_normally_distributed_random_ variables.

Yeremeev, V.S., Goncharov, A.Yu., & Nikulin. M.A. (2023, May 23). Comparative analysis of accuracy of calculation of definite integrals by different methods. *Modern problems of geometrical modeling and information technologies*. Melitopol: MSU. 7-19.

Author Information

Vladimir Yeremeev
Federal State Budgetary Educational Institution of Higher Education «Melitopol State University»
Melitopol, Russia

Andrey Naidysh
Federal State Budgetary Educational Institution of Higher Education «Melitopol State University»
Melitopol, Russia

Oksana Stokan
Federal State Budgetary Educational Institution of Higher Education «Melitopol State University»
Melitopol, Russia
Contact e-mail: *oksana.strokan0@mail.ru*

To cite this article:

Yeremeev, V., Naidysh, A., & Strokan, O. (2024). Constructıon of an integral distribution function of random variables determining the impact of wind power plant on birds in accordance with the predictive analysis by experts. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 30,* 9-16.