UDC 519.2:519.6:519.9 **doi:** https://doi.org/10.20998/2522-9052. 2025.3.04

Svitlana Gadetska[1], Valeriy Dubnitskiy[2], Yuri Kushneruk[3], Yuriy Ponochovnyi[4], Alexander Khodyrev[2]

[1] Kharkiv National Automobile and Highway University, Kharkiv, Ukraine
[2] V. N. Karazin Kharkiv National University, Kharkiv, Ukraine
[3] Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine
[4] Poltava State Agrarian University, Poltava, Ukraine

DETERMINATION OF PARAMETER-LIMITED ESTIMATES OF EXTREME VALUE DISTRIBUTIONS AND MODELING OF CONDITIONS FOR THEIR OCCURRENCE USING STATGRAPHICS AND MATLAB

Abstract. Research objective is to estimate the parameters of extreme value distribution laws constrained by parameters using the maximum likelihood method and to model the conditions of their occurrence with STATGRAPHICS and MATLAB tools. **The subject of the study** includes Anscombe's quartet and the first (Gumbel), second (Fréchet), and third (Weibull) laws of extreme value distribution. **The research method** involves numerical methods for solving systems of equations obtained by the maximum likelihood method, as well as statistical modeling techniques. **The results of research** show, through the example of statistical data analysis of Anscombe's quartet, the necessity of verifying the correspondence between the physical content of the studied processes and the applicability of extreme value distribution laws for their analysis. The linear regression equation, which corresponds to all possible combinations of this quartet's data, is not optimal according to the criterion of maximum coefficient of determination. Using this criterion, it has been established that different data pairs have different nonlinear regression equations. Ignoring this fact may lead to errors in managing the processes they model. It has been shown that the data presented in Anscombe's quartet follow Gumbel's law, although the construction scheme of Anscombe's quartet does not correspond to the conditions of its occurrence. The limiting laws of extreme value distribution are presented in a form convenient for practical application in the design of engineering structures with constraints on the parameters of these distributions, caused by the specifics of the design. Parameter estites of extreme value distribution laws were performed using the maximum likelihood method. A numerical method for solving the corresponding systems of equations was considered. For modeling the scheme of extreme value occurrence, a matrix of random variables was generated, with twelve columns simulating monthly observations of a certain geophysical phenomenon, and one hundred rows representing a century of observations. The possible distributions of the elements of these matrices (initial distributions) were assumed as follows: double exponential distribution, Laplace distribution, lognormal distribution, Rayleigh distribution, normal distribution, Champernowne distribution, maximum extreme value distribution, minimum extreme value distribution, Birnbaum-Saunders distribution, Burr Type XII distribution, generalized extreme value distribution, inverse Gaussian distribution, Weibull distribution. The conditions of invariance of the extreme statistics distribution laws concerning the initial distributions were confirmed in only 12 out of 48 possible cases when modeled with STATGRAPHICS, and only in 5 out of 18 possible cases when modeled with MATLAB. The modeling results revealed a significant difference between the actual and theoretically possible extreme value distribution laws, which may be due to the peculiarities of the random number generation algorithms and the choice of the best-fit distributions used in STATGRAPHICS and MATLAB systems.

Keywords: Anscombe's quartet; extreme value distribution laws; maximum likelihood method; statistical modeling method.

Introduction

Formulation of the problem. Let us assume that we have been given n finite sequences of equally distributed random variables:

$$Y_i = (y_{i1}, y_{i2}, ..., y_{ij}, ..., y_{im}),\ i = \overline{1, n}\,. \qquad (1)$$

For each such sequence, we define the quantities:

$$x_i^{(mn)} = \min_i (y_{i1}, y_{i2}, ..., y_{ij}, ..., y_{im});$$
$$x_i^{(mx)} = \max_i (y_{i1}, y_{i2}, ..., y_{ij}, ..., y_{im}), \quad i = 1, n. \qquad (2)$$

These quantities are usually called the extremes of sequences of the form (1). Their probabilistic and statistical properties are studied in the theory of records and extreme value statistics [1, 2, 3, 19]. We will explore the differences in these approaches using a finite set $x_i^{(mx)}$. To determine the statistical properties of extreme values, the order of the elements in the equations (1), (2) does not matter. In record theory, the values $x_i^{(mx)}$ are arranged in ascending order, i.e., $x_i^{(mx)} < x_{i+1}^{(mx)}$, with the element $x_{i+1}^{(mx)}$ entering the sequence only if the following condition holds:

$$x_{i+1}^{(mx)} > x_i^{(mx)} (1 + \Delta)\,. \qquad (3)$$

For example, the rules of weightlifting competitions state: "A record is a lift that exceeds the previous record by a minimum of one (1) kg" [13]. Sequences of types (1), (2) are used to solve problems in fields such as meteorology, hydrology, actuarial mathematics, and when determining the load limits for engineering structures. Some of these problem-solving methods are described in detail in [3]. The main physical meaning of this approach is based on the assumption that a random variable can take any value with a non-zero probability. Let us assume that the distribution functions of the quantities (2) are known and equal to $\Phi^{(mn)}(x)$ and $\Phi^{(mx)}(x)$, respectively. It is shown in [4, p. 360] that the distribution function of the minimum value $\Phi^{(mn)}(x)$ and the distribution function of the maximum value $\Phi^{(mx)}(x)$ are related by the equation:

$$\Phi^{(mn)}(x) = 1 - \Phi^{(mx)}(-x)\,. \qquad (4)$$

The density functions of these distributions $\varphi^{(mn)}(x)$ and $\varphi^{(mx)}(x)$

$$\varphi^{(mn)}(x) = \varphi^{(mx)}(-x) . \qquad (5)$$

It should be noted that the application of each of these methods should be preceded by a thorough check of their correspondence to the physical essence of the studied process. To illustrate this point, let us consider the results of a statistical analysis of the so-called Anscombe's quartet. In 1973, F. Anscombe in [5] presented an artificial example, later named the "Anscombe's quartet". The data from this example is shown in Table 1. The graphs corresponding to this data are shown in Fig. 1–4.

Table 1 – **Numerical Data of Anscombe's Quartet**

Variables					
$x_1…x_3$	x_4	y_1	y_2	y_3	y_4
10	8	8,04	9,14	7,46	6,58
8	8	6,95	8,14	6,77	5,76
13	8	7,58	8,74	12,74	7,71
9	8	8,81	8,77	7,11	8,84
11	8	8,33	9,26	7,81	8,47
14	8	9,96	8,1	8,84	7,04
6	8	7,24	6,13	6,08	5,25
4	19	4,26	3,1	5,39	12,5
12	8	10,84	9,13	8,15	5,56
7	8	4,82	7,26	6,42	7,91
5	8	5,68	4,74	5,73	6,89

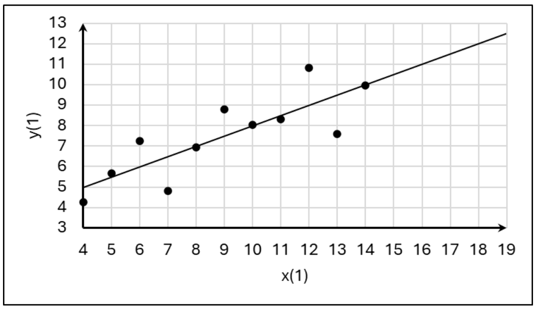

Fig. 1. Addiction $y_1 = f(x_1)$ for the Anscombe's Quartet

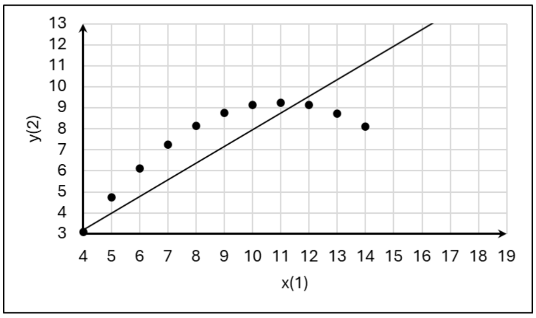

Fig. 2. Addiction $y_2 = f(x_1 = x_2)$ for the Anscombe's Quartet

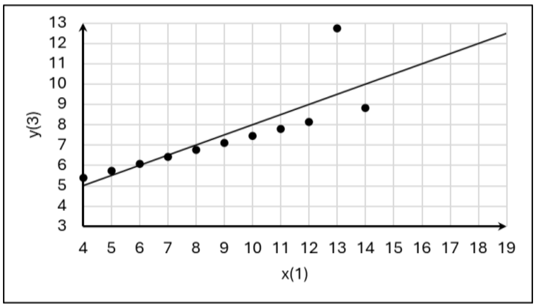

Fig. 3. Addiction $y_3 = f(x_1 = x_3)$ for the Anscombe's Quartet

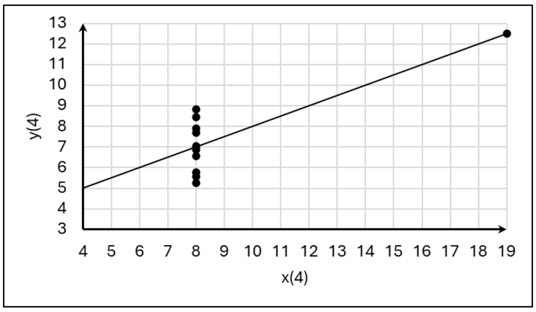

Fig. 4. Addiction $y_4 = f(x_1 = x_4)$ for the Anscombe's Quartet

The main statistical characteristics for the data presented in Anscombe's example are shown in Table 2. Matching data is highlighted in bold and italics.

Table 2 – **Main Statistical Characteristics of Anscombe's Quartet Data**

Statistical Characteristics	Variables					
	$x_1…x_3$	y_1	y_2	y_3	x_4	y_4
Mean (M)	**9**	**7,5**	**7,5**	**7,5**	**9**	**7,5**
Lower 95% bound of M	*6,77*	**6,14**	**6,14**	**6,14**	*6,77*	**6,14**
Upper 95% bound of M	11,23	**8,86**	**8,86**	**8,86**	11,29	**8,86**
Standard Deviation (SD)	*3,32*	**2,03**	**2,03**	**2,03**	*3,32*	**2,03**
Lower 95% bound of SD	*2,32*	**1,42**	**1,42**	**1,42**	*2,32*	**1,42**
Upper 95% bound of SD	*5,82*	**3,56**	**3,56**	**3,56**	*5,82*	**3,56**
Sample Range	10	6,58	6,16	7,35	11	7,25
Skewness Coefficient (As)	0	-0,06	-1,32	1,86	3,7	1,5
Lower 95% bound of As	-0,61	-0,68	-1,92	1,24	2,7	0,89
Upper 95% bound of As	0,61	0,54	-1,7	2,47	3,93	2,12
Kurtosis Coefficient (Ex)	-1,2	-0,53	0,84	4,38	11	3,15
Lower 95% bound of Ex	-1,98	-1,32	0,06	3,59	10,21	2,36
Upper 95% bound of Ex	-0,42	0,25	1,63	5,17	11,78	3,94
Coefficient of Variation (v)	**0,37**	**0,27**	**0,27**	**0,27**	**0,36**	**0,27**
Lower 95% bound of v	0,26	0,18	0,14	0,15	0,16	0,16
Upper 95% bound of v	0,48	0,35	0,4	0,38	0,57	0,37

From table 2, the main feature of Anscombe's quartet becomes clear. The numerical data was selected so that the mean values of the arguments $x_1…x_4$ coincide. The same applies to the mean values of the functions $y_1…y_4$. From Fig. 1... Fig. 4, it is clear that these functions look different, although in [1], the same regression equation was obtained for all the data. Table 3 presents the regression equation provided in [1] and the equations determined using the STATGRAPHICS software.

Table 3 – **Regression Equations Obtained from Anscombe's Quartet Data**

Model form	Model coefficients			Model characteristics		
	a	*b*	*c*	MAE	R^2(%)	R^2_{adf}(%), *d f*= 10
$y_1=a+bx_1$	3,00	0,500	-	0,837	66,654	62,949
$y_1=(a+b/x)^{-1}$	0,057	0,669	-	0,016	71,994	68,883
$y_2=a+bx_2$	3,00	0,500	-	0,967	66,624	62,916
$y_2=a+bx_1+cx_1^2$	-5,995	2,780	0,126	0,01	99,99	92,692
$y_3=a+bx_3$	3,00	0,499	-	0,716	66,632	62,924
$y_3=(a+b/x)^{-1}$	0,217	-0,008	-	0,005	87,869	86,869
$y_4=a+bx_1$ $y_4=a+bx_4$	3,001	0,499	-	0,902	66,671	62,968
$y_4=\sqrt{a+b\ln x}$	-204,10	122,384	-	12,74	78,97	76,63

In Table 3, it is accepted that R^2(%) is the coefficient of determination, and R^2_{adf} (%) is the coefficient of determination adjusted for the number of degrees of freedom. The obtained data show that the equations provided in [4] match, although they describe different processes, as seen from the graphs in Fig. 1...Fig. 4. The assessment of their statistical significance is presented in Table 3. Using these equations as mathematical models for describing real processes may lead to significant errors in managing them.

Before determining the correlation coefficients between the respective pairs of variables, the density of their distribution was calculated. The obtained results are shown in Tables 4 and 5. The STATGRAPHICS system was also used to obtain these results.

Table 4 – **Analytical form of the density distribution functions based on Anscombe's quartet data**

Distribution Type	Distribution density function
Uniform	$f(x)=\dfrac{1}{b-a}, \quad a \le x \le b$
Minimum Value	$f(x)=\dfrac{1}{\sigma}\exp\left[\left(\dfrac{x-\mu}{\sigma}\right)-\exp\left(\dfrac{x-\mu}{\sigma}\right)\right],$ $-\infty < x < \infty, \sigma > 0$
Maximum Value	$f(x)=\dfrac{1}{\sigma}\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right],$ $-\infty < x < \infty, \sigma > 0$

Table 5 – **Numerical values of the density distribution function parameters based on Anscombe's quartet data**

Variables	Distribution Type	Distribution parameters			
		a	*b*	*α*	*β*
$x_1 \ldots x_3$	Uniform	4,0	14,0	-	-
y_1	Uniform	4,26	10,84	-	-
y_2	Minimum Value	-	-	8,337	1,256
y_3	Maximum Value	-	-	6,698	1,257
y_4	Maximum Value	-	-	6,661	1,366

In this case, it is accepted that μ is the location parameter, and σ is the scale parameter. The density distribution functions provided in this table correspond to their definitions in the STATGRAPHICS system.

Due to the fact that the obtained distribution density differs from the normal distribution, in addition to Pearson's correlation coefficients, Spearman's rank correlation was also determined. The calculation results are shown in Table 6.

Table 6 – **Pearson and Spearman correlation coefficients between pairs of variables based on Anscombe's quartet data**

Correlation Coefficients	Pairs of Variables			
	x_1, y_1	x_1, y_2	x_1, y_3	x_1, y_4
Pearson's, *r*	0,8164	0,8162	0,8163	0,8165
Spearman's, *ρ*	0,8182	0,6909	0,9909	0,5

According to the authors of this report, attention should be paid to the appearance of extreme distribution densities (distributions of the maximum and minimum values). Therefore, studying the models of these distributions should be considered as an actual task.

Analysis of recent research and publications. According to the work [7], the following types of distributions are classified as extreme value distributions:

Type 1. Gumbel Distribution (first extreme value distribution law):

$$F_1(x)=\exp\left\{-\exp\left[-\left((x-\mu)/\sigma\right)\right]\right\}, -\infty < x < \infty . \quad (6)$$

Type 2. Fréchet Distribution (second extreme value distribution law):

$$F_2(x)=\begin{cases}\exp\left[-\left((x-\mu)/\sigma\right)^{-k}\right], & for\ x \ge \mu; \\ 0, & for\ x < \mu.\end{cases} \quad (7)$$

Type 3. Weibull Distribution (third extreme value distribution law):

$$F_3(x)=\begin{cases}\exp\left[-\left((\mu-x)/\sigma\right)^{k}\right], & for\ x < \mu, \\ 1, & for\ x \ge \mu.\end{cases} \quad (8)$$

In equations (6), (7), and (8), the parameters μ and σ have the same meaning as in Table 4, and k is the shape parameter. These equations represent the distribution functions for maximum values. According to [4], these distributions can also be applied to determine the distribution functions for minimum values. In [6, 7], a generalized extreme value distribution is proposed, with the function as follows:

$$F_g\left(x;\mu,\sigma,\xi\right)=\exp\left(-\left[1+\xi\left(\left(x-\mu\right)/\sigma\right)\right]^{-(1/\xi)}\right). \quad (9)$$

This equation is valid under the conditions:

$$-\infty < x \le (\mu-\sigma/\xi),\ if\ \ \xi<0, \quad (10)$$

or

$$(\mu-\sigma/\xi) \le x < \infty,\ if\ \ \xi>0. \quad (11)$$

In equations (9) … (11), ξ is the shape parameter. In [7], it is proven that if $\xi=0$ and $-\infty<x<\infty$, then:

$$F_g\left(x;\mu,\sigma,0\right)=\exp\left[-\exp\left(-\left(x-\mu\right)/\sigma\right)\right]. \quad (12)$$

These works also demonstrate how all main extreme value distribution laws, defined by equations (6) … (8), can be derived from equations (9) … (11). In [3], the ways for the extreme value distribution laws shown in (6) … (8) are provided, and they are presented in Tables 7 … 9.

Table 7 – **First Extreme Value Distribution**

Data Type	Distribution Function	Distribution Density
Maximum Values	$G_{mx}(x)=\exp\left\{-\exp\left[-\alpha\left(x-u\right)\right]\right\},\quad -\infty<x<\infty$	$g_{mx}(x)=\alpha\exp\left[-\alpha\left(x-u\right)\right]G_{mx}(x)$
Minimum Values	$G_{mn}(x)=1-\exp\left\{-\exp\left[\alpha\left(x-u\right)\right]\right\},\quad -\infty<x<\infty$	$g_{mn}(x)=\alpha\exp\left[\alpha\left(x-u\right)\right]\left[1-G_{mn}(x)\right]$

Table 8 – **Second Extreme Value Distribution**

Data Type	Distribution Function	Distribution Density
Maximum Values	$F_{mx}(x)=\exp\left[-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{-k}\right],\quad x\ge\varepsilon, u>\varepsilon, k>0$	$f_{mx}(x)=\frac{k}{u-\varepsilon}\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{-k-1}F_{mx}(x)$
Minimum Values	$F_{mn}(x)=1-\exp\left[-\left(\frac{\omega-x}{\omega-u}\right)^{-k}\right],\quad x\ge\omega, u>\omega, k>0$	$f_{mn}(x)=\frac{k}{\omega-u}\left(\frac{\omega-x}{\omega-u}\right)^{-k-1}\left[1-F_{mn}(x)\right]$

Table 9 – **Third Extreme Value Distribution**

Data Type	Distribution Function	Distribution Density
Maximum Values	$W_{mx}(x)=\exp\left[-\left(\frac{\omega-x}{\omega-u}\right)^{k}\right],\quad x>\omega, u>\omega, k>0$	$w_{mx}(x)=\frac{k}{\omega-u}\left(\frac{\omega-x}{\omega-u}\right)^{k-1}W_{mx}(x)$
Minimum Values	$W_{mn}(x)=1-\exp\left[-\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{k}\right],\quad x\ge\varepsilon, u>\varepsilon, k>0$	$w_{mn}(x)=\frac{k}{u-\varepsilon}\left(\frac{x-\varepsilon}{u-\varepsilon}\right)^{k-1}\left[1-F(x)\right]$

The methods for presenting extreme value distribution laws, as shown in these tables, have significant differences from equations (6) … (8), according to the authors. When using equations (6) … (8) for calculations related to determining the safe operation conditions of technical systems, it is necessary to know the parameters of location μ, scale σ, and shape k of the distribution. Presenting these distribution laws in the form shown in Tables 7–9 defines the relationships between parameters, with corresponding limitations and meaningful physical content. Therefore, in this report, we will refer to the distribution laws presented in Tables 7–9 as parameter-limited laws. Statistical modeling of extreme value distribution laws of the form (6) … (8) does not present significant difficulties and is detailed in [10–12]. In [3, 4], it is shown that, with a large number of observations, the distribution law of their extreme values coincides with one of the laws (6) … (8), regardless of the distribution laws of the initial sample. This statement is verified in this report using the STATGRAPHICS and MATLAB systems. A paradoxical result was obtained in [14]. A comparison of the results from the quasi-random number generators in the STATGRAPHICS and MATHCAD systems showed that, for any amount of data, the distributions and their parameters generated by these systems match only for the uniform distribution. The paradoxical result means that the specialized statistical system STATGRAPHICS, in most cases, does not recognize the match to the specified distribution law of random numbers generated by the MATHCAD system. It was established that STATGRAPHICS, in most cases, does not recognize the match to the specified distribution law of random numbers generated by itself. Therefore, to improve the reliability of the modeling results, it is recommended to diversify the software products used, as noted in [15].

Thus, the **goal** of the article is determining the maximum likelihood estimates of the parameters of the parameter-limited extreme value distribution laws and to model their occurrence conditions using STATGRAPHICS and MATLAB.

Main results

The method of maximum likelihood estimation was chosen to calculate the parameters of the extreme value distribution laws. The methodology for its application is detailed in [8, 10]. Without loss of generality, assume that for one of the random sequences defined by equations (1) and (2), the distribution function $\Phi(x;\alpha,\beta,\gamma)$ is known. The likelihood function in this case can be defined as:

$$\Lambda\left(x_1,x_2,...,x_n;\alpha,\beta,\gamma\right)=\prod_{i=1}^{n}\frac{\partial}{\partial x}\Phi\left(x_i;\alpha,\beta,\gamma\right). \quad (13)$$

According to [10], for more convenient transformations, we take the logarithm of (13):

$$L(x_1,x_2,...x_n;\alpha,\beta,\gamma)=\ln\Lambda\left(x_1,x_2,...,x_n;\alpha,\beta,\gamma\right)= \\ =\sum_{i=1}^{n}\ln\frac{\partial}{\partial x}\Phi\left(x_i;\alpha,\beta,\gamma\right). \quad (14)$$

To maximize (14) with respect to the parameters, we need to find the roots of the system:

$$\begin{cases}\frac{\partial}{\partial\alpha}L(x_1,x_2,...,x_n;\alpha,\beta,\gamma)=U(\alpha,\beta,\gamma)=0;\\ \frac{\partial}{\partial\beta}L(x_1,x_2,...,x_n;\alpha,\beta,\gamma)=V(\alpha,\beta,\gamma)=0;\\ \frac{\partial}{\partial\gamma}L(x_1,x_2,...,x_n;\alpha,\beta,\gamma)=W(\alpha,\beta,\gamma)=0.\end{cases} . \quad (15)$$

According to [16], this problem can be reduced to an optimization problem by finding the minimum of the function:

$$Q(\alpha,\beta,\lambda)=U^2(\alpha,\beta,\gamma)+ \\ +V^2(\alpha,\beta,\gamma)+W^2(\alpha,\beta,\gamma)\to\min. \quad (16)$$

Indeed, if $\left(\hat\alpha,\hat\beta,\hat\gamma\right)$ is a solution to problem (16), then:

$$U\left(\hat\alpha,\hat\beta,\hat\gamma\right)=V\left(\hat\alpha,\hat\beta,\hat\gamma\right)=W\left(\hat\alpha,\hat\beta,\hat\gamma\right)=0, \quad (17)$$

therefore, $$Q\left(\hat\alpha,\hat\beta,\hat\gamma\right)=0 \quad (18)$$

Since the function $Q(\alpha,\beta,\gamma)\ge 0$, it reaches its minimum value at point $\left(\hat\alpha,\hat\beta,\hat\gamma\right)$. Conversely, if the function $Q(\alpha,\beta,\gamma)$ reaches its minimum value at point $\left(\hat\alpha,\hat\beta,\hat\gamma\right)$ and $Q\left(\hat\alpha,\hat\beta,\hat\gamma\right)=0$, then equation (17) holds.

Therefore, $\left(\hat\alpha,\hat\beta,\hat\gamma\right)$ is a solution to problem (16). To find the solution to the system of equations (14), it should be noted that the function $Q(\alpha,\beta,\gamma)$ may have several local minima, while the global minimum is required. Therefore, it is recommended to search for the solution either for a justified initial approximation or for different initial values. The methods for determining the initial values necessary for solving equations (15) and (17) are discussed in [3]. The estimates $\left(\hat\alpha,\hat\beta,\hat\gamma\right)$, obtained from solving problem (16) will be normally distributed, unbiased, efficient, and capable [18].

To determine the confidence intervals for the obtained parameter estimates, the Fisher information matrix was calculated as shown in [10]:

$$V^{-1}=\begin{pmatrix}\sigma^2(\alpha) & \mathrm{cov}(\alpha,\beta) & \mathrm{cov}(\alpha,\gamma)\\ \mathrm{cov}(\alpha,\beta) & \sigma^2(\beta) & \mathrm{cov}(\beta,\gamma)\\ \mathrm{cov}(\alpha,\gamma) & \mathrm{cov}(\beta,\gamma) & \sigma^2(\gamma)\end{pmatrix}= \\ =\begin{pmatrix}-\frac{\partial^2 L}{\partial\alpha^2} & -\frac{\partial^2 L}{\partial\alpha\,\partial\beta} & -\frac{\partial^2 L}{\partial\alpha\,\partial\gamma}\\ -\frac{\partial^2 L}{\partial\beta\,\partial\alpha} & -\frac{\partial^2 L}{\partial\beta^2} & -\frac{\partial^2 L}{\partial\beta\,\partial\gamma}\\ -\frac{\partial^2 L}{\partial\alpha\,\partial\gamma} & -\frac{\partial^2 L}{\partial\beta\,\partial\gamma} & -\frac{\partial^2 L}{\partial\gamma^2}\end{pmatrix}^{-1}. \quad (19)$$

It is assumed that the matrix elements are calculated accordingly $(\alpha,\beta,\gamma)=\left(\hat\alpha,\hat\beta,\hat\gamma\right)$. The confidence intervals for the parameter estimates $\left(\hat\alpha,\hat\beta,\hat\gamma\right)$ in the distributions considered in this report were defined using relations of the form:

$$\hat\alpha-K_\alpha\sqrt{Var(\hat\alpha)}\le\alpha\le\hat\alpha+K_\alpha\sqrt{Var(\hat\alpha)}, \quad (20)$$

$$\hat\beta-K_\alpha\sqrt{Var(\hat\beta)}\le\beta\le\hat\beta+K_\alpha\sqrt{Var(\hat\beta)}, \quad (21)$$

$$\hat\gamma-K_\alpha\sqrt{Var(\hat\gamma)}\le\gamma\le\hat\gamma+K_\alpha\sqrt{Var(\hat\gamma)}. \quad (22)$$

The estimates of the variances of these parameters are located on the main diagonal of matrix (19). The other elements of this matrix correspond to the

covariance between the obtained parameter estimates. K_α is the quantile of the two-sided confidence probability at level α. A more detailed analysis of the properties of the Fisher information matrix is beyond the scope of this study.

The logarithm of the likelihood function for the first boundary distribution of maximum values is given by equation (23):

$$L\left(G_{mx}(x;\alpha,u)\right)= \\ =\sum_{i=1}^{n}\ln\left(\alpha e^{-\exp\left(u\alpha-\alpha x_i\right)-\alpha x_i+u\alpha}\right). \quad (23)$$

In equations (24), (25), and onwards, it is assumed that the parameter values are given. The system of equations, whose roots are the estimates of the parameters of this distribution, is presented in (24):

$$\begin{cases}U(G_{mx}(x;\alpha,u))=\frac{\partial}{\partial u}L(G_{mx}(x;\alpha,u))= \\ =n\alpha-\sum_{i=1}^{n}\alpha\exp\left(u\alpha-\alpha x_i\right)=0;\\ V(G_{mx}(x;\alpha,u))=\frac{\partial}{\partial\alpha}L(G_{mx}(x;\alpha,u))= \\ =\sum_{i=1}^{n}(x_i-u)\exp\left(u\alpha-\alpha x_i\right)-\sum_{i=1}^{n}x_i+n\left(u+\frac{1}{\alpha}\right)=0.\end{cases} . \quad (24)$$

Equations (25) … (27) make it possible to calculate the values of confidence intervals for the parameters of this distribution and the covariance between them:

$$\sigma^2(u) = -\frac{\partial^2}{\partial u^2} L(G_{mx}(x;\alpha,u)) = \alpha^2 \sum_{i=1}^{n} \exp(u\alpha - \alpha x_i). \quad (25)$$

$$\begin{aligned}\sigma^2(\alpha) &= -\frac{\partial^2}{\partial \alpha^2} L(G_{mx}(x;\alpha,u)) = \\ &= \sum_{i=1}^{n} (x_i - u)^2 \exp(u\alpha - \alpha x_i) + \frac{n}{\alpha^2};\end{aligned} \quad (26)$$

$$\begin{aligned}\operatorname{cov}(\alpha,u) &= -\frac{\partial^2}{\partial \alpha\, \partial u} L(G_{mx}(x;\alpha,u)) = \\ &= (-1)\sum_{i=1}^{n} \left[(\alpha x_i - u\alpha - 1)\exp(u\alpha - \alpha x_i) + 1 \right].\end{aligned} \quad (27)$$

The logarithm of the likelihood function for the second boundary distribution of maximum values is given by equation (28):

$$L(Fr_{mx}(x;\varepsilon,u,k)) = \\ = \sum_{i=1}^{n} \ln\left\{ k \exp\left[-\left(\frac{x_i - \varepsilon}{u - \varepsilon} \right)^{-k} \right] \left(\frac{x_i - \varepsilon}{u - \varepsilon} \right)^{-k} \Big/ (x_i - \varepsilon) \right\}. \quad (28)$$

The system of equations, whose roots correspond to the estimates of the parameters of this distribution, is presented in equation (29):

$$\begin{cases} U(Fr_{mx}(x;\varepsilon,u,k)) = \dfrac{\partial}{\partial \varepsilon} L(Fr_{mx}(x;\varepsilon,u,k)) = \\ = \displaystyle\sum_{i=1}^{n} \frac{k(x_i - u)\left((x_i - \varepsilon)/(u - \varepsilon)\right)^{-k}}{(u - \varepsilon)(x_i - \varepsilon)} + \\ + \displaystyle\sum_{i=1}^{n} \frac{kx_i - ku - u + \varepsilon}{(\varepsilon - u)(x_i - \varepsilon)} = 0; \\ V(Fr_{mx}(x;\varepsilon,u,k)) = \dfrac{\partial}{\partial u} L(Fr_{mx}(x;\varepsilon,u,k)) = \\ = \displaystyle\sum_{i=1}^{n} \frac{k\left((x_i - \varepsilon)/(u - \varepsilon)\right)^{-k}}{\varepsilon - u} + \frac{nk}{u - \varepsilon} = 0; \\ W(Fr_{mx}(x;\varepsilon,u,k)) = \dfrac{\partial}{\partial k} L(Fr_{mx}(x;\varepsilon,u,k)) = \\ = \displaystyle\sum_{i=1}^{n} \left((x_i - \varepsilon)/(u - \varepsilon)\right)^{-k} \ln\left((x_i - \varepsilon)/(u - \varepsilon)\right) - \\ - \displaystyle\sum_{i=1}^{n} \ln\left((x_i - \varepsilon)/(u - \varepsilon)\right) + \frac{n}{k} = 0. \end{cases} \quad (29)$$

The logarithm of the likelihood function for the third boundary distribution of maximum values is given by equation (30):

$$L(Wb_{mx}(x;\omega,u,k)) = \\ \sum_{i=1}^{n} \ln\left\{ k\left(\frac{x_i - \omega}{u - \omega} \right) \exp\left[(x_i - \omega)^k \right] (\omega - x_i)^{-1} \right\}. \quad (30)$$

The system of equations, whose roots are the estimates of the parameters of this distribution, is presented in equation (31):

$$\begin{cases} U(Wb_{mx}(x;\omega,u,k)) = \dfrac{\partial}{\partial \omega} L(Wb_{mx}(x;\omega,u,k)) = \\ = \displaystyle\sum_{i=1}^{n} \frac{k(x_i - u)\left((x_i - \omega)/(u - \omega)\right)^{k}}{(\omega - u)(x_i - \omega)} + \\ + \displaystyle\sum_{i=1}^{n} \frac{kx_i - ku + u - \omega}{(\omega - u)(x_i - \omega)} = 0; \\ V(Wb_{mx}(x;\omega,u,k)) = \dfrac{\partial}{\partial u} L(Wb_{mx}(x;\omega,u,k)) = \\ = \displaystyle\sum_{i=1}^{n} \frac{k\left((x_i - \omega)/(u - \omega)\right)}{u - \omega} + \frac{nk}{\omega - u} = 0; \\ W(Wb_{mx}(x;\omega,u,k)) = \dfrac{\partial}{\partial k} L(Wb_{mx}(x;\omega,u,k)) = \\ = \displaystyle\sum_{i=1}^{n} \ln\left((x_i - \omega)/(u - \omega)\right) + \frac{n}{k} - \\ - \displaystyle\sum_{i=1}^{n} \left((x_i - \omega)/(u - \omega)\right)^{k} \ln\left((x_i - \omega)/(u - \omega)\right) = 0. \end{cases} \quad (31)$$

The obtained results make it possible not only to calculate the estimates of the parameters of extreme value distribution laws but also to determine the confidence intervals for these estimates.

To model the emergence of extreme value distributions, the following method was applied:

1. In MS Excel, a matrix of uniformly distributed random numbers was generated:

$$R_{n \times m} = (r_{ij}),\ i = \overline{1,100};\ j = \overline{1,12}. \quad (32)$$

This matrix was defined as the initial matrix for all further work stages. The method of obtaining and saving it is shown in Fig. 5.

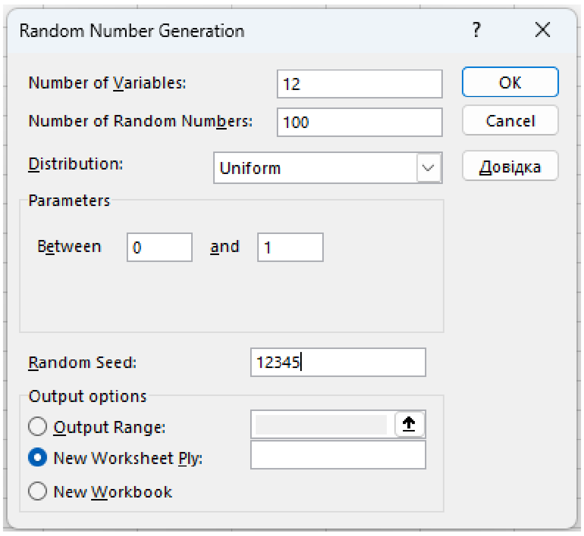

Fig. 5. Random Number Generation function dialog box

In [20] shows that the built-in MS Excel pseudo-random number generator, using the Random Number

Generation function, has good quality when generating numbers distributed over the unit interval according to the uniform distribution law.

The uniformly distributed random numbers were transformed into a matrix of initial data:

$$S = \left(Y \middle| X\right)\left(Y\right)_{n\times m} = \left(y_{ij}\right),\ i=\overline{1,100};\ j=\overline{1,12}\,. \quad (33)$$

In this case, the number of data in the i-th row corresponded to a series of monthly values of some hydrometeorological indicator in the i-th year and its minimum and maximum monthly values in that same year over a hundred years of observations. It should be noted that most of the main hydrometeorological indicators have observation series for longer time intervals. The types of distributions used to model the maximum and minimum observed values using STATGRAPHICS are listed in Table 10.

The properties of these distributions are discussed in detail in [17]. To model the distributions presented in Table 10, uniformly distributed random numbers obtained in step 2 were used. The rules for such transformations are shown in Table 11.

Table 10 – **Types of distributions used to model the maximum and minimum observed values using STATGRAPHICS**

Distribution Type and Notation	Distribution Function
Double exponential D(1)	$F(x;\mu,\lambda)=\exp\left[-\mu\exp(-\lambda x)\right]$
Laplace D(2)	$F(x;\mu,\lambda)=\begin{cases}0,5e^{\lambda(x-\mu)}, if\ x\le\mu;\\ 1-0,5e^{-\lambda(x-\mu)}, if\ x>\mu.\end{cases}$
Lognormal D(3)	$F(x;\mu,\sigma)\ =\Phi\left(\left(\text{n}\,x-\mu\right)/\sigma\right)^{*}$
Rayleigh D(4)	$F(x;a)=1-\exp\left(-x^2\Big/\left(2a^2\right)\right)$
Normal D(5)	$F(x;\mu,\sigma)=\Phi\left(\left(x-\mu\right)/\sigma\right)^{*}$
Champernowne D(6)	$F(x;a,\mu)=\left(2/\pi\right)\cdot arctg\exp\left[\alpha(x-\mu)\right]$
Extreme value D(7)	$F(x;\mu,\sigma)=\exp\left[-\exp\left(-\left(x-\mu\right)/\sigma\right)\right]$
Minimum value D(8)	$F(x;\mu,\sigma)=1-\exp\left[-\exp\left(\left(x-\mu\right)/\sigma\right)\right]$

*) Φ is standard normal distribution function

Table 11 – **Modeling relationships for transforming uniformly distributed random numbers**

Distribution Type	Distribution Parameters	Modeling Relationship	Parameter Estimates
Double Exponential D(1)	$m_x=\frac{1}{\lambda}(\ln\mu+\gamma);\quad \sigma_x=\frac{\pi}{\lambda\sqrt{6}}$	$x_{ij}=\frac{1}{\lambda}\left[\ln\mu-\ln(-\ln r_{ij})\right]$	$\tilde{\mu}=\exp\left(\pi\Big/\left(v_x\sqrt{6}\right)-\gamma\right);$ $\lambda=\pi\Big/\left(\tilde{s}_x\sqrt{6}\right);\quad v_x=s_x/\overline{x}.$
Laplace D(2)	$m_x=\mu;\quad \lambda=\sqrt{2}/\sigma_x$	$x_{ij}=\mu+\lambda^{-1}\cdot\ln\left(r_{ij}/\left(r_i+1\right)\right)$	$\tilde{\mu}=\overline{x};\quad \tilde{\lambda}=\sqrt{2}/s_x.$
Lognormal D(3)	$m_x=me^{a^2/2};\quad \sigma_x=me^{a^2/2}\sqrt{e^{a^2}-1}$	$x_{ij}=m\exp\left(a\,r_{ij}\right)$	$\tilde{m}=\overline{x}\Big/\sqrt{1+v_x^2};\quad \tilde{a}=\sqrt{\ln[1+v_x^2}$
Rayleigh D(4)	$m_x=a\sqrt{\pi/12};\quad \sigma_x=\sqrt{2-(\pi/2)}$	$x_{iji}=a\sqrt{-2\ln r_{ij}}$	$\tilde{a}=\overline{x}\sqrt{2/\pi}$
Normal D(5) *	–	–	–
Champernowne D(6)	$m_x=\mu;\quad \sigma_x=\pi/\left(2\alpha\right)$	$x_{ij}=\mu+\ln tg\left(\pi r_{ij}/2\right)$	$\tilde{\mu}=\overline{x};\quad \tilde{\alpha}=\pi/\left(2s_x\right).$
Extreme Value D(7)**	$m_x=\mu+\gamma\sigma;\quad D_x=(\gamma\pi)^2/6$	$x_{ij}=\mu-\ln\left(-\ln r_{ij}\right)$	$\hat{\mu}=\overline{x}-0,4501s_x;\overline{\lambda}=0,7797s_x$
Minimum Value D(8)	$\{\ m_x=\mu-\gamma\sigma;\quad D_x=(\gamma\pi)^2/6$	$x_{ij}=\mu+\ln\left(-\ln r_{ij}\right)$	$\hat{\mu}=\overline{x}-0,4501s_x;\hat{\sigma}=0,7797s_x$

*) Pseudo-random numbers with normal distribution were obtained using the built-in MS Excel subroutine.
**) γ = 0.57722, Euler's constant.

When testing the invariance of extreme value distribution laws relative to the initial distributions, it was assumed that all of them were obtained at values of mx=100 and σx=10, 20, 30. The accepted standard deviation values were denoted as S10, S20, S30.

To identify the obtained data in this case, the STATGRAPHICS system was used. It provides the opportunity to choose the best distribution law from a built-in list of alternative distributions, shown in Table 12. The best distribution was chosen to use the maximum likelihood criterion. In the list of distributions, the best one (with the highest likelihood) appears first. For example, the results of determining the distribution law of maximum values for a series generated by a double exponential distribution are presented in Table 12.

This table shows that since the D7 distribution from Table 10 matches the A1 distribution from Table 12, the model of the obtained distribution meets the theoretical assumptions. These assumptions suggest that extreme value distribution laws are independent of the initial distribution.

Table 13 presents the results of identifying the extreme value distribution laws obtained using statistical modeling methods. These results are compared with the extreme value distribution laws defined in Table 4.

The highlighted entries in this table represent cases where the STATGRAPHICS system confirmed that the condition of distribution law invariance for extreme statistics relative to the initial distributions was met. This condition was fulfilled in only 12 out of 48 cases.

A similar modeling process was performed using the MATLAB system. The distribution types used for this modeling are listed in Table 14 (data taken from MATLAB documentation [21]).

Table 12 – **Selection of the best distribution law of maximum values for the double exponential distribution and standard deviation S10**

Alternative Distributions and Notation	Log-Likelihood
Largest Extreme Value (A1)	-364,89
Inverse Gaussian A(2)	-366,556
Birnbaum-Saunders (A3)	-366,558
Lognormal A(4)	-366,571
Gamma (A5)	-367,137
Loglogistic (A6)	-368,306
Normal (A7)	-368,536
Logistic A(8)	-369,655
Uniform A(9)	-371,971
Laplace A(10)	-374,669
Weibull A(11)	-378,275
Minimum Extreme Value A(12)	-382,766
Exponential A(13)	-578,036
Pareto A(14)	-734,1

Table 13 – **Results of identifying extreme value distribution laws using STATGRAPHICS**

Initial Distribution Type	Distribution Function Classes					
	Maximum Value Distribution			Minimum Value Distribution		
	S10	S20	S30	S10	S20	S30
Double Exponential	***(A1)***	***(A1)***	***(A1)***	(A1)	(A1)	(A3)
Laplace	***(A1)***	***(A1)***	***(A1)***	***(A12)***	(A11)	A(7)
Lognormal	(A12)	(A12)	(A12)	A(1)	A(11)	A(1)
Rayleigh	A(7)	A(7)	A(5)	A(11)	A(11)	A(11)
Normal	A(2)	A(5)	A(5)	A(2)	A(5)	A(7)
Champernowne	***(A1)***	***(A1)***	***(A1)***	A(7)	A(11)	A(7)
Extreme Value	***(A1)***	***(A1)***	(A2)	A(1)	(A2)	(A2)
Minimum Value	(A12)	(A11)	(A11)	(A11)	***A(12)***	A(7)

Table 14– **Distribution types used for modeling maximum and minimum observed values in MATLAB (the parameters of all distributions are set to one)**

Distribution Type and Notation	Distribution density
Normal (M1)	$y=f(x;\mu,\sigma)=\left(\sigma\sqrt{2\pi}\right)^{-1}\cdot\exp\left(-(x-\mu)^2\Big/\left(2\sigma^2\right)\right),\quad x\in R$
Lognormal (M2)	$y=f(x;\mu,\sigma)=\left(x\sigma\sqrt{2\pi}\right)^{-1}\cdot\exp\left\{(ln\,x-\mu)^2-\Big/\left(2\sigma^2\right)\right\},\quad x>0$
Birnbaum-Saunders (M3)	$y=f(x;\beta,\gamma)=\left(\sqrt{2\pi}\right)^{-1}\cdot\exp\left\{-\left(\sqrt{x/\beta}-\sqrt{\beta/x}\right)^2\Big/\left(2\gamma^2\right)\right\}\left(\left(\sqrt{x/\beta}+\sqrt{\beta/x}\right)\Big/(2\gamma x)\right),\quad x>0$
Burr distribution, type XII (M4)	$f(x;a,c,k)=(kc/a)\cdot(x/a)^{c-1}\Big/\left(1+(x/a)^c\right)^{k+1},\quad x>0$
Extreme value (M5)	$y=f(x;\mu,\sigma)=\sigma^{-1}\exp((x-\mu)/\sigma)\cdot\exp(-\exp((x-\mu)/\sigma)),\ x\in R$
Generalized extreme value (M6)	$y=f(x;k,\mu,\sigma)=\sigma^{-1}\cdot\exp\left(-(1+k\cdot(x-\mu)/\sigma)^{1/2}\right)(1+k\cdot(x-\mu)/\sigma)^{-1-1/k},\quad k\neq 0,\ x\in R$
Inverse Gaussian (M7)	$y=f(x;\lambda,\mu)=\sqrt{\lambda\Big/\left(2\pi x^3\right)}\cdot\exp\left\{-\left(\lambda\Big/\left(2\mu^2x\right)\right)\cdot(x-\mu)^2\right\},\ x>0$
Rayleigh (M8)	$y=f(x;b)=\left(x/b^2\right)\cdot\exp\left(-x^2\Big/\left(2b^2\right)\right),\quad x\in R$
Weibull (M9)	$y=f(x;a,b)=(b/a)\cdot(x/a)^{b-1}\cdot exp\left(-(x/a)^b\right),\ if\ x\geq 0;\quad y=0,\ if\ x<0$

The results of identifying extreme value distribution laws using the MATLAB system are shown in Table 15.

Table 15– **Results of identifying extreme value distribution laws using the MATLAB system**

Initial distribution type	Max. value distribution	Min. value distribution
Normal (M1)	(M8)	(M8)
Lognormal (M2)	**(M5)**	**(M5)**
Birnbaum-Saunders (M3)	(M7)	**(M5)**
Burr, type XII (M4)	(M4)	(M8)
Extreme value (M5)	(M8)	(M8)
Generalized extreme value (M6)	(M6)	(M8)
Inverse Gaussian (M7)	(M3)	Uniform
Rayleigh (M8)	(M9)	(M8)
Weibull (M9)	**(M9)**	**(M5)**

According to the results of statistical modeling, the conditions of invariance were met in 5 out of 18 cases. The modeling result for the maximum value distribution for the initial Weibull distribution (M9) should also be recognized as valid, because, according to (8), the Weibull distribution is the third extreme value limiting distribution. The modeling results revealed a significant difference between actual and theoretically possible extreme value distribution laws. The authors believe this discrepancy may be due to the specifics of the random number generation algorithms and the selection of the best distributions used in the STATGRAPHICS and MATLAB systems.

Conclusions and the directions of further research

1. The need to verify the correspondence between the physical meaning of studied processes and the application of extreme value distribution laws is demonstrated using the statistical analysis of Anscombe's quartet data. The linear regression equation that fits all possible combinations of this quartet's data is not optimal based on the maximum coefficient of determination criterion. Using this criterion, different data pairs produce different non-linear regression

equations. Ignoring this can lead to errors in process management.

2. It is shown that the data in Anscombe's quartet follows Gumbel's law, although the construction of the quartet does not meet the conditions for this law's appearance.

3. Extreme value distribution limit laws are presented in a form suitable for practical use in engineering project design with restrictions on distribution parameters, which are caused by design features. The parameters of extreme value distributions were estimated using the maximum likelihood method. A numerical method for solving the corresponding systems of equations was considered.

4. To model the scheme for the appearance of extreme values, a matrix of random variables was generated. Twelve columns modeled monthly observations of a certain geophysical phenomenon, while 100 rows corresponded to a century-long observation period.

5. The possible distributions of these matrix elements (initial distributions) were as follows: double exponential distribution, Laplace distribution, lognormal distribution, Rayleigh distribution, normal distribution, Champernowne distribution, maximum value distribution, minimum value distribution, Birnbaum-Saunders distribution, Burr type XII distribution, generalized extreme value distribution, inverse Gaussian distribution, and Weibull distribution.

6. Invariance conditions for extreme statistics distribution laws relative to initial distributions were confirmed in only 12 out of 48 possible cases using STATGRAPHICS, and in only 5 out of 18 cases using MATLAB.

7. Modeling results revealed a significant difference between actual and theoretically possible extreme value distribution laws. The reason for this may be in the peculiarities of random number generation algorithms and the selection of the best distributions used in the STATGRAPHICS and MATLAB systems.

REFERENCES

1. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998), *Records*. Hoboken, NJ, USA: John Wiley & Sons, doi: https://doi.org/10.1002/9781118150412
2. Chandler, K. N. (1952), "The Distribution and Frequency of Record Values", *Journal of the Royal Statistical Society Series B*, Vol. 14, pp. 220–228. available at: https://www.scirp.org/reference/referencespapers?referenceid=69144
3. Gumbel, E. J. (1954), *Statistics of extremes*, Columbia Univ. Press, New-York, 400 p. available at: https://www.scirp.org/reference/referencespapers?referenceid=1650874
4. Sarhan, A. E., and Greenberg, B. G. (1962), *Contributions to Order Statistic*, John Wiley & Sons, New York – London, 482 p. doi: https://doi.org/10.1002/BIMJ.19640060221
5. Anscombe, F. J. (1973), "Graphs in statistical analysis", *The American Statistician*, vol. 27(1), 17, doi: https://doi.org/10.2307/2682899
6. Fabozzi, F. J. (1997), *Advances in fixed income valuation, modeling and risk management*, Associates New Hope, Pennsylvania, 391 p. available at: https://www.amazon.com/Advances-Valuation-Modeling-Management-Fabozzi/dp/1883249171
7. Kotz, S. and Nadarajah S. (2000), *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London. 185 p. available at: https://www.goodreads.com/book/show/3721370-extreme-value-distributions
8. Ramos, P. L., Louzada, F., Ramos, E. and Dey, S. (2019), "The Fréchet distribution: Estimation and application - An overview", *Journal of Statistics and Management Systems*, vol. 23(3), pp. 549–578, doi: https://doi.org/10.1080/09720510.2019.1645400
9. El-Sagheer, R. M. (2014), "Inferences for the Generalized Logistic Distribution Based on Record Statistics", *Intelligent Information Management*, vol. 06(04), pp. 171–182, doi: https://doi.org/10.4236/iim.2014.64018
10. (2024) *Life Data Analysis Reference*, ReliaSoft Corporation, South Eastside Loop Tucson, Arizona, USA, 632 p. available at: https://help.reliasoft.com/reference/life_data_analysis/pdfs/lda_ref.pdf
11. Devroye, L. (1986), *Non – uniform Variate Generation*, Springer-Verlag, Berlin, 859 p. available at: https://luc.devroye.org/rnbookindex.html
12. Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events*, Springer, Berlin Heidelberg, doi: https://doi.org/10.1007/978-3-642-33483-2
13. (2024), *IWF technical and competition rules & regulations*, available at: https://iwf.sport/wp-content/uploads/downloads/2024/08/IWF-TCRR-2024.pdf
14. Dubnitsky, V. J. and Protsenko, A. G. (2009), "A comparative analysis of random number generators in statgraphics and mathcad systems", *Radioelectronic and Computer Systems*, vol. 7, pp. 85–88. available at: http://dspace.library.khai.edu/xmlui/handle/123456789/3550
15. Frolov, V., Frolov, O. and Kharchenko, V. (2019), "Classification of Diversity for Dependable and Safe Computing", *CEUR 2362*, *Conference COLINS*, pp. 355–365, available at: https://ceur-ws.org/Vol-2362/paper32.pdf
16. Nocedal, J. and Wright, S. J. (2006), *Numerical Optimization*, Society for Industrial & Applied Mathematics, U.S., New York, 401 p. available at: https://www.amazon.com.be/-/en/Philip-Gill/dp/161197559X
17. Walk, Ch. (2007), *Handbook on Statistical Distributions for experimentalist*, *Internal Report*, SUF-PFY/ 96-01, Stockholm, 310 p., available at: https://www.stat.rice.edu/~dobelman/textfiles/DistributionsHandbook.pdf
18. Soong, T. T. (2004), *Fundamentals of probability and statistics for engineers*, John Wiley & Sons Ltd, Chichester, 391 p. available at: https://www.wiley.com/en-gb/Fundamentals+of+Probability+and+Statistics+for+Engineers-p-9780470868157
19. Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York. doi: https://doi.org/10.1007/978-1-4612-5449-2
20. Dubnytskyi, V., Kobyli, A., Kobylin, O., Kushneruk, Yu. and Khodyrev, O. (2024), "Evaluation of the quality of the uniformly distributed random number generator built into MS EXCEL", *Information Processing Systems*, issue 3(178), pp.17–26, doi: https://doi.org/10.30748/soi.2024.178.02
21. (2024), *Continuous Distributions. Compute, fit, or generate samples from real-valued distributions*. available at: https://www.mathworks.com/help/releases/R2023b/stats/continuous-distributions.html

Received (Надійшла) 12.02.2025
Accepted for publication (Прийнята до друку) 21.05.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Гадецька Світлана Вікторівна – кандидат фізико-математичних наук, доцент, доцент кафедри вищої математики, Харківський національний автомобільно-дорожній університет, Харків, Україна;
Svitlana Gadetska – PhD in Physics and Mathematics, Associate Professor, Associate Professor of Department of Higher Mathematics of Kharkiv National Automobile and Highway University, Kharkiv, Ukraine;
e-mail: svgadetska@ukr.net; ORCID Author ID: https://orcid.org/0000-0002-9125-2363;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57203515764.

Дубницький Валерій Юрійович – кандидат технічних наук, старший науковий співробітник, старший науковий співробітник Харківського національного університету ім. В. Н. Каразіна, Харків, Україна;
Valeriy Dubnitskiy – Candidate of Technical Sciences, Senior Researcher, Senior Researcher of V. N. Karazin Kharkiv National University, Kharkiv, Ukraine;
e-mail: dubnitskiy@gmail.com; ORCID Author ID: https://orcid.org/0000-0003-1924-4104;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=58698004400.

Кушнерук Юрій Іонович – кандидат технічних наук, доцент, старший викладач Інституту цивільної авіації Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна;
Yuri Kushneruk – Candidate of Technical Sciences, Associate Professor, Senior Lecturer of Civil Aviation Institute of Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine;
e-mail: kyshneryk_ui@ukr.net; ORCID Author ID: https://orcid.org/0000-0001-5844-7137;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=36572023100.

Поночовний Юрій Леонідович – доктор технічних наук, професор, професор кафедри інформаційних систем та технологій, Полтавський державний аграрний університет, Полтава, Україна;
Yuriy Ponochovnyi – Doctor of Technical Sciences, Professor, Professor of the Department of Information Systems and Technologies, Poltava State Agrarian University, Poltava, Ukraine;
e-mail: yuriy.ponch@gmail.com; ORCID Author ID: https://orcid.org/0000-0002-6856-2013;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=56446990700.

Ходирєв Олександр Іванович – старший викладач Харківського навчально-наукового інституту "Каразінський банківський інститут" Харківського національного університету ім. В. Н. Каразіна, Харків, Україна;
Alexander Khodyrev – Senior Lecturer of "Karazin Banking Institute" of V. N. Karazin Kharkiv National University, Kharkiv, Ukraine;
e-mail: khodyrevmjk3758@gmail.com; ORCID Author ID: https://orcid.org/0000-0001-9871-9440;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=56446990700.

Визначення оцінок обмежених за параметрами законів розподілу екстремальних величин та моделювання умов їх виникнення засобами STATGRAPHICS та MATLAB

С. В. Гадецька, В. Ю. Дубницький, Ю. І. Кушнерук, Ю. Л. Поночовний, О. І. Ходирєв

Анотація. Мета дослідження: визначення методом максимуму правдоподібності оцінок параметрів законів розподілу екстремальних величин, обмежених за параметрами та моделювання умов їх виникнення засобами STATGRAPHICS і MATLAB. **Предмет дослідження:** квартет Енскомба, перший (Гумбеля), другий (Фреше), третій (Вейбулла) закони розподілу екстремальних значень. **Метод дослідження**: чисельні методи розв'язання систем рівнянь, отриманих за методом максимуму правдоподібності, методи статистичного моделювання. **Отримані результати:** на прикладі статистичного аналізу даних квартету Енскомба показана необхідність перевірки відповідності фізичного змісту процесів, що вивчаються, можливостям застосування для їх аналізу законів розподілу екстремальних значень. Рівняння лінійної регресії, яке відповідає всім можливим комбінаціям даних цього квадрата, не є оптимальним по критерію максимуму коефіцієнта детермінації. При застосуванні цього критерію встановлено, що різні пари даних мають різні рівняння нелінійної регресії. Ігнорування цієї обставини може привести до помилок в управлінні процесами, які вони моделюють. Показано, що дані, які наведено в квартеті Енскомба, розподілені згідно із законом Гумбеля, хоча схема побудови квартету Енскомба не відповідає умовам його появи. Граничні закони розподілу екстремальних значень представлено у вигляді, зручному для застосування в практиці проєктування інженерних споруд з обмеженнями на параметри цих розподілів, які викликано особливостями проєктування. Оцінки параметрів законів розподілу екстремальних значень виконували за методом максимуму правдоподібності. Розглянуто чисельний метод розв'язання відповідних систем рівнянь. Для моделювання схеми появи екстремальних значень отримували матрицю випадкових величин, дванадцять стовпців якої моделювали щомісячні реєстрації спостережень деякого геофізичного явища, сто рядків матриці відповідали сторічному періоду спостережень. Можливі розподіли елементів цих матриць (початкові розподіли) прийняті наступними: подвійний показниковий розподіл, розподіл Лапласа, логнормальний розподіл, розподіл Релея, нормальний розподіл, розподіл Чампернаума, розподіл найбільшого значення, розподіл найменшого значення, розподіл Бірнбаума-Сандерса, розподіл Бурра тип XII, узагальнений розподіл екстремальних величин, обернений розподіл Гаусса, розподіл Вейбулла. Умови інваріантності законів розподілу екстремальних статистик відносно до початкових розподілів при моделюванні засобами STATGRAPHICS підтверджено лише в 12 випадках з 48 можливих, при моделюванні засобами MATLAB – тільки в 5 випадках з 18 можливих. За результатами моделювання встановлена істотна різниця між фактичними та теоретично можливими законами розподілів екстремальних величин, причина якої може бути в особливостях алгоритмів моделювання випадкових чисел та вибору найкращих розподілів, які використано в системах STATGRAPHICS та MATLAB.

Ключові слова: квартет Енскомба; закони розподілу екстремальних значень; метод максимуму правдоподібності; метод статистичного моделювання.