Mathematical Justification for the Information System for Predicting Dangerous Situation

Valentin G. Degtyarev
Emperor Alexander I
St. Petersburg State
Transport University
vdegt@list.ru

Lidiya A. Kuharenko
Emperor Alexander I
St. Petersburg State
Transport University

Ruslan S. Kudarov
Emperor Alexander I
St. Petersburg State
Transport University
r.s.kudarov@gmail.com

Rustem S. Kudarov
Emperor Alexander I
St. Petersburg State
Transport University
r.s.kudarov@gmail.com

Abstract

The work focuses on some of the basics of the theory and practice of applying extreme value statistics and the basics of the information model for predicting dangerous states. The task of predicting dangerous situations is presented and solved on the basis of a statistical description of the characteristics of the object and taking into account the probabilistic characteristics of the extreme external conditions of the object.

1 Introduction. Sustainability and the area of attraction of some distributions

Let's present known distribution properties, using perhaps unusual terms in this context. The composition (or distribution of the sum) of two independent normal distributions leads to a normal distribution. And in general, the sum of any number of normally distributed random values is a normally distributed random value. This property expressed by the solution of the functional equation:[1]

In: A. Khomonenko, B. Sokolov, K. Ivanova (eds.): Selected Papers of the Models and Methods of Information Systems Research Workshop, St. Petersburg, Russia, 4-5 Dec. 2019, published at http://ceur-ws.org.

$$F_n(x) = F(a_n x + b_n) \qquad (1)$$

where - $F_n(x)$ the distribution of the amount of independent normally distributed random values, the corresponding parameters of scale and distribution shift. In this sense, we can talk about the stability (or, if you will, conformity) of the form of normal law regarding the procedure of composition of distributions.

The various distributions, which are likely to be of some known law of distribution, form the area of attraction of this law. In this sense, all distributions that meet the conditions of the Central Limit Theoreme (e.g., Lapunov's conditions) are part of the area of attraction of normal law.

The second property is that for an arbitrary infinite sequence of distributions that meet the conditions of the Central Limit Theorem, the amount of asymptomatically normal in the sense of convergence by probability measure. In other words, many distributions that meet the conditions of the Central Limit Theoreme (e.g. Lapunov's conditions) are part of the scope of the normal law. The simplicity and ease of applying normal law in applications largely determined by these properties.

Naturally, the question is whether there are distributions and procedures with similar properties. For example, it is obvious that an exponential law is sustainable when selecting a minimum for an arbitrary number of random amounts subject to that law. However, the details of the properties of this law, which is a private case of gamma-distribution, discussed below.

In the task of modeling extreme situations, important to note that the two-parametric double exponential law has similar properties: 1) the stability of the form relative to the distribution of extreme value from any number /

equally distributed values; 2) is the limit for a certain class, the "3" of the original distributions in the sense of convergence by probability.

2 About abnormal sampling members

When choosing a probability model based on statistics, the question naturally arises: what to do when there are abnormal observations? How do you distinguish anomalous sampling elements from possible but unlikely elements? There are known different approaches.
Let the sample on the basis of which the appropriate conclusion should be drawn, $\{x_n\}_{n=1}^{n=N}$, $\{x_n\}_{n=1}^{n=N}$- the same sample, the elements of which$x_{(n_0)}$are located in a non-decreasing orderand $x_{(1)}$ - the sample element, suspicious of the anomaly, in particular, the smallest or the largest $x_{(N)}$. Here are the well-known rules of such filtering.
1. Reject a suspicious element $x_{(n_0)}$ as an anomalous, if $|x_{(n_0)} - \bar{x}| > c\sigma$, where $\bar{x}$ - the average value of distribution σ - its standard deviation, with c - a pre-selected constant. Otherwise, if $|x_{(n_0)} - \bar{x}| < c\sigma$, the observation is not rejected.
Strictly speaking, the constant value is determined by the level of significance of some specially chosen criterion, and therefore in the decision-making possible errors known in mathematical statistics 1 and 2 genus.
2. If the anomaly of the observation is confirmed, it is necessary to create an adequate model to plan observations for the use of factor dispersion or, if possible, regression analysis. You can't limit yourself to a one-dimensional sample.
3. As we can see, this rule of abnormality testing, proposed in n 1, solves the issue only at some, satisfying researcher, the level of significance of one or another criterion. Objectively speaking, necessary to recognize the possibility of implementing elements of the sample, anomalous in the sense of p.1, in the same observational conditions.

3 Distribution of extreme values

In any case, it is advisable, on the basis of the final number of one-dimensional samples, located in a non-decreasing order, to turn to the distribution of statistics suspicious of abnormality. We are talking about statistics of extreme values.
The sample is considered $\{x_n\}_{n=1}^{n=N}$, with equally distributed by law with the function of distribution *F(x)* elements x_i,$i = 1,2, \dots, N$. The extreme elements of the sample represented by statisticians:

$$x_{\max} = \max_{1<i<N} x_i, \quad x_{\min} = \min_{1<i<N} x_i . \qquad (2)$$

These statistics have distribution functions:

$$F_{\max N}(x) = (F(x))^N, \quad F_{\min N}(x) = 1-(1-F(x))^N. \qquad (3)$$

The same statistics for the case of independent sampling elements with arbitrary independent distributions:

$$F_{\max N}(x) = \prod_{i=1}^{N}(F_i(x)),$$

$$F_{\min N}(x) = 1-\prod_{i=1}^{N}(1-F_i(x)). \qquad (4)$$

The possibilities of applying these statistics to relevant estimates are significantly limited by the fact that known distributions from which the samples are derived are assumed. Universal in this sense is the approach based on the theorem of B.V. Gnedenko [Gne43] . Here's the following formulation.
Gnedenko's theorem. If the random value distributed by law (4) has a limit distribution, it distributed under one of the following three laws after the corresponding rationing. For maximums:

$$F_I = \exp(-\exp(-x)), \quad -\infty < x < +\infty.$$

$$F_{II}(x) = \begin{cases} \exp(-(-x)^{\alpha}, & x \le 0, \alpha > 0, \\ 1, & x > 0. \end{cases}$$

$$F_{III}(x) = \begin{cases} 0, & x \le 0, \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0. \end{cases}$$

These distributions called extreme distributions of the type I, II and III respectively.
The distribution of the first type, under some additional conditions, caused, for example, by the task of ejecting the values of the stationary random process. The distribution of the second type, as we can see, is the distribution of Weibull, in particular, convenient for describing the strength of the object on the break.
The first distribution, called the Double Exponential Law, describes a wide class of original distributions that make up the area of its attraction. The second and third distributions refer to random values limited to the left and right, respectively. These distributions can brought to the first [Dav79].

4 Some properties of double exponential law

The annexes look at two forms of recording the double exponential law distribution function: for maximum values

$F(x;q,\alpha)=1-\exp(-\exp(\alpha(x-q))$ (5)

and minimum values

$F(x;q,\alpha)=\exp(-\exp(-\alpha(x-q)).$ (6)

One feature is given to another replacement*x*on -*x*. Here *q*- the shear parameter *α*- the scale parameter. The parameters are determined by the first two moments of distribution: mathematical expectation and standard deviation:

$$q=m_1+\frac{\gamma}{\alpha},\quad \alpha=\frac{\pi}{\sigma}, \qquad (7)$$

where*γ* - Euler's constant. The moments of the higher order of this distribution have been received. Analysis of the moments in particular.points out that it is unacceptable to replace these distributions for the sake of simplicity with a normal law. This form obtained accordingly for maximum and minimum values by substitution by / (for simplicity do not change the designations):

$$\begin{aligned}&F(x)=1-\exp(-\exp x),\\&F(x)=\exp(-\exp(-x))\end{aligned} \qquad (8)$$

Distributions (6) have, as mentioned above, known properties: 1) the stability of the form relative to the transition to the distribution of extreme value from any number /same distributions respectively (5) or (6); 2) Distribution is the limit (after appropriate rationing) for a certain class of original distributions in the sense of convergence by probability. There's a little bit more to do with that. The first means that each function (8) is the solution to the functional equation

$$F_n(x)=F(a_nx+b_n) \qquad (9)$$

where $F_n(x)$ - distribution of extreme value from the same distributions (8), corresponding parameters of scale and distribution shift (8).

For the convenience of using the second property, let us give a sufficient condition of belonging to the distribution of the area of attraction of the first limit law $F_I(x)$ [Gne43], [Dav79]. . It is enough that the distribution function, at least for large modules, has the appearance:

$$F(x)=1-\exp(-h(x)), \qquad (10)$$

where the function *h(x)* increases monotonously and indefinitely. This condition is satisfied with symmetrical distributions of exponential type, including - normal law.

5 Extreme Situation Forecast

The task of forecasting strength, taking into account external conditions, can lead to the composition of the two distributions in the following sense.

The strength reserve characterized by a different characteristic of the strength of the study object X and the real load (in the same units of measurement) that occurs during its operation Y. If $X \leq Y$, the object is in working or extreme condition, otherwise, if $X < Y$, the object does not work (suffers failure).

Thus, we will be interested for an arbitrary probability event $X-Y<x$, or composition of distributions $X+(-Y)$.

It is clear that the margin of safety, for the sake of simplicity in practice, determined unequivocally by the ratio of these values X/Y (at best - the ratio of their average values) does not stand up to criticism, as a characteristic of the strength of the object. In this sense, it is natural to turn to the marginal distributions of statistics (2) taking into account the first two points of the original distributions.

Suppose each of the component of the sum $X_1=X, X_2=-Y$ distributed under a double exponential law with the appropriate parameters:

$$F(x;q_1,\alpha_1)=1-\exp(-\exp(\alpha_1(x-q_1))) \qquad (11)$$

$$F(x;q_2,\alpha_2)=1-\exp(-\exp(\alpha_2(x-q_2))) \qquad (12)$$

and let's turn to the distribution of the amount of independent random amounts x_1+x_2. Note at once that the distribution of this amount of distributions (11) and (12) does not retain the original distribution form. This complicates the technique in comparison with the simplest model based on normal distribution of characteristics x, y which is incorrect here.

As you know, it is possible to determine the distribution of the amount of independent values by the characteristic function expressed through characteristic functions component of composition: / In relation to distributions we receive (6):

$$\varphi(t)=e^{i(q_1+q_2)t}\Gamma(1+\frac{it}{\alpha_1})\Gamma(1+\frac{it}{\alpha_2}) \qquad (13)$$

Where $\Gamma(t)$ – gamma function defined by Euler's integrative:

$$\Gamma(z)=\int_0^\infty e^{-x}x^{z-1}dx,\quad \operatorname{Re} z>0. \qquad (14)$$

However, the reverse transition from a characteristic function/ to an appropriate function of the distribution of the amount is very problematic precisely because the law

(8) does not retain forms in the composition of distributions (not stable), as will be shown below. Therefore, we have directly turned to the roll-out of distribution densities and the known formula for the function of distributing the amount x_1+x_2 that can be presented in the form of:

$$H(x;q_1,q_2,\alpha_1,\alpha_2)=\int_{-\infty}^{+\infty}F(z-x,q_2,\alpha_2)dF(x,q_1,\alpha_1) \quad (15)$$

The immediate calculation of this formula leads to results that can formulated in the form of the following theorem.
Theorem 1. The composition of two independent distributions (9) and (10) has a distribution $q_1, q_2, \alpha_1, \alpha_2$:

$$\alpha=\frac{\alpha_2}{\alpha_2},\ a=\frac{1}{\alpha_2},\ b=q_2-q_1.$$

Theorem 2. The function $H_\alpha(ax+b)$ if $a=1$, $b=0$ has a view:

$$H_\alpha(x)=1-\int_0^\infty \exp(-e^x u^{-\alpha}-u)du \quad (16)$$

Private cases: $H_0(x)=1-\exp(-e^x)$.

$$H_1(x)=1-\int_0^\infty \exp(-e^x u^{-1}-u)du.$$

Thus, the prediction of dangerous states of an object can based on the distribution:

$$H_\alpha(ax+b)=1-\int_0^\infty \exp(-e^{ax+b}u^{-\alpha}-u)du \quad (17)$$

Where is $\alpha=\frac{\alpha_2}{\alpha_2}, a=\frac{1}{\alpha_2}, b=q_2-q_1$.

A special standard single-parametric family of distributions, generated by the composition (1): is introduced:

$$F_1(x), F_2(x):$$

$$F(z)=\int_{-\infty}^{+\infty}F_1(z-x)dF_2(x) \quad (18)$$

distribution for which function value tables $H_\alpha(x)$ depending on α, x, a snippet of which is below, should be available. A reference to these existing tables does not mean a complete solution to this issue. We need such tables in a form that corresponds to modern information tools, and are more convenient for the user, who does not have special training.

In general, $\alpha\epsilon[0,+\infty)$. But tables are sufficient for the proposed methodology, $\alpha\epsilon[0,1]$. This follows from the commutativeness of the bundle of distributions. However, the distribution of the option for all $\alpha\epsilon[0,+\infty)$: is associated with assigning indexes 1.2 for the original distributions (11), (12). Therefore, it is preferable for a specially untrained user to complete limit distribution tables $H_\alpha(x), \alpha\epsilon[0,+\infty)$.

6 Table of probabilities of dangerous states

The dangerous state here discussed as approximating the object parameter to an unacceptable value, taking into account external influences during its operation. The operational characteristics of the system are known to be determined by the ratio of different characteristics, among which it is necessary to distinguish 1) the own characteristics of the object, and 2) the characteristics of its operating conditions. Considered, a pair of one-dimensional characteristics 1 and 2 preferably of the same dimension: strength and active load. Their attitude $X_1 < X_2, X_1 = X_2$ or $X_1 > X_2$ really determines the assessment of the quality of the system. Simple Risk Assessment Procedure is based on the probability of the probability of the different characteristics $X_1 - X_2$. In this case, the distribution function $F(x)$ of the relevant criterion $t = X_1 - X_2$ determined by the composition of the respective distributions.

$$K_\alpha(x)=1-\int_0^\infty \exp(-e^x u^{-\alpha}-u)du. \quad (21)$$

Normalized distribution (4) linear transformation with shear and scale parameters (a and b) determines the desired probability distribution (19).
The risk of an error associated with accepting a satisfactory criterion t assessment result justifies the choice [Dav79] of a relatively difficult one for the distribution of extreme performance distributions X_1, X_2. This is a double exponential law, depending on two parameters a, q, which in turn are clearly determined by the first two moments of the respective distributions. This is an interest in the worst-case scenario: the least strength is the greatest load. This complexity is due to the fact that the distribution composition (1) does not retain the shape of the distribution component X_1, X_2. In the case when the distributions of independent random values are subject to the limit distribution of extreme values of type 1, their

composition results after the corresponding rationing $z = ax + b$ to distribution [Kuk75], [Deg03], [Kuk96], depending on the shape setting α:

$$K_\alpha(ax+b) = 1 - \int_0^\infty \exp(-e^x u^{-\alpha} - u)du \quad (19)$$

Here are the distribution options $\alpha; a, b$ are uniquely determined by the parameters $\alpha_1, q_1;\ \alpha_2, q_2$ initial distributions and $\alpha = \frac{\alpha_2}{\alpha_1}$. At the same time α and a not independent, which presents some difficulties for the distribution application (2) in the task of assessing the probability of dangerous states. Overcoming these difficulties in practice provided by the appropriate function tables $H_\alpha(x)$:

$$H_\alpha(x) = \int_0^\infty \exp(-e^x u^{-\alpha} - u)du \quad (20)$$

Developed [Kuk96], [Kry17] Detailed feature table $H_\alpha(x)$: and a probability calculation methodology based on this table. Below is a snippet of the abbreviated version of this table (table.1).

Table1 Distribution table

x	α = 0,1	α = 0,3	α = 0,5	α = 0,7	α = 0,9	α = 1
-4	0.929	0.93	0.931	0.932	0.933	0.933
-3,5	0.897	0.898	0.9	0.901	0.903	0.903
-3	0.853	0.855	0.857	0.859	0.861	0.862
-2,5	0.794	0.797	0.8	0.802	0.805	0.806
-2	0.719	0.722	0.725	0.729	0.732	0.734
-1,5	0.625	0.63	0.634	0.638	0.642	0.643
-1	0.517	0.521	0.526	0.53	0.535	0.537
-0,5	0.398	0.403	0.408	0.412	0.417	0.42
0	0.28	0.284	0.289	0.293	0.298	0.3

Pictured (Figure 1) shows the function $H_\alpha(x)$: for parameter values $\alpha, \alpha\epsilon[0;1]$ with long intervals $\Delta\alpha = 0{,}2$. Given the properties of the commutativeness of the roll-up operation, as well as the fact that determined by the ratio of input variances $\alpha = \sqrt{\frac{\mu_1}{\mu_2}}$, we've developed a table $\alpha \in [0;1]$ is sufficient to calculate probabilities by formula (4) at all values α.

However, for values $\alpha, \alpha > 1$, the proposed method of assessing probabilities is somewhat complicated. This follows from the fact that the parameters of the shift and the scale in formula (4), defined by the first two moments of input distributions, depend on the order of their chosen statistics: $X_1 - X_2$ or $X_2 - X_1$. This complication is surmountable after selecting the appropriate table column and recalculating the scale and shift parameters a, b.

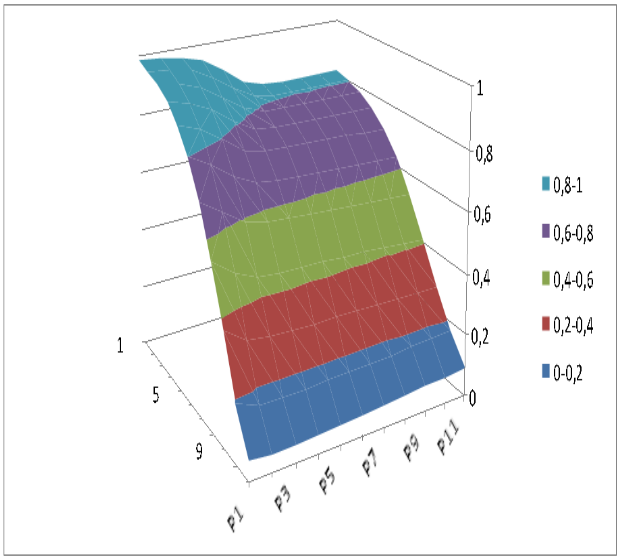

Fig. 1 Schedule Distributions

References

[Gne43] B. V. Gnedenko (1943) Sur la distribution limite du terme maximum d'uneseriealeatoiire//Ann. Math. Sttist. 44.

[Gum65] E. Gumbel (1965) Extreme stats. M.: The World.

[Dav79] G. David (1979) Serial statistics.-M.: Science.

[Kra60]. G. Kramer, Leadbetter (1960) Stationary random processes. Properties of selective features and their applications. M.: Mir.

[Kuk75]. L.A. Kukharenko (1975) Study of electrical strength insulation Monolith / Informelectro 30, sulfur. Electrical materials, 9.

[Deg03]. V.G. Degtyarev, L.A. Kukharenko, V.A. Khodakovsky (2003) Stochastic transport system tasks/ PSUPS Herald. Issue 1 - St. Petersburg.

[Kuk96]. L.A. Kukharenko (1996) On the composition of distributions of extremes. Proceedings of the international scientific and methodological conference Mathematics at the university. St. Petersburg.

[Kuk02]. L.A. Kukharenko, T.S. Moiseenko (2002) About one model of extreme situations. Materials of the scientific and practical conference Topical economic problems. St. Petersburg, IBI.

[Kuk10]. L.A. Kukharenko, T.S. Moiseenko, O.I. Januszewski (2010) Stochastic models in the mathematical training of an engineer. Materials of the Scientific and Methodical Conference Problems of Mathematical Training in Engineering Education. St. Petersburg, PSUPS.

[Kry17]. A.S. Kryukov, V.A. Xenophontova, L.A Kukharenko (2017) Table of probabilities of dangerous states. Mathematics at university and at school. A collection of works of the national scientific and methodological conference. Pskov. 2017. p. 112-113. 96-99

[Kho08]. V.A. Khodakovsky, L.A. Kukharenko, L.P. Pirozerskaya (2008) Probability theory. Ch.2: study. allowance - St. Petersburg, PSUPS.

[Kuk11] L.A. Kukharenko, E.V. Runev, O.I. Januszewski (2011) About strength and variance. Proceedings of the International Scientific and Methodical Conference Mathematics e University. St. Petersburg, PSUPS.

[Bor15] T.P. Borovkova, L.A. Kukharenko, E.V. Runev (2015) About the mathematical model of quality control. Proceedings of the International Scientific and Methodical Conference Mathematics e University and School. St. Petersburg, PSUPS.

[Xen18]. V.A. Xenophontova, L.A. Kukharenka, E.V. Runev (2018) About the moments and instant characteristics of one limiting law of extreme situations. System analysis and analytics. No5 (6). p. 15-23.

[Kuk16]. L.A.Kukharenko, V.A. Xenophontova, A.S. Kryukov (2016) A mathematical model of predicting dangerous situations. A collection of works IV international scientific and methodological conference Problems of mathematical and naturally scientific training in engineering education. St. Petersburg, PSUPS.