5-2015

Modeling Extremal Events Is Not Easy: Why the Extreme Value Theorem Cannot Be As General As the Central Limit Theorem

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Hung T. Nguyen
New Mexico State University - Main Campus, hunguyen@nmsu.edu

Songsak Sriboonchitta
Chiang Mai University, songsakecon@gmail.com

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Modeling Extremal Events Is Not Easy: Why the Extreme Value Theorem Cannot Be As General As the Central Limit Theorem

Vladik Kreinovich[1], Hung T. Nguyen[2,3],
Songsak Sriboonchitta[3], and Olga Kosheleva[4]

[1] Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA, `vladik@utep.edu`
[2] Department of Mathematical Sciences, New Mexico State University
Las Cruces, New Mexico 88003, USA, `hunguyen@nmsu.edu`
[3] Department of Economics, Chiang Mai University
Chiang Mai, Thailand, `songsakecon@gmail.com`
[4] University of Texas at El Paso, 500 W. University,
El Paso, TX 79968, USA, `olgak@utep.edu`

Abstract. In many real-life situations, a random quantity is a joint result of several independent factors, i.e., a *sum* of many independent random variables. The description of such sums is facilitated by the Central Limit Theorem, according to which, under reasonable conditions, the distribution of such a sum tends to normal. In several other situations, a random quantity is a *maximum* of several independent random variables. For such situations, there is also a limit theorem – the Extreme Value Theorem. However, the Extreme Value Theorem is only valid under the assumption that all the components are identically distributed – while no such assumption is needed for the Central Limit Theorem. Since in practice, the component distributions may be different, a natural question is: can we generalize the Extreme Value Theorem to a similar general case of possible different component distributions? In this paper, we use simple symmetries to prove that such a generalization is not possible. In other words, the task of modeling extremal events is provably more difficult than the task of modeling of joint effects of many factors.

1 Sums and Maxima of Independent Factors: Formulation of the Problem

Why normal distributions are ubiquitous: sums of many independent factors. In many practical situations, we have a joint effects of many independent small factors. This happens, e.g., in measurements, when:

– after eliminating all major sources of possible measurement error,
– we end up with a measurement error which results from a joint effect of multiple difficult-to-eliminate independent sources of measurement uncertainty.

In this case, the measurement error – i.e., the difference $\Delta X = \widetilde{X} - X$ between the measurement result $\widetilde{X}$ and the actual value X of the corresponding quantity – can be represented as a sum $\Delta X = \sum_{i=1}^{n} X_i$ of a large number n of small independent random variables X_i.

The description of the resulting probability distribution for this sum ΔX is facilitated by the well-known *Central Limit Theorem*, according to which, under some reasonable conditions, the distribution of such a sum tends to Gaussian as n tends to infinity; see, e.g., [13]. This limit result means that for large n, the distribution of the sum ΔX is close to Gaussian.

This is indeed the case for most measuring instruments: experimental analysis shows that for about 60% of them, the measurement error is normally distributed [10, 11]. The Central Limit Theorem also explains why normal distribution is ubiquitous in many other areas as well: the familiar bell-shaped curve indeed describes many phenomena, from distribution people by height or by weight to distribution of molecules by velocity.

Extremal events: maxima of many independent factors. In many other practical situations, we are interested in describing the *maxima* of several independent factors. For example, in structural engineering, to estimate the structure's stability under catastrophic events such as hurricanes and earthquakes, it is important to estimate the probability that this structure collapses – i.e., that in one of its components, the tension exceeds the stability threshold. The condition that one of the tension values X_i exceeds the threshold x is equivalent to $X \geq x$, where $X \stackrel{\text{def}}{=} \max_i X_i$ is the maximum of several independent components. Thus, to study such extremal events, it is important to know the probability distribution of such maxima.

Similar arguments show the need to study similar maxima in earth sciences, in finances, in hydrology, and in many other areas where rare disastrous events are possible; see, e.g., [1–7, 9, 12].

Limit theorems for extreme events: what is known. Similarly to the Central Limit Theorem that describes the limit of sums, there are the limit theorems that describe the limits of maxima of many independent random variables. The most well-known of these limit theorems is the *Extreme Value Theorem* (also known as the *Fisher-Tippett-Gnedenko Theorem*), according to which, if we have a sequence of independent identically distributed random variables X_i and thee distributions of their maxima $M_n \stackrel{\text{def}}{=} \max(X_1, \ldots, X_n)$ has a limit, then this limit has one of the following forms [5]:

- *Weibull law*, with cumulative distribution function (cdf)
$$F(x) = \exp\left(-\left|\frac{x-b}{a}\right|^{\alpha}\right) \text{ for } x \leq b \text{ and } F(x) = 1 \text{ for } x \geq b;$$

- *Gumbel law* $F(x) = \exp\left(-\exp\left(\frac{b-x}{a}\right)\right)$; and

- *Fréchet law* $F(x) = \exp\left(-\left(\frac{x-b}{a}\right)^{-\alpha}\right)$ for $x > b$ and $F(x) = 0$ for $x \leq b$.

Formulation of the problem: what is available for the central limit theorems but missing for extreme value theorems. In many formulations of the Central Limit Theorem, it is not necessary to require that all components random variables X_i are identically distributed. These theorems are applicable to many situations in which different variables X_i have different distributions.

In contrast, the Extreme Value Theorem is only known for the case when all the component random variables X_i are identically distributed. In practical applications, the distributions of the corresponding random variables X_i – e.g., variables describing stability of different part of the construction – are, in general, different. A natural question arises: can we generalize the Extreme Value Theorem so that it can be applied to the case when we have different distributions X_i?

What we prove in this paper. In this paper, we prove that such a generalization is not possible. In this sense, the task of modeling extremal events is provably harder than the task of modeling a joint effect of several factors.

2 Analysis of the Problem

What do we mean by the desired generalization? Both in case of the Central Limit Theorem and in the case of the Extreme Value Theorem, we have a finite-parametric family of limit distributions such that, under certain reasonable conditions, the distribution of the corresponding sum or maxima tends to one of the distributions from this class.

From this viewpoint, when we say that we are looking for a generalization of the Extreme Value Theorem – which would be similar to the current (generalized) versions of the Central Limit Theorem – we mean that:

- we are looking for a finite-dimensional family $\mathcal{F}$ of probability distributions,
- such that, that, under certain reasonable conditions, the distribution of the corresponding maxima tends to one of the distributions from the class $\mathcal{F}$.

Let us describe which properties of the class $\mathcal{F}$ follow from this desired description.

First desired property: the class $\mathcal{F}$ should be invariant under shifts and scalings. We are interested in the distributions of physical quantities X_i. Of course, in all data processing, we deal with the numerical values of the corresponding physical quantities. The numerical value of a quantity depends on the choice of a measuring unit and on the choice of a starting point. For example, we can measure time in years or days or seconds, we can start measuring time with year 0 or with year 1950 (as is sometimes done in astronomical computations), etc.

When we change a starting point for measuring X by a new starting point which is smaller by b units, then the numerical value X changes to $X' = X + b$. This shift changes the numerical expression for the cdf: instead of the original

probability $F(x) = \text{Prob}(X \leq x)$ that $X \leq x$, for $X' = X + b$, we have

$$F'(x) = \text{Prob}(X' \leq x) = \text{Prob}(X + b \leq x) = \text{Prob}(X \leq x - b) = F(x - b).$$

When we change a measuring unit to a new one which is $a > 0$ times smaller, then the numerical value X changes to $X' = a \cdot X$. For example, if instead of meters, we use centimeters, a unit which is $a = 100$ times smaller, then all numerical values of length are multiplied by 100: e.g., $X = 2$ m becomes $X' = 200$ cm. This scaling changes the numerical expression for the cdf: instead of the original probability $F(x) = \text{Prob}(X \leq x)$ that $X \leq x$, for $X' = a \cdot X$, we have

$$F'(x) = \text{Prob}(X' \leq x) = \text{Prob}(a \cdot \leq x) = \text{Prob}\left(X \leq \frac{x}{a}\right) = F\left(\frac{x}{a}\right).$$

In general, if we change both the starting point and the measuring unit, we get a new cdf

$$F'(x) = F\left(\frac{x - b}{a}\right).$$

If we perform this transformation, then all the values X_i are replaced by the new values $X'_i = a \cdot X_i + b$. For the maximum, we similarly have

$$M'_n = \max(X'_1, \ldots, X'_n) = \max(a \cdot X_1 + b, \ldots, a \cdot X_n + b) =$$

$$a \cdot \max(X_1, \ldots, X_n) + b = a \cdot M_n + b.$$

Thus, if in the original units, we had a limit distribution $F(x)$, in the new units, we will have a limit distribution $F'(x) = F\left(\dfrac{x - b}{a}\right)$.

The desired limit theorem should not depend on the choice of the starting point or on the choice of a measuring unit. Thus, it is reasonable to require that if the class $\mathcal{F}$ of limit distributions contains a cdf $F(x)$, then it should also contain a re-scaled and shifted distribution $F'(x) = F\left(\dfrac{x - b}{a}\right)$.

Second desired property: the class $\mathcal{F}$ should be closed under multiplication of cdfs. Let us assume that $F(x)$ and $F'(x)$ are two cdfs from the desired class $\mathcal{F}$. By definition of the class $\mathcal{F}$, all distributions from this class are limits of distributions of the maxima. In particular:

- the cdf $F(x)$ is the limit of the distributions $F_n(x)$ of $M_n = \max(X_1, \ldots, X_n)$ for some sequence of independent random variables X_i, and
- the cdf $F'(x)$ is the limit of the distributions $F'_n(x)$ of $M'_n = \max(X'_1, \ldots, X'_n)$ for some sequence of independent random variables X'_i.

Then, for a combined sequence $X''_i \overset{\text{def}}{=} X_1, X'_1, \ldots, X_n, X'_n, \ldots$, the corresponding maxima will have the form

$$M''_{2n} = \max(X_1, X'_1, \ldots, X_n, X'_n) =$$

$$\max(\max(X_1, \ldots, X_n), \max(X_1', \ldots, X_n')) = \max(M_n, M_n').$$

The distribution of M_n is close to $F(x)$, the distribution of M_n' is close to $F'(x)$. The cdf $F_n''(x)$ for the maximum M_{2n}'' can be thus described as follows:

$$F_{2n}''(x) = \mathrm{Prob}(M_n'' \leq x) = \mathrm{Prob}(\max(M_n, M_n') \leq x) =$$

$$\mathrm{Prob}(M_n \leq x \,\&\, M_n' \leq x).$$

Since the variables X_i and X_i' are independent, their maxima M_n and M_n' are also independent, so

$$F_{2n}''(x) = \mathrm{Prob}(M_n \leq x \,\&\, M_n \leq x) = \mathrm{Prob}(M_n \leq x) \cdot \mathrm{Prob}(M_n' \leq x),$$

i.e., $F_{2n}''(x) = F_n(x) \cdot F_n'(x)$. In the limit, the distribution of M_n tends to $F(x)$ and the distribution of M_n' tends to $F'(x)$, so the distribution for the new sequence tends to the product $F(x) \cdot F'(x)$.

Thus, with every two cdfs $F(x)$ and $F'(x)$, the class $\mathcal{F}$ should also contain their product $F(x) \cdot F'(x)$.

The class $\mathcal{F}$ should be finite-dimensional. The previous property is easier to describe if we consider logarithms $\ln(F(x))$: the logarithm of product of the cdfs is the sum of their logarithms, so the class $\mathcal{L}$ of such logarithms should be closed under addition.

One can easily check that if the original class $\mathcal{F}$ is closed under shifts and scalings, then the class $\mathcal{L}$ of logarithms of functions $F(x) \in \mathcal{F}$ should also be similarly closed.

Each such class can be naturally extended to a linear space. We can show that this space should also be closed under shift and scaling.

The fact that the original set is finite-dimensional (= finite-parametric) implies that this space should also be finite-dimensional, i.e., all its functions should have the form

$$\ell(x) = C_1 \cdot e_1(x) + \ldots + C_m \cdot e_m(x),$$

where m is the dimension of this space, $e_i(x)$ are given functions, and C_i are arbitrary real values.

Now, we are ready to formulate and prove our main result.

3   Definitions and the Main Result

Definition. *We say that a finite-dimensional linear space $\mathcal{L}$ of differentiable functions is* shift- and scale-invariant *if with every function $\ell(x)$ and for every two real numbers $a > 0$ nd b, this class also contains the function*

$$\ell'(x) = \ell\left(\frac{x-b}{a}\right).$$

Proposition. *For every shift- and scale-invariant finite-dimensional linear space $\mathcal{L}$ of differentiable functions, all its elements are polynomials.*

Comment. All the proofs are given in the following section.

Corollary. *If $F(x)$ is a cdf, then its logarithm $\ln(F(x))$ cannot be an element of a shift- and scale-invariant finite-dimensional linear space.*

Discussion. This result shows that a finite-dimensional limit class $\mathcal{F}$ is not possible. Thus, the Extreme Value Theorem indeed cannot be extended to the general case when variables X_i are not necessarily identically distributed.

4   Proofs

Proof of the Proposition. The main ideas of this proof can be found in [8].

1°. The fact that the linear space $\mathcal{L}$ is shift-invariant means, in particular, that for every basis function $e_i(x)$ and for every real value b, the shifted function $e_i(x + b)$ also belongs to this linear space. Since all the function from the linear space are linear combinations of the basis functions $e_1(x), \ldots, e_m(x)$, this means that for every b, there exist values $C_{i,j}(b)$ for which

$$e_i(x + b) = C_{i,1}(b) \cdot e_1(x) + \ldots + C_{i,m}(b) \cdot e_m(x). \tag{1}$$

For each b, we can take m different values $x_1, \ldots, x_m$, and get m resulting equalities:

$$e_i(x_1 + b) = C_{i,1}(b) \cdot e_1(x_1) + \ldots + C_{i,m}(b) \cdot e_m(x_1);$$

$$\ldots$$

$$e_i(x_j + b) = C_{i,1}(b) \cdot e_1(x_j) + \ldots + C_{i,m}(b) \cdot e_m(x_j); \tag{2}$$

$$\ldots$$

$$e_i(x_m + b) = C_{i,1}(b) \cdot e_1(x_m) + \ldots + C_{i,m}(b) \cdot e_m(x_m).$$

We thus get a system of m linear equations for m unknowns $C_{i,1}(b), \ldots, C_{i,m}(b)$. By using Cramer's rule, we can describe the values $C_{i,j}(b)$ as ratios of polynomials in terms of the coefficients $e_j(x_k)$ and the right-hand sides $e_i(x_j + b)$. Since the functions $e_i(x)$ are differentiable, we can conclude that the dependence $C_{i,j}(b)$ on b is differentiable as well.

2°. We can now combine the equations (1) corresponding to different functions $e_i(x)$. As a result, we get the following system of m equations:

$$e_1(x + b) = C_{1,1}(b) \cdot e_1(x) + \ldots + C_{1,m}(b) \cdot e_m(x);$$

$$\ldots$$

$$e_i(x + b) = C_{i,1}(b) \cdot e_1(x) + \ldots + C_{i,m}(b) \cdot e_m(x); \tag{3}$$

$$\ldots$$

$$e_m(x + b) = C_{m,1}(b) \cdot e_1(x) + \ldots + C_{m,m}(b) \cdot e_m(x).$$

Differentiating both sides of these equations by b and taking $b = 0$, we get the following system of differential equations:

$$\frac{de_1(x)}{dx} = c_{1,1} \cdot e_1(x) + \ldots + c_{1,m} \cdot e_m(x);$$

$$\ldots$$

$$\frac{de_i(x)}{dx} = c_{i,1} \cdot e_1(x) + \ldots + c_{i,m} \cdot e_m(x); \tag{4}$$

$$\ldots$$

$$\frac{de_m(x)}{dx} = c_{m,1} \cdot e_1(x) + \ldots + c_{m,m} \cdot e_m(x),$$

where we denoted $c_{i,j} \stackrel{\text{def}}{=} \dfrac{dC_{i,j}(b)}{db}\Big|_{b=0}$.

We have a system (4) of linear differential equations with constant coefficients. It is known that a general solution to this system is a linear combination of expressions of the type $x^k \cdot \exp(\lambda \cdot x)$, where:

- the value λ is a (possible complex) eigenvalue of the matrix $c_{i,j}$, and
- the value k is a natural number; thus number should be smaller than the multiplicity of the corresponding eigenvalue.

Thus, each function $e_i(x)$ is such a linear combination.

3°. Let us now use scale-invariance. The fact that the linear space $\mathcal{L}$ is scale-invariant means, in particular, that for every basis function $e_i(x)$ and for every real value b, the shifted function $e_i(a \cdot x)$ also belongs to this linear space. Since all the function from the linear space are linear combinations of the basis functions $e_1(x), \ldots, e_m(x)$, this means that for every a, there exist values $A_{i,j}(a)$ for which

$$e_i(a \cdot x) = A_{i,1}(a) \cdot e_1(x) + \ldots + A_{i,m}(a) \cdot e_m(x). \tag{5}$$

For each a, we can take m different values $x_1, \ldots, x_m$, and get m resulting equalities:

$$e_i(a \cdot x_1) = A_{i,1}(a) \cdot e_1(x_1) + \ldots + A_{i,m}(a) \cdot e_m(x_1);$$

$$\ldots$$

$$e_i(a \cdot x_j) = A_{i,1}(a) \cdot e_1(x_j) + \ldots + A_{i,m}(a) \cdot e_m(x_j); \tag{6}$$

$$\ldots$$

$$e_i(a \cdot x_m) = A_{i,1}(a) \cdot e_1(x_m) + \ldots + A_{i,m}(a) \cdot e_m(x_m).$$

We thus get a system of m linear equations for m unknowns $A_{i,1}(a), \ldots, A_{i,m}(a)$. By using Cramer's rule, we can describe the values $A_{i,j}(a)$ as ratios of polynomials in terms of the coefficients $e_j(x_k)$ and the right-hand sides $e_i(a \cdot x_j)$. Since the

functions $e_i(x)$ are differentiable, we can conclude that the dependence $A_{i,j}(a)$ on a is differentiable as well.

$4°$. We can now combine the equations (5) corresponding to different functions $e_i(x)$. As a result, we get the following system of m equations:

$$e_1(a \cdot x) = A_{1,1}(a) \cdot e_1(x) + \ldots + A_{1,m}(a) \cdot e_m(x);$$

$$\ldots$$

$$e_i(a \cdot x) = A_{i,1}(a) \cdot e_1(x) + \ldots + A_{i,m}(a) \cdot e_m(x); \tag{7}$$

$$\ldots$$

$$e_m(a \cdot x) = C_{m,1}(a) \cdot e_1(x) + \ldots + A_{m,m}(a) \cdot e_m(x).$$

Differentiating both sides of these equations by a and taking $a = 1$, we get the following system of differential equations:

$$x \cdot \frac{de_1(x)}{dx} = a_{1,1} \cdot e_1(x) + \ldots + a_{1,m} \cdot e_m(x);$$

$$\ldots$$

$$x \cdot \frac{de_i(x)}{dx} = a_{i,1} \cdot e_1(x) + \ldots + a_{i,m} \cdot e_m(x); \tag{8}$$

$$\ldots$$

$$x \cdot \frac{de_m(x)}{dx} = a_{m,1} \cdot e_1(x) + \ldots + a_{m,m} \cdot e_m(x),$$

where we denoted $a_{i,j} \stackrel{\text{def}}{=} \dfrac{dA_{i,j}(a)}{da}\bigg|_{a=1}$.

$5°$. To solve this new system of equations, we can introduce a new variable $t \stackrel{\text{def}}{=} \ln(x)$, for which $\dfrac{dx}{x} = dt$. Here, $x = \exp(t)$, so for the new functions $E_i(t) \stackrel{\text{def}}{=} e_i(\exp(t))$, the system (8) takes the following form:

$$\frac{dE_1(t)}{dt} = a_{1,1} \cdot E_1(t) + \ldots + a_{1,m} \cdot E_m(t);$$

$$\ldots$$

$$\frac{dE_i(t)}{dt} = a_{i,1} \cdot E_1(t) + \ldots + a_{i,m} \cdot E_m(t); \tag{9}$$

$$\ldots$$

$$\frac{dE_m(x)}{dx} = a_{m,1} \cdot E_1(t) + \ldots + a_{m,m} \cdot E_m(t).$$

This is a system of linear differential equations with constant coefficients; so, each function $E_i(t)$ is linear combination of the expressions of the type $t^k \cdot \exp(\lambda \cdot t)$, Thus, for $e_i(x) = E_i(\ln(x))$, we conclude that each function $e_i(x)$ is a linear combination of the expressions

$$(\ln(x))^k \cdot \exp(\lambda \cdot \ln(x)) = (\ln(x))^k \cdot x^\lambda.$$

$6°$. We have proven that:

- on the one hand, each function $e_i(x)$ is a linear combination of the expressions $x^k \cdot \exp(\lambda \cdot x)$, where k is a natural number;
- on the other hand, each function $e_i(x)$ is a linear combination of the expressions $(\ln(x))^k \cdot x^\lambda$.

One can easily see that the need to be represented in the second form excludes the possibility of $\lambda \neq 0$. Thus, each function $e_i(x)$ is a linear combination of the expressions of the type x_k with natural k – i.e., a polynomial. Every function from the linear space $\mathcal{L}$ is a linear combination of the basis functions $e_i(x)$ and is, thus, also a polynomial.

The proposition is proven.

Proof of the Corollary. Let us prove this result by contradiction. Let us assume that for some cdf $F(x)$, its logarithm $\ln(F(x))$ belongs to a shift-and scale-invariant linear space $\mathcal{L}$. Due to Proposition, this implies that this logarithm is a polynomial $P(x)$: $\ln(F(x)) = P(x)$ and thus, $F(x) = \exp(P(x))$.

When $x \to -\infty$, we have $F(x) \to 0$, so we should have

$$P(x) = \ln(F(x)) \to -\infty.$$

For the corresponding polynomial $P(x) = a_0 \cdot x^k + a_1 \cdot x^{k-1} + \ldots$, this means that:

- either k is even and $a_0 < 0$,
- or k is odd and $a_0 > 0$.

When $x \to +\infty$, then:

- in the first case, we have $P(x) \to -\infty$, while
- in the second case, we have $P(x) \to +\infty$.

However, we should have $F(x) \to 1$ and thus, $P(x) = \ln(F(x)) \to \ln(1) = 0$.

This contradiction shows that our assumption was wrong, and logarithms for cdfs cannot belong to shift-and scale-invariant linear spaces.

Acknowledgments

References

1. J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels, *Statistics of Extremes: Theory and Applications*, Wiley, New York, 2004.
2. E. Castillo, A. S. Hadi, N. Balakrishnan, and J. M. Sarabia, *Extreme Value and Related Models with Applications in Engineering and Science*, Wiley, New York, 2004

3. S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Verlag, London, 2001.
4. L. de Haan and A. Ferreira, *Extreme Value Theory: An Introduction*, Springer Verlag, New York, 2006.
5. P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Spring Verlag, Berlin 1997.
6. E. J. Gumbel, *Statistics of Extremes*, Dover, New York, 2013.
7. S. Kotz and S. Nadarajah, *Extreme Value Distributions: Theory and Applications*, Imperial College Press, 2000.
8. H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.
9. S. Y. Novak, *Extreme Value Methods with Applications to Finance*, Chapman & Hall/CRC Press, London, 2011.
10. P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991 (in Russian).
11. A. I. Orlov, "How often are the observations normal?", *Industrial Laboratory*, 1991, Vol. 57. No. 7, pp. 770–772.
12. S. I. Resnick, "Extreme Values, Regular Variation and Point Processes", Springer Verlag, New York, 2008.
13. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2011.