Extreme distributions on intervals

M. Chiangpradit[1], W. Panichkitkosolkul[1],
H. T. Nguyen[2], V. Kreinovich[3]

[1]*King Mongkut's University of Technology North Bangkok, Thailand*
[2]*New Mexico State University, Las Cruces, USA*
[3]*University of Texas at El Paso, USA*
e-mail: {monchaya_c,wararit_tu}@hotmail.com,
hunguyen@nmsu.edu, vladik@utep.edu

One of the main tasks of interval computation is to analyze situations in which we only know the lower and upper bounds on the desired quantity, i.e., we only know an interval that contains this quantity. One of the objectives of such analysis is to make decisions. According to decision theory, a consistent decision making procedure is equivalent to assigning probabilities to different values within each interval. Thus, we arrive at the problem of describing a natural probability distribution on an interval. In this paper, we describe such a distribution for the practically important case of the "weakest link" arrangement, when the collapse of each link is catastrophic for a system. This situation occurs in fracture mechanics, when a fracture in one of the areas makes the whole plane inoperable, in economics, when the collapse of one large bank or one country can have catastrophic consequences, etc.

Keywords: extreme distributions, interval, decision making, symmetries.

Introduction

Need to make decisions. One of the main practical objectives of science and engineering is to make decisions, i.e., to select an alternative which is the best for the decision maker.

How to describe preferences of a decision maker: the notion of utility. A standard way to describe preferences of a decision maker is to use the notion of *utility*; see, e.g., [7, 8, 10, 12, 14]. To describe the utility of an outcome A, we need to select two extreme outcomes: a very unfavorable alternative A_- and a very favorable outcome A_+.

We assume that all outcomes A in which we are interested are better than A_- and worse than A_+. If we denote the relation "the decision maker prefers A' to A" by $A \leq A'$, then we can describe this assumption as $A_- \leq A \leq A_+$.

Then, for each probability $p \in [0, 1]$, we can consider a *lottery* $L(p)$ in which we have A_+ with probability p and A_- with the remaining probability $1 - p$.

For $p = 1$, the lottery $L(p)$ coincides with A_+, so we have $A \leq L(1)$. For $p = 0$, the lottery $L(p)$ coincides with A_-, so we have $L(0) \leq A$. The larger p, i.e., the larger the probability of a beneficial event A_+, the more beneficial is the lottery $L(p)$ for the decision maker. So, if $p < q$, then $L(p) < L(q)$.

Let p_0 be the infimum (greatest lower bound) of the set of all the values p for which $A \leq L(p)$. Then:

- When $p < p_0$, then for $\widetilde{p} = (p + p_0)/2$, we have $\widetilde{p} < p_0$ and thus, by definition of the infimum, we cannot have $A \leq L(\widetilde{p})$. Thus, we have $L(\widetilde{p}) \leq A$. Since $p < \widetilde{p}$, we have $L(p) < L(\widetilde{p}) \leq A$ and thus, $L(p) < A$.

- When $p > p_0$, then, since p_0 is the greatest lower bound, p is not a lower bound, i. e., there exists a value $\widetilde{p}$ for which $A \leq L(\widetilde{p})$ and $\widetilde{p} < p$. Since $\widetilde{p} < p$, we have $L(\widetilde{p}) < L(p)$ hence $A < L(p)$.

Thus, we have the value p_0 that has the following property:

- when $p < p_0$, the corresponding lottery is worse than the event A:

$$L(p) < A;$$

- when $p > p_0$, the corresponding lottery is better than the event A:

$$L(p) > A.$$

This threshold value p_0 is called the *utility* of the event A. The utility is usually denoted by $u(A)$.

We can simplify the above somewhat complicated relation between A and p_0 by saying that the event $L(p_0)$ is *equivalent* to A. We will denote this equivalence by $A \sim L(p_0)$.

How to describe the utility of an action with uncertain consequences. In practice, we can rarely predict the exact consequences of each decision. The consequences depend on the circumstances. For example, if we decide whether to take an umbrella or not, the consequences of this decision depend on whether it will rain or not. In the ideal situation, we know the probabilities $p_1, \ldots, p_n$ of different possible consequences $E_1, \ldots, E_n$. In other words, the action leads to E_1 with probability p_1, to E_2 with probability p_2, $\ldots$, and to E_n with probability p_n.

By definition of the utility, the event E_1 is equivalent to a lottery $L(u(E_1))$ in which we get A_+ with probability $u(E_1)$, the event E_2 is equivalent to a lottery $L(u(E_2))$ in which we get A_+ with probability $u(E_2)$, etc. Thus, the original action is equivalent to the composite lottery, in which:

- with probability p_1, we get a lottery that results in A_+ with probability $u(E_1)$, and in A_- otherwise;

- with probability p_2, we get a lottery that results in A_+ with probability $u(E_2)$, and in A_- otherwise;

- ...

In this composite lottery, we get either A_+ or A_-, and the probability of getting A_+ can be easily computed as

$$u \stackrel{\text{def}}{=} p_1 \cdot u(E_1) + p_2 \cdot u(E_2) + \ldots + p_n \cdot u(E_n).$$

Thus, the original action is equivalent to the lottery $L(u)$. By definition of the utility, this means that the utility of the action is equal to u.

From the mathematical viewpoint, u is the expected value of the utility of different consequences, so we can conclude that the utility of an action is the expected value of utilities of its consequences.

What if we do not know the probabilities of different consequences? In many practical situations, we do not know the exact values of the probabilities of different consequences. For each event E_i, we can estimate its *subjective probability* $ps(E_i)$ as the probability p_i for which the lottery $L(p_i)$ (in which we get A_+ with probability p_i, otherwise we

get A_-) is equivalent to the new "lottery" $L(E_i)$ in which we get A_+ if E_i occurs, otherwise we get A_-.

In other words, we determine the subjective probability $ps(E_i)$ as the utility $u(L(E_i))$ of the new lottery $L(E_i)$.

In practice, it is sometimes difficult to ask experts. The traditional utility theory approach — that we described above — is to elicit, from the experts, all the information about their preferences and their subjective probabilities.

In some applications, e. g., when we undertake a large project, this is possible and reasonable. For example, decision theory has been used to select a location for a major airport. With such a long-term billion-dollar investment that affects many potential users, it makes sense to spend a certain amount of time and resources to get a clear picture of user preferences.

However, often, we face decisions which need to be made fast and which are not that critical. In such situations, we do not have time to elicit all the values of subjective probabilities, and, even when we have some time for such an elicitation, we may end up spending more resources on this elicitation than we gain from knowing these probability values.

Need to get reasonable probability distributions. A typical situation is when the consequences of an action depend on some quantity a, we do not know the actual probabilities of different values of this quantity, and we have no time and/or resources to elicit subjective probabilities of different values a.

In this case, we need to come up with a reasonable probability distribution for a, a distribution that will be used in decision making.

Need to assign distributions on an interval. Usually, we know some bounds on each quantity, i. e., we know that the value a is always larger than or equal to some value $\underline{a}$ and always smaller than or equal to some value $\overline{a}$. In other words, we know that the value a belongs to the interval $[\underline{a}, \overline{a}]$.

In this case, we need a natural way to assign probabilities on an interval.

What we do: consider weakest link case. In this paper, we consider a practically important case of the "weakest link" arrangement.

Informally, this means that we have a multi-link system, and the collapse of each link is catastrophic for a system. Such situations are typical in economics, when the collapse of one large bank or one country can have catastrophic consequences. They are also typical in fracture mechanics, when a fracture in one of the areas makes the whole structure (e. g., an airplane wing) inoperable.

1. Analysis of the problem

Weakest link: a usual mathematical description. The weakest link situation is usually described as follows: the quality of each link i is characterized by a value v_i, and the quality of a system as a whole is determined by the smallest of the corresponding values v_i: $v \stackrel{\text{def}}{=} \min_i v_i$.

It is reasonable to assume that the values v_i are independent random variables. In mathematical terms, this means that we are looking for a distribution of the minimum $\min_i v_i$ of several independent variables. When n is a large, we get close to the limits of such distributions.

Limit distributions of minimum or maximum of independent identically distributed random variables (properly centered and normalized) are usually called *extreme value distri-*

butions; see, e. g., [1, 5, 6, 9, 15]. There is a known description of such distributions; they include Frechet, Gumbel, and Weibull distributions.

A usual mathematical description of the weakest link distributions: limitation. Alas, none of the known extreme value distributions describe a random variable which is located on a finite interval.

What we plan to do. Our objective is to describe extreme value distributions located on an interval. To describe such distributions, we will use the symmetries approach.

In order to be able to do that, let us first show that the known extreme value distributions can also be derived from the appropriate symmetries. We will then show how this derivation can be extended to the new case — when a random variable is limited to an interval.

Known extreme value distributions: main ideas behind derivation based on symmetries. We want to find the probability distribution of the extreme values. Traditionally, a probability distribution is described by a cumulative distribution function $F(v_0) \stackrel{\text{def}}{=} \text{Prob}(v \leq v_0)$ that describes the probability that a random variables does not exceed a given number v_0. However, from the practical viewpoint, we are interested in probabilities of rare events, i. e., in the probabilities that v *exceeds* the given value v_0. Thus, for extreme value distributions, it is more convenient to use the corresponding function $G(v_0) = 1 - F(v_0) = \text{Prob}(v > v_0)$.

In deriving the types of distributions, it is usually taken into account that the numerical value of a physical quantity v depends:

- on the choice of a measuring unit $v \to a \cdot v$ (e. g., 1.7 m = 170 cm), and
- on the choice of the starting point $v \to v+b$ (e. g.: A.D. or since the French Revolution).

Under these transformations, the original function $G(v_0)$ turns into a new re-scaled function $G(a \cdot v_0 + b)$. It is therefore reasonable, instead of looking for a single function $G(v_0)$, to look for a *family* $\mathcal{G}$ of distributions $\{G(a \cdot v_0 + b)\}_{a,b}$ obtained from some function $G(v_0)$.

By definition of an extreme distribution as a minimum, if n independent identically distributed variables v_i are distributed according to the extreme value distribution, then their minimum $v' \stackrel{\text{def}}{=} \min v_i$ is the minimum of minimums, so it should also be distributed according to the extreme value distribution. Clearly, $v' > v_0 \Leftrightarrow v_1 > v_0 \& \ldots \& v_n > v_0$; so, since v_i are independent, we conclude that

$$G'(v_0) = \text{Prob}(v > v_0) = \prod_{i=1}^{n} \text{Prob}(v_i > v_0) = (G(v_0))^n.$$

Thus, the desired family $\mathcal{G}$ should contain, with each function $G(v_0)$, also a function $G'(v_0) = (G_i(v_0))^n$.

Similarly, for the maximum v'' of $\alpha \cdot n$ values, we conclude that the function $G''(v_0) = (G(v_0))^{\alpha \cdot n}$ belongs to the family $\mathcal{G}$, where $G''(v_0) = (G'(v_0))^\alpha$.

It is therefore reasonable to conclude that if $G(v_0) \in \mathcal{G}$, then $G^\alpha(v_0) \in \mathcal{G}$ for all α. By definition of the family $\mathcal{G}$, this means that for every α, there exist $a(\alpha)$ and $b(\alpha)$ such that $G^\alpha(v_0) = G(a(\alpha) \cdot v_0 + b(\alpha))$.

Extreme value distributions: symmetry-based derivation of the known formulas. The above functional equation can be simplified if we consider an auxiliary function $g(v_0) \stackrel{\text{def}}{=} -\ln(G(v_0))$. For this auxiliary function, the above formula takes the form $\alpha \cdot g(v_0) = g(a(\alpha) \cdot v_0 + b(\alpha))$.

When $\alpha = 1$, we have $a(\alpha) = 1$, and $b(\alpha) = 0$. Differentiating both sides of the above formula by α and taking $\alpha = 1$, we get $g = \dfrac{dg}{dv_0} \cdot (a \cdot v_0 + b)$, i.e., $\dfrac{dg}{g} = \dfrac{dv_0}{a \cdot v_0 + b}$, where we denoted $a \stackrel{\text{def}}{=} a'(1)$.

When $a = 0$, integration leads to $\ln(g) = \dfrac{v_0}{b} + c$, so $g(v_0) = \exp\left(\dfrac{v_0}{b} + c\right)$ and $G(v_0) = \exp\left(-\exp\left(\dfrac{v_0}{b} + c\right)\right)$. When $a \neq 0$, for $v \stackrel{\text{def}}{=} v_0 + \Delta v$, with $\Delta v = b/a$, we get $\dfrac{dg}{g} = \dfrac{dv}{a \cdot v}$ hence $\ln(g) = a \cdot \ln(v) + c$, so $g = c \cdot v^a = c \cdot (v_0 - \Delta v)^a$, hence

$$G(v_0) = \exp\left(-c \cdot (v_0 - \Delta v)^a\right).$$

Comment. Actually, we get two different types of distributions depending on whether $a > 0$ or $a < 0$.

Not all linear transformations are physically meaningful. The above derivations are based on the assumption that we have linear symmetries $v_0 \to a \cdot v_0 + b$.

For some quantities like time or temperature, all values are possible, so we have both shift- and scale-invariance.

For other quantities, only some values are possible. For example, height can only take non-negative values, i.e., possible values are limited to the set $[0, \infty)$. In this case, only linear transformations that preserve this set make physical sense. In other words, we only consider scalings $v_0 \to a \cdot v_0$.

How to extend this analysis to distributions on an interval: discussion. For the quantities whose values are limited to a fixed interval $[\underline{v}, \overline{v}]$, it also makes sense to restrict ourselves to linear transformations that preserve this set $[\underline{v}, \overline{v}]$ of possible values. However, the only linear transformation that preserves this interval is identity.

Our solution: to go beyond linear symmetries, to more general (non-linear) symmetries.

Basic nonlinear symmetries: reminder. Sometimes, a system also has *nonlinear* symmetries. How can we describe the set of such symmetries?

If a system is invariant under transformations $f(x)$ and $g(x)$, then:

• it is invariant under their composition $f \circ g$, and
• it is invariant under the inverse transformation f^{-1}.

In mathematical terms, this means that symmetries form a *group*.

In general, we may have transformation groups that require infinitely many parameters: for example, the group of all possible transformations, i.e., all possible one-to-one functions from the real line to itself. However, in practice, at any given moment of time, we can only store and describe finitely many parameters. Thus, it is reasonable to restrict ourselves to transformation groups whose elements can be described by finitely many parameters, i.e., to *finite-dimensional* groups.

Thus, we arrive at the following problem: describe all finite-dimensional transformation groups that contain all linear transformations. This question was first formulated by N. Wiener, the father of cybernetics, in [17]. For an Euclidean space of arbitrary dimension n, such Lie groups have been classified in [16]. In particular, for our case $n = 1$, the only such groups are the group of linear mappings and the group of all fractionally-linear mappings

$$f(x) = \frac{a \cdot x + b}{c \cdot x + d}. \tag{1}$$

Since we are interested in non-linear re-scalings, we should therefore consider re-scalings of the type (1).

Resulting idea. In the linear case, we required that $G^\alpha(v_0)$ is equal to the result $G(a(\alpha) \cdot v_0 + b(\alpha))$ of applying a *linear* transformation $v_0 \to a(\alpha) \cdot v_0 + b(\alpha)$ to v_0.

Now, we similarly require that $G^\alpha(v_0)$ is equal to the result of applying a *fractionally-linear* transformation, i.e., that for every α, there exist $a(\alpha)$, $b(\alpha)$, $c(\alpha)$, and $d(\alpha)$ for which

$$G^\alpha(v_0) = G\left(\frac{a(\alpha) \cdot v_0 + b(\alpha)}{c(\alpha) \cdot v_0 + d(\alpha)}\right)$$

for some transformation

$$v_0 \to \frac{a(\alpha) \cdot v_0 + b(\alpha)}{c(\alpha) \cdot v_0 + d(\alpha)}$$

that preserves the set $[\underline{v}, \overline{v}]$ of possible values.

Side observation: symmetries explain the basic formulas of Neural Networks. Fractionally-linear transformations have been actively used; for example, they were used to explain the empirically successful form $f(x) = 1/(1 + e^{-x})$ of the *activation function*, i.e., a function that is used in describing how the output y of a neuron is related to its inputs $x_1, \ldots, x_n$: $y = f(x)$, where $x \stackrel{\text{def}}{=} \sum_{i=1}^{n} w_i \cdot x_i - w_0$, for some real numbers w_i.

The details of this explanation are given, e.g., in [13]. The main idea behind this explanation is as follows. The input x is only determined modulo starting point. If we change the starting point for measuring x_i, then the original value x changes into $x + s$. When we apply the activation function $f(x)$ to this changed input, we get the value $f(x + s)$.

In other words, changing the starting point means that we replace the original activation function $f(x)$ with a new activation function $f(x + s)$. It is reasonable to require that the new output $f(x + s)$ is equivalent to the original output $f(x)$ modulo an appropriate transformation. We have already shown that all appropriate transformations are fractionally linear. Thus, we conclude that for every s, there exist values $a(s)$, $b(s)$, $c(s)$, and $d(s)$ for which we have:

$$f(x + s) = \frac{a(s) \cdot f(x) + b(s)}{c(s) \cdot f(x) + d(s)}.$$

Differentiating both sides of this functional equation by s and equating s to 0, we get a differential equation for $f(x)$. Its known solution is the above activation function — which can thus be explained by symmetries.

2. Extreme distributions on an interval: derivation and the main result

Reduction to $[0, 1]$. Before we start our derivation, let us observe that every interval can be linearly reduced to the interval $[0, 1]$. Thus, it is sufficient to consider the case when $[\underline{v}, \overline{v}] = [0, 1]$.

Fractionally-linear transformations that preserve the interval $[0, 1]$. According to our idea, we must describe all fractionally-linear transformations

$$f(x) = \frac{a \cdot x + b}{c \cdot x + d}$$

that preserve the interval $[0, 1]$.

First, dividing both numerator and denominator of the fractionally-linear formula by d, and using a/d, b/d, and c/d instead of the original values of a, b, and c, we get a simplified expression

$$f(x) = \frac{a \cdot x + b}{1 + c \cdot x}.$$

For a monotonic transformation to preserve $[0, 1]$, we must have $f(0) = 0$ and $f(1) = 1$. Substituting the above formula for $f(x)$ into the equation $f(0) = 0$, we can conclude that $b = 0$. Substituting the above expression (with $b = 0$) into the equation $f(1) = 1$, we conclude that $\dfrac{a}{1+c} = 1$, hence $c = a - 1$ and

$$f(x) = \frac{a \cdot x}{1 + (a-1) \cdot x}.$$

Resulting reformulation of our problem. Now, we can reformulate our problem as follows: for every α, there exists $a(\alpha)$ for which

$$G^{\alpha}(v_0) = G\left(\frac{a(\alpha) \cdot v_0}{1 + (a(\alpha) - 1) \cdot v_0}\right).$$

Derivation of the formula. By taking logarithms of both sides of the above formula, we get

$$\alpha \cdot g(v_0) = g\left(\frac{a(\alpha) \cdot v_0}{v_0 + (a(\alpha) - 1)}\right).$$

When $\alpha = 1$, there is no transformation, so $a(\alpha) = 1$.

Differentiating both sides of the above equation by α and substituting $\alpha = 1$, we get the differential equation

$$g = \frac{dg}{dv_0} \cdot (a \cdot v_0 - a \cdot v_0^2).$$

Moving all the terms depending on g to the left-hand side and all the terms depending on v_0 to the right-hand side, we conclude that

$$\frac{dg}{g} = \frac{dv_0}{a \cdot v_0 - a \cdot v_0^2}.$$

The fraction in the right-hand side can be represented as the sum of two simpler fractions:

$$\frac{dg}{g} = \frac{1}{a} \cdot \left(\frac{1}{v_0} + \frac{1}{1 - v_0}\right).$$

Now, we can explicitly integrate both sides. As a result, we get the following formula:

$$\ln(g) = \frac{1}{a} \cdot (\ln(v_0) - \ln(1 - v_0)) + c = \frac{1}{a} \cdot \ln\left(\frac{v_0}{1 - v_0}\right) + c,$$

hence

$$g(v_0) = \beta \cdot \left(\frac{1 - v_0}{v_0}\right)^C$$

for some parameter C.

For a general interval $[\underline{v}, \overline{v}]$, we get

$$g(v_0) = \beta \cdot \left(\frac{\overline{v} - v_0}{v_0 - \underline{v}} \right)^C .$$

Exponentiating, we get $G(v_0) = \exp(-g(v_0))$, hence we arrive at the following result:

Result. For variables on an interval $[\underline{v}, \overline{v}]$, the general extreme distribution has the following form

$$G(v_0) = \exp \left(-\beta \cdot \left(\frac{\overline{v} - v_0}{v_0 - \underline{v}} \right)^C \right) .$$

Discussion. These distributions were empirically found in fracture mechanics by A. Chudnovsky and B. Kunin [2–4, 11].

For the specific case of $C = 0$, we get a uniform distribution — a usual distribution on an interval.

Acknowledgments

References

[1] Beirlant J., Goegebeur Y., Segers J. et al. Statistics of Extremes: Theory and Applications. N.Y.: Wiley, 2004.

[2] Chudnovsky A., Botsis J., Kunin B. The role of microdefects in fracture propagation process // Cracking and Damage. Strain Localization and Size Effect / Eds. J. Mazais, Z. Bažant. Amsterdam: Elsevier, 1988. P. 140–149.

[3] Chudnovsky A., Kunin B. A probabilistic model of brittle crack formation // J. Appl. Phys. 1987. Vol. 62(25). P. 4124–4129.

[4] Chudnovsky A., Kunin B. On applications of probability in fracture mechanics // Computational Mechanics of Probabilistic and Reliability Analysis / Eds. W.K. Liu, T. Belytschko. Lausanne, Switzerland: Elmepress Intern., 1989. P. 396–415.

[5] Coles S. An Introduction to Statistical Modeling of Extreme Values. Berlin, Heidelberg, N.Y.: Springer, 2001.

[6] Embrechts P., Klueppelberg C., Mikosch T. Modelling Extremal Events: For Insurance and Finance. Berlin, Heidelberg, N.Y.: Springer, 2010.

[7] Fishburn P.C. Utility Theory for Decision Making. N.Y.: John Wiley & Sons Inc., 1969.

[8] Fishburn P.C. Nonlinear Preference and Utility Theory. Baltimore, Maryland: The John Hopkins Press, 1988.

[9] Gumbel E.J. Statistics of Extremes. N.Y.: Dover, 2004.

[10] Keeney R.L., Raiffa H. Decisions with Multiple Objectives. N.Y.: John Wiley and Sons, 1976.

[11] Kunin B. A Probabilistic Model for Predicting Scatter in Brittle Fracture. PhD Dissertation. Department of Mathematics, Statistics, and Computer Science. Univ. of Illinois at Chicago, 1992.

[12] Luce R.D., Raiffa R. Games and Decisions: Introduction and Critical Survey. N.Y.: Dover, 1989.

[13] Nguyen H.T., Kreinovich V. Applications of Continuous Mathematics to Computer Science. Dordrecht: Kluwer, 1997.

[14] Raiffa H. Decision Analysis. Reading, Massachusetts: Addison-Wesley, 1970.

[15] Resnick S.I. Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Berlin, Heidelberg, N.Y.: Springer, 2006.

[16] Singer I.M., Sternberg S. Infinite groups of Lie and Cartan. Part 1 // J. d'Analyse Math. 1965. Vol. XV. P. 1–113.

[17] Wiener N. Cybernetics, or Control and Communication in the Animal and the Machine. Cambridge, Massachusetts: MIT Press, 1962.