Fuzzy Techniques, Laplace Indeterminacy Principle, and Maximum Entropy Approach Explain Lindy Effect and Help Avoid Meaningless Infinities in Physics

1st Julio C. Urenda
Departments of Mathematical Sciences and Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
jcurenda@utep.edu

2nd Sean Aguilar
Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
sraguilar4@miners.utep.edu

3rd Olga Kosheleva
Department of Teacher Education
University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu

4th Vladik Kreinovich
Department of Computer Science and Computational Science Program
University of Texas at El Paso
El Paso, Texas 79968, USA
vladik@utep.edu

***Abstract*—In many real-life situations, the only information that we have about some quantity S is a lower bound $T < S$. In such a situation, what is a reasonable estimate for S? For example, we know that a company has survived for T years, and based on this information, we want to predict for how long it will continue surviving. At first glance, this is a type of a problem to which we can apply the usual fuzzy methodology – but unfortunately, a straightforward use of this methodology leads to a counter-intuitive infinite estimate for S. There is an empirical formula for such estimation – known as Lindy Effect and first proposed by Benoit Mandelbrot – according to which the appropriate estimate for S is proportional to T: $S = c \cdot T$, where, with some confidence, the constant c is equal to 1. In this paper, we show that a deeper analysis of the situation enables fuzzy methodology to lead to a finite estimate for S, moreover, to an estimate which is in perfect accordance with the empirical Lindy Effect. Interestingly, a similar idea can help in physics, where also, in some problems, straightforward computations lead to physically meaningless infinite values.**

***Index Terms*—Lindy Effect, fuzzy techniques, Laplace Indeterminacy Principle, probabilistic techniques, Maximum Entropy approach, infinities in physics**

I. Formulation of the Problem

What is Lindy Effect. In this paper, we analyze a phenomenon known as *Lindy Effect.* Its main idea is intuitively clear, but its formal description is not that well known, so let us start by describing what is the Lindy Effect.

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes). It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

Lindy Effect: intuitive idea. If we have a company that has been in successful existence for many decades and another company which is a recent startup, what are the chances that each of these companies will survive for another decade? Intuitively, it is clear that the company that has been successful for many years, that have successfully survived many crises, will probably survive for another decade (and probably even longer), while a start-up has a high risk of not surviving – as most startups do.

This is an important issue if we plan a long-term investment: the stocks of which of the two companies shall we mostly buy?

If we have a building that has been standing since the 19 century, and another modernist experimental building built a few years ago, which of them has a better chance of survival? Clearly, the one that has been standing for more than 100 years will probably stand some more, while an experimental building, built by using not-yet-fully-tested technology, is at risk of needing repairs soon.

If we have a family that has recently celebrated its 50th anniversary and another family whose marriage has just been announced – who has a bigger chance of not divorcing?

In all these cases, it is quite possible that an old company will crumble while a startup will turn into a new Microsoft, that an old building will catch fire and collapse while the new one will persist, that the old couple will divorce after 50 years of marriage while the newlyweds will live happily even after – but in all these cases, the opposite is much more frequent.

Why is this called Lindy Effect? This name came from New York's *Lindy's Delicatessen*, which in the 1960s was a favorite gathering place for New York comedians – and in those days, this meant the majority of top US comedians. Once in a while, a new comedian would burst into the stage, so a natural question was: will he (it was usually a he) last for long? Young people may have believed in every single newcomer's success, but more experience folks – who remembered that many new promising comedians did not last long – would cool down the younger folks' optimism.

Lindy Effect: towards formalization. In all the above situations – and in many similar ones:

- we know that some object has already survived for T years, and
- we are trying to predict the amount of time t during which it will most probably survive in the future as well.

Alternatively, we can say that we want to predict the overall survival time $S \stackrel{\text{def}}{=} T + t$.

If the value T is all we know, then we need to estimate the future value t based only on this information. Let us denote the corresponding estimate by $t = f(T)$. Which function $f(T)$ should we use for this estimation?

Lindy Effect: qualitative idea. The above informal discussion enables us to conclude that the larger survival-so-far time T, the largesr should be our estimate $t = f(T)$. In other words, the desired estimation function $f(T)$ should be increasing.

However, there are many increasing functions. Which one should we choose?

Lindy Effect: precise formulation(s). The first person who tried to come up with a precise formula for the Lindy Effect was Benoit Mandelbrot – the father of fractals. By considering several actual situations, he concluded that the desired dependence is linear: there exists a constant $c > 0$ such that if a system survived for T years, it will, with high probability, survive for another $t = c \cdot T$ years; see, e.g., [8].

Later, Nassim Nicholas Taleb analyzed even more cases and concluded that we can safely take $c = 1$ and $t = T$; see, e.g., [13]. In plain Englisjh, this means that if a company survived for 100 years, it is reasonable to expect that it will survive for another 100 years.

Weak and Strong Lindy Effect. We have two versions of Lindy Effect:

- The first version – that $t = c \cdot T$ for some $c > 0$ – is somewhat more accurate, since we have a parameter here that we can adjust to make a better fit.
- The second version – that $t = T$ – is somewhat less accurate but stronger.

To distinguish between these two formulations, we will call the dependence $t = c \cdot T$ a *weak* Lindy Effect and the dependence $t = T$ the *strong* Lindy effect.

Why? Both formulations seem to be consistent with data, so they are real. The fact that they are ubiquitous, that they cover all kinds of phenomena, seems to indicate that there must be a general first-principles explanation for this effect.

What we do in this paper. In this paper, we will try to come up with this explanations.

All this is very imprecise ("fuzzy"), so a natural idea is to try to use fuzzy techniques; see, e.g., [2], [6], [9]–[11], [15]. On the complications side, we will see that in the process of these tries, we will encounter a need to somewhat modify the way such problems are usually described by fuzzy techniques.

The resulting complications will not be fully in vain: they will enable us to come up with a natural way to avoid meaningless infinities in computations related to physics.

II. Let Us Use Fuzzy Techniques: A Straightforward Approach and Why It Does Not Work in This Case

Starightforward approach: idea. At first glance, we have a typical problem of the type solved by fuzzy techniques – e.g., in fuzzy control. We have rules which are imprecise – in the sense that by themselves, they do not lead to an exact answer.

- In the control case, we may have rules like "if x is small, then control should be small" – which allow many different control values (as long as they are small).
- In our case, all we know is that the overall survival time S should be larger than the survival-so-far time T. This also allows many different values S – as long as they are larger than T.

In fuzzy control, the fuzzy methodology means that:

- we describe the knowledge in terms of fuzzy degrees,
- we come up with a fuzzy recommendation, and then
- we apply an appropriate defuzzification procedure to come up with the numerical recommendation.

Let us try to apply the same idea to our problem.

Straightforward approach: let us try. If all we know is the value T, and the only thing that we know about the desired value S is that $S > T$, then the corresponding membership function $\mu(S)$ describing this knowledge is straightforward:

- it assigns $\mu(S) = 1$ to all the values $S > T$, and
- it assigns $\mu(S) = 0$ to all other values.

So far so good, but the problem starts when we try to apply defuzzification.

The most natural idea is to select the value in which we have most confidence, i.e., for which the corresponding value of the membership function is the largest. In our case, this does help at all: the largest value $\mu(S) = 1$ is attained for *all* numbers $S > T$, so this idea does not allow us to select any specific value at all.

OK, this happens in fuzzy control as well. To avoid this non-uniqueness, fuzzy control applications usually use *centroid defuzzification*, i.e., transform a membership function $\mu(x)$ into a value

$$\overline{x} = \frac{\int x \cdot \mu(x)\, dx}{\int \mu(x)\, dx}.$$

Of course, we cannot directly apply this formula to our membership function $\mu(S)$, since for this function, both integrals – in the numerator and in the denominator – are infinite. However, what we *can* do is to consider our function $\mu(S)$ as the limit of functions $\mu_n(S)$ which coincide with $\mu(S)$ up to $S = T+n$ and are equal to 0 after that. In the limit $n \to \infty$, the functions $\mu_n(S)$ tend to the desired function $\mu(S)$. So, it makes sense:

- first, to apply defuzzification to each of these functions $\mu_n(S)$, resulting in values $\overline{S}_n$, and
- then use the limit $\overline{S} = \lim\limits_{n\to\infty}$ of the resulting values $\overline{S}_n$ as the desired estimate for $\overline{S}$.

Unfortunately, this does not work either: for each function $\mu_n(S)$, centroid defuzzification leads to $\overline{S}_n = T + (n/2)$, and thus, the limit $\overline{S} = \lim\limits_{n\to\infty} = \infty$. Mathematically, it is correct, but it does not convey the meaning that we want: instead of saying that a company will survive for $c \cdot T$ more years, this conclusion says that the company will last forever.

We know that this is not true: many companies do survive for a long time, but most of them eventually stop functioning. There are not that many companies that have survived for many centuries: maybe Lloyd insurance is the only one.

III. Let Us Add Common Sense to Mathematics

So what can we do? At first glance, the above negative results may sound like a paradox that shows limitations of the fuzzy approach. But a deeper analysis shows that nothing wrong with fuzzy approach, it is that we relied too much on mathematics and did not use enough common sense.

Specifically, we naively assumed that $\mu(S) = 1$ for all $S > T$. Mathematically, it makes sense, but do we really believe – with confidence 1 – that a company that survived for 100 years will survive for 1000 years more? If you believe this, how about 1 million years? 1 billion years? Clearly not.

From the viewpoint of common sense, the value of the membership function $\mu(S)$ describing a seemingly crisp property $S > T$ should not stay constant, but should instead decrease as S increases.

What is an adequate membership function: analysis of the problem. We are interested in designing, for each T, a membership function $\mu_T(S)$ that describes our degree of belief that, once the system has survived for time T, it will survive for a longer time $S \ge T$.

What should be reasonable properties of these functions?

First, we know for sure that the system has survived for time T, so we should have $\mu_T(T) = 1$.

Second, the longer the time S, the smaller is our belief that the system will survive for this time. Thus, for each T, the function $\mu_T(S)$ should be decreasing. We will call this property *monotonicity*.

Third, if we originally observed the system surviving for time T, and then later, it turns out that it has survived for time $T' > T$, this means that from the original function $\mu_T(S)$, we should only consider values $S \ge T'$. Of course, since the function $\mu_T(S)$ is decreasing, the largest remaining value is the value $\mu_T(T')$ which is smaller than $\mu_T(T) = 1$. In fuzzy techniques, we usually consider *normalized* membership functions, i.e., functions whose maximum is 1. So, to obtain the appropriate function $\mu_{T'}(S)$, we need to normalize the resulting restriction of the original function $\mu_T(S)$ to values $S \ge T'$. Normalization is usually performed by dividing all the membership degrees by the largest one – which, is in this case, is equal to $\mu_T(T')$. Thus, we must have

$$\mu_{T'}(S) = \frac{\mu_T(S)}{\mu_T(T')}$$

for all $S \ge T'$. We will call this property *consistency*.

Finally, since we are trying to understand the phenomenon of Lindy Effect, which is reasonable universal, we want the expressions $\mu_T(S)$ to be *universal*. In particular, it means that this effect should be the same whether we consider micro-objects or macro-objects or mega-objects (how long will the Sun continue to shine?). The corresponding membership degrees should thus not change is we simply change the units in which we measure time. If we replace the original unit of time with the one which is λ times smaller, then numerical values of both T and S and multiplied by λ: we get $\lambda \cdot T$ instead of T and $\lambda \cdot S$ instead of S. In these terms, universality means that $\mu_{\lambda\cdot T}(\lambda \cdot S) = \mu_T(S)$.

Definitions and the main result. Now, we are ready to formulate our first result.

Definition 1. *By a* family of membership functions corresponding to $>$, *we mean a family of membership functions $\mu_T(S)$ with parameter $T > 0$ each of which is defined for all $S \ge T$ and which satisfy the following properties:*

- *for each T, we have $\mu_T(T) = 1$;*
- *for each T, the function $\mu_T(S)$ is decreasing with S (*monotonicity*);*
- *for each $T < T' \le S$, we have*
$$\mu_{T'}(S) = \frac{\mu_T(S)}{\mu_T(T')}; \quad \text{(consistency), and}$$
- *for each $T \le S$ and for each $\lambda > 0$, we have*
$$\mu_{\lambda\cdot T}(\lambda \cdot S) = \mu_T(S) \quad \text{(universality).}$$

Proposition 1. *Every family of membership functions corresponding to $>$ has the form $\mu_T(S) = \left(\dfrac{T}{S}\right)^\alpha$ for some $\alpha > 0$.*

Proof. For $T = 1$ and $\lambda \ge 1$, universality implies that

$$\mu_\lambda(\lambda \cdot S) = \mu_1(S).$$

On the other hand, due to consistency, with $T = 1 < T' = \lambda$, we have

$$\mu_\lambda(\lambda \cdot S) = \frac{\mu_1(\lambda \cdot S)}{\mu_1(\lambda}).$$

Equating the resulting two expressions for the same value $\mu_\lambda(\lambda \cdot S)$, we conclude that

$$\mu_1(S) = \frac{\mu_1(\lambda \cdot S)}{\mu_1(\lambda)},$$

i.e., equivalently,

$$\mu_1(\lambda \cdot S) = \mu_1(\lambda) \cdot \mu_1(S) \tag{1}$$

In particular, for $S=\lambda^{-1}$, we get

$$1=\mu_1(1)=\mu_1(\lambda)\cdot\mu_1(\lambda^{-1}),$$

hence

$$\mu_1(\lambda^{-1})=\frac{1}{\mu_1(\lambda)}. \quad (2)$$

For each $\lambda<1$ and S, and for $S'=\lambda\cdot S$ and $\lambda'=1/\lambda>1$, the formula (1) leads to $\mu_1(\lambda'\cdot S')=\mu_1(\lambda')\cdot\mu_1(S')$, i.e., $\mu_1(S)=\mu_1(1/\lambda)\cdot\mu_1(\lambda\cdot S)$, and thus, due to (2), to the formula (1).

For $\lambda=1$, the property (1) is trivially true. Thus, the property (1) is satisfied for all $\lambda>0$ and for all S.

Functions that satisfy this property are known as *multiplicative*, and it is known that every monotonic multiplicative function has the form $\mu_1(x)=x^{-\alpha}$ for some real value α; see, e.g., [1]. Since all membership functions $\mu_T(S)$ are decreasing, we must have $\alpha>0$.

For each $T\le S$, we can then use the universality property with $\lambda=T^{-1}$ and get $\mu_T(S)=\mu_1(S/T)$, thus $\mu_T(S)=(S/T)^{-\alpha}$. The proposition is proven.

This explains (weak) Lindy Effect. To make sure that for the membership function $\mu_T(S)=\left(\dfrac{T}{S}\right)^\alpha$, both numerator and denominator integrals in the formula for centroid defuzzification are finite, we must have $\alpha>2$. In this case,

$$\int_T^\infty S\cdot\left(\frac{T}{S}\right)^\alpha dS=T^\alpha\cdot\frac{1}{\alpha-2}\cdot T^{2-\alpha}=\frac{1}{\alpha-2}\cdot T^2$$

and

$$\int_T^\infty \left(\frac{T}{S}\right)^\alpha dS=T^\alpha\cdot\frac{1}{\alpha-1}\cdot T^{1-\alpha}=\frac{1}{\alpha-1}\cdot T,$$

thus

$$\overline{S}=\frac{\int S\cdot\mu_T(S)\,dS}{\int\int\mu_T(S)\,dS}=\frac{\alpha-1}{\alpha-2}\cdot T.$$

Thus, the remaining time $t=S-T$ is indeed proportional to T, which is exactly what we called weak Lindy Effect.

IV. What About Probabilistic Case

Probabilistic case: (almost) the same result. In the previous section, we considered the case when we use fuzzy logic to describe the corresponding uncertainty. What if instead we use probabilities?

In this case, for each T, we have the probability $p_T(S)$ that the system will survive for time S once it has survived for time T. The same arguments as in the fuzzy case show that this function:

- should also satisfy the condition $p_T(T)=1$,
- should also be decreasing as S increases, and
- should also not depend on the choice of the measuring unit, i.e., we should have $p_{\lambda\cdot T}(\lambda\cdot S)=p_T(S)$ for all $T\le S$ and $\lambda>0$.

And if we have already observed the system for time $T'>T$ and the system survived during this time, then the new probabilities $p_{T'}(S)$ should be computed by using the formulas for conditional probability: $p_{T'}(S)=\dfrac{p_T(S)}{p_T(T')}$.

Thus, the new functions should satisfy the same conditions as described in Definition 1, and thus, by Proposition 1. it should have the same form $p_T(S)=\left(\dfrac{T}{S}\right)^{-\alpha}$ for some $\alpha>0$.

In the probabilistic case, a natural numerical estimate is the mean value $\overline{S}=\int S\cdot\rho_T(S)\,dS$, where the probability density function $\rho_T(S)$ can be obtained by differentiating the function $p_T(S)$ – which is, in effect, equal to 1 minus the cumulative distribution function; see, e.g., [12]. In this case, we get $\overline{S}=\dfrac{\alpha-1}{\alpha}$. So, in this case, we also get the weak Lindy Effect.

Why do fuzzy and probabilistic approaches lead, in effect, to the same formula? The fact that by using such different techniques as fuzzy and probabilistic, we get the exact same result – that the expected remaining survival time t is proportional to the survival-so-far time T – is a good indication that there is an even more fundamental reason behind this dependence, reason not depending on which technique we use to describe uncertainty.

And indeed, such a reason is easy to describe: the reason is what we called *unversality*, that the result should not depend on the choice of the measuring unit. Our original problem was to find the estimate $t=f(T)$. In terms of the estimating function $f(x)$, universality means that if we have $t=f(T)$ in the original units, then the same relation $t'=f(T')$ should hold if we describe the times in the new units, i.e., if we take $t'=\lambda\cdot t$ and $T'=\lambda\cdot T$.

Formulating the problem in precise terms. Let us describe this requirement in precise terms.

Definition 2. *We say that the function $t=f(T)$ is* universal *if for all t, T, and $\lambda>0$, the equality $t=f(T)$ implies that $t'=f(T')$, where $t'=\lambda\cdot t$ and $T'=\lambda\cdot T$.*

Proposition 2. *Every universal function has the form $f(T)=c\cdot T$ for some constant c.*

Proof. Let us denote $f(1)$ by c, so that $c=f(1)$. Then, for each T, if we take $\lambda=T$, then universality enables us to imply that $T\cdot c=f(T\cdot 1)$, i.e., that indeed $f(T)=c\cdot T$. The proposition is proven.

Discussion. So, indeed, universality implies the weak Lindy Effect.

V. Why Strong Lindy Effect

Reminder. In the above text, we explained the *weak* Lindy Effect, according to which the remaining survival time t is related to the survival-so-far time T by the formula $t=c\cdot T$, for some constant t. However, as we have mentioned, there is strong evidence that this constant c is equal to 1, i.e., that we have what we called the *strong* Lindy Effect $t=T$.

How can we explain this?

A simplified (somewhat naive) explanation. A simplified explanation comes from Laplace Indeterminacy Principle (see, e.g., [5]), according to which if we have no reason to believe that two quantities are different, it makes sense to assume that they are equal.

From this viewpoint, since we do not have any reason to believe that the remaining survival time t is smaller or larger than the survival-so-far time T, so it makes sense to take

$$t = T.$$

A better explanation: fuzzy case. In our problem, we know the value T, and know that $T < S$. In this case, as we have mentioned earlier, the straightforward fuzzy approach does not lead to any meaningful estimate for S.

But what if we *reverse* the problem: what is we assume that S is known, and the only information that we have about T is that $0 \le T \le S$. In this case, the corresponding (crisp) knowledge leads to the following membership function: $\mu_S(T) = 1$ when $0 \le T \le S$ and $\mu_S(T) = 0$ otherwise. For this membership function, centroid defuzzification leads to $\overline{T} = S/2$.

So, if we know S, then we should take $T = S/2$. It is therefore natural to conclude that if we know T, then we should take S for which $T = S/2$. For this S, we have $S = 2T$, so the remaining survival time is $t = S - T = T$, which is exactly the strong Lindy Effect.

Probabilistic case. We can apply the same reversal idea to the case of probabilistic uncertainty.

Suppose that we know the value S, and the only information that we have about T is that T is between 0 and S. In this case, the maximum likelihood approach – a natural formalization of the Laplace Indeterminacy Principle – implies that the corresponding probability distribution on the interval $[0, S]$ is uniform [5]. For this uniform distribution, the mean value is $\overline{T} = S/2$, which also prompts us to use the estimate $S = 2T$ and thus, $t = T$.

Comment. In [4], [7], a similar idea was used to explain why in engineering, after we get an estimate of uncertainty based on known factors, practitioners usually double this estimate to take into account possible unknown factors as well.

This leads, e.g., to doubling the safety margins computed based only on the known factors.

VI. Application to Physics: How to Avoid Physically Meaningless Infinite Values

Problem: reminder. It is known that in physics, some computations lead to meaningless infinite values. The simplest example of such a phenomenon is computing the overall energy of an electron's electric field; see, e.g., [3], [14] for detail.

An electron is an elementary particle, which means that it has no independent parts. According to special relativity, all velocities are bounded by the speed of light. Thus, if the electron was not point-wise, if it had at least two spatially separated points, then it would take some time for these points to influence each other – and therefore, during this time, these two points would act independently. So, an electron has to be a point-wise particle.

For a point-wise particle, the value of its electric field $\vec{E}$ at any point x is determined by the Coulomb Law, as proportional to the r^{-2}, where r is the distance between this point and the location of the electron.

It is known that the energy density $\rho(x)$ is proportional to the square of the electric field, i.e., to r^{-4}. The overall energy E can be computed by integrating this density over the whole 3-D space: $E = \int \rho(x), dx$. The problem is the resulting integral is infinite:

$$E = \int r^{-4}\,dx = \int_0^\infty r^{-4} \cdot 4\pi \cdot r^2\,dr = 4\pi \cdot \int_0^\infty r^{-1}\,dr = 4\pi \cdot r^{-1}|_\infty^0 = \infty.$$

So, we get a physically meaningless value for a physically meaningful quantity – the overall energy of the electron's electric field.

How can we made the corresponding estimate physically meaningful – i.e., finite?

Comment. There are many such infinities in classical physics – the existence of such infinities was one of the main reasons why quantum physics was discovered in the first place. However, in contrast to many other cases when the answer become finite in the quantum case, for the overall energy of the electron's electric field remains infinite in the quantum cases as well.

Known idea. A previously proposed possible way to solve this problem is to take into account that measurements are always imprecise, that at any given moment of time, there is a limit on how accurately we can measure, e.g., the distance – and probably there is a fundamental limit.

So, instead of the actual distance r, we can only conclude that the actual distance is between $r - \varepsilon$ (to be more precise, $\max(0, r - \varepsilon)$, since the distance cannot be negative) and $r + \varepsilon$ for some ε. Thus, the value of the electric field at any point x is somewhere between $(r+\varepsilon)^{-2}$ and $(\max(0, r-\varepsilon))^{-2}$, and, correspondingly, the overall energy is between

$$\underline{E} = \int (r+\varepsilon)^{-4}\,dx \text{ and } \overline{E} = \int (\max(0, r-\varepsilon))^{-4}\,dx.$$

One can check that the first integral $\underline{E}$ is finite – for small r, the integrated function $(r+\varepsilon)^{-4}$ is bounded from above by the value ε^{-4}. However, the second integral is clearly infinite – since for $r \le \varepsilon$, we have $\max(0, r - \varepsilon) = 0$ and thus,

$$(\max(0, r-\varepsilon))^{-4} = \infty.$$

So, instead of the infinite value for the total energy E of the electron's electric field, we have a semi-infinite interval of possible values $[\underline{E}, \infty)$. In other words, the only information that we have about the overall energy is that it is larger than or equal to $\underline{E}$.

Lindy Effect helps. The situation when the only information that have about an unknown quantity S is that it is larger than or equal to some known quantity T is exactly the situation described by the Lindy Effect.

According to the Lindy Effect – which we explained in this paper – in such a situation, the appropriate estimate for the unknown value E is a finite estimate $E = c \cdot \underline{E}$ (where it is highly probable that $c = 1$).

So, we have a finite estimate for the overall energy – thus avoiding the meaningless infinity.

References

[1] J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.

[2] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.

[3] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[4] A. Gholamy and V. Kreinovich, "Safety factors in soil and pavement engineering: theoretical explanation of empirical data", *Abstracts of the 23rd Joint UTEP/NMSU Workshop on Mathematics, Computer Science, and Computational Sciences*, El Paso, Texas, November 3, 2018.

[5] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

[6] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[7] J. Lorkowski and V. Kreinovich, "How to gauge unknown unknowns: a possible theoretical explanation of the usual safety factor of 2", *Mathematical Structures and Modeling*, 2014, Vol. 32, pp. 49–52.

[8] B. B. Mandebrot, *The Fractal Geometry of Nature*, Henry Holt & Co., New York, 1983.

[9] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.

[10] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.

[11] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.

[12] D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures*, Chapman & Hall/CRC, London, UK, 2011.

[13] N. N. Taleb, *Antifragile: Things That Gain from Disorder*, Random House, New York, 2012.

[14] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.

[15] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.