# To Properly Reflect Physicists' Reasoning about Randomness, We Also Need a Maxitive (Possibility) Measure

Andrei M. Finkelstein
Institute of Applied Astronomy
Russian Academy of Sciences
St Petersburg, Russia

Olga Kosheleva,
Vladik Kreinovich,
Scott A. Starks
Pan-American Center for Earth and Environmental Studies
University of Texas at El Paso
El Paso, TX 79968, USA
Email: vladik@cs.utep.edu

Hung T. Nguyen
Dept. of Mathematical Sciences
New Mexico State University
Las Cruces, NM, 88003, USA
Email: hunguyen@nmsu.edu

***Abstract*— According to the traditional probability theory, events with a positive but very small probability can occur (although very rarely). For example, from the purely mathematical viewpoint, it is possible that the thermal motion of all the molecules in a coffee cup goes in the same direction, so this cup will start lifting up.**

**In contrast, physicists believe that events with extremely small probability cannot occur. In this paper, we show that to get a consistent formalization of this belief, we need, in addition to the original probability measure, to also consider a maxitive (possibility) measure.**

## I. Physicists Assume That Initial Conditions and Values of Parameters Are Not Abnormal

To a mathematician, the main contents of a physical theory is the equations. The fact that the theory is formulated in terms of well-defined mathematical equations means that the actual field must satisfy these equations. However, this fact does *not* mean that *every* solution of these equations has a physical sense. Let us give three examples:

**Example 1.** At any temperature greater than absolute zero, particles are randomly moving. It is theoretically possible that all the particles start moving in one direction, and, as a result, a person starts lifting up into the air. The probability of this event is small (but positive), so, from the purely mathematical viewpoint, we can say that this event is possible but highly unprobable. However, the physicists say plainly that such an abnormal event is *impossible* (see, e.g., [3]).

**Example 2.** Another example from statistical physics: Suppose that we have a two-chamber camera. The left chamber is empty, the right one has gas in it. If we open the door between the chambers, then the gas would spread evenly between the two chambers. It is theoretically possible (under appropriately chosen initial conditions) that the gas that was initially evenly distributed would concentrate in one camera. However, physicists believe this abnormal event to be impossible. This is an example of a "micro-reversible" process: on the atomic level, all equations are invariant with respect to changing the order of time flow ($t \to -t$). So, if we have a process that goes from state $A$ to state $B$, then, if while at $B$, we revert all the velocities of all the atoms, we will get a process that goes from $B$ to $A$.

However, in real life, many processes are clearly irreversible: an explosion can shatter a statue but it is hard to imagine an inverse process: an implosion that glues together shattered pieces into a statue. Boltzmann himself, the 19th century author of statistical physics, explicitly stated that such inverse processes "may be regarded as impossible, even though from the viewpoint of probability theory that outcome is only extremely improbable, not impossible." [1].

**Example 3.** If we toss a fair coin 100 times in a row, and get heads all the time, then a person who is knowledgeable in probability would say that it is possible – since the probability is still positive. On the other hand, a physicist (or any person who uses common sense reasoning) would say that the coin is not fair – because if it is was a fair coin, then this abnormal event would be impossible.

In all these cases, physicists (implicitly or explicitly) require that the actual values of the physical quantities must not only satisfy the equations but they must also satisfy the additional condition: that the initial conditions should *not* be *abnormal*.

*Comment.* In all these examples, a usual mathematician's response to physicists' calling some low-probability events "impossible", is just to say that the physicists use imprecise language.

It is indeed true that the physicists use imprecise language, and it is also true that in the vast majority of practical applications, a usual probabilistic interpretation of this language perfectly well describes the intended physicists' meaning. In other words, the probability language is perfectly OK for most physical applications.

However, there are some situations when the physicists' intuition seem to differ from the results of applying traditional probability techniques:

- From the probability theory viewpoint, there is no fundamental difference between such low-probability events as a person winning a lottery and the same person being lifted up into the air by the Brownian motion. If a person plays the lottery again and again, then – provided that this person lives for millions of years – he will eventually win. Similarly, if a person stands still every morning, then – provided that this person lives long enough – this person will fly up into the air.
- On the other hand, from the physicist viewpoint, there is a drastic difference between these two low-probability events: yes, a person will win a lottery but no, a person will never lift up into the air no matter how many times this person stands still.

We have just mentioned that the traditional mathematical approach is to treat this difference of opinion as simply caused by the imprecision of the physicists' language. What we plan to show is that if we take this difference more seriously and develop a new formalism that more accurately captures the physicists' reasoning, then we may end up with results and directions that are, in our opinion, of potential interest to foundations of physics. In other words, what we plan to show is that if we continue to use the traditional probability approach, it is perfectly OK but if we try to formalize the physicists' opinion more closely, we may sometimes get even better results.

## II. A SEEMINGLY NATURAL FORMALIZATION OF THIS IDEA

The above-mentioned property of being "not abnormal" ("typical") has a natural formalization: if a probability $p(E)$ of an event $E$ is small enough, i.e., if $p(E) \leq p_0$ for some very small threshold $p_0$, then this event cannot happen.

In other words, there exists the "smallest possible probability" $p_0$ such that:

- if the computed probability $p$ of some event is larger than $p_0$, then this event can occur, while
- if the computed probability $p$ is $\leq p_0$, the event cannot occur.

For example, the probability that a fair coin falls heads 100 times in a row is $2^{-100}$, so, if the threshold probability $p_0$ satisfies the inequality $p_0 \geq 2^{-100}$, then we will be able to conclude that such an event is impossible.

## III. THE ABOVE FORMALIZATION OF THE NOTION OF "TYPICAL" IS NOT ALWAYS ADEQUATE

In the previous section, we described a seemingly natural formalization of the notion "typical" ("not abnormal"): if a probability of an event is small enough, i.e., $\leq p_0$ for some very small $p_0$, then this event cannot happen.

The problem with this approach is that *every* sequence of heads and tails has exactly the same probability. So, if we choose $p_0 \geq 2^{-100}$, we will thus exclude all possible sequences of 100 heads and tails as physically impossible. However, anyone can toss a coin 100 times, and this proves that some such sequences are physically possible.

*Historical comment.* This problem was first noticed by Kyburg under the name of *Lottery paradox* [5]: in a big (e.g., statewide) lottery, the probability of winning the Grand Prize is so small that a reasonable person should not expect it. However, some people do win big prizes.

## IV. KOLMOGOROV'S IDEA: USE COMPLEXITY

Crudely speaking, the main problem arises because we select the same threshold $p_0$ for all events. For example, if we toss a fair coin 100 times then a sequence consisting of all heads should not be possible, and it is a reasonable conclusion because the probability that tossing a fair coin will lead to this sequence is extremely small: $2^{-100}$.

On the other hand, whatever specific sequence of heads and tails we get after tossing a coin, this sequence also has the same small probability $2^{-100}$. In spite of this, it does not seem to be reasonable to dismiss such sequences.

Several researchers thought about this, one of them A. N. Kolmogorov, the father of the modern probability theory. Kolmogorov came up with the following idea: the probability threshold $t(E)$ below which an event $E$ is dismissed as impossible must depend on the event's complexity. The event $E_1$ in which we have 100 heads is easy to describe and generate; so for this event, the threshold $t(E_1)$ is higher. If $t(E_1) > 2^{-100}$ then, within this Kolmogorov's approach, we conclude that the event $E_1$ is impossible. On the other hand, the event $E_2$ corresponding to the actual sequence of heads and tails is much more complicated; for this event $E_2$, the threshold $t(E_2)$ should be much lower. If $t(E_2) < 2^{-100}$, we conclude that the event $E_2$ is possible.

The general fact that out of $2^n$ equally probable sequences of $n$ 0s and 1s some are "truly random" and some are not truly random was the motivation behind Kolmogorov and Martin-Löf's formalization of randomness (and behind the related notion of Kolmogorov complexity; the history of this discovery is described in detail in [6]).

This notion of Kolmogorov complexity was introduced independently by several people: Kolmogorov in Russia and Solomonoff and Chaitin in the US. Kolmogorov defined complexity $K(x)$ of a binary sequence $x$ as the shortest length of a program which produces this sequence. Thus, a sequence consisting of all 0s or a sequence 010101... both have very small Kolmogorov complexity because these sequences can be generated by simple programs; on the other hand, for a sequence of results of tossing a coin, probably the shortest program is to write `print(0101...)` and thus reproduce the entire sequence. Thus, when $K(x)$ is approximately equal to the length $\text{len}(x)$ of a sequence, this sequence is random, otherwise it is not. (The best source for Kolmogorov complexity is a book [6].)

However, the existing Kolmogorov complexity theory does not yet lead to a formalism describing when low-probability events do not happen; we must therefore extend the original Kolmogorov's idea so that it would cover this case as well.

## V. Formalization and the Main Result

Let us start with motivations. We have mentioned that we cannot consistently claim that an event $E$ is possible if and only its probability $p(E)$ exceeds a certain threshold $p_0$; instead, we must take into consideration that "complexity" $c(E)$ of an event, and claim, e.g., that an event $E$ is possible if and only if $p(E) > p_0 \cdot c(E)$, i.e., equivalently, $m(E) > p_0$, where we denoted $m(E) \stackrel{\text{def}}{=} p(E)/c(E)$.

*Comment.* To handle events with 0 probability, we must extend the ratio $m(E)$ to such events – otherwise, e.g., for the uniform distribution on the interval $[0,1]$, we would have $p(\{x\}) = 0 \le p_0 \cdot c(\{x\})$ hence no point $x$ would be possible.

We would like to characterize the "ratio measures" $m(E)$ for which this definition is, in some reasonable sense, consistent for all possible thresholds $p_0$. In order to do that, let us first find out how to formalize the notion of consistency.

Let $X$ be the set of all possible outcomes. An *event* is then simply a subset $E$ of the set $X$, and $p$ is a probability measure on a $\sigma$-algebra of sets from $X$.

Let $T \subseteq X$ be the set of all outcomes that are actually possible. Then, an event $E$ is possible if and only if there is a possible outcome that belongs to the set $E$, i.e., if and only if $E \cap T \neq \emptyset$.

Now, we are ready for the main definition:

**Definition 1.** *Let $X$ be a set, and let $p$ be a probability measure on a $\sigma$-algebra $\mathcal{A} \subseteq 2^X$ of subsets of the set $E$. By a* ratio measure *$m$ we mean a mapping from $\mathcal{A}$ to the set of non-negative real numbers (and, possibly a value $+\infty$) such that for every real number $p_0 > 0$, there exists a set $T(p_0)$ for which*

$$\forall E \in \mathcal{A}\,(m(E) > p_0 \leftrightarrow E \cap T(p_0) \neq \emptyset). \tag{1}$$

To describe our main result, we need to recall the definition of a maxitive (possibility) measure [2], [7], [8]:

**Definition 2.** *A mapping $m$ from sets to real numbers (and possibly a value $+\infty$) is called a* maxitive *(*possibility*) measure if for every family of sets $X_\alpha$ for which $m(X_\alpha)$ and $m(\cup X_\alpha)$ are defined, we have*

$$m\left(\bigcup_\alpha X_\alpha\right) = \sup_\alpha m(X_\alpha).$$

**Theorem 1.** *For a given probability measure $p(E)$, a function $m(E)$ is a ratio measure if and only if it is a maxitive (possibility) measure.*

*Comment.* Since $m(E) = p(E)/c(E)$ is a possibility measure, we thus have $c(E) = m(E)/p(E)$. In other words,

$$\text{complexity} = \frac{\text{possibility}}{\text{probability}}.$$

**Proof.** Let us first prove that every ratio measure $m(E)$ is indeed a maxitive measure. By definition of a maxitive measure, we need to prove that if $X_\alpha$ is a family of sets from the $\sigma$-algebra $\mathcal{A}$ for which the union $X = \cup X_\alpha$ also belongs to $\mathcal{A}$, we have $m(X) = \sup\limits_\alpha m(X_\alpha)$.

Let us prove this inequality by reduction to a contradiction. Let us assume that $m(X) \neq \sup\limits_\alpha m(X_\alpha)$. In this case, we have two options:

- $m(X) > \sup\limits_\alpha m(X_\alpha)$ and
- $m(X) < \sup\limits_\alpha m(X_\alpha)$.

Let us show that in both cases, we have a contradiction.

Indeed, by definition of a ratio measure, for every $p_0$, there exists a set $T(p_0)$ such that for every set $E$, we have $m(E) > p_0$ if and only if $E \cap T(p_0) > 0$.

If $m(X) < \sup\limits_\alpha m(X_\alpha)$, let us select $p_0$ for which

$$m(X) < p_0 < \sup_\alpha m(X_\alpha).$$

Since $m(X) < p_0$, we conclude that the event $X$ is not possible, i.e.,

$$X \cap T(p_0) = \emptyset. \tag{2}$$

On other hand, since $p_0 < \sup\limits_\alpha m(X_\alpha)$, there exists a value $\alpha_0$ for which $p_0 < m(X_{\alpha_0})$. For this $\alpha_0$, by definition of a ratio measure, the event $X_{\alpha_0}$ is possible, so there exists an outcome $x$ from $X_{\alpha_0}$ that also belongs to the set $T(p_0)$ of possible events. However, since $X = \cup X_\alpha$, we have $x \in X$, so $x \in X \cap T(p_0)$ – which contradicts our previous conclusion (2). This contradiction shows that the inequality $m(X) < \sup\limits_\alpha m(X_\alpha)$ is impossible.

If $m(X) > \sup\limits_\alpha m(X_\alpha)$, let us select $p_0$ for which

$$m(X) > p_0 > \sup_\alpha m(X_\alpha).$$

Since $m(X) > p_0$, we conclude that the event $X$ is possible, i.e., there exist an outcome $x$ that belongs both to $X$ and to $T(p_0)$. Since $X$ is the union of the set $X_\alpha$, this event $x$ belongs to one of the sets $X_{\alpha_0}$. Thus, $X_\alpha \cap T(p_0) \neq \emptyset$, so by definition of a complexity measure, we should have

$$m(X_{\alpha_0}) > p_0 \tag{3}$$

for this $\alpha_0$. However, from our assumption $m(X) > \sup\limits_\alpha m(X_\alpha)$ and from the fact that $\sup\limits_\alpha m(X_\alpha) \ge m(X_{\alpha_0})$, we conclude that $m(X_{\alpha_0}) < p_0$ – a contradiction with our previous conclusion (3). This contradiction shows that the inequality $m(X) > \sup\limits_\alpha m(X_\alpha)$ is also impossible.

Thus, every ratio measure $m(E)$ is indeed a maxitive measure.

To complete the proof of Theorem 1, we must now prove that if $m(E)$ is a maxitive measure, then it is a ratio measure. To prove this, we will show that for every positive real number

$p_0$, there exists a set $T(p_0)$ that satisfies the condition (1). We will show that as such a set, we can take a complement to the union of all sets $S \in \mathcal{A}$ for which $m(S) \le p_0$, i.e.,

$$T(p_0) = - \cup \{S \in \mathcal{A} \,|\, m(S) \le p_0\}. \quad (4)$$

We must prove that for every $E \in \mathcal{A}$, $E \cap T(p_0) \neq \emptyset$ if and only if $m(E) > p_0$. Actually, we will prove an equivalent statement: that for every $E \in \mathcal{A}$, $E \cap T(p_0) = \emptyset$ if and only if $m(E) \le p_0$.

If $m(E) \le p_0$, then $E$ is completely contained in the union $\cup\{S \in \mathcal{A} \,|\, m(S) \le p_0\}$, thus, $E$ cannot have common points with the complement $T(p_0)$ to this union.

Vice versa, let us assume that for some event $E \in \mathcal{A}$, we have $E \cap T(p_0) = \emptyset$. This means that the set $E$ is completely contained in the complement to $T(p_0)$, i.e., that

$$E \subseteq \cup\{S \in \mathcal{A} \,|\, m(S) \le p_0\}.$$

Thus,

$$E = \cup\{S \cap E \,|\, S \in \mathcal{A} \,\&\, m(S) \le p_0\}. \quad (5)$$

If the set $S$ and $E$ belongs to a $\sigma$-algebra, then their intersection and their difference also belong to the $\sigma$-algebra. From $S = (S \cap E) \cup (S - E)$ and the definition of a maxitive measure, we thus conclude that $m(S) = \max(m(S \cap E), m(S - E))$ hence $m(S \cap E) \le m(S)$. So, if $m(S) \le p_0$, we have $m(S \cap E) \le m(S) \le p_0$ hence $m(S \cap E) \le p_0$.

Applying the definition of a maxitive measure to the formula (5), we can now conclude that $m(E) = \sup m(S \cap E)$, where supremum is taken over all $S \in \mathcal{A}$ for which $m(S) \le p_0$. We have already shown that for all such $S$, we have $m(S \cap E) \le p_0$. Thus, $m(E)$ is the supremum of a set of numbers each of which is $\le p_0$. We can therefore conclude that $m(E) \le p_0$.

The theorem is proven.

## VI. Auxiliary result

Our definition of complexity depends on the choice of the probability measure. In other words, complexity of an event depends on the problem that we are trying to solve. This makes sense because what we are looking for is complexity relevant to the problem.

However, a natural question is: is it possible to have a "universal" complexity measure, i.e., a complexity measure that will serve all possible probability measures $p(E)$? The answer is "no", even if, instead of all possible thresholds $p_0$, we just consider a single one. This result is true even for $X$ equal to the standard interval $[0, 1]$.

Let us describe this result in precise terms.

**Definition 3.** *Let $X = [0, 1]$, and let $\mathcal{A} \subseteq 2^X$ be a $\sigma$-algebra of all Lebesgue-measurable sets. By a* universal complexity measure *$c$ we mean a mapping from $\mathcal{A}$ to the interval $[0, 1]$ for which $0 < c([a, b]) < 1$ for every interval $[a, b]$, and for every probability measure $p$ on $\mathcal{A}$, there exists a set $T(p)$ for which*

$$\forall E \in \mathcal{A}\, (p(E) > 0 \to (p(E) > c(E) \leftrightarrow E \cap T(p) \neq \emptyset)).$$

**Theorem 2.** *A universal complexity measure is impossible.*

**Proof.** To prove this theorem, we will assume that a universal complexity measure $c(E)$ exists, and from this assumption, we will deduce a contradiction.

First, let us show that if $A \subset B$ are two sets from the $\sigma$-algebra $\mathcal{A}$ for which $B \neq X$, then $c(A) \ge c(B)$. Indeed, let us prove that if $c(A) < c(B)$, then we get a contradiction.

If $c(A) < c(B)$, then we can set up a probability measure $p$ for which $p(A) = c(B) > 0$ and $p(B - A) = 0$. For this probability measure, $p(B) = p(A) + p(B - A) = c(B)$, hence $p(B) \le c(B)$ and $p(B) > 0$. By definition of a universal complexity measure, this means that the set $B$ has no common points with $T(p)$. Since $A$ is a subset of $B$, it also has no common points with $T(p)$; due to $p(A) > 0$, we should have $p(A) \le c(A)$. However, $p(A) = c(B) > c(A)$ – a contradiction shows that the case $c(A) < c(B)$ is impossible.

Let us now show if $A \subset B$ and $c(B) > 0$, then $c(A) = c(B)$. We already know that $c(A) \ge c(B)$. Thus, it is sufficient to show that if $c(A) > c(B)$, then we get a contradiction.

Indeed, since $c(B) > 0$ and $B - A \subseteq B$, we have

$$c(B - A) \ge c(B) > 0.$$

Let $c(A) > c(B)$, then we can set up a probability measure $p$ for which $p(A) = c(A)$ and

$$0 < p(B - A) \le c(B - A).$$

For this probability measure, $p(A) \le c(A)$ and $p(A) > 0$, hence the set $A$ cannot have any common points with $T(p)$. Similarly, since $0 < p(B - A) \le c(B - A)$, the set $B - A$ cannot have any common points with $T(p)$. Since neither the set $A$ nor the set $B - A$ can have common points with $T(p)$, their union $B = A \cup (B - A)$ also cannot have any common points with $T(p)$. According to the definition of a universal complexity measure and the fact that $p(B) > 0$, this would mean that $p(B) \le c(B)$, but $p(B) > c(A) > c(B)$. The contradiction shows that the case $c(A) > c(B)$ is also impossible.

So, $A \subseteq B$ and $c(B) > 0$ imply that $c(A) = c(B)$.

Let $[a, b]$ be an arbitrary interval $\neq [0, 1]$. Then, by definition, $c([a, b]) > 0$, so, for every set $E \subseteq [a, b]$, we have $c(E) = c([a, b])$. Let us select an integer $n > 1/c([a, b])$ and divide the interval $[a, b]$ into $n$ subintervals of equal size. For the uniform distribution on the interval $[a, b]$, the probability $p(E)$ of each subinterval $E$ is equal to $1/n$. Since $n > 1/c([a, b])$, we thus conclude that $p(E) = 1/n \le c([a, b])$, i.e., $p(E) < c(E) = c([a, b])$. Thus, none of these $n$ subintervals can contain elements from $T(p)$. On the other hand, $p([a, b]) = 1 > c([a, b])$ hence the union $[a, b]$ of these $n$ subintervals does contain elements from $T(p)$ – a contradiction.

The theorem is proven.

## Acknowledgments

This work was supported by NASA grant NCC5-209, by USAF grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, by Army Research Laboratories grant DATM-05-02-C-0046, and by the NIH grant 3T34GM008048-20S1.


## References


[1] L. Boltzmann, "Bemrkungen über einige Probleme der mechanischen Wärmtheorie", *Wiener Ber. II*, 1877, Vol. 75, pp. 62–100.

[2] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.

[3] R. P. Feynman, *Statistical Mechanics*, W. A. Benjamin, 1972.

[4] V. Kreinovich and A. M. Finkelstein, "Towards Applying Computational Complexity to Foundations of Physics", *Notes of Mathematical Seminars of St. Petersburg Department of Steklov Institute of Mathematics* (to appear).

[5] H. E. Kyburg, Jr., *Probability and the logic of rational belief*, Wesleyan Univ. Press, 1961.

[6] M. Li and P. M. B. Vitanyi, *An Introduction to Kolmogorov Complexity*, Springer, N.Y., 1997.

[7] A. Puhalskii, *Large Deviations and Idempotent Probability*, Chapman and Hall/CRC, Boca Raton, 2001.

[8] L. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems*, 1978, Vol. 1, pp. 3–28.