Toward Formalizing Non-Monotonic Reasoning in Physics: the Use of Kolmogorov Complexity to Formalize the Notions "Typically" and "Normally"

Vladik Kreinovich

Department of Computer Science, University of Texas at El Paso,
El Paso, TX 79968, USA, vladik@cs.utep.edu

Abstract. When a physicist writes down equations, or formulates a theory in any other terms, he usually means not only that these equations are true for the real world, but also that the model corresponding to the real world is "typical" among all the solutions of these equations. This type of argument is used when physicists conclude that some property is true by showing that it is true for "almost all" cases. There are formalisms that partially capture this type of reasoning, e.g., techniques based on the Kolmogorov-Martin-Löf definition of a random sequence. The existing formalisms, however, have difficulty formalizing, e.g., the standard physicists' argument that a kettle on a cold stove cannot start boiling by itself, because the probability of this event is too small.
We present a new formalism that can formalize this type of reasoning. This formalism also explains "physical induction" (if some property is true in sufficiently many cases, then it is always true), and many other types of physical reasoning.

In the current mathematical formalizations of physics, physically impossible events are sometimes mathematically possible. From the physical and engineering viewpoints, a cold kettle placed on a cold stove will never start boiling by itself. However, from the traditional probabilistic viewpoint, there is a positive probability that it will start boiling, so a mathematician might say that this boiling event is rare but still possible.

In the current formalizations, physically possible indirect measurements are often mathematically impossible. In engineering and in physics, we often cannot directly measure the desired quantity; instead, we measure related properties and then use the measurement results to reconstruct the measured values. In mathematical terms, the corresponding reconstruction problem is called the *inverse problem*. In practice, this problem is efficiently used to reconstruct the signal from noise, to find the faults within a metal plate, etc. However, from the purely mathematical viewpoint, most inverse problems are *ill-defined* meaning that we cannot really reconstruct the desired values without making some additional assumptions.

What we are planning to do. A physicist would explain that in both situations, the counter-examples like a kettle boiling on a cold stove or a weird configuration that is mathematically consistent with the measurement results are *abnormal*. In this paper, we show that if we adequately formalize this notion of abnormality, we will be able to weed out these counterexamples and thus, make the formalization of physics better agreeing with common sense and with the physicists' intuition.

Our approach uses the notion of Kolmogorov complexity.

What is Kolmogorov complexity. This notion was introduced independently by several people: Kolmogorov in Russia and Solomonoff and Chaitin in the US. Kolmogorov used it to formalize the notion of a random sequence. Probability theory describes most of the physicist intuition in precise mathematical terms, but it does not allow us to tell whether a given finite sequence of 0's and 1's is random or not. Kolmogorov defined a 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 0's or a sequence 010101... have a very small Kolmogorov complexity because these sequences can be generated by simple programs, while for a sequence of results of tossing a coin, probably the shortest program is to write print(0101...) and then reproduce the entire sequence. Thus, when $K(x)$ is approximately equal to the length $\mathrm{len}(x)$ of a sequence, this sequence is random, otherwise it is not. The best source for Kolmogorov complexity is a book [10].

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, the chair that I am sitting on 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 if 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, but physicists believe this abnormal event to be impossible. This is a general example of what physicists call *irreversible processes*: 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 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 19 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 flip 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, while an engineer (and 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 the above cases, we knew something about probability. However, there are examples of this type of reasoning in which probability does not enter into picture at all. For example, in general relativity, it is known that for almost all initial conditions (in some reasonable sense) the solution has a singularity point. From this, physicists conclude that the solution that corresponds to the geometry of the actual world has a singularity (see, e.g., [11]): the reason is that the initial conditions that lead to a non-singularity solution are abnormal (atypical), and the actual initial conditions must be not abnormal.

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

How is all this connected with the existing work in NMR, in particular, in probability-based NMR. There is a massive body of work by J. Pearl, H. Geffner, E. Adams, F. Bacchus, Y. Halpern, and others on probability-based non-monotonic reasoning; many of them are cited in a recent book [5] that also describes other existing approaches to non-monotonic reasoning, approaches found in the AI and Knowledge Representation communities. The existing approaches have shown that many aspects of non-monotonic reasoning can indeed be captured by the existing logic-related and probability-related ideas. In this paper, we consider aspects of non-monotonic reasoning expert that are not captured by the previous formalisms, and we produce a new probability-related formalism for capturing these aspects.

In the future, it is desirable to combine our new approach with the existing logic-based and probability-based NMR into a single NMR technique.

The notion of "not abnormal" is difficult to formalize. At first glance, it looks like in the probabilistic case, this property has a natural formalization: if a probability of an event is small enough (say, $\leq p_0$ for some very small p_0), then this event cannot happen. For example, the probability that a fair coin falls heads 100 times in a row is 2^{-100}, so, if we choose $p_0 \geq 2^{-100}$, then we will be able to conclude that such an event is impossible. 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 heads and tails as physically impossible. However, anyone can toss a coin 100 times, and this prove that some sequences are physically possible.

Historical comment. This problem was first noticed by Kyburg under the name of *Lottery paradox* [9]: in a big (e.g., state-wide) lottery, the probability of winning the Grand Prize is so small, then a reasonable person should not expect it. However, some people do win big prizes.

How to formalize the notion of "not abnormal": idea. "Abnormal" means something unusual, rarely happening: if something is rare enough, it is not typical ("abnormal"). Let us describe what, e.g., an abnormal height may mean. If a person's height is ≥ 6 ft, it is still normal (although it may be considered abnormal in some parts of the world). Now, if instead of 6 pt, we consider 6 ft 1 in, 6 ft 2 in, etc, then sooner or later we will end up with a height h such that everyone who is higher than h will be definitely called a person of abnormal height. We may not be sure what exactly value h experts will call "abnormal", but we are sure that such a value exists.

Let us express this idea is general terms. We have a *Universe of discourse*, i.e., a set U of all objects that we will consider. Some of the elements of the set U are abnormal (in some sense), and some are not. Let us denote the set of all elements that are *typical* (*not abnormal*) by T. On the set U, we have a decreasing sequence of sets $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ with the property that $\cap A_n = \emptyset$. In the above example, U is the set of all people, A_1 is the set of all people whose height is ≥ 6 ft, A_2 is the set of all people whose height is ≥ 6 ft 1 in, A_2 is the set of all people whose height is ≥ 6 ft 2 in, etc. We know that if we take a sufficiently large n, then all elements of A_n are abnormal (i.e., none of them belongs to the set T of not abnormal elements). In mathematical terms, this means that for some n, we have $A_n \cap T = \emptyset$.

In case of a coin: U is the set of all infinite sequences of results of flipping a coin; A_n is the set of all sequences that start with n heads but have some tail afterwards. Here, $\cap A_n = \emptyset$. Therefore, we can conclude that there exists an n for which all elements of A_n are abnormal. According to mechanics, the result of flipping a coin is uniquely determined by the initial conditions, i.e., on the initial positions and velocities of the atoms that form our muscles, atmosphere, etc. So, if we assume that in our world, only not abnormal initial conditions can happen, we can conclude that for some n, the actual sequence of results of flipping a coin cannot belong to A_n. The set A_n consists of all elements that start with n heads and a tail after that. So, the fact that the actual sequence does not belong to A_n means that if an actual sequence has n heads, then it will consist of all heads. In plain words, if we have flipped a coin n times, and the results are n heads, then this coin is biased: it will always fall on heads.

Let us describe this idea in mathematical terms [4, 8]. To make formal definitions, we must fix a formal theory: e.g., the set theory ZF (the definitions and results will not depend on what exactly theory we choose). A set S is called

definable if there exists a formula $P(x)$ with one (free) variable x such that $P(x)$ if and only if $x \in S$.

Crudely speaking, a set is definable if we can *define* it in ZF. The set of all real numbers, the set of all solutions of a well-defined equations, every set that we can describe in mathematical terms is definable.

This does not means, however, that *every* set is definable: indeed, every definable set is uniquely determined by formula $P(x)$, i.e., by a text in the language of set theory. There are only denumerably many words and therefore, there are only denumerably many definable sets. Since, e.g., there are more than denumerably many set of integers, some of them are thus not definable.

Definition 1. *A sequence of sets $A_1, \ldots, A_n, \ldots$ is called* definable *if there exists a formula $P(n, x)$ such that $x \in A_n$ if and only if $P(n, x)$.*

Definition 2. *Let U be a universal set.*

- *A non-empty set $T \subseteq U$ is called a* set of typical (not abnormal) elements *if for every definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ and $\cap A_n = \emptyset$, there exists an N for which $A_N \cap T = \emptyset$.*
- *If $u \in T$, we will say that u is* not abnormal.
- *For every property P, we say that* "normally, for all u, $P(u)$" *if $P(u)$ is true for all $u \in T$.*

Relation to Kolmogorov complexity. Kolmogorov complexity enables us to define the notion of a *random sequence*, e.g., as a sequence s for which there exists a constant $c > 0$ for which, for every n, the (appropriate version of) Kolmogorov complexity $K(s_{|n})$ of its n-element subsequence $s_{|n}$ exceeds $n - c$. Crudely speaking, c is the amount of information that a random sequence has.

Random sequences in this sense do not satisfy the above definition, and are not in perfect accordance with common sense – because, e.g., a sequence that starts with 10^6 zeros and then ends in a truly random sequence is still random. Intuitively, for "truly random" sequences, c should be small, while for the above counter-example, $c \approx 10^6$. If we restrict ourselves to random sequences with fixed c, we satisfy the above definition.

There are many ways to define Kolmogorov complexity and random sequences [10]; it is therefore desirable to aim for results that are true in as general case as possible. In view of this desire, in the following text, we will not use any specific version of these definitions; instead, we will assume that Definition 2 is true.

It is possible to prove that abnormal elements do exist [4]; moreover, we can select T for which abnormal elements are as rare as we want: for every probability distribution p on the set U and for every ε, there exists a set T for which the probability $p(x \notin T)$ of an element to be abnormal is $\leq \varepsilon$:

Proposition 1. *[8] For every probability measure μ on a set U (in which all definable sets are measurable), and for every $\varepsilon > 0$, there exists a set T of typical elements for which $\mu(T) > 1 - \varepsilon$.*

Application: restriction to "not abnormal" solutions leads to regularization of ill-posed problems. An ill-posed problem arises when we want to reconstruct the state s from the measurement results r. Usually, all physical dependencies are continuous, so, small changes of the state s result in small changes in r. In other words, a mapping $f : S \to R$ from the set of all states to the set of all observations is continuous (in some natural topology). We consider the case when the measurement results are (in principle) sufficient to reconstruct s, i.e., the case when the mapping f is 1-1. That the problem is ill-posed means that small changes in r can lead to huge changes in s, i.e., that the inverse mapping $f^{-1} : R \to S$ is *not* continuous.

We will show that if we restrict ourselves to states S that are not abnormal, then the restriction of f^{-1} will be continuous, and the problem will become well-posed.

Definition 3. *A definable metric space (X, d) is called* definably separable *if there exists a definable everywhere dense sequence $x_n \in X$.*

Proposition 2. *Let S be a definably separable definable metric space, T be a set of all not abnormal elements of S, and $f : S \to R$ be a continuous 1-1 function. Then, the inverse mapping $f^{-1} : R \to S$ is continuous for every $r \in f(T)$.*

In other words, if we know that we have observed a not abnormal state (i.e., that $r = f(s)$ for some $s \in T$), then the reconstruction problem becomes well-posed. So, if the observations are accurate enough, we get as small guaranteed intervals for the reconstructed state s as we want.

Proof. It is known that if a set K is compact, then for any 1-1 continuous function $K \to R$, its inverse is also continuous. Thus, to prove our result, we will show that the closure $\overline{T}$ of the set T is compact.

A set K in a metric space S is compact if and only it is closed, and for every positive real number $\varepsilon > 0$, it has a finite ε-net, i.e., a finite set $K(\varepsilon)$ with the property that every $s \in K$, there exists an element $s(\varepsilon) \in K(\varepsilon)$ that is ε-close to s.

The closure $K = \overline{T}$ is clearly closed, so, to prove that this closure is compact, it is sufficient to prove that it has a finite ε-set for all $\varepsilon > 0$. For that, it is sufficient to prove that for every $\varepsilon > 0$, there exists a finite ε-net for the set R.

If a set T has a ε-net $T(\varepsilon)$, and $\varepsilon' > \varepsilon$, then, as one can easily see, this same set $T(\varepsilon)$ is also a ε'-net for T. Therefore, it is sufficient to show that finite ε-nets for T exist for $\varepsilon = 2^{-k}, k = 0, 1, 2, \ldots$

Let us fix $\varepsilon = 2^{-k}$. Since the set S is definably separable, there exists a definable sequence $x_1, \ldots, x_i, \ldots$ which is everywhere dense in S. As A_n, we will now take the complement to the union U_n of n closed balls $B_\varepsilon(x_1), \ldots, B_\varepsilon(x_n)$ of radius ε with centers in $x_1, \ldots, x_n$.

Clearly, $A_n \supseteq A_{n+1}$. Since x_i is an everywhere dense sequence, for every $s \in S$, there exists an n for which $s \in B_\varepsilon(x_n)$ and for which, therefore, $s \in U_n$ and $x \notin A_n = S \setminus U_n$. Hence, the intersection of all the sets A_n is empty.

Therefore, according to the definition of a set of typical elements, there exists an N for which $T \cap A_N = \emptyset$. This means that $T \subseteq U_N$. This, in its turn, means that the elements $x_1, \ldots, x_N$ form an ε-net for T. So, the set T has a finite ε-net for $\varepsilon = 2^{-k}$.

Comment. To actually use this result, we need an *expert* who will tell us what is abnormal, and whose ideas of what is abnormal satisfies the (natural) conditions described in Definition 2.

Application: justification of physical induction. From the viewpoint of an experimenter, a physical theory can be viewed as a statement about the results of physical experiments. If we had an infinite sequence of experimental results $r_1, \ldots, r_n, \ldots$, then we will be able to tell whether the theory is correct or not. So, a theory can be defined as a set of sequences $r_1, r_2, \ldots$ that are consistent with its equations, inequalities, etc. In real life, we only have finitely many results $r_1, \ldots, r_n$, so, we can only tell whether the theory is *consistent* with these results or not, i.e., whether there is an infinite sequence $r_1, r_2, \ldots$ that starts with the given results that satisfies the theory.

It is natural to require that the theory be *physically meaningful* in the following sense: if all experiments confirm the theory, then this theory should be correct. An example of a theory that is not physically meaningful is easy to give: assume that a theory describes the results of tossing a coin, and it predicts that at least once, there should be a tail. In other words, this theory consists of all sequences that contain at least one tail. Let us assume that actually, the coin is so biased that we always have heads. Then, this infinite sequence does not satisfy the given theory. However, for every n, the sequence of the first n results (i.e., the sequence of n heads) is perfectly consistent with the theory, because we can add a tail to it and get an infinite sequence that belongs to the set $\mathcal{T}$. Let us describe this idea in formal terms.

Definition 4. *Let a definable set R be given. Its elements will be called* possible results of experiments. *By S, we will denote the set of all possible sequences $r_1, r_n, \ldots$, where $r_i \in R$. By a* theory, *we mean a definable subset $\mathcal{T}$ of the set of all infinite sequences S. If $r \in \mathcal{T}$, we say that a sequence r* satisfies *the theory $\mathcal{T}$, or, that for this sequence r,* the theory $\mathcal{T}$ is correct.

Comment. A theory is usually described by its axioms and deduction rules. The theory itself consists of all the statements that can be deduced from the axioms by using deduction rules. In most usual definitions, the resulting set is r.e. – hence definable. We therefore define a theory as a definable set.

Definition 5. *We say that a finite sequence $(r_1, \ldots, r_n)$ is* consistent *with the theory $\mathcal{T}$ if there exists an infinite sequence $r \in \mathcal{T}$ that starts with $r_1, \ldots, r_n$ and that satisfies the theory. In this case, we will also say that the* first n experiments confirm the theory.

Definition 6. *We say that a theory $\mathcal{T}$ is* physically meaningful *if the following is true for every sequence $r \in S$:*

If for every n, the results of first n experiments from r confirm the theory $\mathcal{T}$, then, the theory $\mathcal{T}$ is correct for r.

In this case, the universal set consists of all possible infinite sequence of experimental results, i.e., $U = S$. Let $T \subseteq S$ be the set of all typical (not abnormal) sequences.

Proposition 3. *For every physically meaningful theory $\mathcal{T}$, there exists an integer N such that if a sequence $r \in S$ is not abnormal and the first N experiment confirm the theory $\mathcal{T}$, then this theory $\mathcal{T}$ is correct.*

Idea of the proof: as A_n, we take the set of all the sequences r for which either the first n experiments confirm $\mathcal{T}$ or $\mathcal{T}$ is not correct for r.

This result shows that we can *confirm* the theory based on finitely many observations. The derivation of a general theory from finitely many experiments is called *physical induction* (as opposed to *mathematical induction*). The general physical induction is difficult to justify, to the extent that a prominent philosopher C. D. Broad has called the unsolved problems concerning induction a *scandal of philosophy* [2]. We can say that the notion of "not abnormal" justifies physical induction by making it a provable theorem (and thus resolves the corresponding scandal).

Acknowledgments. This work was supported in part by the NASA grant NCC5-209, by the AFOSR grant F49620-00-1-0365, by NSF grants EAR-0112968, EAR-0225670, and EIA-0321328, and by the ARL grant DATM-05-02-C-0046. The author is thankful to the anonymous referees for valuable comments.

References

1. L. Boltzmann, "Bemrkungen über einige Probleme der mechanischen Wärmtheorie", *Wiener Ber. II*, vol. 75, pp. 62–100, 1877.
2. C.D. Broad, *Ethics and the history of philosophy*, Routledge and Kegan Paul, 1952.
3. R.P. Feynman, *Statistical Mechanics*, W.A. Benjamin, 1972.
4. A.M. Finkelstein, V. Kreinovich. "Impossibility of hardly possible events: physical consequences," *Abstr. 8th Int'l Congr. Log., Methodology & Philosophy of Science*, Moscow, vol. 5, pt. 2, pp. 23–25, 1987.
5. J.Y. Halpern, *Reasoning about uncertainty*, MIT Press, 2003.
6. V. Kreinovich, I.A. Kunin, "Kolmogorov Complexity and Chaotic Phenomena", *Int'l J. Engin. Science*, vol. 41, no. 3–5, pp. 483–493, 2003.
7. V. Kreinovich, I.A. Kunin, "Kolmogorov Complexity: How a Paradigm Motivated by Foundations of Physics Can Be Applied in Robust Control", In: A.L. Fradkov and A.N. Churilov (eds.), *Proc. Int'l Conf. "Physics and Control" PhysCon'2003*, St. Petersburg, Russia, August 20–22, 2003, pp. 88–93.
8. V. Kreinovich, L. Longprè, M. Koshelev, "Kolmogorov complexity, statistical regularization of inverse problems, and Birkhoff's formalization of beauty", In: A. Mohamad-Djafari (ed.), *Bayesian Inference for Inverse Problems*, Proc. SPIE, vol. 3459, San Diego, CA, pp. 159–170, 1998.
9. H.E. Kyburg, Jr., *Probability and the logic of rational belief*, Wesleyan Univ., 1961.
10. M. Li, P.M.B. Vitanyi, *An Introduction to Kolmogorov Complexity*, Springer, 1997.
11. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation*, W.H. Freeman, 1973.