Towards a "Generic" Notion of Genericity: From "Typical" and "Random" to Meager, Shy, etc.

Ali Jalal-Kamali, Ondrej Nebesky, Michael H. Durcholz
Vladik Kreinovich,* Luc Longpré
Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA

Abstract

In many application areas, it is important to study "generic" properties, i.e., properties which hold for "typical" examples. For example, if we know the probabilities of different events, we can consider a "random" object – i.e., an object that, crudely speaking, does not belong to any class of "unusual" events (i.e., to any class with a small probability). In other cases, "typical" may mean not belonging to an "unusual" subset which is small in some other sense – e.g., a subset of a smaller dimension. The corresponding notion of "typicalness" has been formalized for several cases, including the case of random events. In this case, the known Kolmogorov-Martin-Löf definition of randomness captures the idea that properties with probability 0 are impossible. In our previous papers, we modified this definition to take into account that from a practical viewpoint, properties with very small probabilities are often considered impossible as well. In this paper, we extend this definition to a general notion of "generic".
©2012 World Academic Press, UK. All rights reserved.

1 Formulation of the Problem

Need for studying generic properties. Let us show that in many application areas, it is important to study "generic" properties, i.e., properties which hold for "typical" examples.

Need for studying random objects. For example, in situations in which we know the probabilities of different events, it is usually assumed that the actually observed event should be "random" – in sense that it should not satisfy any property whose probability is 0 (or even whose probability is very small). When we flip a fair coin, we do not expect this process to result in an infinite sequence consisting only of heads – because the probability of this event is 0. We also do not expect that we will observe a sequence consisting of a 1,000 heads – because, in essence, the probability of this event is extremely low: it is equal to 2^{-1000}.

Similarly, from the purely mathematical viewpoint, it is theoretically possible that the random heat-related motions of all the molecules in an object accidentally go in the same direction – and this object will simply float away. However, the probability of this event happening is so low that any physicist will consider this event to be practically impossible.

Need for studying generic objects. In other applications, we similarly claim that a "typical" event does not have a certain property, because the class C of all the events that satisfy this property is "small" in some non-probabilistic sense: e.g., C is a subset of smaller dimension in the class of all possible events.

For example, in theoretical cosmology (see, e.g., [12]), it is usually believed that the actual space-time not only satisfies the corresponding equations (namely, Einstein's equations), but it is also believed that the metric field corresponding to the space-time is *generic* among all such solutions – in this case, generic in the sense that does not belong to any meager set. (For readers who are not familiar with this notion, the definition of a meager set will be given later in this paper.) For example, when a cosmologist says that Einstein's equations imply that a cosmological solution must have a singularity, he or she does not mean that *every* such solution has a singularity – some solutions actually don't, it means that a theorem was proven that *almost all* solutions

---

*Corresponding author. Email: vladik@utep.edu (V. Kreinovich).

have a singularity, and that we can therefore conclude that the *actual* space-time has a singularity as well. (This theorem was proven in the 1960s by S. Hawking and R. Penrose [2, 12].)

How randomness is described now: a brief reminder. The formal description of a generic object started with a definition of a random object by Kolmogorov and Martin-Löf; see [11] for details. This definition is based on the following idea: in statistics, if a sequence is considered to be random, then it must satisfy all the laws of probability. For example, coin flipping can be naturally described as a (potentially infinite) sequence of independent events in each of which we have Head or Tail with probability $1/2$. For such sequences, there is a Law of Large Numbers, according to which, with probability 1, for each sequence, the ratio of Heads among the first n experiments tends to $1/2$ as $n \to \infty$. If this property is not satisfied, then we can reasonably conclude that the corresponding sequence is not random (at least not random with respect to this distribution). Similarly, there is a Central Limit Theorem, according to which, with probability 1, the distribution of the actual ratio of Heads tends to normal (Gaussian) when $n \to \infty$. This property must also be satisfied for a random sequence. If a sequence does not satisfy this property, then we can safely conclude that it is not random.

All known tests of randomness are based on laws like this. So, if a sequence survives all these tests, it is reasonable to call it random.

A "law" is a property that is satisfied with probability 1. Thus, we can say that a sequence is random if it satisfies all the properties which hold with probability 1. Equivalently, instead of checking whether a sequence satisfies a property, we can check whether it satisfies the opposite property (e.g., in the example of coin flipping, the property of ratios (frequencies) *not* tending to $1/2$). For a property $P(x)$ which is satisfied with probability $\text{Prob}(P(x) = 1)$, its opposite $\neg P(x)$ is satisfied with probability $\text{Prob}(\neg P(x)) = 1 - \text{Prob}(P(x)) = 1 - 1 = 0$. Thus, we can say that a sequence is random if it does not satisfy any property which holds with probability 0.

Instead of properties, it is reasonable to talk about sets. Every property $P(x)$ defines a set, namely, the set $\{x : P(x)\}$ of all the objects that satisfy this property. However, not all sets correspond to what we intuitively mean by properties. Indeed, in statistics, properties must be well defined and well described in an appropriate formal language. The corresponding sets are called *definable*. A set is definable if it can be uniquely described by a formula in an appropriate language. Since there are more than continuum many sets and only countably many formulas (and thus, only countably many definable sets), it follows easily that not every set is definable.

In terms of sets, the Kolmogorov-Martin-Löf definition thus takes the following form:

Definition 1. *Let (X, P) be a probability space, i.e., a set X with a probability measure defined on its subsets. An element $x \in X$ is called* random *if it does not belong to any definable subset $S \subseteq X$ of probability measure 0 (i.e., to any definable set S for which $P(S) = 0$).*

We can use different languages to describe properties, and, depending on the choice of a language, we will have different versions of the notion of a random sequence [11].

Need for a more physically adequate notion of a random object. The above notion of a random object captures the intuitive idea that events with probability 0 cannot happen, but it does not capture a similar (and equally intuitive) idea that events with a very small probability cannot happen either – an idea that is actively used in practical applications of statistics and in physical reasoning.

This idea cannot be simply captured by setting a probability threshold p_0 below which all events are impossible. Indeed, in this case, for a number N for which $2^{-N} < p_0$, all sequences of Heads and Tails of length N are impossible because they have the same probability 2^{-N}. However, in reality, we can flip a coin N times an get a random sequence. So, we have to come up with a more complex definition, where a threshold depends on the complexity of the property.

Specifically, once we have a definable decreasing sequence of events $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ with the decreasing probability $p(A_n) \to 0$, then when N is large enough, the probability $p(A_N)$ of the event A_N becomes so small that the event A_N becomes impossible. Thus, we arrive at the following definition (see, e.g., [1, 5, 6, 7, 8, 9, 10]):

Definition 2.   *Let (X, P) be a probability space, i.e., X is a set with a probability measure P defined on its subsets. A set $R \subseteq X$ is called* a set of random elements *if it has the following property: for every definable sequence of sets A_n for which $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ and $p(A_n) \to 0$, there exists an integer N for which $A_N \cap R = \emptyset$.*

Please note that the set R is *not* uniquely defined by the probability space (X, P). For the same probability space, we can have *different* sets R satisfying the above definition.

What if we do not know the probabilities?   To take care of the cases when we do not have the probability distributions, we proposed the following modification of this definition [1, 5, 6, 7, 8, 9, 10]:

Definition 3.   *Let X be a set. A set $T \subseteq X$ is called* a set of typical elements *if it has the following property: for every definable sequence of sets A_n for which $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ and $\bigcap_n A_n = \emptyset$, there exists an integer N for which $A_N \cap T = \emptyset$.*

Similarly to the probabilistic case, for the same set X, we can have several *different* sets T satisfying this property.

Relation between random objects and typical objects.   There is a known relation between possible sets of random elements R and possible sets of typical elements T:

Proposition.   [5] *Let (X, P) be a probability space, i.e., X is a set with a probability measure P defined on its subsets, and let R_K be the set of all the elements of X which are random in the sense of Kolmogorov-Martin-Löf.*

- *Every set of random elements R (in the sense of Definition 2) is also a set of typical elements (in the sense of Definition 3).*

- *If T is a set of typical elements (in the sense of Definition 3), then the intersection $T \cap R_K$ is a set of random elements (in the sense of Definition 2).*

In other words, an element is random (in the sense of Definition 2) if and only if it is typical and KML-random. So, random elements can be described as typical elements which are also random in the sense of Kolmogorov-Martin-Löf. Thus, to describe all possible sets of random elements, it is sufficient to describe all possible sets of typical elements.

Remaining problem.   It is desirable to extend the above definitions from the probabilistic "smallness" to other types of smallness.

What we do in this paper.   In Sections 2 and 3, we give examples of different types of smallness; reader familiar with these notions can skip these sections. In Section 4, we extend the above definition of "typicality" to different types of smallness, and analyze the relation between the new definitions and the existing ones.

2   Examples of Other Notions of "Typicality": Meager Sets

Intuitive idea.   The notion of meager sets comes from mathematics, where it is often stated that some properties are *generic* in the sense that they hold for "typical" examples.

Examples.   In some cases, this notion is intuitively clear. For example, if we take two straight lines on a plane, then, in general, these lines will not be parallel. There is a "degenerate" case when they are parallel, but intuitively, the lines from a "generic" pair of straight lines are not parallel.

Another example: if we pick a number, there may be a case when this number is equal to 0, but intuitively, a "generic" number should be different from 0. Similarly, a "generic" number should not be equal to 1, to π, or to any other definable number.

Comment. At first glance, this may sound like a contradiction, since a "generic" number cannot be equal to any number that we can explicitly define. However, in reality, there is no contradiction: similarly to what we said about definable sets, there are only countably many definable numbers; thus, since there are continuum many real numbers, there are continuum many real numbers which are not definable.

Similarly, a generic polynomial $f(x)$ does not have a root at $x = 0$ – because this would mean that the value $f(x)$ is equal to 0, and we have already argued that a generic number should not be equal to 0.

Another example is that a generic matrix A is invertible. Indeed, a matrix is invertible if and only if its determinant $\det(A)$ is different from 0. A "typical" matrix A should have all its properties "typical". In principle, any real number can appear as a determinant of a matrix. Intuitively, for a "typical" matrix A, the determinant should also be "typical" (generic). Indeed, if a determinant of a certain matrix is not "typical", if it is a special number, then this would make this matrix somewhat special – and not "typical" – too. Since a generic real number is different from 0, we can therefore conclude that the determinant $\det(A)$ of a generic matrix should also be different from 0 – i.e., that a generic matrix is invertible.

Similar intuitive statements can be made about typical functions, but these statements are often not that easy to formalize.

Towards formalization. The notion of a meager set (also called *Baire first category* set, or a *thin* set) was defined to formalize some of these statements. Originally, this definition was proposed for *complete metric spaces*.

A *metric space* is a pair (X, d) consisting of a set X and a function $d : X \times X \to \mathbb{R}$ (called a *distance function*, or a *metric*) that satisfies the following properties for all x, y, and z:

- $d(x, y) \geq 0$,

- $d(x, y) = 0$ if and only if $x = y$,

- $d(x, y) = d(y, x)$, and

- $d(x, z) \leq d(x, y) + d(y, z)$.

For every point $x \in X$ and for every real number $\varepsilon > 0$, the *open ball* $B_\varepsilon(x)$ of radius ε centered in x is defined as the set $B_\varepsilon(x) \stackrel{\text{def}}{=} \{y : d(x, y) < \varepsilon\}$ of all the points y whose distance to x is smaller than ε.

In each metric space, we can define a limit $x = \lim_n x_n$ of a sequence x_n in a usual way, expressing the informal meaning that points x_n get closer and closer to x as n increases. Formally, a sequence x_n tends to a limit x if for every $\varepsilon > 0$, there exists an N such that for every $n \geq N$, we have $d(x_n, x) \leq \varepsilon$. A sequence of elements $x_n \in X$ is called a *Cauchy* sequence if for every $\varepsilon > 0$ there exists an N such that for every $n, n' \geq N$, we have $d(x_n, x_{n'}) \leq \varepsilon$. A metric space is called *complete* if every Cauchy sequence has a limit. For example, the set $\mathbb{R}$ of all real numbers is complete, as well as the n-dimensional space $\mathbb{R}^n$.

A set $S \subseteq X$ is called *closed* if for every sequence $\{s_n\} \subseteq S$ that has a limit x, this limit x also belongs to the set S. For every set S, the set consisting of all the limit points of all the sequences $\{x_n\} \subseteq S$ is called a *closure* of the set S and denoted by $\overline{S}$.

A set $S \subseteq X$ is called *open* if for every point s from S, an entire open ball is contained in this set S, i.e., $B_\varepsilon(s) \subseteq S$ for some $\varepsilon > 0$.

Comment. It is known that a set S is open if and only if its complement $-S$ is closed.

A set $S \subseteq X$ is called *nowhere dense* if its closure does not contain any open balls at all.

Resulting definition. A set is called *meager* if it can be represented as a union of countably many nowhere dense sets.

Properties of meager sets. The properties of meager sets are similar to properties of sets of measure 0; for example:

- every subset of a meager set is meager;

- a union of countably many meager sets is meager.

3   Examples of Other Notions of "Typicality": Shy Sets

Main idea.   For classes of functions, the above notion of a meager set often does not capture the intuitive idea of "typical". For such classes, a new notion of *shy* sets was proposed in [3, 4].

Metric linear spaces.   This notion is defined for *complete metric linear spaces*, i.e., for complete metric spaces X which are also *linear spaces*, i.e., spaces on which there is a fixed element called 0, and two operations are defined:

- a binary operation $x + y$ called *addition* which has the following properties for all x, y, and z:

    - $x + y = y + x$ (*commutativity*),
    - $x + (y + z) = (x + y) + z)$ (*associativity*), and
    - $x + 0 = x$;

- an operation $\lambda \cdot x$ with real numbers λ called *multiplication* by a real number for which $1 \cdot x = x$ and which is *distributive* ($\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$ and $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$) and associative ($\lambda \cdot (\mu \cdot x) = (\lambda \cdot \mu) \cdot x$).

For every two elements $x, y \in X$, the element $x + y$ is also called the result of *shifting* x by y. This notion of shift can be naturally extended from elements $x \in X$ to subsets $S \subseteq X$ as follows: a shift $S + y$ of the set S by an element y is defined as the collection of the results $s + y$ of shifting all elements $s \in S$ by y: $S + y \stackrel{\text{def}}{=} \{s + y : s \in S\}$.

The distance $d(x, y)$ on a linear space should be consistent with addition and multiplication in the following sense: $d(x + z, y + z) = d(x, y)$ and $d(\lambda \cdot x, \lambda \cdot y) = |\lambda| \cdot d(x, y)$ for all x, y, z, and λ.

For example, a usual vector space $\mathbb{R}^n$ consisting of all the tuples $x = (x_1, \ldots, x_n)$ with the usual component-wise addition $(x_1, x_2, \ldots) + (y_1, y_2, \ldots) = (x_1 + y_1, x_2 + y_2, \ldots)$, component-wise multiplication by a real number $\lambda \cdot (x_1, x_2, \ldots) = (\lambda \cdot x_1, \lambda \cdot x_2, \ldots)$, and the standard metric $d((x_1, x_2, \ldots), (y_1, y_2, \ldots)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots}$ is a complete metric linear space.

Borel sets.   On each such space X – just like in every other metric space (or every other topological space), we can define *Borel sets* – as the smallest collection of sets that contains all open sets and that is closed under complement and countable union. For example:

- every open set S is a Borel set;

- every complement to an open set S (i.e., every closed set) is a Borel set; we will denote the complement to a set S by $-S$;

- every countable union $\bigcup_n S_n$ of open or closed sets S_n is a Borel set;

- every complement $-\bigcup_n S_n$ to such a union is a Borel set.

Measures: definition.   A *measure* is usually defined as a mapping μ that maps every Borel set S into either a non-negative real number or $+\infty$ in such as way that $\mu(\emptyset) = 0$, $\mu(X) > 0$, and μ is *additive*, i.e., if we have a sequence of pairwise disjoint Borel sets S_n, then $\mu\left(\bigcup_n S_n\right) = \sum_n \mu(S_n)$.

In particular, for a finite sequence of pairwise disjoint sets $S_1, \ldots, S_n$, we have $\mu(S_1 \cup \ldots \cup S_n) = \sum_n \mu(S_n)$.

Measure of a complement.   For every set S, the space X can be represented as the union of two disjoint sets: the set S itself and its complement $-S$. Thus, we have $\mu(X) = \mu(S) + \mu(-S)$. So, if the measure $\mu(X)$ of the whole space X is finite, then we have $\mu(-S) = \mu(X) - \mu(S)$.

Measures: monotonicity.   If $S \subseteq S'$, then we can represent the set S' as $S' = S \cup (S' - S)$, where the sets S and $S' - s$ are mutually disjoint. Thus, $\mu(S') = \mu(S) + \mu(S' - S)$. Since the values of the mapping μ are always non-negative, we have $\mu(S' - S) \geq 0$ and thus, $\mu(S') \geq \mu(S)$.

Measures: subadditivity. In the following text, we will also need to describe the measure $\mu(U)$ of the union $U \stackrel{\text{def}}{=} S_1 \cup \ldots \cup S_n$ in the general case, where the sets S_n are not necessarily disjoint. In this case, an element $s \in U$ can, in principle, belong to several of the sets S_i. By definition of the union, each element $s \in U$ belongs to one of the sets S_i. For each $s \in U$, among all indices i for which $s \in S_i$, there is the smallest index. We will denote this smallest index by $f(s)$. Under this notation, $f(s) = i$ means that s belongs to the set S_i and does not belong to any of the sets with smaller indices, i.e., s does not belong to S_1, ..., S_{i-1}. In other words, $f(s) = i$ if and only if $s \in S_i - (S_1 \cup \ldots \cup S_{i-1})$. Every element $s \in U$ has some value $i = f(s)$, so it belongs to one of the sets $S_i' \stackrel{\text{def}}{=} S_i - (S_1 \cup \ldots \cup S_{i-1})$. Thus, U is a union of the sets S_i'.

For $i \neq j$, the sets S_i' and S_j' are mutually disjoint, because for elements $s \in S_i'$, we have $f(s) = i$, while for elements $s \in S_j'$, we have $f(s) = j \neq i$. Thus, the union U can be represented as a union of pairwise disjoint sets S_i': $U = S_1' \cup \ldots \cup S_n'$. Due to additivity, we have $\mu(U) = \sum_{i=1}^{n} \mu(S_i')$. For each i, we have $S_i' \subseteq S_i$ hence $\mu(S_i') \leq \mu(S_i)$, so $\sum_{i=1}^{n} \mu(S_i') \leq \sum_{i=1}^{n} \mu(S_i)$. Thus,

$$\mu(S_1 \cup \ldots \cup S_n) = \mu(U) = \sum_{i=1}^{n} \mu(S_i') \leq \sum_{i=1}^{n} \mu(S_i),$$

i.e.,

$$\mu(S_1 \cup \ldots \cup S_n) \leq \sum_{i=1}^{n} \mu(S_i).$$

A similar inequality can be similarly proven if we have an infinite sequence of sets:

$$\mu\left(\bigcup_i S_i\right) \leq \sum_{i=1}^{\infty} \mu(S_i).$$

This inequality is called *subadditivity*.

Probability measures. Measures for which $\mu(X) = 1$ are called *probability measures*. For such measures, $\mu(S)$ is the probability of the event S. For probability measures, the meaning of additivity is as follows:

- If we have two exclusive events A and B, i.e., if A and B cannot occur at the same time, then the probability that we have either A or B is equal to the sum of the probabilities of A and B. For example, the probability that it is either raining or snowing is equal to the sum of the probability $\mu(A)$ that it is raining and the probability $\mu(B)$ that it is snowing.

- Similarly, if we have a sequence of mutually exclusive events S_n, then the probability $\mu\left(\bigcup_n S_n\right)$ that one of these events occurs is equal to the sum of the probabilities $\mu(S_n)$ that one of them occurs.

Not all measures are probability measures: e.g., the area $\mu(S)$ of sets S in a plane $S \subseteq \mathbb{R}^2$ and the volume $\mu(S)$ of sets S in 3-D space $S \subseteq \mathbb{R}^3$ are examples of non-probability measures, since in both cases, the area (volume) of the whole space $X = \mathbb{R}^n$ is infinite: $\mu(X) = +\infty$.

Measures on non-Borel sets: an example. Measures (in particular, probability measures) can be defined for non-Borel sets S as well. A standard example of a measure in a finite-dimensional space $\mathbb{R}^n$ is a *Lebesgue* measure.

The Lebesgue measure on a unit cube $X = [0,1]^n$ can be defined as follows. The measure $\mu(B)$ of a closed box

$$B = [\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_n, \overline{x}_n] \subseteq [0,1]^n$$

is defined as its volume

$$\mu(B) = (\overline{x}_1 - \underline{x}_1) \cdot \ldots \cdot (\overline{x}_n - \underline{x}_n).$$

In particular, for the space X itself, we have $\mu(X) = 1$.

For a union $\cup B_n$ of the boxes, we have $\mu(\bigcup B_n) \le \sum \mu(B_n)$. For every set S which is contained in such a union $\cup B_n$, we have $\mu(S) \le \mu(\bigcup B_n)$ and thus, $\mu(S) \sum \mu(B_n)$. Hence, the measure $\mu(S)$ of each set S is bounded by the smallest of the sums $\sum \mu(B_n)$ for all sequences of boxes $\{B_n\}$ for which $S \subseteq \cup B_n$. This smallest value is called the *outer Lebesgue measure* of the set S; it is denoted by $\mu^*(S)$. Formally,

$$\mu^*(S) \overset{\text{def}}{=} \inf \left\{ \sum \mu(B_n) : S \subseteq \cup B_n \right\}.$$

Thus, we always have $\mu(S) \le \mu^*(S)$.

For a complement $-S$, we similarly have $\mu(-S) \le \mu^*(-S)$. Since $\mu(-S) = \mu(X) - \mu(S) = 1 - \mu(S)$, we thus have $1 - \mu(S) \le 1 - \mu*(-S)$, therefore $\mu(S) \ge 1 - \mu^*(-S)$. The corresponding lower bound $1 - \mu^*(-S)$ is called an *inner Lebesgue measure* of the set S. It is usually denoted by $\mu_*(S) \overset{\text{def}}{=} 1 - \mu^*(-S)$.

A set S is called *Lebesgue measurable* if its inner and outer Lebesgue measures coincide $\mu_*(S) = \mu^*(S)$. This common value of inner and outer measures is called the *Lebesgue measure $\mu(S)$* of the set S. It can be shown that this definition indeed leads to an additive measure.

Compact sets. To define shy sets, we also need to define *compact* sets. A closed set S is called *compact* if every time S is contained in a union of a family $\{S_\alpha\}_{\alpha \in A}$ of open sets $S \subseteq \bigcup_{\alpha \in A} S_\alpha$, there exist finitely many open sets from this family whose union contain S: $S \subseteq S_{\alpha_1} \cup \ldots \cup S_{\alpha_n}$ for some $\alpha_1, \ldots, \alpha_n \in A$. For a finite-dimensional space $\mathbb{R}^n$ with the usual metric, a closed set is compact if and only if it is bounded. For a general linear space, not every bounded closed set is compact.

Shy sets. Finally, we are ready to define the notion of a shy set.

A subset S of a linear space is called *shy* if there exists a measure μ and a compact set U for which $\mu(U) > 0$ (and $\mu(U) < +\infty$) and for which $\mu(S + x) = 0$ for all $x \in X$.

Comment. For $X = \mathbb{R}^n$, a set is shy if and only if it has Lebesgue measure 0. Thus, the notion of shyness is a natural extension of the "probability 0" notion to general linear spaces.

Properties of shy sets. The properties of shy sets are similar to properties of sets of measure 0 and meager sets; for example:

- every subset of a shy set is shy;

- a union of countably many shy sets is shy.

Prevalent sets. A set is called *prevalent* if its complement is shy.

If some property holds on a prevalent set, it is a reasonable indication that this property is "almost always" (or "in general") true. Examples of such properties are [3, 4]:

- a generic polynomial does not have a root at zero;

- a generic matrix is invertible;

- if $f : M \to N$ is a smooth function between smooth manifolds, then a generic point of N is not a critical value of the function f.

4   Kolmogorov-Martin-Löf-Type Definition of a Generic Element

Main idea. The original Kolmogorov-Martin-Löf definition of a random sequence is as follows: a sequence is random if and only if it does not belong to any definable set of measure 0. So, it is reasonable to extend it to other notions of small sets (meager, shy, etc.) in the following way:

Definition 4. *Let X be a set. A non-empty definable family $\mathcal{S}$ of subsets of X is called* a family of small subsets of X *if it satisfies the following two properties:*

- *if a set S belongs to the family $\mathcal{S}$, then every subset $S' \subseteq S$ of the set S also belongs to this family;*

- *a countable union $\bigcup_n S_n$ of sets from the family $\mathcal{S}$ also belongs to the family $\mathcal{S}$.*

Sets from the family $\mathcal{S}$ will be called $\mathcal{S}$-small.

Comment. Since the family $\mathcal{S}$ is non-empty, it contains some set S. An empty set is a subset of S and thus, an empty set belongs to the class $\mathcal{S}$. Thus, the empty set is $\mathcal{S}$-small.

Definition 5. *Let X be a set, and let $\mathcal{S}$ be a family of small subsets of X. An element $x \in X$ is called* KML-$\mathcal{S}$-generic *if it does not belong to any definable set $S \subseteq \mathcal{S}$.*

Comment. In particular, if the class $\mathcal{S}$ consists of all sets of measure 0, then KML-$\mathcal{S}$-generic elements are exactly elements which are random in the sense of Kolmogorov-Martin-Löf.

Limitations. This definition has the same limitation as the original Kolmogorov-Martin-Löf notion of randomness.

Indeed, according to the Kolmogorov-Martin-Löf definition, if a binary sequence $\omega_1 \omega_2 \ldots$ is random, then it remains Kolmogorov-Martin-Löf random if we add, in front of it, a large number N of zeros, i.e., consider a new sequence $0 \ldots 0 \omega_1 \omega_2 \ldots$

Mathematically, this may sound reasonable, but intuitively, "random" means something that can appear in real tossing of a fair coin. From this viewpoint, a sequence that starts with one million zeros means that we toss a fair coin million times and get tail every time – this is not right. The notion of randomness that fully captures the intuitive understanding should take into account that, intuitively, not every finite sequence $\omega_1 \omega_2 \ldots \omega_n$ can appear as a beginning of truly "random" infinite sequence.

As we have mentioned, in some physical applications, the intuitive notion of "generic" is better captured by meager sets, not by sets of measure 0. However, one can show that for all the above definitions of a family $\mathcal{S}$, if we add a large number of zeros at the beginning of an $\mathcal{S}$-generic sequence, the resulting sequence will still be $\mathcal{S}$-generic.

In other words, no matter what are the first N bits of a sequence, this sequence may still turn out to be Kolmogorov-Martin-Löf random, or generic, or KML-$\mathcal{S}$-generic for some other concept $\mathcal{S}$. So, if we stick to the above definition, there is no way to detect a KML-random (or an KML-$\mathcal{S}$-generic) sequence based on the observations. To overcome this limitation, it is desirable to modify the above definition, so that we will be able to detect generic elements based on the observations.

Discussion. We want to be able to detect genericity based on the observation. Maybe we do not detect genericity based on a single observation because the measurement inaccuracy is too high, but as we increase the measurement accuracy more and more, we should be eventually be able to detect genericity. Actual genericity means not belonging to a definable small set S. We should be able to tell, based on some approximate observation, that an element does not belong to S.

Let us assume that we have a sequence of observation procedures of higher and higher accuracy. Let A_n denote the set of all the elements which, due to the inaccuracy of the n-th observation procedure, lead to the same results as some elements from the set S.

For small n, when inaccuracy is large, we may have many elements that lead to the same observations results as elements from the set S. The larger n, the more accurate the measurements are, so fewer elements can lead to the same observation results as the elements of S. In mathematical terms, this means that $A_n \supseteq A_{n+1}$ for every n.

By definition of the set S, every element of the set S belongs to the sets A_n for all n. Thus, every element from the set S also belongs to the intersection $\cap A_n$:

$$S \subseteq \cap A_n.$$

If an element s does not belong to the set S, then, when the accuracy is good enough, we should be able to detect this fact. In other words, there should exists a level n for which $s \notin A_n$. Thus, we conclude that if an element s does not belong to the set S, then this element should not belong to the intersection $\cap A_n$:

$$-S \subseteq - \cap A_n.$$

From $S \subseteq \cap A_n$ and $-S \subseteq - \cap A_n$, we conclude that the intersection $\cap A_n$ coincides with the set S: $S = \cap A_n$.

Our requirement is that we should be able to tell whether an element is generic based on some sufficiently accurate observation. In other words, the set G of "truly generic" sequences must satisfy the following property. Let us take an arbitrary definable non-generic property $S \in \mathcal{S}$. As we have mentioned, this means that we have a definable sequence A_n for which $A_n \supseteq A_{n+1}$ and $S = \cap A_n \in \mathcal{S}$. The meaning of A_n is that A_n consists of all the objects which, based on the first n observations, are indistinguishable from objects $s \in S$. Our claim is that that there should be an observation level N on which we should be able to detect this non-genericity. In other words, there should be a level N for which none of the truly generic objects $s \in G$ will display the same results as elements of S. Objects which display the same results as elements of S form the set A_N; thus, the above property takes the form $A_N \cap G = \emptyset$.

Thus, we arrive at the following definition.

Definition 6. *Let X be a set, and let $\mathcal{S}$ be a family of small subsets of X. A set $G \subseteq X$ is called a set of $\mathcal{S}$-generic elements if it has the following property: for every definable sequence of sets A_n for which $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ and $\bigcap_n A_n \in \mathcal{S}$, there exists an integer N for which $A_N \cap G = \emptyset$.*

Let us now describe the relation between this definition and the notion of set of typical elements.

Theorem. *Let X be a set, let $\mathcal{S}$ be a family of small subsets of X, and let G_K denote the set of all KML-$\mathcal{S}$-generic elements of X. Then*

- *Every set of $\mathcal{S}$-generic elements T (in the sense of Definition 6) is also a set of typical elements (in the sense of Definition 3).*

- *If T is a set of typical elements (in the sense of Definition 3), then the intersection $T \cap G_K$ is a set of $\mathcal{S}$-generic elements (in the sense of Definition 6).*

In other words, an element is $\mathcal{S}$-generic (in the sense of Definition 6) if and only if it is typical and KML-$\mathcal{S}$-generic. So, $\mathcal{S}$-generic elements can be described as typical elements which are also KML-$\mathcal{S}$-generic. Thus, to describe all possible sets of $\mathcal{S}$-generic elements, it is sufficient to describe all possible sets of typical elements.

Proof. Let us first assume that G if a set of $\mathcal{S}$-generic elements (in the sense of Definition 6). Let us prove that it is also a set of typical elements, i.e., that it satisfies Definition 3. Indeed, let A_n be a $\mathcal{L}$-definable sequence of sets for which $A_n \supseteq A_{n+1}$ for all n and $\bigcap_n A_n = \emptyset$. By definition of the family of small sets, it contains an empty set, and since $\emptyset \in \mathcal{S}$, we have $\bigcap_n A_n \in \mathcal{S}$. Thus, by Definition 6, there exists an N for which $A_N \cap G = \emptyset$. So, G indeed satisfies Definition 3.

Let us prove that if T is a set of typical elements, then the intersection $G \stackrel{\text{def}}{=} T \cap G_K$ is indeed a set of $\mathcal{S}$-generic elements in the set of Definition 6. Indeed, let A_n be a $\mathcal{L}$-definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $A \stackrel{\text{def}}{=} \bigcap_n A_n \in \mathcal{S}$. Since the sequence A_n is definable, its intersection A is also definable. Thus, by definition of KML-$\mathcal{S}$-genericity, we have $G_K \cap A = \emptyset$.

The auxiliary sequence $A'_n \stackrel{\text{def}}{=} A_n - A$ is also definable. It is easy to check that $A'_n \supseteq A'_{n+1}$, and that $\bigcap_n A'_n = \emptyset$. Thus, by Definition 3, there exists an N for which $A'_N \cap T = \emptyset$.

Let us show that for the same index N, we have $A_N \cap G = \emptyset$. Indeed, since the sequence A_n is decreasing, we have $A \subseteq A_N$ and thus, $A_N = A'_N \cup A$. Here, the set A'_N does not have any common elements with the set T – hence with the intersection $G = T \cap G_K$; the set A also does not have any common elements with the set G_K – hence with the intersection $G = T \cap G_K$. Thus, the entire union $A_N = A'_N \cup A$ does not have any common points with the intersection G, i.e., indeed, $A_N \cap G = \emptyset$. The theorem is proven.

Acknowledgments

References

[1] Finkelstein, A.M., and V. Kreinovich, Impossibility of hardly possible events: physical consequences, *Abstracts of the 8th International Congress on Logic, Methodology, and Philosophy of Science*, Moscow, vol.5, part 2, pp.23–25, 1987.

[2] Hawking, S., and R. Penrose, *The Nature of Space and Time*, Princeton University Press, Princeton, New Jersey, 1996.

[3] Hunt, B.R., The prevalence of continuous nowhere differentiable functions, *Proceedings of the American Mathematical Society*, vol.122, no.3, pp.711–717, 1994.

[4] Hunt, B.R., T. Sauer, and J.A. Yorke, Prevalence: a translation-invariant 'almost every' on infinite-dimensional spaces, *Bulletin of American Mathematical Society*, vol.27, pp.217–238, 1992.

[5] Kreinovich, V., Toward formalizing non-monotonic reasoning in physics: the use of Kolmogorov complexity, *Revista Iberoamericana de Inteligencia Artificial*, vol.41, pp.4–20, 2009.

[6] Kreinovich, V., 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*, vol.316, pp.63–110, 2004.

[7] Kreinovich, V., and A.M. Finkelstein, Towards applying computational complexity to foundations of physics, *Journal of Mathematical Sciences*, vol.134, no.5, pp.2358–2382, 2006.

[8] Kreinovich, V., and I.A. Kunin, Kolmogorov complexity and chaotic phenomena, *International Journal of Engineering Science*, vol.41, nos.3-5, pp.483–493, 2003.

[9] Kreinovich, V., and I.A. Kunin, Kolmogorov complexity: how a paradigm motivated by foundations of physics can be applied in robust control, *Proceedings of the International Conference "Physics and Control" PhysCon'2003*, St. Petersburg, Russia, pp.88–93, 2003.

[10] Kreinovich, V., L. Longprè, and M. Koshelev, Kolmogorov complexity, statistical regularization of inverse problems, and Birkhoff's formalization of beauty, *Bayesian Inference for Inverse Problems*, edited by A. Mohamad-Djafari, vol.3459, pp.159–170, 1998.

[11] Li, M., and P.M.B. Vitanyi, *An Introduction to Kolmogorov Complexity*, Springer-Verlag, Berlin, Hiedelberg, New York, 2008.

[12] Misner, C.W., K.S. Thorne, and J.A. Wheeler, *Gravitation*, W.H. Freeman and Company, New York, 1973.

[13] Oxtoby, J.C., *Measure and Category: A Survey of the Analogies between Topological and Measure Spaces*, Springer Verlag, New York, Heidelberg, Berlin, 1980.