Kolmogorov Complexity Leads to a Representation Theorem for Idempotent Probabilities (σ-Maxitive Measures)

Vladik Kreinovich and Luc Longpré*
Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu longpre@cs.utep.edu

Abstract

In many application areas, it is important to consider maxitive measures (idempotent probabilities), i.e., mappings m for which $m(A \cup B) = \max(m(A), m(B))$. In his papers, J. H. Lutz has used Kolmogorov complexity to show that for constructively defined sets A, one maxitive measure – fractal dimension – can be represented as $m(A) = \sup\limits_{x \in A} f(x)$. We show that a similar representation is possible for an arbitrary maxitive measure.

1 Introduction

1.1 Probabilities: Reminder

σ-additive measures. In probability theory, a probability measure on the set X is defined as as follows. First, a *σ-algebra* over the set X is defined as a family $\mathcal{A} \subseteq 2^X$ of subsets of the set X that is closed under complement and countable union. A function $P : \mathcal{A} \to R_0^+$ from a σ-algebra $\mathcal{A}$ to the set R_0^+ of all non-negative numbers is called *σ-additive* if $P(A \cup B) = P(A) + P(B)$ when $A \cap B = \emptyset$ and, more generally, for every countable family of sets $\{A_i\}$ for which $A_i \cap A_j = \emptyset$ for $i \neq j$, we have $P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} P(A_i)$. A *probability measure* on the set X is then defined as a σ-additive measure P for which $P(X) = 1$.

Why we need a representation theorem. From the purely mathematical viewpoint, the above definition is a correct standard definition of a probability measure. From a computational viewpoint, we may need to compute $P(A)$ for an arbitrary set A. Storing a collection of values $P(A)$ for all possible sets A is prohibitive. Even when X is a finite set of reasonable size n, if $\mathcal{A} = 2^X$, we would need to store 2^n values, which is not feasible.

From the computational viewpoint, it is therefore desirable to produce an alternative representation of probability measures – ideally, a representation that requires, for sets of size n, only linearly (or at most polynomially) many values to store. For probability measures, such representations are well known.

Representation theorem for the finite case. When the universal set X is finite, it is possible to define $P(A)$ for all subsets $A \subseteq X$, i.e., take $\mathcal{A} = 2^X$. In this case, we have $A = \bigcup\limits_{x \in A} \{x\}$, hence, due to additivity, $P(A) = \sum\limits_{x \in A} P(\{x\})$. Thus, in order to describe the values $P(A)$ for all 2^n sets $A \subseteq X$, it is sufficient to describe n values $p(x) \stackrel{\text{def}}{=} P(\{x\})$. The following formula ("representation theorem") enables us to reconstruct all the values $P(A)$ from the known values $p(x)$: $P(A) = \sum\limits_{x \in A} p(x)$.

Representation theorem for the general case. In the continuous case, e.g., for $X = R$, many important probabilities measures can be represented in a similar way, with an integral instead of the sum:

$$P(A) = \int_{x \in A} \rho(x)\,dx$$

for some continuous function $\rho(x)$. The corresponding function $\rho(x)$ – a continuous analogue of point probabilities $p(x)$ – is called a *probability density function*.

For some probability measures P, such an integral representation with a continuous function $\rho(x)$ is not possible: e.g., this representation is not possible for a point measure P_{x_0} for which $P_{x_0}(A) = 1$ if $x_0 \in A$ and $P_{x_0}(A) = 0$ otherwise.

For such measures, we can still have a similar integral representation if, instead of continuous functions $\rho(x)$, we also allow "functions" $\rho(x)$ defined as limits of continuous functions – with appropriately defined limits as integrals. Such limits are called *generalized functions*, or *distributions*; see, e.g., [2].

For example, the above point measure $P_{x_0}(A)$ can be represented in the integral form if we use a *delta-function* density $\rho(x) = \delta(x - x_0)$

1.2 Idempotent Probabilities, a.k.a. Maxitive Measures

What they are. In many application areas, it is important to consider "idempotent probabilities" (maxitive measures), i.e., functions m from sets into real numbers for which $m(A \cup B) = \max(m(A), m(B))$.

Let us give three examples of maxitive measures.

Example 1: rare events. Many examples of maxitive measures come from the analysis of rare events; see, e.g., [7]. Rare events – such as unusually large deviations – are extremely important, because they account for catastrophic failures of technical systems, for natural disasters such as earthquakes and floods, etc. Since they are rare, we do not have a large number of observed events of this type and therefore, we cannot use traditional engineering statistical techniques for processing such events. As an alternative, statisticians have developed *asymptotic* techniques, in which instead of describing the probability $p(L)$ of a specific large deviation L, we try to describe an asymptotic expression $p_a(L)$ that describes how the probability $p(L)$ of a deviation of size $\geq L$ depends on L. By definition of asymptotic, when L is large, the actual probability is close to this asymptotic expression, and the larger L (i.e., the more important the deviation), the closer this asymptotic estimate $p_a(L)$ to the actual value $p(L)$.

In many cases, for large deviations L, the dependence of p on L is *scale-invariant*, i.e., crudely speaking, leads to $p(L) \sim C \cdot L^{-\alpha}$ for some real number α. In this case, we have two parameters to characterize this dependence: C and α. If we want to characterized the rarity of an event by a single parameter, then which of these two parameters should we choose? A small change in α leads to a much faster decrease in $p(L)$ than a small change in C, so it is natural to select α as a measure of rarity.

In this case, if we have two rare events with rarities $\alpha(A)$ and $\alpha(B)$, what is the rarity of $A \cup B$? In other words, how can we estimate the probability of the event that either A or B will lead to a large deviation $\geq L$? One can easily see that if, say, $\alpha(A) < \alpha(B)$, then $P_A(L) \gg P_B(L)$, moreover, $P_B(L)/P_A(L) \to 0$ and therefore, asymptotically, the total probability $P_{A\,\&\,B}(L)$ is equal to $p_A(L)$. In other words, our newly defined measure of rarity satisfies the condition $\alpha(A \cup B) = \max(\alpha(A), \alpha(B))$ – i.e., it is a maxitive measure. It is easy to see that the same property holds if $A \cap B \neq \emptyset$.

Usually, σ-additive probability measures lead to σ-maxitive measures m, i.e., functions $m : \mathcal{A} \to R$ for which $m(A \cup B) = \max(m(A), m(B))$ and also $m\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sup\limits_i m(A_i)$.

Example 2: Hausdorff (fractal) dimension. It is known that many processes in nature are not smooth, many natural sets are not traditional sets with smooth or piece-wise smooth boundaries. To characterize such sets, researchers have invented a notion of *Hausdorff dimension* – also known as *fractal* dimension; see, e.g., [5].

To illustrate this notion (and to show why the resulting function is a maxitive measure), let us use the following example. Suppose that a fast food company wants to place its restaurants in a region S. The company has decided on the largest distance ε that a person has to travel from any point to reach the nearest restaurant. What is the smallest number of restaurants that we need to build to satisfy this requirement?

In mathematical terms, we want to find a finite set $S' \subseteq S$ such that for every $s \in S$, there exists $s' \in S'$ for which $D(s, s') \leq \varepsilon$, where $D(s, s')$ denotes the distance between the points s and s'. The smallest possible number of elements in such a set S' is called the ε-*entropy* of the set S and denoted by $N_\varepsilon(S)$.

Let us first consider the case when a region S is simply a straight line segment $[0, L]$ (e.g., a region along a road). In this case, every restaurant covers at most a stretch of length 2ε (ε on one side of the restaurant and ε on the other side). So, to cover the whole segment, we need to place at least $N(\varepsilon) \geq L/(2\varepsilon)$ restaurants. Vice versa, if we have exactly $L/(2\varepsilon)$ restaurants, we can place them at points ε, 3ε, ..., $(2i + 1) \cdot \varepsilon$, ..., and cover the entire segment. In general, for a smooth curve, $N_\varepsilon(S) \sim c/\varepsilon$ for some constant c (proportional to the length of the curve).

If a region S is a square, then each restaurants covers a subregion of area $\sim \varepsilon^2$, so we need $\geq c/\varepsilon^2$ restaurants; vice versa, if we place the restaurants on a grid, we will have an arrangement with $N \le c/\varepsilon^2$. In general, for a regular-shaped 2-D domain S (with piece-wise smooth boundaries), we have $N_\varepsilon(S) \sim c/\varepsilon^2$.

For a 3-D domain (e.g., if we want to place restaurants in space), we have $N_\varepsilon(S) \sim c/\varepsilon^3$. In general, for regular sets for which the notion of a dimension is well defined, $N_\varepsilon(S) \sim c/\varepsilon^d$, where d is the dimension of the region S.

However, e.g., for a trajectory S of a Brownian motion (which is known to be very non-smooth), with probability 1, we have $N_\varepsilon(S) \sim c/\varepsilon^{1.5}$. It is natural to say that this trajectory has dimension 1.5. In general, if for a set S, we have $N_\varepsilon(S) \sim c/\varepsilon^\alpha$, then we say that the Hausdorff dimension of S is equal to α.

For some sets S, the asymptotic of N_ε is more complex; e.g., similarly to computational complexity, we may have logarithmic factors like $N_\varepsilon(S) \sim c/(\varepsilon^\alpha \cdot \log(\varepsilon))$. To cover such cases, we need a slightly more complex definition of the Hausdorff dimension.

Why is this dimension maxitive? Let A and B be two sets with dimensions $\alpha(A) < \alpha(B)$; this means that to cover A, we can use $c_A/\varepsilon^{\alpha_A}$ elements, and to cover B, we need $c_B/\varepsilon^{\alpha_B}$ elements. Thus, overall, to cover $A \cup B$, we need $c_A/\varepsilon^{\alpha_A} + c_B/\varepsilon^{\alpha_B} \sim c_B/\varepsilon^{\alpha_B}$ elements, so $N_\varepsilon(A \cup B) \preceq c_B/\varepsilon^{\alpha_B}$. On the other hand, to cover both A and B, we need at least as many elements as to cover B, so $N_\varepsilon(A \cup B) \succeq c_B/\varepsilon^{\alpha_B}$, hence $N_\varepsilon(A \cup B) \sim c_B/\varepsilon^{\alpha_B}$ and $\alpha(A \cup B) = \alpha(B)$. Similar arguments cover the case when $\alpha(A) = \alpha(B)$, so, in general, we have $\alpha(A \cup B) = \max(\alpha(A), \alpha(B))$ – i.e., Hausdorff dimension is indeed a maxitive measure.

It can also be proven that it is a σ-maxitive measure.

Example 3: possibility measures. Maxitive measures are also actively used to describe the degree of possibility in human reasoning; see, e.g., [1, 6]. The corresponding maxitive measures are also called *possibility measures*.

Representation theorem for the finite case. When the universal set X is finite, it is possible to define $m(A)$ for all subsets $A \subseteq X$. In this case, we have $A = \bigcup\limits_{x \in A} \{x\}$, hence, due to maxitivity, $m(A) = \max\limits_{x \in A} m(\{x\})$. Thus, in order to describe the values $m(A)$ for all 2^n sets $A \subseteq X$, it is sufficient to describe n values $f(x) \stackrel{\text{def}}{=} m(\{x\})$. The following formula ("representation theorem") enables us to reconstruct all the values $m(A)$ from the known values $f(x)$: $m(A) = \max\limits_{x \in A} f(x)$.

Representation theorem for the continuous case: a problem. In the continuous case, such a representation is no longer possible: e.g., when $m(A)$ is the Hausdorff (fractal) dimension of a set A, we have $f(x) = m(\{x\}) = 0$ for all points x, but, e.g., for an interval A, $m(A) = 1 > 0 = \max f(x)$.

Lutz's result. In [4, 3], Jack H. Lutz used Kolmogorov complexity to show that for constructively defined sets A, the fractal dimension $m(A)$ can be represented as $m(A) = \sup\limits_{x \in A} f(x)$.

What we do: in brief. In this paper, we show that a similar representation is possible for an arbitrary σ-maxitive measure.

2 Result

Let us start by describing what we mean by a *constructive* set. Intuitively, a set is constructive if there exists a constructive procedure for producing elements of this set. Every procedure has to be described by a finite sequence of instructions, i.e., by a finite sequence of symbols in some alphabet used to describe these instructions. Since there are countably many such sequences, there can only be countably many constructive sets. We thus arrive at a following definition:

Definition 1. *Let X be a set, and let $\mathcal{F} \subseteq 2^X$ be a countable family of subsets of X. Elements of $\mathcal{F}$ will be called* constructive sets.

When we define a maxitive measure, it is reasonable to demand that it is defined, in particular, for all constructive sets.

Definition 2. *By a σ-*maxitive measure *on X, we mean a mapping $m : \mathcal{A} \to R$, where $\mathcal{A} \subseteq 2^X$ is a σ-algebra that contains all constructive sets ($\mathcal{F} \subseteq \mathcal{A}$), and for every sequence of sets $A_i \in \mathcal{A}$, $m\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sup\limits_i m(A_i)$.*

Comment. In some situations, it is necessary to use an alternative definition of a maxitive measure, where the value $m(A)$ can be infinite. For example, for the Hausdorff dimension, we can have infinite-dimensional sets. In this paper, we only consider real-valued maxitive measures, for which $m(A) < +\infty$ for every set A.

Representation theorem. *For every σ-maxitive measure on X, there exists a function $f : X \to R$ such that for every constructive set A, we have $m(A) = \sup\limits_{x \in A} f(x)$.*

Proof. In this proof, we use a known monotonicity property of a maxitive measure: that if $A, B \in \mathcal{A}$ and $A \subseteq B$, then $m(A) \le m(B)$. Indeed, since $A \cup B = B$, we have $m(B) = m(A \cup B) = \max(m(A), m(B))$, hence $m(B) \ge m(A)$.

Let us show that the statement of the theorem holds for $f(x) \stackrel{\text{def}}{=} \inf\{m(S) : x \in S \in \mathcal{F}\}$. Let A be a constructive set. For every point $x \in A$, from the definition of $f(x)$, it follows that $f(x) \le m(S)$ for all constructive sets S that contain this point x. In particular, we have $f(x) \le m(A)$. Since $m(A) \ge f(x)$ for every $x \in A$, we thus conclude that $m(A) \ge \sup\limits_{x \in A} f(x)$.

Let us now take an arbitrary $\varepsilon > 0$ and prove that $m(A) \le \sup\limits_{x \in A} f(x) + \varepsilon$. If we prove this for every ε, then we will be able to conclude that $m(A) \le \sup\limits_{x \in A} f(x)$ hence $m(A) = \sup\limits_{x \in A} f(x)$.

Indeed, for every x, by definition of $f(x)$ as the infimum, there exists a constructive set $S_x \ni x$ for which $m(S_x) \le f(x) + \varepsilon$. Since every point $x \in A$ belongs to one of the sets S_x, we conclude that $A \subseteq \bigcup\limits_{x \in A} S_x$.

Since the family of constructive sets is countable, the family of all the constructive sets S_x corresponding to all the points $x \in A$ is also countable. Thus, since m is a σ-maxitive measure, we conclude that $m\left(\bigcup_{x\in A} S_x\right) = \sup_{x\in A} m(S_x)$. Since $m(S_x) \le f(x)+\varepsilon$, we thus conclude that $m\left(\bigcup_{x\in A} S_x\right) = \sup_{x\in A}(f(x)+\varepsilon) = \sup_{x\in A} f(x)+\varepsilon$. Due to the monotonicity property of maxitive measures and the fact that $A \subseteq \bigcup_{x\in A} S_x$, we thus get $m(A) \le \sup_{x\in A} f(x)+\varepsilon$. The theorem is proven.

Comment. The above representation holds not only for constructive sets, but for more general sets as well – e.g., as one can easily show by using σ-maxitivity, for countable unions of constructive sets.

For example, let X be a separable metric space, i.e., a metric space with a metric d that has a countable everywhere dense sequence $\{x_1, x_2, \ldots, x_n, \ldots\}$. The standard real line or Euclidean space are separable: we can take points with rational coordinates as x_i.

Let $\mathcal{F}$ be a family of all the open balls of rational radii with centers in x_i. Then, every open set is a union of such "constructive" sets, and thus, the above representation theorem holds not only for constructive sets, but for all open sets as well.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, NSF grant EAR-0225670, NIH grant 3T34GM008048-20S1, and Army Research Lab grant DATM-05-02-C-0046.

The authors are thankful to the participants of the Conference on Computability and Complexity in Analysis CCA'2003 (Cincinnati, Ohio, USA, August 28-30, 2003) for valuable discussions. The authors are especially grateful to Jack H. Lutz for his guidance and help.

References

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

[2] R. E. Edwards, *Functional analysis: theory and applications*, Dover Publ., N.Y., 1995.

[3] J. H. Lutz, "Effective fractal dimensions", *Mathematical Logic Quarterly*, 2005, Vol. 51, pp. 62–72.

[4] J. H. Lutz, "The dimensions of individual strings and sequences", *Information and Computation*, 2003, Vol. 187, pp. 49–79.

[5] B. B. Mandelbrot, *The fractal geometry of Nature*, Freeman, San Francisco, 1982.

[6] J. Pearl, *Causality*, Cambridge University Press, Cambridge, MA, 2000.

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