

io-port 00107774

Li, Ming; Vitányi, Paul

An introduction to Kolmogorov complexity and its applications.

Texts and Monographs in Computer Science. Berlin: Springer-Verlag. xx, 546 p. DM 108.00; öS 842.40; sFr 108.00 /hc (1993).

This is the first encyclopedic book on Kolmogorov complexity and related topics. It starts with the very basics (definitions of a Turing machine, algorithmic complexity, and probabilities in Chapter 1), and goes into the most complicated and advanced of applications. All major results are given with proofs, and historical sections are both comprehensive and a delight to read. It also includes lots of exercises, ranked by the authors from simple ones to open problems. Crudely speaking, a Kolmogorov complexity $C(x)$ of a word x is the length of the shortest program that generates x (“crudely speaking”, because this length depends on the programming language; it turns out, however, that all universal languages lead to asymptotically the same definition). The original idea behind Kolmogorov complexity was to formalize the notion (supported by intuition and used in applied statistics) that some (finite) sequences of observations are random, and some are not. A sequence that can be generated by a simple algorithm (e.g., 010101 . . .) is clearly not random, and a sequence that can only be generated by enumerating all its 0’s and 1’s (i.e., for which the shortest program is $\text{print}(x)$ of length greater than the length $l(x)$ of x) is clearly not random. So, a sequence is random iff $C(x) \approx l(x)$. Formalizing this \approx is not straightforward: namely, it turns out that among algorithmic tests for randomness (i.e., tests that succeed on sequences of measure 1), there is a universal test that succeeds iff all algorithmic tests succeed, and so, we can define an infinite sequence ω to be random iff it satisfies the universal test. This definition can be reformulated as follows: we want to call a sequence random if it satisfies all the laws of probability theory (like large number law). A “law” can be defined as a set of measure 1 that is constructive in some reasonable sense. So, a sequence is random iff it belongs to the intersection of all constructive sets of measure 1 (there are denumerably many of them, and therefore, this intersection is also of measure 1). The relationship between this definition and Kolmogorov complexity $C(\omega_{|n})$ of fragments (of length n) is not straightforward because $C(\omega_{|n})$ is a complexity of a task to generate exactly $\omega_{|n}$ and to stop at exactly n . A better description of randomness occurs when we use prefix complexity $K(x)$: the smallest length of a program that generates a sequence that starts with x . Another notion to formalize is “ x is random with respect to y ” (as opposed to “ y has some information about x ”). It can be described by using a conditional complexity $C(x|y)$, defined as the smallest length of the program that generates x using y as an oracle. The information $l(x : y)$ that is contained in x about y is then measured by the decrease $C(y) - C(y|x)$ in the program’s length that we can achieve if we are allowed to use x . Similarly to the statistical notion of information, this algorithmic information is symmetric ($l(x : y) \approx l(y : x)$). These basic definitions, their origins, and their properties are described in Chapters 2 and 3. In Chapter 4, $K(x)$ is related to a priori probabilities. In many statistical techniques, we assume some a priori probabilities, and then update them using Bayes’ Rule. The problem is that if initially the probability was 0, it will stay 0. So, it is desirable to find an a priori measure for which as few sets have measure 0 as possible. Intuitively, there is a hope that such a measure exists, because there are only denumerably many algorithmically defined probability measures. It turns out that there only exists a universal semi-measure $\mathbf{M}(x)$ (with sub-additivity instead of real additivity) and that the semi-measure of all sequences that start with x is $\approx 2^{-K(x)}$. In Chapter 5, conditional universal a priori probability $\mathbf{M}(x|y)$ is described. If we have a potentially infinite sequence of observations y_1, \dots, y_n, \dots , then when $n \rightarrow \infty$, $\mathbf{M}(x|y_1 \dots y_n)$ tends to the original distribution for y_i . This result is the basis of R. Solomonoff’s 1960 Induction Theory that started the entire field. For finite n , we can extract from this result a recommendation to choose hypotheses h with small $K(h)$ (this recommendation is thus a formalization of Occam’s razor). One of the big applications of Kolmogorov complexity is to estimate the average complexity of algorithms: instead of analyzing all possible cases, we can analyze cases of high Kolmogorov complexity (and, e.g., for a given length n , the ratio of words x with $C(x) < n - C$ is $\leq 2^{-C}$, so “almost all” words do have high $C(x)$). Such applications are described in Chapter 6. In these applications, it is often difficult to apply the original Kolmogorov complexity, because it counts only the length of a program, without bothering about whether a program is feasible (e.g., polynomial-time), or runs forever (e.g., superexponential time). Therefore, a natural idea (presented in Chapter 7) is to define a resource-bounded complexity as the smallest length of a program that computes x within given resources (e.g., within given time). Universality results can also be proved for such cases, including L. Levin’s construction of an (asymptotically) best algorithm for solving NP problems (so, the question $P = ?NP$ can be reduced to checking whether this particular algorithm is polynomial time). The last Chapter 8 describes applications to statistical physics, where arguments that some sequence is not random and therefore impossible have been used (without proper formalization) for a long time: this is, e.g., how physics explains why a kettle placed on a cold stove does not start boiling, although, mathematically speaking, Brownian motion can

cause it to boil. Kolmogorov complexity provides a consistent language to describe this type of reasoning, and there are holes that it will help in new applications of physics: to biology, to physics of computation, etc. *V.Ya.Kreinovich (El Paso)*

Keywords: Kolmogorov complexity; algorithmic information theory; foundations of statistical methods in physics; inductive reasoning