Research Project

Teturo Kamae

In this research, I'll study the meaning of randomness and quantities representing the degree of randomness. To answer to the philosophical question what is random straightforwardly, the logical point of view is necessary. In fact, randomness is defined using algorithmic notion, Kolmogorov-Chaitin complexity. But, the random numbers thus defined have a fatal weak point, that is, they do exist with probability 1, but none of them are not constructible. On the other hand, pseudo-random numbers are indispensable tool for the computer simulations and they have to be constructed algorithmically. In this research, forgetting the philosophical aspect, we put emphasis on the practical aspect of randomness.

Normal numbers are defined as infinite sequences realizing the law of large number for the uniform i.i.d. random variables in the sense that the relative frequency of blocks occurring in them coincide with their expectations. In the 1970's, it was asked to what extend the normal numbers are considered to be random from the point view of absence of winning strategy (Von Mises' notion of collective). We choose a subsequences of an infinite sequence looking it up to the $(n-1)$-th place to decide whether to put the n-th place into the subsequence or not. If we use a finite automaton for this decision, then it was known that the infinite subsequence obtained from any normal number is again a normal number. Thus, the normality is preserved. I generalized this result for a wider class of infinite automata. I want to revisit this problem to find a necessary and sufficient condition for to preserve the normality.

To be a normal number is a minimum requirement to be a random number, but it is too week. We need a stronger requirement. For this purpose, I introduced a criterion $\Sigma(x)$ for finite sequences $x = x_1 \cdots x_n \in \mathbb{A}^n$ over an alphabet $\mathbb{A}$ $(2 \leq d := \#\mathbb{A} < \infty)$ to measure the degree of randomness. It is defined as the sum over all finite blocks ξ so that $\Sigma(x) = \sum_\xi |x_1 \cdots x_n|_\xi^2$, where $|x_1 \cdots x_n|_\xi$ is the number of occurrences of ξ in x. Comparing this value among finite sequences of the same length, smaller values imply more random. It is proved that $\lim_{n\to\infty}(1/n^2)\Sigma(X_1\cdots X_n) = \frac{d+1}{2(d-1)}$ holds with probability 1 for the uniform i.i.d. random variables $X_1, X_2, \cdots$ over $\mathbb{A}$. I call the infinite sequences $x_1x_2\cdots$ satisfying this almost all property Σ-random numbers. This notion is stronger than the normal numbers. We know an algorithm to construct a Σ-random number starting from any finite sequences. I want to study the properties of the Σ-random numbers to see how they are good as pseudo-random numbers.

The degree of randomness contained in a random variable is the quantity of indefiniteness which turn out to be the quantity of information obtained by the observation since the indefiniteness vanishes by the observation, which is formalized as the entropy. For infinite sequences, the entropy per one letter is discussed, which implies the exponent of the increasing order of the information contained in the subsequence as the length increases. That is, if the information in the subsequence of length n is c^n, then $\log c$ is the entropy per letter. This quantity is the main term of the order of increase of information, but we often need the next term. That is, the polynomial degree of the increase of the information of the subsequence of length n. In fact, in the field of machine learning, the polynomial degree for the 0 entropy case is important and called VC-dimension. I want to study this more precise quantity of information than the entropy. It has just started, which I'll try to develop in this research.