Some definitions of Random Sequence

吕优 518030910427

June 28,2020

In this document, I want to introduce some definitions about whether a binary sequence is random. Let S be a binary sequence. Consider the following S and judge whether it is random.

- $S = 000...000...$
- $S = 010101...01...$
- $S = 0110101000101.....$

All of them are not random, the third S is: $S_i = 1 \Leftrightarrow i$ is a prime. So how to define randomness of a sequence ?

1 A simple idea

Recall the law of large numbers, if we flip a coin infinitely many times, then we have

$$\lim_{n\to\infty} \frac{h_n}{n} = \frac{1}{2} \tag{1}$$

where h_n is the number of head occurs. It is normal to think that if a sequence satisfies the law of large numbers, then it is a random sequence. But it is not that easy. Look at the second S above. It follows the law of large numbers, but it is not random at all.
Recall the definition of normal number in base 2. Let $N(w, n)$ be the number of times the string w as a substring in the first n digits of the sequence S. S is normal if for all finite strings w,

$$\lim_{n\to\infty} \frac{N(w,n)}{n} = \frac{1}{2^{|w|}} \tag{2}$$

We can define that a sequence is random if it is a normal number. However, this is also not a proper definition. Consider this sequence: $S = 011011100101....$ This sequence can be constructed by this way: we write the natural number $0, 1, 2, 3, 4, 5, ...$ in binary and concatenate them together in order. This sequence is called *Champernowne constant* and it is a normal number. So this definition is not good enough.

2 Kolmogorov complexity

The idea of Kolmogorov complexity is how to describe something. A sequence is random if it can not be described in a easy way. For example, the first S above we can say "0 repeats infinite times", and the second is "01 repeats infinite times" and the third is "a sequence that the prime number positions are 1 and others are 0"
If we can describe a given object, then we can judge the complexity of the object by the number of words we use. But we must be careful about the description. For example, we define a natural number as "the least natural number that cannot be described in fewer than 20 words." But we know this sentence words are fewer than 20 words. So Kolmogorov uses Turing Machine to describe. We can take Turing Machine as any program language like C++, Java or Python. Let P is a program with no input and P's output is S. Define $b(P)$ is the binary codes of P. $K(S) = min\{|b(P)|\}$. So the $K(S)$ is the the minimum length of a computer program's binary codes that takes no input and will output S

Observation 1. $K(S) \leq |S| + c$

c is a constant which is determined by the language we use. If we use C++, we can just write printf("0001010...")

Definition 2. *For each constant c we say that a sequence S is c-incompressible if $K(S) \geq |S| - c$*

We may ask if there is a c-incompressible sequence S. There are 2^n binary sequence of length n, but only $\sum_{i=0}^{n-1} 2^i = 2^n - 1$ possible shorter descriptions. Therefore, there is at least one binary sequence S of length n such that $K(S) \geq n$. Using the same counting argument, we find that the number of sequences of length n that are c-incompressible is at least $2^n - 2^{n-c} - 1$. So we can see for each constant $c > 1$, the majority of all strings of length n with $n > c$ are c-incompressible.

One can consider those objects as nonrandom in which one can find sufficiently many regularities. In other words, we would like to identify incompressibility with randomness. This is proper if the strings that are incompressible can be shown to possess the various properties of randomness (stochasticity) known from the theory of probability. That this is possible is the substance of the celebrated theory developed by the Swedish mathematician Per Martin-Lof.

3 Martingale

It is a natural idea that if we are given the first i numbers and we can guess the $(i+1)$th number, the sequence is not random. It is similar to the definition of randomness by Martingale or betting.
Let us consider a fair betting process. The rules are the same as we learned in the class.

- We have one unit of money in the beginning.
- Then we split our money between betting on the current bit value.
- We can look the current bit of S now. If it is 0, the casino pays out twice our bet on 0. If it is 1 the casino pays out twice our bet on 1.
- We again split our new money between betting on the next bit value.
- We can look the next bit of S and get our money.
- We can keep playing for as long as we like.

If the sequence S has a clear pattern on it, it is easy to predict the next bit and make a lot of money. On the other hand, if the sequence is generated by flipping a coin, we can hardly to make large profits. So a natural definition is

Definition 3. *A sequence is random if and only if there is no betting strategy which guarantees arbitrarily large profits when betting on x*

Using mathematical words, If martingale is viewed as a betting strategy, then we can use a function d to describe our betting strategy. $d : \{0,1\}^* \to [0,\infty)$ such that for all finite strings w, $d(w) = \frac{1}{2}(d(w \cdot 0) + d(w \cdot 1))$, where $a \cdot b$ is the concatenation of the strings a and b This limit is to make sure it is a fair betting process. $d(w)$ is the amount of money we have after seeing the string w. If $\lim\limits_{n\to\infty} sup\{d(S_n)\} = \infty$, where S_n is the first n bits of S, then we say d

succeed on S. So we can define a sequence is random if no betting strategy can succeed on the sequence.

4 Reference and My Contribution

The main material I referred is Algorithmic Random Sequence in Wikippedia. Here are my contributions.

- Read the first three chapters of the book An Introduction to Kolmogorov Complexity and Its Application and try to give a definition that is easy to understand of Kolmogorov Complexity.
- Think some other definitions by myself and find it is not very proper. (mainly discussed in part one "a simple idea")
- Try my best to make the discussion interesting and easy to understand.