WHAT EXACTLY DO APPLIED MATHEMATICIANS MEAN BY RANDOM?

Aidan Rocke
aidanrocke@gmail.com

January 27, 2021

ABSTRACT

For applied mathematicians and applied physicists, a definition of randomness that is both general and computable has so far proven elusive due to contradictions between theory and practice. In theory, mathematicians have a definition of algorithmic randomness which is uncomputable whereas in practice engineers use pseudo-random number generators. Here I show that if we start from the international cryptographic standard for random number generators, a strong definition arises naturally from the notion of epistemic uncertainty combined with a combinatorial definition of the Shannon entropy.

1 NIST's definition of randomness

If you turn to page 14 of the international standard for random number generators defined by NIST you may find the following precisions:

Random Number Generator(RNG) :A RNG uses a non-deterministic source(i.e. entropy source) along with some processing function(i.e. the entropy distillation process) to produce randomness.

Entropy source: The entropy source typically consists of some physical quantity such as the noise in an electronic circuit. . . or the quantum effects in a semiconductor.

The authors then add the clarification: 'For cryptographic purposes, the output of RNGs needs to be unpredictable.' From the NIST document we may infer that when scientists and engineers speak of randomness they are implicitly talking about entropy production and uncertainty/predictability. While statistical and epistemic uncertainty are well-defined, the same can't be said of randomness.

2 What is the role of randomness in a computable universe?

If the notion of randomness has any scientific merit then it must correspond to an observable physical process. However, given that every computational process is a mathematical abstraction of a physical process, the observable universe must be a computable universe. In particular, since all physical theories are used to make quantitative predictions they are all effectively computable. This includes the Schrödinger equation and Quantum Field Theory of course.

Furthermore, given that fundamental physics depends upon the assumption that the universe has a compositional structure, for any physical process we may conjecture the existence of a digital or analog computer which may be designed to simulate that process. From this correspondence, we may deduce that algorithmically random(and therefore uncomputable) processes are physically impossible and so any usage of the term 'random' must be short-hand for pseudo-random(i.e. computable).

If the argument elaborated in the last paragraph is not completely clear, the reader may consider that an algorithmically random data generating process implies an incompressible dataset. It follows that it would be void of any structure and so science would be impossible and therefore we would have discovered 'magic'.

3 Good random number generators produce entropy, not randomness

Having expressed doubts on the notion of randomness, we may clarify what is meant by random variable. Within the context of RNGs, a random variable is a deterministic process that serves as an entropy source. This requires uncertainty, a probability distribution and a method for computing entropy.

Let's first focus on uncertainty. If we are speaking of pseudo-random number generators(PRNG) then our source of uncertainty is an information asymmetry between the generator and a potential hacker. This falls under the category of statistical uncertainty, which may be undone with a sufficiently powerful computer. On the other hand, if we have a true random number generator(TRNG) then we are in possession of a physical device which is also a source of epistemic uncertainty. Assuming that we live in a computable universe, this source of uncertainty is due to limitations in humanity's knowledge of physics. Due to the much greater strength of TRNGs, I shall focus the discussion on this category from this point onwards.

Now, if we consider probabilities from a frequentist point of view they are measurable observables, and if you are Bayesian probabilities are still data points. As for a well-defined notion of entropy, we may define the Shannon entropy from a combinatorial perspective where each probability p_i is a useful piece of relative information.

4 Shannon entropy from a combinatorial point of view

Let's suppose you have a small deck of cards in your hands with four kings, three aces, and two queens. If you'd like to consider the number of rearrangements of this deck you might want to know how many unique permutations are possible assuming that cards of the same type are indistinguishable from each other: In principle, financiers may rely upon portfolio theory which may be considered a rational approach to gambling. Given a finite number of assets, a financier will construct a portfolio by taking a weighted average of these assets $W = \frac{9!}{4!3!2!}$, and in general this multinomial coefficient is a solution of the general case:

$$W = \frac{N!}{\prod_i N_i!} \tag{1}$$

where N is the total number of cards and N_i is the number of cards of a particular type.

Now, we may observe that all permutations may be enumerated and assigned a number(in binary) from 0 to W-1 so the string of $\log_2(W)$ bits may be used to encode each permutation. This allows us to define the Combinatorial Entropy S_c as the average number of bits per permutation:

$$S_c = \frac{\log_2(W)}{N} = \frac{1}{N}\log_2\big(\frac{N!}{\prod_i N_i!}\big) \tag{2}$$

and the reader may infer that S_c is in some sense a measure of diversity as the magnitude of S_c varies inversely with the number of cards of the same type $\frac{\partial S_c}{\partial N_i} = -\frac{1}{N}\cdot\frac{1}{N_i}$ and using Stirling's log-factorial approximation $\ln N! \approx N\ln N - N$, we have:

$$S_c \approx \frac{1}{N}\big(\sum_i N_i \log_2 N - \sum_i N_i \log_2 N_i\big) \tag{3}$$

since $\sum_i N_i = N$. Now, we find:

$$S_c \approx -\big(\sum_i \frac{N_i}{N}\log_2\frac{N_i}{N}\big) \tag{4}$$

and if we define the frequencies $p_i = \frac{N_i}{N}$ we recover the usual Shannon entropy $H(\cdot)$:

$$H(X) = -\sum_i p_i \log_2 p_i \tag{5}$$

where X refers to the card deck and $P = \{p_i\}_{i=1}^n$ is a discrete probability distribution.

5 Discussion

If we summarise what has been demonstrated here, we may argue that a necessary and sufficient definition of randomness requires three components:

1. Uncertainty, which may be either epistemic or statistical, which is the source of an information asymmetry between potential hackers and the generator.

2. Physical processes whose statistical behaviour is well-understood, which we use as an entropy source.

3. A method for computing entropy, which allows us to quantify the strength of an entropy source in statistical terms.

Regarding the second criterion, it is worth noting that even deterministic pseudo-random number generators are physical processes since digital computer programmers implicitly orchestrate a large number of electromagnetic interactions in a very precise manner. Now, given that the Shannon entropy is additive we may note that for a sequence of states $\{a_i\}_{i=1}^n$ we have:

$$H(a_1, ..., a_n) = H(a_1, ..., a_{n-1}) + H(a_n) \tag{6}$$

and the process is uninformative if:

$$\forall i, j, k \in \mathbb{N}, H(a_i, a_j) = H(a_i, a_k) \tag{7}$$

and this holds true if, modulo the hacker's epistemic uncertainty of the system's behaviour, this hacker empirically observes a statistically uniform distribution over the state-space. As a corollary, the system's behaviour appears ergodic.

References

[1] Andrew Rukhin et al. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST. 2010.

[2] G. E. Volovik. The Universe in a Helium Droplet. Oxford Science Publications. 2003.

[3] Pablo Arrighi and Gilles Dowek. The physical Church-Turing thesis and the principles of quantum theory. 2011.

[4] Edwin Jaynes. Information Theory and Statistical Mechanics I. The Physical Review. 1957.

[5] Edwin Jaynes. Information Theory and Statistical Mechanics II. The Physical Review. 1957.

[6] Marcus Hutter. Algorithmic information theory. Scholarpedia. 2007.