Randomness

Qcamp 2019, Experimental Session 20/06/2019

1 Introduction

A main ingredient of a secure QKD system is a random number generator. In this exercise, we will explore some simple aspects of quantifying randomness.

1.1 In Our Experiment

In our experiment, the random numbers (for polarisation choices) were generated with the "Entropy" library [1]. This library derives random numbers from the random (entropic) behaviour in the timing (jitter) of the different timers in Arduino: one generated by the crystal oscillator (clock), and another one from an RC circuit. Figure 1 shows the scatter plot for the generated random numbers.

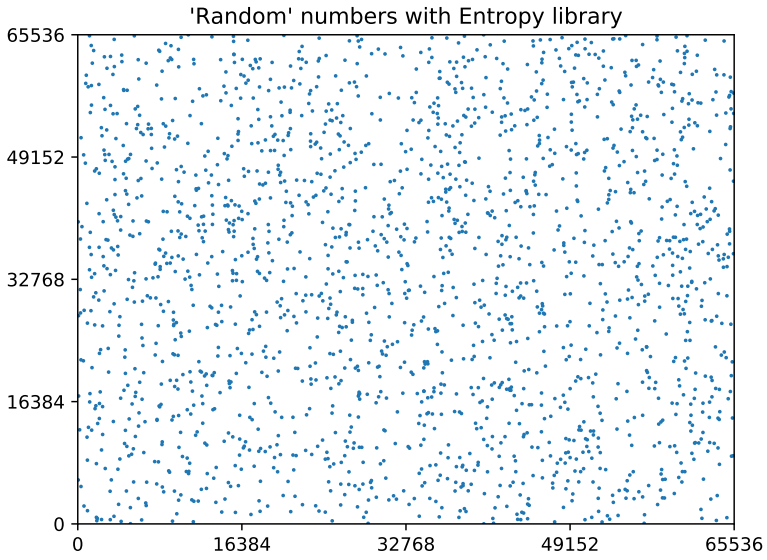

Figure 1: Scatter plot of the random numbers generated with the Entropy library.

One can see that the values generated with the Entropy library are pretty random, but is it? How do you quantify it? How to compare the randomness between, say the Entropy library and the random numbers generated by humans? This question is very hard to answer, and this exercise is designed to give a bit of flavour and perspective in this problem.

1.2 Randomness And Entropy

The concept of randomness is probably related to the concept of disorder, and as physicists (and also computational scientists) we have a quantity for the degree of disorder: entropy [2]. There are many proposed

[1] https://sites.google.com/site/astudyofentropy/

[2] https://en.wikipedia.org/wiki/Entropy_(information_theory)

methods on how to measure entropy and the degree of disorder.

Let us first begin with an overview of the concept of entropy. You may be more familiar with the macroscopic description of entropy as the energy that cannot be used to do work. A more modern understanding of entropy S relates it to the probability distributions of microstates p_i the system has:

$$S = -\sum_i p_i \log p_i \tag{1}$$

Here, we have introduced the term "microstate", which simply means the nitty-gritty details of the system. In contrast, there is also "macrostate" which refers to the big picture of the system. For example, we toss two coins and attempt to predict the result. We know that there will be a total of four outcomes:

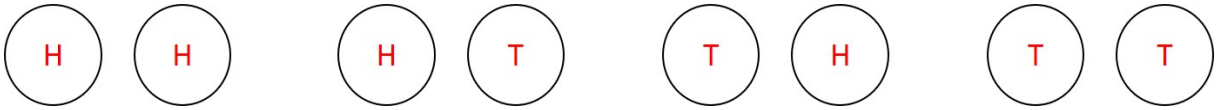

Figure 2: Four different microstates for the same macrostate of two tossed coins.

We say that the for a macrostate of two coins, there are four microstates.

Now let's ask ourselves a simple question: can we predict the most likely outcome of the toss? Suppose we have a fair coin, then each of the four microstates are equally likely to happen and we do not have a clear idea of which outcome gets realised. On the other hand, if we have a biased coin such that the probability of getting a head is much higher than that of a tail, then we are most likely to get HH and least likely to get TT. We can tell quite quickly that we have a biased coin by noting the relative frequency of the HH outcome, which informs us that the situation is not quite random anymore.

With this qualitative picture in mind, we shall now explore the quantification of the results. Using Equation (1), the entropy for two fair coins is

$$\begin{aligned} S &= -\sum_{i=1}^{4} p_i \log_2 p_i \\ &= -4\left(\frac{1}{4}\log_2 \frac{1}{4}\right) \\ &= 2 \end{aligned}$$

where we have chosen the logarithm base to be two, for simplicity. What happens in the case of a biased coin? Let's take, for example, a biased coin with the probability of a head to be 0.9. Then, we have which

HH	HT	TH	TT
0.81	0.09	0.09	0.01

Table 1: Probabilites of the microstates of a two-coin toss.

gives an entropy of $S = 0.938$. The entropy has decreased in going from a distribution with uniform probability to one where certain occurrences are more likely. This means that the entropy is the highest whenever the randomness is the highest. We can essentially see this happening when we plot the entropy S against the probability of a head p:

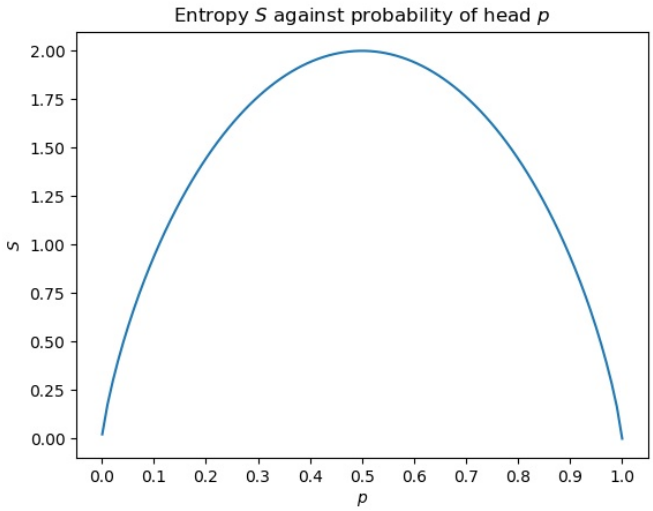

Figure 3: Distriubtion of entropy as the bias of a coin changes

Certainly we see that the maximum is achieved when $p = 0.5$, which is that of a fair coin. At any other values, the entropy decreases towards zero at $p = 0$ and $p = 1$. The reason for this is clear: there is absolutely no doubt about the outcomes as we only get heads or tails. Thus, there is no randomness, and hence no entropy.

2 Assignment

Task 1 [2 pts] Generate 50 random numbers, with each number ranging from 0 to 9! Every member in the group has to contribute at least 5 random numbers.

Task 2 [2 pts] Calculate the entropy for a system of 50 truly random numbers and compare it with the entropy of your generated random numbers.

Task 3 [2 pts] Let us denote the term $\log p_i$ in Equation (1) as X_i. Offer an alternative interpretation of entropy.

Task 4 [2 pts] Prove that the entropy is a maximum only when the probabilities are uniform for the case of tossing two biased coins. Warning: If you want to challenge yourself, try the general case. However, this requires mathematics beyond the scope of JC!

Task 5 [2 pts] Calculate the entropy for the words 'cat', 'car', and 'cqt'. What kind of trend do you notice?

A	B	C	D	E	F	G	H	I	J	K	L	M
8.2%	1.5%	2.8%	4.2%	12.7%	2.2%	2.0%	6.1%	7.0%	0.1%	0.8%	4.0%	2.4%
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
6.7%	7.5%	1.9%	0.1%	6.0%	6.3%	9.0%	2.8%	1.0%	2.4%	0.1%	2.0%	0.1%

Table 2: Relative frequencies of the letters.