Title : Random number generator: testing and whitening.

Co-Encadrants : Andrei ROMASHCHENKO and Alexander SHEN (LIRMM)
contact for more detail : `andrei.romashchenko@lirmm.fr`, `alexander.shen@lirmm.fr`

Keywords : random number generators, statistical tests

Prerequisites : The candidate should have programming skills and some knowledge in probability theory.

Abstract : Generation of random bits is a classical problem known in the context of pseudo-random generators and also in connection with of truly random physical processes (there exist electronic devices that produce random bits using an unpredictable physical noise or intrinsically nondeterministic quantum phenomena). However, the quality of physical generators of random bits remains badly founded and poorly tested. The first objective of this project is an experimental study of the validity and quality of several physical random numbers generators.

When we talk about the quality of random or pseudo-random generators, we have to use randomness tests. The second objective of the project is an inventory and revision of statistical tests for random and pseudo-random generators. We suggest to improve the quality of statistical tests and develop new techniques of "whitening" that improves the quality of non-ideal sources of random bits. Another axis of the project is a conversion of various probabilistic proofs into unconventional randomness tests.

Some more detail : Randomness (in a form of sequences of random bits, random numbers, and so on) is widely used in computer science — in cryptography, in randomized algorithms, in various simulations, etc. So the question arises : where can we obtain necessary random digits suitable for randomized computations and communication protocols ? In some applications even a simple pseudo-random generator would cope with a task. But in more delicate cases we would like to have a source of "truly random" digits, so that all techniques of probability theory would apply properly. So, how to produce a stream of truly random numbers ?

On the face of this problem, it seems pretty simple. Indeed, randomness is ubiquitous — from coin tossing and cosmic rays to thermal noise in audio and video recordings, Brownian motion, quantum measurements and radioactive decay. Some randomness is present in a physical noise of any nature, and

noise sources are cheap and easy to find. In practice, we can produce random digits from a tailor-made physical device generating random noise (a classical example is a Zener diode which costs few cents ; the noise generated by it is enough to be captured by an inexpensive mixer that has microphone inputs). Alternatively, we can use more professional gadgets available on the market (some of these gadgets also retrieve randomness from a Zener diode or from similar classical physical schemes ; the other obtain more refined randomness by measurements of quantum mechanical systems, e.g., with a photon polarization experiments).

So one may think that constructing a good randomness generator is an easy task. However, if we require that the output distribution is guaranteed with high precision, the problem becomes much more difficult. Even tossing a coin is not an ideal experiment. A coin may be biased, the independence between two consecutive coin tosses may be not absolute. The things are even more complicated if we retrieve randomness from a noisy physical process.

Evaluation of the quality of the resulting random bits is a rather delicate question. In practice people use *randomness tests*, which should detect an apparent or hidden deficiency of randomness in the outcome of the generator. The early history of randomness tests (as well as pseudorandom number generators) is described by Knuth in [1]. For a recent survey of this area we address the reader to [3].

Thus, whatever physical device provides us with random digits, we face two problems :

(i) How to process the physical random bits in order to improve their "quality" if they are initially not perfectly unbiased and not perfectly independent ?

(ii) to test and evaluate the resulting random bits and how to distinguish between bad and good sources of randomness ?

In the proposed internship project we suggest to study a collection of physical generators of random bits (simple electronic devices made in our laboratory and several gadgets purchased on the market) and address the two problems mentioned above. More specifically, we propose to test the produced random bits with statistical tests (the standard battery of tests from `Dieharder`, see [2], and its extension) and try different techniques of post-processing ("whitening") of the raw physical bits with *randomness extractors* that may help to improve the quality of "weak" sources of randomness.

This projects requires programming skills (adjusting the existing tests and

implementing new one) as well as a general mathematical culture (to understand the mathematics behind the programming code implementing statistical tests). Thus, programming competencies are mandatory ; basic knowledge of probability and statistics is highly desirable. Though this project studies the quality and properties of electronic devices (several specific physical generators of random digits), we do *not* require any knowledge of physics or microelectronics.

Bibliography :

[1] Donald Knuth, The Art of Computer Programming, Volume 2 : Seminumerical Algorithms, 3rd edition (Addison-Wesley, Boston, 1998).

[2] Robert G. Brown, DieHarder : A Gnu Public License Random Number Generator, version 3.31.1. (2006–2018)
http://www.phy.duke.edu/∼rgb/General/dieharder.php

[3] Alexander Shen, Randomness tests : theory and practice.
http://www.lirmm.fr/∼ashen/racaf/2019-preliminary-report.pdf