entropy

MDPI

Article

Advanced Statistical Testing of Quantum Random Number Generators

Aldo C. Martínez [1], Aldo Solis [2], Rafael Díaz Hernández Rojas [3], Alfred B. U'Ren [2], Jorge G. Hirsch [2] and Isaac Pérez Castillo [4,5,*]

[1] Department of Physics, Center for Research in Photonics, University of Ottawa, 25 Templeton St, Ottawa, ON K1N 6N5, Canada; mbac@ciencias.unam.mx

[2] Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Apdo. Postal 70-543, Cd. Mx., C.P. 04510 Mexico, Mexico; aldo.solis@correo.nucleares.unam.mx (A.S.); alfred.uren@correo.nucleares.unam.mx (A.B.U.); hirsch@nucleares.unam.mx (J.G.H.)

[3] Dipartimento di Fisica, Sapienza University of Rome, P.le Aldo Moro 5, I-00185 Rome, Italy; rafael.diaz.hr@gmail.com

[4] Departamento de Física Cuántica y Fótonica, Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, Cd. Mx., C.P. 04510 Mexico, Mexico

[5] London Mathematical Laboratory, 8 Margravine Gardens, London W6 8RH, UK

* Correspondence: isaacpc@fisica.unam.mx

Received: 20 October 2018; Accepted: 14 November 2018; Published: 17 November 2018

Abstract: Pseudo-random number generators are widely used in many branches of science, mainly in applications related to Monte Carlo methods, although they are deterministic in design and, therefore, unsuitable for tackling fundamental problems in security and cryptography. The natural laws of the microscopic realm provide a fairly simple method to generate non-deterministic sequences of random numbers, based on measurements of quantum states. In practice, however, the experimental devices on which quantum random number generators are based are often unable to pass some tests of randomness. In this review, we briefly discuss two such tests, point out the challenges that we have encountered in experimental implementations and finally present a fairly simple method that successfully generates non-deterministic maximally random sequences.

Keywords: Bell inequalities; algorithmic complexity; Borel normality; Bayesian inference; model selection; random numbers

1. Introduction

Monte Carlo methods are one of the essential staples of the basic sciences in the modern age. Although these gained prominence during the early 1940s, thanks to secret research projects carried out in Los Alamos Scientific Laboratory by Ulam and von Neumann [1,2], their origins may be traced back to the famous Buffon's needle problem, posed by Georges-Louis Leclerc, Comte de Buffon, in the 18th century. In the present day, Monte Carlo "experiments" are seen as a broad class of computational algorithms that use repeated random sampling to obtain numerical estimates of a given natural or mathematical process. In order to use these methods efficiently, fully random sequences of numbers are needed. Back in the 1940s, this was a tall order, and various methods to generate random sequences were used (some of them literally using roulettes), until von Neumann pioneered the concept of computer-based random number generators. During the following years, these became the standard tool in Monte Carlo methods and are still generally well-suited for many applications. However, these computer-based methods generate pseudo random numbers [3], which means that the generated sequence can be determined given an algorithmic program and an initial seed, two ingredients which are hardly random. Thus, in order to achieve a truly unpredictable source of random numbers, we must eliminate these two deterministic aspects. The former is easy to overcome using, for example, a pattern of keystrokes typed on a computer keyboard as a random seed. On the other hand, the algorithmic program could be replaced, for instance, by a classical chaotic system [4]. Examples of the latter abound in the area of weather prediction and climate sciences.

In recent years, however, the community has been moving towards using the fundamental laws dictating the behaviour of the quantum realm for the generation of sequences of truly random numbers. This seems, at a first glance, to be at odds with the following rather naïve thought: if the natural laws of the microscopic world are considered to be a computer program under which a system evolves from an initial state (a seed), should not its corresponding generated sequence also be predictable? It turns out that Quantum Mechanics, in its current standard view, related to the Copenhagen interpretation, has a special ingredient that makes the random sequence inherently unpredictable for both the generator and the observer. Such a strange behaviour has been eloquently recast over the years in various forms, famously by the quote "spooky action at a distance" due to Einstein, or mathematically by the celebrated work of Bell [5,6]. The application of quantum randomness in cryptography has given rise to the concept of device independent randomness certification, which, in a nutshell, corresponds to those processes that violate Bell's inequalities [7,8]. However, there seems to be some confusion in the literature regarding two different properties of a given sequence of random numbers. The first one, rather important as we have argued above, is whether the sequence is truly random, meaning that it is unpredictable. In contrast, the second one is related to the issue of assessing whether or not it is biased. It is crucial to keep in mind that these two properties are independent, as evidenced by the random number generator Quantis [9], which is based on a quantum system and is able to pass the standard tests of randomness (NIST (National Institute of Standards and Technology) suite) [10] but has difficulties with other tests [11].

Due to recent advances in quantum technologies, and since the NIST suite has been examined in other works [12], together with a critical view on the use of p-values on which the NIST suite relies [13], it becomes necessary to consider other criteria for measuring the performance of quantum random number generators. Thus, we focus solely on two recently introduced approaches: the first one is based on algorithmic complexity theory evaluating incompressibility and bias at the same time, since an incompressible sequence is necessarily an unbiased one [14], while the second one relies on Bayesian model selection. Both methods are based on solid structures which lead to a definition of randomness that is very intuitive and which arises independently of the development of random number generators. We apply them to analyze sequences of random bits generated in our laboratory using quantum systems. We also address the issue about the origin of the biases observed when utilizing these types of devices.

2. Tests of Randomness

A simple criterion for assessing the predictability of a sequence is the presence of patterns in it. For example, for the sequence 01010101..., we can ask ourselves whether the next number is either 1 or 0. The natural answer is 0 based on the pattern observed in the previous bits. In general, we would like to find any possible regularity that helps us to predict the next bit. Within the framework of algorithmic information theory, it is possible to address this problem by noting that any sequence which exhibits regularity can be compressed using a short algorithm which can produce as output precisely such patterns. Thus, a sequence of this type could be reproduced using fewer bits than the ones contained in its original form. Therefore, whenever a sequence lacks regularity, we refer to it as "algorithmically" random.

We now introduce a remarkable result from algorithmic information theory: the Borel-normality criterion due to Calude [14], which allows us to asymptotically check whether a sequence is not "algorithmically" random. Assuming we are given a string $\ell = \{1001010110110\cdots\}$ of $|\ell| = n$ bits (We will only consider binary sequences, but our results are easily generalizable to other alphabets), the idea of the Borel-normality criterion consists primarily of dividing the original sequence ℓ into consecutive substrings of length i and then computing the frequencies of occurrence of each of them. For brevity and later use, let us define $\Omega^{(i)}$ as the set of 2^i substrings that can be formed with i characters, let ℓ_i be the sequence obtained after dividing it into substrings, and $|\ell|_i \equiv [|\ell|/i]$. Additionally, let $N_i^j(\ell)$ be the number of times the j-th substring of length i appears in ℓ. For example, when considering substrings of length $i = 1$, we are looking at the frequencies of the symbols $\Omega^{(1)} = \{0, 1\}$ that conform the original string ℓ, while for $i = 2$, we have to consider the frequencies of four substrings, namely $\Omega^{(2)} = \{00, 01, 10, 11\}$. According to Calude, a necessary condition for a sequence to be maximally random is that the deviations of these frequencies with respect to the expected values in the ideal random case should be bounded as follows [14,15]:

$$\left| \frac{N_i^j(\ell)}{|\ell|_i} - \frac{1}{2^i} \right| < \sqrt{\frac{\log_2(n)}{n}}, \qquad j = 0, \ldots, 2^i - 1 \,. \tag{1}$$

This condition must be satisfied for all substrings of lengths from $i = 1$ up to $i_{\max} = \log_2(\log_2(n))$. Intuitively, this criterion "compresses" the original sequence by reading i bits at a time and tests whether the substrings appear with a frequency that differs from what would be expected in the random case, thus indicating the presence of some regularity. We emphasize that since Borel-normality is not a sufficient criterion for randomness, it can only be used to assess whether a given sequence is not random. In other words, even if a sequence satisfies Equation (1) for all substrings and allowed values of i, the Borel-normality condition cannot guarantee that it is indeed random.

Recently, a Bayesian criterion has been introduced [16,17] by some of the authors of the present article to test, from a purely probabilistic point of view, whether a sequence is maximally random as understood within information theory [18]. The method works by exploiting the Borel-normality compression scheme and then recasting the problem of finding possible biases in the sequence as an inferential one in which Bayesian model selection can be applied. Specifically, for a fixed value of i, we need to consider all the possible probabilistic models, henceforth denoted as $\{\mathcal{M}_\alpha^{(i)}\}_\alpha$, that could have generated the sequence ℓ. Each such model determines a unique probability assignation to the elements of $\Omega^{(i)}$, which depends on a set of prior parameters $\boldsymbol{\theta}$. For these parameters, the Jeffreys' prior, $P_{\text{Jeff}}(\boldsymbol{\theta})$, turns out to be a convenient choice of prior parameter distribution, as it entails the "Occam Razor principle" in which more complex models are penalized, as well as being mathematically convenient for the case at hand; some other advantages are pointed out in [16,17].

Next, the question of finding all the generative models that can produce a sequence ℓ is ultimately solved by noticing that all the possible probabilities assignations are in a one-to-one correspondence with all possible partitions of $\Omega^{(i)}$. Since obtaining the partitions of any set is a straightforward combinatorial task [19], we are able to determine all the relevant models when searching for possible biases in the generation of ℓ. For instance, when $i = 1$, there are two possible models: one in which the two elements of $\Omega^{(1)}$ are equiprobable, corresponding to the partition $\{\{0, 1\}\}$ of $\Omega^{(1)}$ into one subset—i.e., the same set—and another model with probabilities $p_0 = \theta$, $p_1 = 1 - \theta$ corresponding to the partition $\{\{0\}, \{1\}\}$ of $\Omega^{(1)}$ into two subsets. Even though it might seem that the first model is just a particular case of the second one (by letting $\theta = 1/2$), we should keep in mind that the prior distributions are different in both cases, $\delta(\theta - 1/2)$ and $\frac{1}{\pi\sqrt{\theta(1-\theta)}}$, respectively, thus yielding two different models. Analogously, for $i = 2$, there is a single unbiased model, which corresponds to the partition of $\Omega^{(2)}$ into one subset (with probabilities $p_j = 1/4$, for $j = 00, 01, 10, 11$), and 14 additional models associated with the different ways of dividing $\Omega^{(2)}$ into subsets. The latter are related to the number of ways of distinguishing among the elements of $\Omega^{(2)}$ during the assignation of probabilities, and thus any of these models would entail some bias when generating a sequence. Note that, in general, for any value of i, we will face a similar situation in which a single model can produce an unbiased, and hence maximally random, sequence by means of a uniform distribution, while the rest of them will be some type of categorical distribution.

Once all the models have been identified, the remaining part is the computation of the posterior distribution $P\left(\mathcal{M}_\alpha^{(i)}\middle|\ell\right)$, which from an inferential point of view is the most relevant distribution as it gives the probability that the model $\mathcal{M}_\alpha^{(i)}$ has indeed produced the given sequence ℓ. Note that, since a generative approach was adopted, we have direct access to the distribution $P\left(\ell\middle|\boldsymbol{\theta}, \mathcal{M}_\alpha^{(i)}\right)$, which can be combined with the parameters' prior $P_{\text{Jeff}}(\boldsymbol{\theta})$ to obtaining the distribution $P\left(\ell\middle|\mathcal{M}_\alpha^{(i)}\right) = \int \mathrm{d}\boldsymbol{\theta}\, P\left(\ell\middle|\mathcal{M}_\alpha^{(i)}, \boldsymbol{\theta}\right) P_{\text{Jeff}}(\boldsymbol{\theta})$. One of the most important results of [16] is that this marginalization can be accomplished exactly for all the models and any value of i. Therefore, we can obtain the posterior distribution by a simple application of Bayes' rule:

$$P\left(\mathcal{M}_\alpha^{(i)}\middle|\ell\right) = \frac{P\left(\ell\middle|\mathcal{M}_\alpha^{(i)}\right) P\left(\mathcal{M}_\alpha^{(i)}\right)}{\sum_\gamma P\left(\ell\middle|\mathcal{M}_\gamma^{(i)}\right) P\left(\mathcal{M}_\gamma^{(i)}\right)}, \qquad (2)$$

with $P\left(\mathcal{M}_\gamma^{(i)}\right)$ being the prior distribution in the space of models that can generate sequences of strings of length i. Therefore, the best model $\alpha^\star$ that describes the dataset ℓ is quite simply given by

$$\alpha^\star = \arg\max_\alpha P\left(\mathcal{M}_\alpha^{(i)}\middle|\ell\right). \qquad (3)$$

If the best model $\mathcal{M}_{\alpha^\star}^{(i)}$ turns out to be the unbiased one for all possible lengths i of the substrings, then we can say that the process that generated that dataset was maximally random. However, it remains to discuss how large the length i of substrings can be for a given dataset of n bits. To answer this, we first note that, for any set containing N elements, the possible number of partitions is given by the N-th Bell number B_N [19]. Thus, for a given i, all possible partitions of the set $\Omega^{(i)}$ will result in B_{2^i} models to be tested, and, therefore, it is expected for them to be sampled at least once when observing ℓ. This means that $B_{2^{i_{\max}}} \sim n$, which, for sufficiently large n, yields $i_{\max} = \log_2(\log_2(n))$, precisely as in the Borel-normality criterion.

Randomness characterization through Bayesian model selection has some clear and natural advantages, as already pointed out in [16], but, unfortunately, it has an important drawback: the number of all possible models for a given length i, given by B_{2^i}, grows supra-exponentially with i: indeed, for $i = 1$, we have two possible models, for $i = 2$, we have 15 possible models, for $i = 3$, we have instead 4140 possible models, while, for $i = 4$, we have 10,480,142,147 models. Thus, even if we are able to acquire data for the evaluation of these many models, it becomes computationally impractical to estimate the posterior for all of them using Equation (2). There is an elegant strategy to overcome this difficulty: one can derive bounds similar to those provided by the Borel-normality criterion, by comparing the log-likelihood ratio between the maximally random model and the maximally biased one. This yields the following bound for the frequencies of occurrence [17]:

$$\sqrt{\sum_{j\leq j'=1}^{2^i-1} \left(\frac{N_i^j(\ell)}{|\ell|_i} - \frac{1}{2^i}\right)\left(\frac{N_i^{j'}(\ell)}{|\ell|_i} - \frac{1}{2^i}\right)} < \sqrt{\frac{i^2}{n^2\psi_1\left(\frac{1}{2}+\frac{n}{i2^i}\right)} \ln\left(\frac{2^{-n}\Gamma^{2^i}\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2^{1-i}}+\frac{n}{i}\right)}{\Gamma\left(\frac{1}{2^{1-i}}\right)\Gamma^{2^i}\left(\frac{1}{2}+\frac{n}{i2^i}\right)}\right)}, \qquad (4)$$

where ψ_1 is the polygamma function of order 1. Note that, unlike Calude's bound given by Equation (1), this new Borel-type bound couples all frequencies, and, moreover, results in highly restrictive bounds.

3. Ideal Random Number Generation

While intuition dictates that quantum random number generators (QRNG) should be superior to their classical counterparts, such a comparison was carried out in [11] and very recently in [20], with rather disappointing results. For the classical case, the authors used three Pseudo-Random Number Generators (PRNG): the generators included in the software packages Mathematica and Maple, and the digits of π expressed in base 2. For QRNG, they used two devices: (i) Quantis, developed by IDQ [9], a quantum random number generator interfaced with a common computer, and (ii) an experiment from a quantum optics group in Vienna. The latter experiment consists of a very weak light source, attenuated to the single photon level, a beam splitter, and two single photon detectors. Leaving aside the question of which QRNG performs better, the real surprise was that the PRNGs come out in this test with a superior performance, by far, as compared to their quantum counterparts. This result appears to be at odds with the natural randomness associated with quantum phenomena. Why is it that the inherent quantum randomness does not translate into better performance with respect to classical systems? Does randomness, as discussed in this paper, have no impact on the performance of the generators? Is this a fundamental or a technical problem?

In order to explore this apparent paradox, we will discuss the different technical and design difficulties associated with quantum random number generation using light. These days, it is straightforward to detect single photons using avalanche photo-diodes (APD), devices capable of detecting up to a few million single photons per second with $> 60\%$ detection efficiencies employing relatively simple electronics. With this simple design in mind, we only need a single photon source, a beam splitter (BS), and a couple of single-photon detection devices in order to set up a QRNG device. This minimalistic design is sketched in Figure 1.

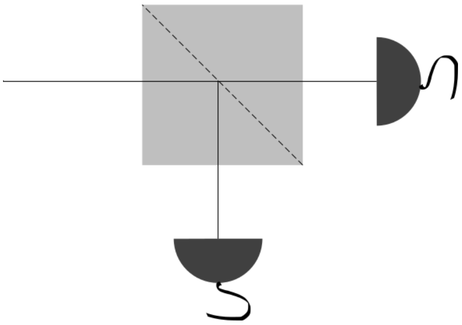

Figure 1. Ideal experimental setup for a naïve QRNG (quantum random number generators) based on an individual photon source and a beam splitter (BS). Neglecting the possible losses, a photon will activate only one of the two detectors, therefore producing a random bit per photon.

4. Experimental Challenges in Random Number Generation

Suppose now that we have a single-photon source and we want to generate a sequence of bits to be tested against the bounds given by Equation (1). Let us start by first focusing on the so-called Borel level, the word length i, which can take a maximum value of $i_{\max} = \log_2(\log_2(n))$. In Table 1, we report how $i_{\max}$ grows with n, up to a value of $i_{\max} = 6$. In order to achieve a Borel level $i_{\max} = 6$, we will require a dataset of length 10^{18} events. Assuming that our single-photon generator and detectors can cope with around a million of events per second, we would then require on the order of 600 years to generate a sequence of that length! It turns out that $i_{\max} = 5$ is a more realistic value, since it leads to a required dataset ℓ of size 4.3×10^9 events, which can be realistically produced in a couple of hours.

Table 1. Necessary data lengths for maximum Borel level $i_{\max} = \log_2(\log_2(n))$. The double exponential relation grows so quickly that it is not possible get to level 6.

Maximum Borel Level $i_{\max} = \log_2(\log_2(n))$	Data Length n
1	4
2	16
3	256
4	65,536
5	4,294,967,296
6	18,446,744,073,709,551,616

The bound on the right-hand-side of Equation (1) implies that the frequencies of occurrence $\frac{N_i^j(\ell)}{|\ell|_i}$ for substrings of length $i \leq i_{\max}$ cannot deviate from the ideal random one, $1/2^i$, by more than 8.6×10^{-5}, which constitutes an extremely tight tolerance. Hence, in practice, any part of the naïve experimental setup that gives rise to some bias will unfortunately make the dataset ℓ unable to pass the Calude criterion. The first component that we must be wary about is the BS. A regular BS usually has an error figure in the region of 1%, which is very high with respect to the stringent tolerance ℓ would need in order fulfill Borel normality. Is it plausible to correct this using a Polarizing Beam Splitter (PBS), instead of the BS, with an active control through feedback of the state of polarization so as to compensate for any bias in the PBS? In what follows, we investigate this question through a simple experiment. The state of polarization of a single photon entering the PBS can be written as

$$|\psi\rangle = a\,|V\rangle + e^{i\phi} b\,|H\rangle\,, \tag{5}$$

where $|V\rangle$ and $|H\rangle$ refer to the vertical and horizontal polarization components, respectively. We can approach the state in Equation (5) by transmitting the laser beam trough a half wave plate (HWP) so as to achieve arbitrary rotation of the linear polarization. Assuming a perfect, unbiased BS, we would need an incoming polarization state with $a = b = 1/\sqrt{2}$ so that the resulting sequence of bits is unbiased. If, on the other hand, the PBS exhibits biases (e.g., due to manufacturing error), we can adjust the orientation of the above-mentioned half wave plate so as to adjust precisely the value of our coefficients a and b to compensate for the PBS bias.

Our experimental setup, shown in Figure 2, can be regarded as the minimal realistic device for the implementation of a QRNG. The main questions which we wish to address are: (i) how good are the sequences of bits generated by such a device? In addition, (ii) do they pass the Borel-normality criterion?

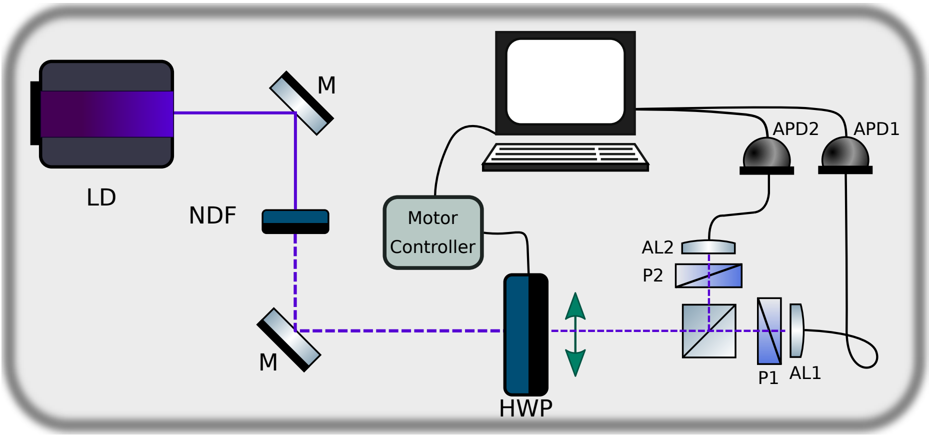

Figure 2. Experimental setup. The relative angle of the half wave plate is controlled in order to reduce bias.

The input state is prepared using the beam from a laser diode (LD). The beam is transmitted through a set of neutral density filters (NDF) with a combined optical density 7.3 for attenuation to a level compatible with the maximum recommended count rate of our single photon detectors. The beam is then transmitted through a half wave plate (HWP) mounted on a motorized rotation stage so as to control its orientation angle relative to the PBS axes. The PBS splits the beam into two spatial modes according to the H and V polarizations, each of which is coupled with the help of an aspheric lens (AL1 and AL2) into a multimode fiber leading to an avalanche photodiode (APD1 and APD2). We include a polariser (P1 and P2), with an extinction ratio, defined as the ratio of the maximum to the minimum transmission of a linearly polarized input, of 100,000:1 prior to each of the aspheric lenses (AL1 and AL2) for a reduction of the non-polarized intensity reaching the detectors.

Suppose now that we prepare our system so that the average relative power $P_i/(P_0 + P_1)$ of each APD detector $i = 0, 1$ starts ideally at $1/2$. In Figure 3, we show how this average relative power evolves with time (see curve labelled "without feedback"). Note that, even though the system starts in a perfectly balanced state, it rapidly deviates from this condition. The slow change of this curve can be attributed to thermal drift while the oscillatory component with a period of approximately half an hour is related to the air conditioning system in the laboratory. These effects can be effectively compensated by rotating the HWP. After some study of the response function of our experimental setup, a correction every minute with a proportional controller was sufficient to correct for all these effects leading to a steady response (see curve labelled "with feedback") [21].

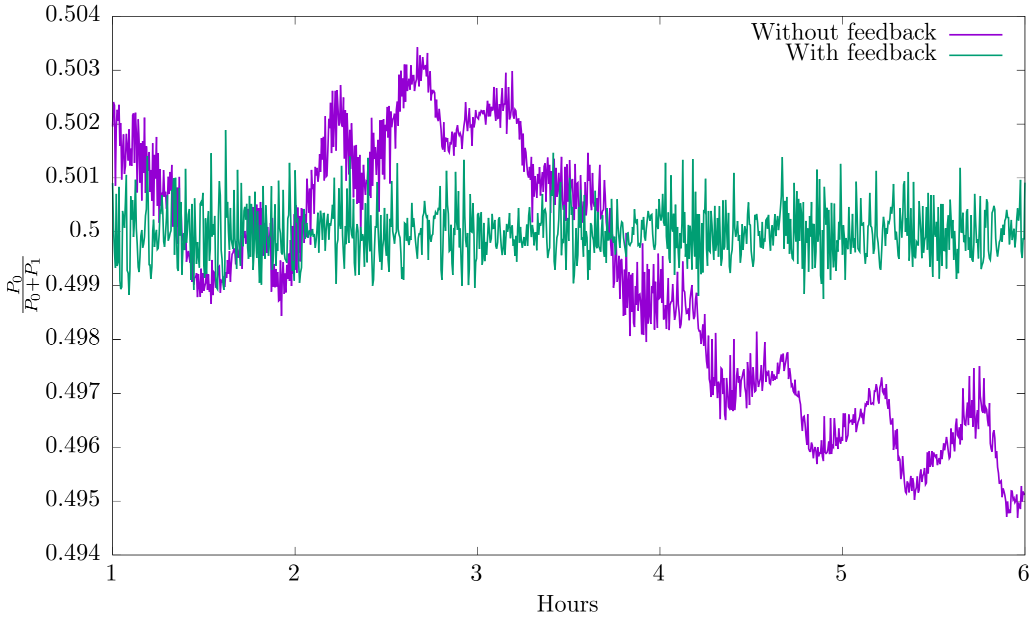

Figure 3. Evolution in time of the normalized power. The device starts in a perfect balanced state but quickly deviates from this condition. By using a feedback mechanism, we can obtain stability of the normalized power within an error of 0.004.

5. First Battery of Results

We have used the experimental setup described in the previous section to generate a sequence of 4,294,967,296 bits, allowing us to test the Borel–Normality criterion up to level $i_{\max} = 5$. The results of this analysis are depicted in Figure 4. In the plot, bars represent the deviations from the ideal value for all the strings at each Borel level. For instance, in the first part of the analysis (purple bars), there are only two bars corresponding to the frequency of occurrences of substrings "1" and "0". As our initial setup is very fine-tuned and stable, the bars have practically zero height, with value 5×10^{-6}. The green bars represent the second part of the analysis, or Borel level two, corresponding to frequencies of occurrences of symbols $\{00, 01, 10, 11\}$, and so on. In the same figure, the horizontal lines represent the bound given by the right-hand-side of Equation (1). Our first battery of results are a clear disappointment: only the first set for substrings of length one clearly passes the test, while, for higher lengths, our QRNG fails miserably to pass Calude's criterion.

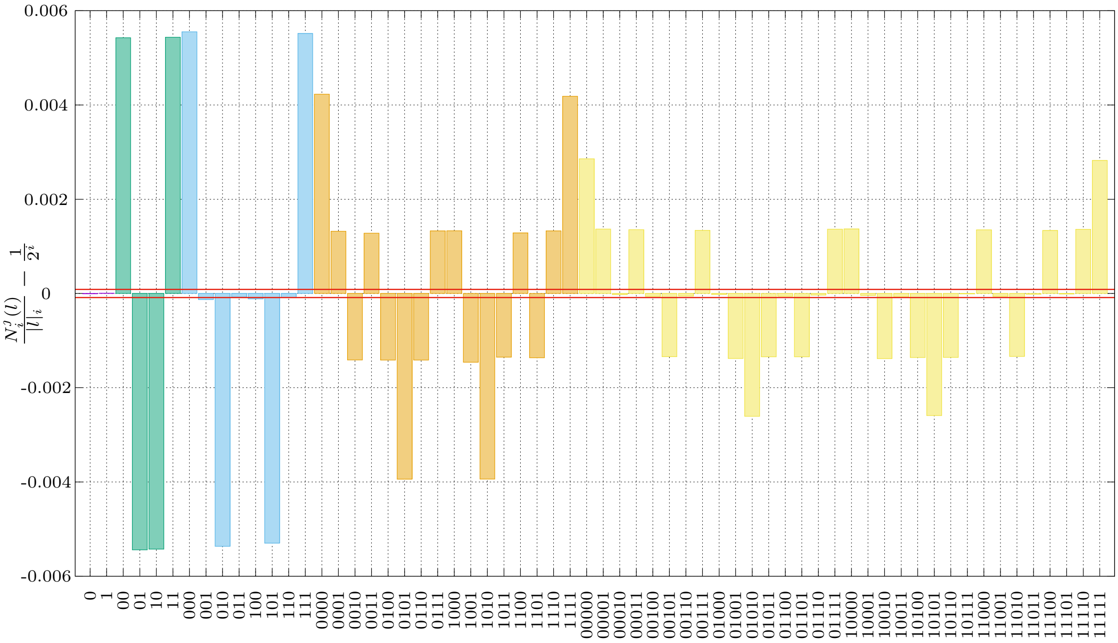

Figure 4. Results from Borel analysis. The first two boxes correspond to the deviations from the mean at the first Borel level; these exhibit the same height but opposite signs. The next four bars (green) represent the deviations for level two, i.e., "00", "01", "10", "11". The blue and yellow boxes represent the deviations for level three and four, respectively. The red lines correspond to Borel's bound, which turns out to be much smaller than the deviations. Only the first level passes the test.

Furthermore, a closer look at the green bars shows that events 00 and 11 appear more frequently than expected, by about 0.005%, (while events 01 and 10 appear less frequently than expected by the same margin). This effect reveals a correlation between "equal events", that is, the same digit appearing twice. In terms of our experiment, this means that it is more probable to observe an event in a detector once a previous event has already been recorded. Other parts of our test validate this: for Borel level three, the yellow bars indicate that events with alternate zeroes and ones (010 and 101) appear less frequently than expected, also by about 0.005%. At Borel levels four and five, the larger deviations appear for events 0101 and 1010, clearly in accordance with the previous results. This indicates that certain parts of our experimental setup are introducing unwanted correlations between bits, which results in the magnitude of some of the deviations to be 50 times larger than expected. Our experimental effort clearly does not suffice for our sequences to pass the Borel–Normality criterion. How is this possible?

6. APD Effects on Introducing Correlations

The two main effects in the behavior of our APDs which can introduce undesired correlations in the resulting sequences of bits are called after-pulsing and dead time [22,23]. The first effect, roughly speaking, corresponds to a false detection event due to the residual effects of an avalanche triggered by a previous event, while the dead time is the time period after each event during which the system is not able to record a subsequent incoming optical signal. In this case, we have a typical dead time of 22 ns, a maximum after-pulsing probability of 1%, and a dark counts rate of 100 counts/s.

While the device exhibits a linear behaviour up to 5 × 10^6 counts/s, the detection rates used in our experiments are an order of magnitude lower.

The mechanism by which dead time introduces correlations in our data, particularly in experimental arrangements with two or more APDs as in our case, is as follows: suppose that we have an event in one of our detectors. During its dead time, it will have zero probability of recording another incoming event in that detector, while the other detector still exhibits a non-zero probability of recording an event. On the other hand, the way after-pulsing introduces correlations in our data is by increasing the probability of observing consecutively the same event, resulting in the observation of an excess of the events $\{00\}$ and $\{11\}$ in Figure 4.

These two effects can somewhat be corrected either by re-designing our experimental setup and/or modifying the software. Instead of following this route to generate maximally random sequences of bits, let us pursue a rather simple solution as discussed below.

7. Random Number Generation Using Time Measurements

We now follow a method introduced in [3]. Suppose that $\rho(x)$ is the probability density function of a continuous random random variable X on an interval $x \in (a, b)$. Let us further assume that its real value x is represented up to a given precision so that we assign a parity to x according to the parity of its least significant digit. Next, we divide the interval (a, b) into an even number $2L$ of bins and introduce

$$x_i = a + \frac{i(b-a)}{2L}, \qquad i = 0, \ldots, 2L\,. \tag{6}$$

Suppose that $2L$ and the precision has been chosen so that x_i is even for i even and odd for i odd. It follows that

$$1 = \int_a^b dx\rho(x) = \sum_{i=0}^{2L-1} \int_{x_i}^{x_{i+1}} dx\rho(x) \equiv \mathcal{N}_{\text{even}} + \mathcal{N}_{\text{odd}}\,, \tag{7}$$

with

$$\mathcal{N}_{\text{even}} \equiv \sum_{i=0}^{L-1} \int_{x_{2i}}^{x_{2i+1}} dx\rho(x)\,, \qquad \mathcal{N}_{\text{odd}} \equiv \sum_{i=0}^{L-1} \int_{x_{2i+1}}^{x_{2i+2}} dx\rho(x)\,. \tag{8}$$

Approximating the integrals by the left sum rule, we can write that

$$\mathcal{N}_{\text{even}} \sim \frac{b-a}{2L} \sum_{i=0}^{L-1} \rho(x_{2i})\,, \qquad \mathcal{N}_{\text{odd}} \sim \frac{b-a}{2L} \sum_{i=0}^{L-1} \rho(x_{2i+1})\,, \tag{9}$$

which implies that, roughly, the probability that the least significant digit is odd can be expressed as

$$\mathcal{N}_{\text{odd}} \sim \frac{1}{2} + \frac{1}{2} \sum_{i=0}^{L-1} \rho'(x_{2i}) \left(\frac{b-a}{2L}\right)^2 , \tag{10}$$

where the bias term can be fine-tuned by increasing either the number of bins or through a smooth density $\rho(x)$, or both.

This method can be very easily implemented in the lab as follows. Suppose that the random variable X is the time difference between two consecutive photon arrivals to the detector. In our case, these times are of the order of 500 ns to 10 µs. A typical sequence of these time differences look like:

$$592\,342 ps,$$
$$595\,634 ps,$$
$$593\,645 ps,$$
$$592\,342 ps,$$
$$595\,634 ps.$$
$$\vdots$$

We can then look at the parity of the least significant digit and assign, for instance, a 0-bit to even parity and 1-bit to odd parity, thus generating a dataset ℓ of n bits. In Figure 5, we show the results of testing such a sequence using the Borel-normality bounds and Bayesian bounds, given by Equations (1) and (4), respectively, up to Borel level $i_{\text{max}} = 5$. The colored bins correspond to the deviations from the ideal value of relative frequencies for all the possible subsequences. These are ordered using its length $i = 1$ (purple bins), $i = 2$ (green bins), $i = 3$ (blue bins), $i = 4$ (orange bins), and $i_{\text{max}} = 5$ (yellow bins). The solid red line corresponds to the Borel bound, 8.6×10^{-5}. In the same graph, the green line is the Bayesian bound given the right-hand-side of Equation (4) that depends on i and therefore it is not a constant, as is the case for the Borel bound. Finally, the height of the various background colored boxes correspond to the values given by the left-hand-side expression of the Bayesian bound.

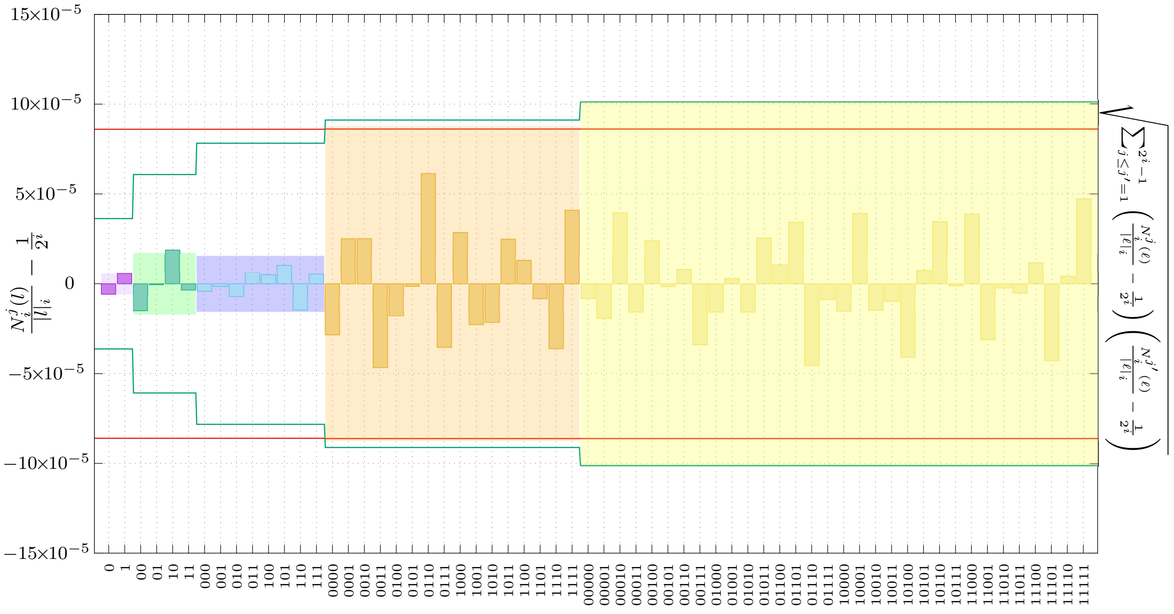

Figure 5. Results for generation using the least significant bits of time tags. In this case, the deviations are very small, so this generation scheme is excellent. The solid red lines represent the maximum deviations allowed by the Calude test, while the solid green lines correspond to the bounds given the Bayesian approach.

As we can see, this extremely simple QRNG passes the Borel-normality criterion up to $i = 5$, and nearly passes the Bayesian criterion (passes it for $i \leq 4$ and slightly exceeds the bound for $i = 5$). Notice that, while the previous experimental setup required an accurate balance between zeroes and ones, in the present case we already have very small deviations at Borel level $i = 1$, less than 10^{-5}, showing the convenience of this method. For $i = 2$ (green bins), the deviations are much larger, almost three times the value for $i = 1$, but nevertheless they pass the test again by a considerable margin. These results show the lack of correlations between consecutive events, which is the main drawback of the previous approach. It is important to note that, in the results at Borel level $i = 4$, there is a substantial increase in the deviations compared with $i = 1, 2, 3$. While this increase may indicate the presence of some as yet unidentified correlations, these are of an insufficient magnitude to reach the bounds. On the other hand, this experimental setup fails to pass some of the requirements of the Bayesian scheme. The deviations derived from the Bayesian criterion are shown in Table 2, and also in Figure 5. As we can see, all Borel levels pass the Bayesian test, except for the last one, albeit by a small margin.

Table 2. Comparison between the left-hand-side and the right-hand-side of the Borel-type bounds given by Equation (4), applied to the sequence of bits generated in our lab. As we can observe, the Bayesian bounds are satisfied for the first four Borel levels, but not the last one by a slight margin.

Borel Level	LHS of Equation (4)	RHS of Equation (4)	LHS < RHS
1	5.719×10^{-6}	3.62956×10^{-5}	Yes
2	1.7129×10^{-5}	6.08097×10^{-5}	Yes
3	1.54974×10^{-5}	7.82572×10^{-5}	Yes
4	8.74186×10^{-5}	9.11726×10^{-5}	Yes
5	1.01138×10^{-4}	1.01069×10^{-4}	No

We can also look at the value of the posterior distribution given by Equation (2) for the maximally random model. The value of the posterior for the four word lengths is reported in Table 3. For the first three Borel levels, the posterior distribution of the maximally random model α_{sym} is very close to one, indicating that, given the dataset, this is the most likely model to have generated such data. For Borel level $i = 4$, we are only able to analyse those models which are in the vicinity, in parameter space, to the maximally random model. These models correspond to partitioning $\Omega^{(5)}$ into two subsets, resulting in 32,767 models, giving a total of 32,768, including the maximally random one. In this case, it turns out that the most likely model is not the maximally random one. Actually, using the value of the posterior probability, this model is ranked in the position 9240 out of all the explored models, and therefore the sequence of bits fails to pass the Bayesian criterion already at Borel level $i = 4$. Note that $i = 5$ was not included in Table 3 because we lack the computational power to address this Borel level.

Table 3. Value of the posterior distribution $P\left(\mathcal{M}^{(i)}_{\text{sym}} \middle| \ell\right)$ given by Equation (2) for the maximally random model. Note that the prior distribution for each Borel model is a flat distribution along all the models tested. This means, for instance, that, at level $i = 3$, while we do not have any observational bias to choose among any particular model, that is $P(\mathcal{M}^{(i)}_{\alpha}) = \frac{1}{4140}$, after observing the data, the maximally random model is the most plausible.

Borel Level i	Number of Models Analysed	$P\left(\mathcal{M}^{(i)}_{\text{sym}} \middle| \ell\right)$
1	$B_{2^i} = 2$	0.999984
2	$B_{2^i} = 15$	0.999634
3	$B_{2^i} = 4140$	0.995476
4	32,768 models considered out of $B_{2^i} =$ 10,480,142,147	9.2179×10^{-42}

8. Conclusions

A vast amount of literature exists which claims that QRNGs are superior when compared to their classical counterparts, based on purely theoretical arguments. Indeed, randomness in Quantum Mechanics is usually justified by the unpredictability of individual measurement outcomes given some initial conditions. More concretely, quantum unpredictability is based on no-go theorems, such as Bell's theorem, that simply tells us that, given some initial conditions, it is impossible to predict the outcome of a single measurement. However, in the present review, we have shown that QRNGs actually perform rather poorly in tests of randomness as compared to classical PRNGs. The reason is fairly simple: unpredictability has nothing to do with bias, and while experimental devices based on Quantum Mechanics may produce a truly unpredictable random signal, they also tend, more often than not, to introduce correlations. In particular, for QNRGs based on optical devices, we have been able to account for two, perhaps amongst the many, effects that introduce bias in our data. While in our own experimental work involving a QNRG we have failed to obtain sequences which obey the Borel and Bayesian criteria, we were able to show that extracting sequences from the least significant digits of times of arrival represents a promising strategy.

Author Contributions: J.G.H. and A.B.U. conceived and designed the experiments; A.C.M. and A.S. performed the experiments; I.P.C. and R.D.H.R. developed the method based on Bayesian Inference; A.C.M, A.S. and R.D.H.R. analysed the data. All authors contributed to writing the paper.

Funding: Financial support of UNAM-DGAPA-PAPIIT IA103417 and IN109417 is acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Metropolis, N.; Ulam, S. The Monte Carlo method. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341. [CrossRef] [PubMed]
2. Von Neumann, J. Various techniques used in connection with random digits. *J. Res. Nat. Bur. Stand. Appl. Math. Ser.* **1951**, *12*, 36–38.
3. Isida, M.; Ikeda, H. Random number generator. *Ann. Inst. Stat. Math.* **1956**, *8*, 119–126. [CrossRef]
4. Stojanovski, T.; Kocarev, L. Chaos-based random number generators-part I: Analysis [cryptography]. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **2001**, *48*, 281–288. [CrossRef]
5. Einstein, A.; Podolsky, B.; Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **1935**, *47*, 777–780. [CrossRef]
6. Bell, J.S. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*; Cambridge University Press: Cambridge, UK, 2004.
7. Pironio, S.; Acín, A.; Massar, S.; de La Giroday, A.B.; Matsukevich, D.N.; Maunz, P.; Olmschenk, S.; Hayes, D.; Luo, L.; Manning, T.A.; et al. Random numbers certified by Bell's theorem. *Nature* **2010**, *464*, 1021–1024. [CrossRef] [PubMed]
8. Acín, A.; Masanes, L. Certified randomness in quantum physics. *Nature* **2016**, *540*, 213–219. [CrossRef] [PubMed]
9. Quantis Random Number Generator. Available online: https://www.idquantique.com/random-number-generation/products/quantis-random-number-generator (accessed on 9 October 2018).
10. Rukhin, A.; Soto, J.; Nechvatal, J.; Smid, M.; Barker, E. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*; Technical Report; Booz-Allen and Hamilton Inc.: Mclean, VA, USA, 2001.
11. Calude, C.S.; Dinneen, M.J.; Dumitrescu, M.; Svozil, K. Experimental evidence of quantum randomness incomputability. *Phys. Rev. A* **2010**, *82*, 022102. [CrossRef]
12. Pareschi, F.; Rovatti, R.; Setti, G. Second-level NIST randomness tests for improving test reliability. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2007), New Orleans, LA, USA, 27–30 May 2007; pp. 1437–1440.
13. Wasserstein, R.L.; Lazar, N.A. The ASA's statement on p-values: Context, process, and purpose. *Am. Stat.* **2016**, *70*, 129–133. [CrossRef]
14. Calude, C.S. Borel normality and algorithmic randomness. *Dev. Lang. Theory* **1993**, *355*, 113–129.
15. Calude, C.S. *Information and Randomness: An Algorithmic Perspective*, 2nd ed.; Springer Publishing Company: New York, NY, USA, 2010.
16. Díaz Hernández Rojas, R.; Solís, A.; Angulo Martínez, A.M.; U'ren, A.B.; Hirsch, J.G.; Marsili, M.; Pérez Castillo, I. Improving randomness characterization through Bayesian model selection. *Sci. Rep.* **2017**, *7*, 3096. [CrossRef] [PubMed]
17. Díaz Hernández Rojas, R. Mejora en la Caracterización de la Aleatoriedad Usando Selección Bayesiana de Modelos. Master's Thesis, Posgrado en Ciencias Físicas UNAM, Mexico, Mexico, 2017. (In Spanish)
18. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.

19. Pemmaraju, S.; Skiena, S.S. *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*; Cambridge University Press: Cambridge, UK, 2003.
20. Abbott, A.A.; Calude, C.S.; Dinneen, M.J.; Huang, N. Experimentally Probing the Incomputability of Quantum Randomness. *arXiv* **2018**, arXiv:1806.08762.
21. Martínez Becerril, A.C. Un Generador Cuántico de Números Al Azar: De un PBS a Etiquetas Temporales. Bachelor's Thesis, Facultad de Ciencias, UNAM, Mexico, Mexico, 2017. (In Spanish)
22. Horoshko, D.B.; Chizhevsky, V.N.; Kilin, S.Y. Afterpulsing model based on the quasi-continuous distribution of deep levels in single-photon avalanche diodes. *J. Mod. Opt.* **2017**, *64*, 191–195. [CrossRef]
23. Kang, Y.; Lu, H.X.; Lo, Y.H.; Bethune, D.S.; Risk, W.P. Dark count probability and quantum efficiency of avalanche photodiodes for single-photon detection. *Appl. Phys. Lett.* **2003**, *83*, 2955–2957. [CrossRef]