Improving randomness characterization through Bayesian model selection

Rafael Díaz Hernández Rojas¹, Aldo Solís², Alí M. Angulo Martínez², Alfred B. U’Ren², Jorge G. Hirsch², Matteo Marsili³ & Isaac Pérez Castillo¹,4,*

¹Instituto de Física, Universidad Nacional Autónoma de México. Apdo. Postal 20-364, Cd. Mx., Mexico, C.P. 04510
²Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Apdo. Postal 70-543, Cd. Mx., Mexico, C.P. 04510
³The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy
⁴London Mathematical Laboratory, 14 Buckingham Street, London WC2N 6DF, United Kingdom

Random number generation plays an essential role in technology with important applications in areas ranging from cryptography to Monte Carlo methods, and other probabilistic algorithms. All such applications require high-quality sources of random numbers, yet effective methods for assessing whether a source produce truly random sequences are still missing. Current methods either do not rely on a formal description of randomness (NIST test suite) on the one hand, or are inapplicable in principle (the characterization derived from the Algorithmic Theory of Information), on the other, for they require testing all the possible computer programs that could produce the sequence to be analysed. Here we present a rigorous method that overcomes these problems based on Bayesian model selection. We
derive analytic expressions for a model’s likelihood which is then used to compute its posterior distribution. Our method proves to be more rigorous than NIST’s suite and Borel-Normality criterion and its implementation is straightforward. We applied our method to an experimental device based on the process of spontaneous parametric downconversion to confirm it behaves as a genuine quantum random number generator. As our approach relies on Bayesian inference our scheme transcends individual sequence analysis, leading to a characterization of the source itself.

Random numbers have acquired an essential role in our daily lives because of our close relationship with communication devices and technology. There are also numerous scientific techniques and applications that rely fundamentally on our ability for generating such numbers and typically pseudo-random number generators (pRNGs) suffice for those purposes. A new alternative has been proposed by exploiting the inherently probabilistic nature of quantum mechanical systems. These Quantum Random Number Generators (QRNGs) are in principle superior to their classical counterparts and recent experiments have shown that they can reach the same quality as commercial pRNGs. However, the natural question of how to assess whether a sequence is truly random is not yet fully established. Pragmatically, the NIST test suite has become the standard method for analysing sequences coming from a RNG. The suite is based on testing certain features of random sequences that are hard to reproduce algorithmically, such as its power spectrum, longest string of consecutive 1’s, and so on. Even though it constitutes an easily applicable procedure, recent findings show that its reliance on $P$-values is a drawback, while its lack of formality is a major disadvantage. On the other hand, although no definition of randomness is deemed ab-
solute, a rigorous characterization is presented by the Algorithmic Theory of Information (ATI) but it is unfortunately inapplicable in real cases\(^2\). An alternative which overcomes both formal and applicability issues is the Borel-normality criterion\(^3\) (BN). Intuitively, this approach works by successively compressing a given dataset, e.g. \(\hat{s} = \{010101001010101010110101010101010 \cdots\}\) of \(M\) bits, by taking strings of \(\beta\) consecutive bits and computing the frequency of occurrences \(\gamma_i^{(\beta)}\) of each of those \(i = 0, 1, \ldots, 2^{\beta} - 1\) possible strings. For example, \(\beta = 1\) corresponds to looking for the frequencies of the strings \(\{0, 1\}\) in the dataset \(\hat{s}\), while \(\beta = 2\) corresponds to analysing the frequencies of the strings \(\{00, 01, 10, 11\}\), and so on. The whole sequence is said to be Borel-normal if the frequencies are bounded individually according to

\[
\left|\gamma_i^{(\beta)} - \frac{1}{2^\beta}\right| < \sqrt{\frac{\log_2 M}{M}},
\]

and with \(\beta\) an integer ranging from 1 to \(\beta_{\text{max}} = \log_2 \log_2 M\). It is important to mention that BN criterion is a (nearly) necessary condition for a sequence to be considered random\(^2\). Note that this test is restricted to a-single-sequence classification, so it cannot determine the random character of the generating source.

In the present work, we show that randomness characterization can also be addressed using a Bayesian inference approach for model selection\(^7\), borrowing the compression scheme of BN. For simplicity, for a fixed \(\beta\) we denote each string with its decimal base representation \(j \in \{0, 1, \ldots, 2^\beta - 1\} \equiv \Xi_\beta\). The first step consists in identifying the models which could have generated a compressed dataset \(\hat{s}\). For instance if \(\beta = 1\), we can describe it as \(M\) realizations of a Bernoulli process, leading to two possible models: with and without bias. Similarly, for \(\beta = 2\), a model represents a way of constructing \(\hat{s}\) with bias in some of the \(2^2\) possible strings. A simple
combinatorial counting reveals that all the possible bias assignations correspond to all partitions of the four strings of $\Xi_2$.

Thus, in general, given the set $\Xi_\beta$, let $\mathcal{P}_{\Xi_\beta}$ denote the family of its $B_2^\beta = \sum_{K=1}^{2^\beta} \{ \begin{array}{c} 2^\beta \\ K \end{array} \}$ possible partitions, with $B_2^\beta$ the Bell’s numbers and $\{ \begin{array}{c} 2^\beta \\ K \end{array} \}$ the Stirling numbers of the second kind, which counts the different ways of grouping $2^\beta$ elements into $K$ sets. Formally, $\alpha^{(K)}_{\ell} = \{\omega^{(1)}_{\ell}, \ldots, \omega^{(K)}_{\ell}\} \in \mathcal{P}_{\Xi_\beta}$ would refer to the $\ell$-th partition into $K$ subsets, but for notational simplicity we will omit henceforth the index $\ell$. To each partition $\alpha^{(K)}$ there corresponds a unique model $\mathcal{M}_{\alpha^{(K)}}$ which assigns a probability $p_j$ to string $j \in \Xi_\beta$ according to the following rule:

$$\mathcal{M}_{\alpha^{(K)}} = \left\{ p_j = \frac{\theta_r}{|\omega^{(r)}|}; \quad \forall r = 1, \ldots, K; \quad \forall j \in \omega^{(r)} \right\}. \quad (2)$$

This means that all strings contained in a given subset $\omega^{(r)}$ are deemed equiprobable within the specified model. Thus, keeping $\beta$ fixed, the likelihood of observing the given dataset $\hat{s}$ in a model $\mathcal{M}_{\alpha^{(K)}}$ is:

$$P\left(\hat{s}|\mathcal{M}_{\alpha^{(K)}}, \{\theta_r\}_{r=1}^K\right) = \prod_{r=1}^{K} \left(\frac{\theta_r}{|\omega^{(r)}|}\right)^{k_{\omega^{(r)}}}, \quad (3)$$

where $k_j^{(\beta)}$ is the frequency of string $j \in \Xi_\beta$ and we have defined $k_{\omega^{(r)}} = \sum_{j \in \omega^{(r)}} k_j^{(\beta)}$ as the aggregate frequencies of the strings in the subset $\omega^{(r)}$. (For further use, we also introduce the relative aggregate frequencies $\gamma_{\omega^{(r)}} = \frac{\beta}{\mathcal{M}} k_{\omega^{(r)}}$.) From this perspective, only the model that is symmetric under any reordering of the possible strings is identified with a complete random source, because any other model entails biases assignations according to the strings’ grouping represented by the corresponding partition. This symmetry only exists when the partition is the set $\Xi_\beta$ itself, hence we denote $\mathcal{M}_{\alpha^{(1)}} = \mathcal{M}_{\text{sym}}$. 

4
Consider now that when characterising randomness the only essential feature is whether bias for or against some strings is present, but the degree of bias is irrelevant. We can eliminate the dependence on the bias parameters by multiplying with a prior for \( \{\theta_r \}_{r=1}^K \) and derive the so called evidence for a given model\(^9\). Following\(^10\), we use the Jeffreys prior for it yields a model’s probability distribution invariant under reparametrization and provides a measure of a model’s complexity, thus giving a mathematical representation of Occam’s Razor principle\(^10–12\). After integrating in the parameter space, we arrive at (see Supplementary Information (SI), Sec. 2)

\[
P(\hat{s}|M_{\alpha(K)}) = \frac{\Gamma\left(\frac{K}{2}\right)}{\Gamma^K} \prod_{r=1}^K \left( \frac{1}{|\omega(r)|} \right) \frac{\prod_{r=1}^K \Gamma\left(\frac{1}{2} + \frac{M}{\beta} \gamma_{\omega(r)}\right)}{\Gamma\left(\frac{K}{2} + \frac{M}{\beta}\right)}.
\]

Eq. (4) is our main result, for it will let us perform the model selection straightforwardly. For \( M_{\text{sym}} \), its evidence is fairly intuitive:

\[
P(\hat{s}|M_{\text{sym}}) \equiv P(\hat{s}|M_{\alpha(1)}) = 2^{-M}.
\]

Finally, we want to infer the model that best describes our source, after a dataset \( \hat{s} \) is given. Using Bayes’ theorem the posterior distribution \( P(M_{\alpha(K)}|\hat{s}) \) reads:

\[
P(M_{\alpha(K)}|\hat{s}) = \frac{P(\hat{s}|M_{\alpha(K)})P_0(M_{\alpha(K)})}{\sum_{\gamma} P(\hat{s}|M_{\gamma})P_0(M_{\gamma})}.
\]

Henceforth we will consider a uniform prior over models (which is justified in SI), so the model’s posterior is simply proportional to its evidence.

Suppose now we want to assess whether a source can be considered truly random. This is performed in two steps. As the first step, we need a model ranking procedure based on the posterior distribution. The second step consists in quantifying the goodness of our choice of model.
As a decision rule for the ranking process we use the Bayes Factor perspective,

\[ BF_{\alpha,\alpha'} = \frac{P(M_\alpha | \hat{s})}{P(M_{\alpha'} | \hat{s})} = \frac{P(\hat{s} | M_\alpha)}{P(\hat{s} | M_{\alpha'})}. \] (7)

Thus, we will choose \( M_\alpha \) over \( M_{\alpha'} \) whenever \( BF_{\alpha,\alpha'} > 1 \). It has been shown that \( BF_{\alpha,\alpha'} \) provides a measure of goodness of fit and \( \lim_{M \to \infty} BF_{\alpha,\alpha'} = \infty \) if \( M_\alpha \) is the true model.\(^\text{14}\)

To implement the second step, which is nothing more than a hypothesis testing problem, we have two alternatives: either we check whether \( \log_{10} BF_{\alpha,\alpha'} \geq 2 \) which is considered decisive in favour of model \( M_\alpha \)\(^\text{13}\), or we compute the ratio between the posterior and the prior of a given model to assess how certain the posterior has become under the information provided by the dataset.

From a computational point of view notice that the evaluation of the posterior requires to being able to compute the normalization factor \( \sum_{\gamma} P(\hat{s} | M_\gamma) P_0(M_\gamma) \) that appears in (6). When the number of models is very large we can choose either to work with a subspace of models or use the logarithm of the Bayes Factor, as in this case the normalisation factor cancels out.

It is clear that a full test of randomness requires different values of \( \beta \) to be used for the same dataset, while the strings should be short enough so that the \( M \) bits allow for each of the possible models to be sampled at least once. Thus, heuristically, \( B_{2^{\beta_{\text{max}}}} \sim M \) whence we can reproduce the BN limit\(^\text{3}\), \( \beta_{\text{max}} \sim \log_2 \log_2(M) \), after using an asymptotic expansion for the Bell number.

Note that by fixing \( \beta \) we have the set of parameters \( \left\{ \gamma_j \right\}_{j=0}^{2^\beta-1}, M \), whose space can be divided into regions identifying the likeliest model according to Eq. (4). As illustrative cases, in Fig. 1 we show a phase-type diagram for \( \beta = 1 \) and \( \beta = 2 \) (upper and lower panel, respectively), where the orange-filled area delimits the parameters values that renders \( M_{\text{sym}} \) the likeliest model.
The top panel includes the bounds according to the BN criterion (green curves) given by Eq. (1), and shows that for any sequence length, $M$, our method allows for considerably smaller variations of $\gamma_0$. This is a significant improvement, since only necessary criteria exist for testing randomness. The lower panel depicts the analogous regions when $\beta = 2$, for which there are fifteen models (see a list in the SI) and we have fixed two frequencies: $\gamma_1 = 1/6$ and $\gamma_2 = 1/4$. The complete models distribution can be deduced from the structure of this graph, by distinguishing, a posteriori, the equiprobable strings for which the corresponding model is the likeliest. Thus more information than complete randomness classification can be readily obtained from our method.

Also in Fig. 1, the red curves of the $\beta = 1$ case are bounds obtained by comparing the likelihood of $M_{\text{sym}}$ with models involving partitions into $K = 2$ subsets. Agreement with the regions boundary is excellent. Our choice of $K = 2$ is justified as we would expect that models corresponding to partitions into two subsets to be the closest ones to the model $M_{\text{sym}}$. An explicit expression for these bounds is derived in SI, Sec. 3, and Extended Data Figures 2 and 3 depict that they also bound considerably well the region in which $M_{\text{sym}}$ is the likeliest for $\beta = 2$.

For further benchmarking, we have compared our method against the NIST test suite. The result is depicted in Fig. 2 as a function of the sequence length $M$ and bias $b$ employed to generate $a_0$. The upper panel on Fig. 2 shows the averaged number of tests passed when employing the NIST suite, while the lower one shows the frequency of $M_{\text{sym}}$ being the likeliest, for $\beta = 1, 2$ and 3. We believe that our technique can contribute to test the quality of RNG in a more stringent form, since by applying a single test thrice (once for each value of $\beta$), we determined more precisely the
Figure 1: **Phase diagram of Randomness Characterisation.** Division of the parameter space into regions according to the likeliest model. The top figure corresponds to $\beta = 1$ in terms of the frequency $\gamma_0$ of the string 0 and the sample size $M$. The green curves corresponds to Borel’s normality criterion, while the red curves are Borel-type bounds obtained by an approximation obtained from Eq. (4) (see Sec. 3 of SI). The bottom plot corresponds to $\beta = 2$ where each coloured area identifies the likeliest model in that region. Here we fixed the frequencies $\gamma_1 = 1/6$ and $\gamma_2 = 1/4$ and varied the frequency $\gamma_0$ of the string 00 and the sample size $M$. 
random character of the sample of sequences.

As an application, we have tested our method in a bit sequence obtained experimentally from the differences in time detection in the process of spontaneous parametric down conversion (SPDC). Sequences generated via a SPDC photon-pair source have been shown to fulfil with ease the BN criterion, and to pass comfortably the NIST’s suite. In the SPDC process a laser pump beam illuminates a crystal with a $\chi^{(2)}$ nonlinearity, leading to the annihilation of pump photons and the emission of photon pairs, typically referred to as signal and idler. Our experimental setup is shown in Extended Figure 1 and we explain how to construct a 0 or 1 symbol from the detection signals in Section 1 of SI. We generated a $4 \times 10^9$ bits sequence, so $\beta_{\text{max}} \sim 4$. When $1 \leq \beta \leq 3$, we used all the possible models in the comparison, while, for computational ease, when $\beta = 4$, we restricted the model space to the 32,768 models corresponding to $K = 1$ and $K = 2$ subsets (consider that $B_{24} = 10^{10}$). Our inference showed that $M_{\text{sym}}$ was the likeliest model for every value of $\beta$.

As explained above, to achieve a full characterization of our QRNG as a random source, we need to go further from the model ranking based on the Bayes Factor and measure our certainty that $M_{\text{sym}}$ is the true model governing the source. This (un)certainty quantification is the hallmark of Bayesian statistics, since $P(M_{\text{sym}}|\hat{s})$ represents the probability that modelling our QRNG as a random source is correct. Computing this posterior distribution directly from Bayes’ Theorem, Eq. 6, we arrive at the values shown in Table I for each $\beta$. The first three values are at least 0.95, but the corresponding to $\beta = 4$ is about 0.32, considerably smaller. However, this represents
Figure 2: **Comparison with NIST Suite test.** Comparison of the bias allowed on a given sequence for it to be considered random using the NIST suite (upper panel) and our Bayesian method for randomness characterisation (lower panel).
an improvement of order $10^4$ when compared with the initial value for the prior, $P_0(M_{sym}) = 1/32,768 \approx 3.1 \times 10^{-5}$. Alternatively, we computed $\log_{10} BF_{sym,\alpha'}$ for each value of $\beta$. The values reported in Table I correspond to the comparison of $M_{sym}$ and the second likeliest model, hence the inequality for $\beta > 2$. These two criteria combined lead us to conclude that there is decisive evidence for our hypothesis that $M_{sym}$ is the underlying model driving our source, thus verifying that the photonic RNG is strictly random in the sense described in the article.

Table 1: Posterior $P(M_{sym}|\hat{s})$ calculated for a dataset of $4 \times 10^9$ bits.

| $\beta$ | $P(M_{sym}|\hat{s})$ | $\log_{10} BF_{sym,\alpha'}$ |
|---------|---------------------|-----------------------------|
| 1       | 0.999965            | 4.45                        |
| 2       | 0.999562            | $\geq 3.72$                |
| 3       | 0.968353            | $\geq 2.01$                |
| 4       | 0.46718             | $\geq 3.46$                |

From a more general perspective, we propose that $P(M_{\alpha(K)}|\hat{s})$ quantifies our certainty on the hypothesis that a sequence $\hat{s}$ was generated using the biases on strings associated with $\alpha^{(K)}$. Because Bayesian methods entails a model’s generalizability\cite{9,10}, the likeliest model provides a characterization of the source of $\hat{s}$. All partitions can be identified with standard computational packages, although it can be computationally demanding for sequences of $\sim 10^{10}$ bits. In any case, once a partition is given, its model’s likelihood is easily found using Eq. (4). A simplified analysis can be performed with the BN-type bounds given in Section 3 of the SI, which also leads to more stringent criteria than other approaches.
1. Rukhin, A. et al. *Statistical test suite for random and pseudorandom number generators for cryptographic applications, nist special publication* (Citeseer, 2010).

2. Calude, C. S. *Information and Randomness: An Algorithmic Perspective* (Springer Publishing Company, Incorporated, 2010), 2nd edn.

3. Calude, C. Borel normality and algorithmic randomness. In *Developments in Language Theory*, vol. 355, 113–129 (Citeseer, 1993).

4. Solis, A. et al. How random are random numbers generated using photons? *Physica Scripta* **90**, 074034 (2015).

5. Pareschi, F., Rovatti, R. & Setti, G. Second-level NIST randomness tests for improving test reliability. In *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, 1437–1440 (IEEE, 2007).

6. Wasserstein, R. L. & Lazar, N. A. The ASA’s statement on p-values: context, process, and purpose. *The American Statistician* 129–133 (2016).

7. Haimovici, A. & Marsili, M. Criticality of mostly informative samples: a bayesian model selection approach. *Journal of Statistical Mechanics: Theory and Experiment* **2015**, P10013 (2015).

8. Pemmaraju, S. & Skiena, S. S. *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica®* (Cambridge university press, 2003).

9. MacKay, D. J. Bayesian interpolation. *Neural computation* **4**, 415–447 (1992).
10. Myung, I. J., Balasubramanian, V. & Pitt, M. A. Counting probability distributions: Differential geometry and model selection. Proceedings of the National Academy of Sciences 97, 11170–11175 (2000).

11. Balasubramanian, V. Statistical inference, occam’s razor, and statistical mechanics on the space of probability distributions. Neural computation 9, 349–368 (1997).

12. Balasubramanian, V. A geometric formulation of occam’s razor for inference of parametric distributions. arXiv preprint adap-org/9601001 (1996).

13. Robert, C. The Bayesian choice: from decision-theoretic foundations to computational implementation (Springer Science & Business Media, 2007).

14. Verdinelli, I., Wasserman, L. et al. Bayesian goodness-of-fit testing using infinite-dimensional exponential families. The Annals of Statistics 26, 1215–1241 (1998).

15. Burnham, D. C. & Weinberg, D. L. Observation of simultaneity in parametric production of optical photon pairs. Physical Review Letters 25, 84 (1970).

16. Vicent, L. E. et al. Design of bright, fiber-coupled and fully factorable photon pair sources. New Journal of Physics 12, 093027 (2010).

17. Mastromatteo, I. On the typical properties of inverse problems in statistical mechanics. PhD Thesis, Scuola Internazionale Superiore di Studi Avanzati (2013).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature
Acknowledgements IPC and RDHR thank hospitality to the Abdus Salam ICTP. RDHR also thanks Susanne Still, Valerio Volpati, and Aaron King for helpful discussions regarding the choice of models priors. This work has received partial economical support from Consejo Nacional de Ciencia y Tecnología (Conacyt): SEP-Conacyt and RedTC-Conacyt, Mexico, PAPIIT-UNAM project IN109417, and PAPIIT-UNAM project IA103417. We also want to thank Mark Buchanan for his helpful feedback for writing the manuscript.

Author Contributions I.P.C., R.D.H.R. and M.M. developed the Bayesian approach for the current application and derived the analytic expressions for the evidence of models. A.S., J.G.H., A.U., and A.M.A.M. furnished our work as a randomness characterization and provided the experimental datasets. The comparison with the NIST test suite and BN criterion was done by R.D.H.R. and A.S. All authors discussed the results and commented the manuscript.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Isaac Pérez Castillo. (email: isaacpc@fisica.unam.mx).
Supplementary Information

1 Experimental Setup and conversion to a sequence of random bits.

The quantum state of the emitted photon pairs can be written as $|\Psi\rangle = |\text{vac}\rangle + \eta |\Psi_2\rangle$ in terms of the vacuum $|\text{vac}\rangle$, the two-photon component $|\Psi_2\rangle$, and of a constant $\eta$ related to the conversion efficiency. Under the assumptions a continuous-wave, plane-wave pump $|\Psi_2\rangle$ may be expressed as

$$|\Psi_2\rangle = \int d\omega \int d\mathbf{k}^\perp F(\omega, \mathbf{k}^\perp)|\omega, \mathbf{k}^\perp\rangle_s|\omega_p - \omega, -\mathbf{k}^\perp\rangle_i,$$

written in terms of a joint amplitude function $F(\omega, \mathbf{k}^\perp)$, and where $|\omega, \mathbf{k}^\perp\rangle_\mu$ represents a single-photon Fock state with frequency $\omega$ and transverse wavevector $\mathbf{k}^\perp$ for mode $\mu$, with $\mu = s, i$ for the signal ($s$) and idler ($i$). In writing the two-photon state, we have assumed that the parametric down-conversion process is in the spontaneous regime, so that the appearance of multiple-pair events can be neglected. This assumption is valid if the parametric gain is sufficiently low; experimentally, we restrict the pump power so that the process remains spontaneous. In all likelihood, a similar experiment and analysis carried out in the high-gain, stimulated regime would yield different results from those presented on this paper.

The state in Eq. (8) is entangled since it cannot be factored into a direct product of separate states $|S\rangle$ (signal) and $|I\rangle$ (idler) as $|\Psi\rangle = |S\rangle|I\rangle$. While in many works based on SPDC photon pairs entanglement is the key resource, in our case we exploit instead the random times of emission (and detection) of signal and idler photons.
We have used a pump beam from a diode laser (DL407) centred at 407nm with \( \sim 60 \text{mW} \) power, and as nonlinear medium a \( \beta \) barium borate (BBO) crystal of 1mm length; see Extended Data Figure 1. The BBO crystal, which is negative uniaxial, was cut so that the angle subtended by the optic axis with respect pump beam axis is \( \theta_{\text{pm}} = 29.2^\circ \) which yields phase matching for the generation of frequency-degenerate, non-collinear photon pairs. Signal and idler photons are emitted on diametrically opposed portions of an emission cone centred on the pump beam axis, with a 3.6° half opening angle. Pump photons are suppressed by transmitting the signal and idler modes through a long-pass filter which transmits wavelengths \( \lambda > 488 \text{nm} \) (F1), followed by a bandpass filter centred at 800nm with a 40nm bandwidth (F2).

A halfwaveplate (HWP2) and a polarising beam splitter (PBS) are placed on the signal arm so that the signal photon is transmitted or reflected with 50/50 probability. Each of the idler, reflected signal and transmitted signal collection modes is defined by an \( f = 8 \text{mm} \) focal length aspheric lens (L1, L2 and L3) which focuses incoming light into the core of a multi-mode fibre with a 50\( \mu \text{m} \) diameter (MMF1, MMF2 and MMF3). The plane defined by the collection fibres is chosen for convenience to be parallel to the optical table. By monitoring coincidences between the reflected signal and idler modes, on the one hand, and between the transmitted signal and idler modes, on the other hand, we are able to probabilistically exclude double (and multiple) pair events.

Each of the three photon-collection fibres leads to a silicon-based avalanche photodiode (APD1, APD2 and APD3), which emits an electronic TTL pulse for each detection event. The
times of arrival of these pulses are monitored with a time to digital converter (TDC; id800 from IdQuantique), or time-tagger, with a resolution of 81 ps. The TDC produces three time series containing the time of arrival data for each of the idler ($i_n$), and transmitted ($s^t_n$) and reflected ($s^r_n$) signal channels. We generate by post-processing the two time series defined as $c^t_n = s^t_n \times i_n$, and $c^r_n = s^r_n \times i_n$, corresponding to those bins for which there are coincident detection events between the (reflected or transmitted) signal and idler channels. A sequence of bits is generated by comparing the differences in time detection with a fully regular time series with the same number of events per second. A value of 1 is assigned if the time of detection is smaller than the corresponding time in the regular time series, and a value of 0 otherwise.

We have checked on the efficiency of our QRNG in our experimental setup. According to our data, the efficiency based on the SPDC is 240 kilocounts per second in each channel. If only those events in which the signal and the idler photon are detected in coincidence are registered, the efficiency of random number generation is reduced to 27 kilocounts per second. Moreover, our experimental setup is such that we are able to discriminate four-photon versus two-photon events. This is achieved by noticing that, first of all, we have used a pump power such that the rate of four-photon generation is essentially negligible: less than 0.2% according to our data. Secondly, in one of the SPDC arms we have placed a beamsplitter so that by discarding those events in which both APD’s in that arm click, we can eliminate all the events in which events are detected in same time bin in the three detectors.
Extended Data Figure 1: **Experimental Setup.** A pump laser beam centred at 407nm (DL407) incides into nonlinear BBO crystal. The signal and idler generated photons are emitted at diametrically opposed portions of an emission cone which yields phase matching for frequency-degenerate non-collinear photon pairs. A polarising beam splitter (PBS) and a Half wavelength plate (HWP2) are placed at the signal portion of the cone so this photon can be transmitted or reflected with a 50/50 probability, the reflected and transmitted signal and idler photons are collected into multimode fibers that lead to avalanche photodiodes(APD1,2,3) which emit a TTL pulse for each detection event.
2 Derivation of Jeffreys Prior and Model’s evidence

The idea of the Jeffreys prior is to take into account model indistinguishability from a point of view of a statistical sample. Based on Sanov’s theorem\textsuperscript{17} we know that the volume of models which are indistinguishable is inversely proportional to the square root of the determinant of the Fisher information matrix. This idea of measuring relevant volumes across models, but using a graining approach has also been explored previously\textsuperscript{10,11} in a rigorous geometric treatment. Note that in this case, our parameters are the \( \theta \)’s of which only the (say) first \( K - 1 \) are independent due to the normalization requirement. Then, considering a model \( \mathcal{M}_\alpha(K) \) – also obviating the index \( \ell \) in the partition, as we did in the main text – we have the following minus log-likelihood for a string \( s \) assigned by \( \mathcal{M}_\alpha(K) \) to partition \( \omega^{(a)} \)

\[
- \log P(s|\mathcal{M}_\alpha(K), \{\theta_r\}) = - \log \left( \frac{\theta_a}{|\omega^{(a)}|} \right)
\]

From here we derive the Fisher information matrix \( J_{ab} \) for \( a, b = 1, \ldots, K \)

\[
J_{ab}(\theta) = -E \left[ \frac{\partial^2}{\partial \theta_a \partial \theta_b} \log P(s|\mathcal{M}_\alpha(K), \{\theta_r\}) \right] \propto \frac{1}{\theta_a} \delta_{a,b},
\]

where \( E[\cdots] \) denotes the expected value. Its determinant is simply \( \det[J_{ab}(\theta)] \propto \frac{1}{\prod_{r=1}^{K} \theta_r} \). The proportionality constants will cancel out, once we normalize our expression for \( P_{\text{Jeff}} \). From here we have the following expression for Jeffreys prior:

\[
P_{\text{Jeff}}(\theta) = \frac{\Gamma \left( \frac{K}{2} \right)}{\Gamma \left( \frac{K}{2} \right)} \prod_{r=1}^{K} \theta_r^{-1/2}, \tag{9}
\]

where the normalization factor comes from:

\[
\int \left[ \prod_{r=1}^{K} d\theta_r \right] \left[ \prod_{r=1}^{K} \theta_r^{-1/2} \right] \delta \left( \sum_{r=1}^{K} \theta_r - 1 \right) = \frac{\Gamma^K \left( \frac{1}{2} \right)}{\Gamma \left( \frac{K}{2} \right)}.
\]
Notice that in this case the Jeffreys prior always behaves as a proper one, that is, it is normalizable.

Finally, a similar integration shows that the model’s evidence is given by

\[
P(\hat{s}|\mathcal{M}_{\alpha^{(K)}}) = \int \prod_{r=1}^{K} d\theta_r P_{\text{det}}(\theta) P(\hat{s}|\mathcal{M}_{\alpha^{(K)}}, \{\theta_r\})
\]

\[
= \frac{\Gamma \left( \frac{K}{2} \right) \prod_{r=1}^{K} \left( \frac{1}{|\omega^{(r)}|} \right)^{k_{\omega^{(r)}}} \prod_{r=1}^{K} \Gamma \left( \frac{1}{2} + k_{\omega^{(r)}} \right)}{\Gamma \left( \frac{K}{2} + \frac{M}{2} \right)} .
\]

This allows us to identify the terms \( \left( \frac{1}{|\omega^{(r)}|} \right)^{k_{\omega^{(r)}}} \) as the maximum likelihood estimators, and the ones involving the gamma functions as a measure of the relevant volume occupied in the parameter space, related to the model’s complexity.

3 Borel-normality-type (BN-type) bounds

Suppose we are interested in discerning whether a given sequence is completely random or not. This means that we must look for the region in the parameter space \((\{\gamma_j\}_{j \in \Xi_\beta}, M)\) in which the evidence of the symmetric model –corresponding to the partition of \(\Xi_\beta\) into one subset– is bigger than the rest of the models. As the empirical frequencies \(\{\gamma_j\}_{j \in \Xi_\beta}\) are grouped into \(K\) subsets for a given partition \(\alpha^{(K)}\), then the corresponding model has in effect \(K - 1\) free parameters \(\{\gamma_{\omega^{(r)}}\}_{r=2}^{K}\). Recalling that we used the Bayes Factor as a decision rule in the main text, we can explore the conditions such that \(\mathcal{M}_{\text{sym}}\) is the likeliest by the behaviour of the log-likelihood ratio,

\[
\log \left( \frac{P(\hat{s}|\mathcal{M}_{\text{sym}})}{P(\hat{s}|\mathcal{M}_{\alpha^{(K)}})} \right).
\]

To obtain a BN-type bound, we do the following: i) look for the values \(\{\gamma_{\omega^{(r)}}^*\}_{r=2}^{K}\) which extremize the log-likelihood ratio; ii) do an expansion around those values up to second order. We
eventually obtain:

\[
\log \left( \prod_{r=1}^{K} \left( |\omega^{(r)}| \right)^{\frac{M}{2} \gamma^{*}_{\omega^{(r)}}} \Gamma^{K} \left( \frac{1}{2} \Gamma \left( \frac{K}{2} + \frac{M}{2} \gamma^{*}_{\omega^{(r)}} \right) \right) \right) \frac{1}{2} \left( \frac{M}{\beta} \right)^{2} \prod_{r=1}^{K} \Gamma \left( \frac{1}{2} + \frac{M}{\beta} \gamma^{*}_{\omega^{(r)}} \right) \]
\[
= \frac{1}{2} \left( \frac{M}{\beta} \right)^{2} \sum_{r,r'=2}^{K} \left( \gamma_{\omega^{(r)}} - \gamma^{*}_{\omega^{(r)}} \right) \left( \gamma_{\omega^{(r')}} - \gamma^{*}_{\omega^{(r')}} \right) \times \left[ \delta_{r,r'} \psi_{1} \left( \frac{1}{2} + \frac{M}{\beta} \gamma^{*}_{\omega^{(r)}} \right) + \psi_{1} \left( \frac{1}{2} + \frac{M}{\beta} \left( 1 - \sum_{r=2}^{K} \gamma^{*}_{\omega^{(r)}} \right) \right) \right] ,
\]

where the $\gamma^{*}$-unknowns obey the following set of equations

\[
\psi \left( \frac{1}{2} + \frac{M}{\beta} \gamma^{*}_{\omega^{(r)}} \right) - \psi \left( \frac{1}{2} + \frac{M}{\beta} \left( 1 - \sum_{r=2}^{K} \gamma^{*}_{\omega^{(r)}} \right) \right) = \log \left[ \frac{\omega^{(r)}}{\omega^{(1)}} \right] , \quad r = 2, \ldots, K .
\]

Here the function $\psi_{n}(x)$ is the polygamma function of order $n$, with $\psi(x) \equiv \psi_{0}(x)$. As the symmetric model is the one that corresponds to no-free parameters, one could reasonable assume that the models which are closer to $M_{\text{sym}}$ are those which correspond a single free parameter.

This, in turn, corresponds to subfamilies of partitions into two subsets of lengths $\{2^\beta - q, q\}$ for $q = 1, \ldots, 2^\beta / 2$, which will have aggregate frequencies $1 - \gamma_{|q|}$ and $\gamma_{|q|}$ respectively. This is also justified by the lower panel of Figure 1 in the main text, which shows that the transition from $K = 1$ to a bigger value should necessarily go through a region where a model with $K = 2$ is likelier than $M_{\text{sym}}$. Applying this to the set of Eqs. (11) and (12) we obtained that $\left| \gamma_{q} - \gamma_{|q|} \right| \leq \frac{\sqrt{\gamma_{\beta}}}{M} W(\gamma_{|q|}^{*})$ with the function $W(\gamma_{|q|}^{*})$ defined as

\[
W(\gamma_{|q|}^{*}) \equiv \left[ \log \left( \frac{\Gamma^{2}(1/2)\Gamma(1+M/\beta)(2^\beta - q) \frac{M}{2} (1-\gamma_{|q|}^{*}) \frac{M}{2} \gamma_{|q|}^{*}}{2^M \Gamma \left( \frac{1}{2} + \frac{M}{\beta} \gamma_{|q|}^{*} \right) \Gamma \left( \frac{1}{2} + \frac{M}{\beta} (1-\gamma_{|q|}^{*}) \right) \right) \right] \psi_{1} \left( \frac{1}{2} + \frac{M}{\beta} \gamma_{|q|}^{*} \right) + \psi_{1} \left( \frac{1}{2} + \frac{M}{\beta} \left( 1 - \gamma_{|q|}^{*} \right) \right) ,
\]

where $\gamma_{|q|}^{*}$ are the aggregated frequencies of a subset of size $q$ satisfying the extremisation condition

\[
\psi \left( \frac{1}{2} + \frac{M}{\beta} \gamma_{|q|}^{*} \right) - \psi \left( \frac{1}{2} + \frac{M}{\beta} \left( 1 - \gamma_{|q|}^{*} \right) \right) = \log \left( \frac{q}{2^\beta - q} \right) ,
\]

\[
\frac{21}{72}
\]
for \( q = 1, \ldots, 2^\beta / 2 \).

In particular, for \( \beta = 1 \), there is only one model to compare to \( M_{\text{sym}} \), which precisely corresponds to \( K = 2 \). Here, the solution of (14) is exactly \( \gamma^*_1 = 1/2 \), which provides the following bound:

\[
\left| \gamma_1 - \frac{1}{2} \right| \leq \frac{1}{M} \left\| \log \left( \frac{2^{-M} \Gamma(1+M)}{\Gamma^2 \left( \frac{1}{2} + \frac{M}{2} \right)} \right) \right\| \psi_1 \left( \frac{1}{2} + \frac{M}{2} \right),
\]

(15)

This is the formula we used to draw the red curves in the top panel of Figure 1 of the main text together with the exact diagram. Agreement for this simple bound is excellent compared to the exact formulas, and rather different as compared to the one of BN. For the case \( \beta = 2 \), one must solve the set of equations numerically to evaluate the bounds. They work reasonably well in the parameter space and much better than the BN bounds as shown in Extended Data Figure 2. Notice that in these figures we only depict two regions in the parameter space: the orange one corresponds to the region in which the symmetric model is likeliest, while the grey-filled area in which it is not.

These previous bounds have the disadvantage of needing to solve the system (14) numerically. However, looking at the set of Eqs. (12) we notice that there is a particular set of partitions for which its solution is particularly simple, namely when the system is solved using only equi-partitions, that is, partitions into subsets of the same size. With this restriction, it is possible to find simpler, less restrictive bounds, yet tighter than the ones derived from other methods. Suppose that we look at partitions into \( K \) subsets. Within this family (and of course for even \( K \)) we will have a subfamily of equi-partitions. For them we have that \( |\omega(r)| = |\omega(1)| = 2^\beta / K \) and therefore \( k_{\omega(r)} = M / (\beta K) \) and \( \gamma^*_r = 1/K \). In particular, for the model corresponding to a partition into
Extended Data Figure 2: **BN-type bounds.** Phase diagram of model selection for the 15 models for $\beta = 2$ and various fixed values of $\gamma_1$ and $\gamma_2$. Here the orange filled area represents the region in which model $M_{\text{sym}}$ is the likeliest, while the grey filled area represents the region in the parameter space in which any other model is the likeliest. Solid red lines represent the BN-type bounds. We also compare with the BN bounds (green filled region). Notice that for the second and the third case, the BN bounds also provides a bound for $M$ given by the solution of $|1/4 - 1/5| = \sqrt{\frac{\log_2(M)}{M}}$ and $|1/4 - 1/6| = \sqrt{\frac{\log_2(M)}{M}}$, respectively.
Extended Data Figure 3: **BN-type bounds.** Phase-type diagram for model selection for $\beta = 2$ and comparison between the bounds given by the simple formula (13) (solid red line) and the Borel-normality bounds (solid green line).

For $K = 2^\beta$ subsets, the formula (11) becomes:

$$
\sum_{i \leq j = 1}^{2^\beta - 1} \left( \gamma_i - \frac{1}{2^\beta} \right) \left( \gamma_j - \frac{1}{2^\beta} \right) = \frac{\beta^2 \log \left( \frac{2^{-M \Gamma^{2^\beta} \left( \frac{1}{2} \right) \Gamma \left( 2^{\beta - 1} + \frac{M}{2} \right)} \Gamma^{2^{\beta - 1}}} \Gamma \left( 2^{\beta - 1} \right)^{2^\beta} \left( \frac{1}{2} + \frac{M}{2^\beta} \right) } M \psi_1 \left( \frac{1}{2} + \frac{M}{2^\beta} \right) \right),
$$

a bound which, unlike the one of Borel-normality, couples all the empirical frequencies. Results of these broader bounds are plotted in Extended Data Figure 3.

### 4 Some examples for the evidence

In this section, we illustrate, with some specific examples, the formulae Eq. (4) for the particular case of $\beta = 2$. Because explicit reference to specific partitions is made, we will use the full notation $\alpha_{\ell}^{(K)}$, although there is no natural order to assign the index $\ell$. In this case we have the
following partitions of $\Xi_{d=2}$, corresponding to 15 models: a partition into $K = 1$ subset (symmetric model) which corresponds to $\alpha_1^{(1)} = \{\{0, 1, 2, 3\}\}$. There are $\binom{4}{2} = 7$ ($\binom{a}{b}$ denotes the Stirling number of second kind) partitions with $K = 2$ subsets, which are: $\alpha_1^{(2)} = \{\{0\}, \{1, 2, 3\}\}$, $\alpha_2^{(2)} = \{\{0, 1\}, \{2, 3\}\}$, $\alpha_3^{(2)} = \{\{0, 2, 3\}, \{1\}\}$, $\alpha_4^{(2)} = \{\{0, 1, 2\}, \{3\}\}$, $\alpha_5^{(2)} = \{\{0, 3\}, \{1, 2\}\}$, $\alpha_6^{(2)} = \{\{0, 1, 3\}, \{2\}\}$, $\alpha_7^{(2)} = \{\{0, 2\}, \{1, 3\}\}$. We have $\binom{4}{1} = 6$ partitions into $K = 3$ subsets: $\alpha_1^{(3)} = \{\{0\}, \{1\}, \{2, 3\}\}$, $\alpha_2^{(3)} = \{\{0\}, \{1, 2\}, \{3\}\}$, $\alpha_3^{(3)} = \{\{0\}, \{1, 3\}, \{2\}\}$, $\alpha_4^{(3)} = \{\{0, 1\}, \{2\}, \{3\}\}$, $\alpha_5^{(3)} = \{\{0, 2\}, \{1\}, \{3\}\}$, $\alpha_6^{(3)} = \{\{0, 3\}, \{1\}, \{2\}\}$. And, finally, one partition $\alpha_1^{(4)} = \{\{0\}, \{1\}, \{2\}, \{3\}\}$ into $K = 4$ subsets.

An example of the evidence, of the model associated to partition e.g. $\alpha_1^{(3)}$ is

$$P\left(\hat{s}|\mathcal{M}_{\alpha_1^{(3)}}\right) = \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma^3\left(\frac{1}{2}\right)} \left(\frac{1}{2}\right)^{k_{\omega(3)}} \frac{\Gamma\left(\frac{1}{2} + k_{\omega(1)}\right) \Gamma\left(\frac{1}{2} + k_{\omega(2)}\right) \Gamma\left(\frac{1}{2} + k_{\omega(3)}\right)}{\Gamma\left(\frac{3}{2} + \frac{M}{2}\right)}$$

$$= \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma^3\left(\frac{1}{2}\right)} \left(\frac{1}{2}\right)^{k_2 + k_3} \frac{\Gamma\left(\frac{1}{2} + k_0\right) \Gamma\left(\frac{1}{2} + k_1\right) \Gamma\left(\frac{1}{2} + k_2 + k_3\right)}{\Gamma\left(\frac{3}{2} + \frac{M}{2}\right)},$$

where $k_{\omega(1)}$ ($k_{\omega(2)}$) is the number of occurrences of string $\{0\} = \{00\}$ (resp. $\{1\} = \{01\}$), and $k_{\omega(3)}$ is the added number of occurrences of the strings $\{2\} = \{10\}$ and $\{3\} = \{11\}$ in the sequence of bits. An equivalent expression with the individual frequencies $k_j$ of the $j$-th string is also given for clarity.

5 On the choice for the Prior of models

Since in this work our particular goal is to assess the randomness of a given sequence with a general applicable method, it would be convenient to obtain a criterion as sharp as possible when
no previous knowledge of the source producing the data is given. Moreover, another desirable
property would be that no particular type of sequence is preferred over the rest, or in other words,
we would like to reproduce a distribution on datasets that resembles closely a uniform prior dis-
tribution over them. As we will justify here, those two features can be achieved by choosing a
uniform prior distribution on the models, that is, for a fixed $\beta$, $P_0(M_\alpha) = \frac{1}{B_{2^\beta}}$, with $B_n$ the $n$-th
Bell number. Indeed, this results in a distribution on sequences for which the unbiased ones are
the most unlikely.

Indeed, first of all, we need to relate the prior distribution on models $P_0(M_\alpha)$ with the
prior distribution on sequences $P_0(\hat{s})$. This can be done by computing the marginal of their joint
distribution, $P_0(\hat{s}) = \sum_\alpha P(\hat{s} | M_\alpha) P_0(M_\alpha)$. We want to show that a uniform prior on models
results into an expression of $P_0(\hat{s})$ that penalizes unbiased sequences. To be specific, let us analyse
the case of $\beta = 1$, for which there are only two possible models, and hence $P_0(M_\alpha) = \frac{1}{2}$. Using
Eqs. (4) and (5) from the main text to calculate the above marginal, we obtain the following
\begin{equation}
    P_0(\hat{s}) = \frac{1}{2} \left[ \frac{1}{2^M} + \frac{\Gamma(1/2) \Gamma(k_0 + 1/2) \Gamma(k_1 + 1/2)}{\Gamma(M + 1) \Gamma^2(1/2)} \right].
\end{equation}

From this expression, we can see that under the assumption of uniform prior distributions over
models, we obtain two terms for the prior distribution on datasets: the first one is independent
on the frequency of strings, while the second term adds a non-negative contribution that depends
explicitly on such frequencies. However, this second term is just the $B$ function, whose global
minimum is achieved when $k_0 = k_1 = M/2$. Thus unbiased sequences for which presumably
$k_0 \approx k_1$ are unfavored with this assumption.
An analogous argument follows straightforwardly for larger values of $\beta$. It is also worth mentioning that were we to assume directly that $P_0(\hat{s}) = \frac{1}{2\pi}$, the only compatible prior over models would be $P_0(\mathcal{M}_\alpha) = \delta_{\text{sym,}\alpha}$.