# HOW RANDOM IS RANDOM? A TOPOLOGICAL PERSPECTIVE[*]

ZOLTAN KOCSIS [†] AND MARWAN FAYED [‡]

**Abstract.** Non-randomness in seemingly random numbers can bias results in scientific simulation, gambling, and cryptography. Random number sequences produced by physical generators are expensive, irreproducible, and have a limited output rate.Computational pseudo-random sequences have the appearance of being random despite being determined by mathematical rules.The quality of individual sequences is assessed using statistical tests.However, there is a shortage of numerical measures for comparing pseudo-random sequences against each other.Here, we show that this gap can be filled by using topological data analysis to extract spatial structure from any number sequence. We calculate the Persistent Homologyof known random sequences and find a specific common spatial structure present on every scale. Crucially, when evaluated against this structure, all pseudo-random sequences will differ at some scale. We use the scale at which the first departure occurs as a numerical indicator of quality. Results reveal that the inversive congruential generators EICG1 and EICG7 have detectable regularities in low dimensions, demonstrating that lattice resultsare insufficient for ruling out other spatial structures. Our work establishes a link between topological data analysis and pseudo-random number generation, laying a foundation for future advancements in the reliability of random sampling and its applications.

**Key words.** example, LaTeX

**AMS subject classifications.** 68Q25, 68R10, 68U05

## 1. Introduction.

Random number generation has become/is a bedrock of a safe, secure, and resilient contemporary society. Non-randomness in seemingly random numbers can bias clinical trials, jury selection, and computer games or simulations. Non-random generation can also render modern operating and networked systems vulnerable to attack [2]. Confidence in the randomness of a sequence is a prerequisite for the confidence in the application of the sequence.

Across a similarly diverse spectrum of applications and environments, data sets and data sampling opportunities are increasing in size and number. Analysis tools are improving and emerging alongside, revealing previously unobservable patterns and structures/connections. In this context, scrutiny of random number generators (RNGs) must keep pace to ensure the high degrees of confidence that they currently provide to their applications.

Number sequences that appear to be random may be generated by either physical or algorithmic methods. Physical generators are expensive, and output irreproducible sequences with limited output rate [13]. Sequences produced by computational pseudo-random number generators (PRNGs) have the appearance of being random despite being deter- mined by mathematical rules [8]. Irrespective of the method of generation, the quality of the output sequence is assessed using statistical tests [1, 10, 14].

When used to evaluate number sequences, statistical tests come with inherent drawbacks. For example, test stringency is indicated by sampling sizes, typically large in number. In addition, rather than assessing ‘randomness,’ statistical tests are an indication of “good enough,” given some measure or usage. This pass-or-fail characteristic provides no reliable means to compare or rank among sequences, an

[*] Submitted to the editors DATE.

**Funding:** This work was funded by XYZ.

[†] (Corresponding Author) School of Mathematics, University of Manchester, UK (zoltan.kocsis@postgrad.manchester.ac.uk)

[‡] Department of Computing Science and Mathematics, University of Stirling, Stirling, UK (mmf@cs.stir.ac.uk).

otherwise highly desirable property for which there is a shortage of measures [5]. These observations are a reminder that all generators are flawed in some fashion, prompting the need for new empirical tests to ensure confidence [6].

In this paper, we explore empirical testing of PRNGs by using topological methods of analysis [4, 15]. To our knowledge this is the first work to establish a link between topological data analysis (TDA) and random number generator (RNG), more generally. Specifically, we make three main contributions. First, we show that persistent homology can be used to identify topological features of RNG sequences at multiple resolutions and dimensions. (to do: In support we construct a new proof of CSEH theorem that connects... U01, etc.) We calculate the persistent homology of *known* random sequences to find specific common spatial structures present on every scale. Crucially, when evaluated against this structure, the persistent homology of *all* non- or pseudo-random sequences must differ at some scale.

The topological separation of pseudo- from true-random motivates our second contribution: We design two topological RNG tests. The tests are designed with orthogonal goals in direct response to drawbacks in statistical testing, alone. Our Multi-scale Equidistribution Test is an entirely new class of test, with output that can be used to rank the performance of PRNGs. Orthogonally, the Matrix Rank Homology Test offers a reduction of the sampling complexity of its known counterpart. We also provide remarks and insights that suggest topological tests might be designed for any metric of interest.

Each test is then used to assess... (to do: 3rd contrib: ranking (obvious from above) and novel EICG result... We use the scale at which the first departure occurs as a numerical indicator of quality. Results reveal that the inversive congruential generators EICG1 and EICG7 have detectable regularities in low dimensions, demonstrat- ing that lattice results [12] are insufficient for ruling out other spatial structures.)

—-

Our work establishes a link between topological data analysis and pseudo-random number generation, laying a foundation for future advancements in the reliability of random sampling and its applications.

**2. Technical Background.** In advance of our test designs, it is instructive to give a brief overview of the technical foundations used in their construction, and review some work in each area.

**2.1. Two-level Testing Protocol.** Multiple levels of empirical RNG tests ensure the generality of test results over small and large sample sizes. In the context of RNGs, the null hypothesis is always that the generator produces a sequence of samples from independent identically distributed uniform random variables.

Normally, a hypothesis test fixes a significance level before performing the test, and computes a rejection region for the null hypothesis. However, establishing an a priori significance level for studying random number generators is inappropriate. This is because sample sizes obtained from an RNG are extremely large; for any sample size and corresponding significance level, the sample number can be increased at will.

As an alternative, L'Ecuyer [9] proposes a *two-level* protocol. The two-level protocol first dictates that any given test is replicated multiple times on disjoint subsequences, yielding a corresponding set of $p$-values. Under the null hypothesis, the distribution of $p$-values should be uniform. The second-level test compares the observed distribution to the expected distribution using any standard goodness-of-fit test; in our work we use Kolmogorov-Smirnov [**?**]. The null hypothesis, then, can be be accepted or rejected based on the single $p$-value returned by the goodness-of-fit test.

The two-level procedure reflects the local behaviour of generators better than their one-level counterparts, and prompted its adoption in modern RNG test suites such as `TestU01` [10]. (to do: Some statement about 2-level criticism?) We adopt the same strategy in our topological tests.

**2.2. $U(0,1)$ Implies Persistent Homology.** Unconstrained sample sizes from RNGs prompt the need for two-level testing. The large number of samples also suggests an application of algebraic topology, wherein samples could be projected onto multidimensional spaces to reveal structure and invariants.

Motivated by this observation, we prove that the persistent homology of a closed set in a metric space can be recovered from a sufficiently large uniform random point samples. The following definitions and theorems can be seen as a de-categorification of (to do: Incomplete statement – presumably adding citations for more details?)

Let $\mathfrak{M}$ be a closed set in some metric space taken from its original closure set$M$(to do: Is this an accurate stmt?). Given a point sample $\mathfrak{p}$, we can approximate the original set $M$ by drawing an open ball of radius $\varepsilon$ around each point of $\mathfrak{p}$. To recover the homology of $\mathfrak{M}$, we have to consider how the homology groups of the approximations *vary* with the parameter $\varepsilon$. We begin by first capturing the notion of a topological object varying with an arbitrary real parameter $\varepsilon$ in Definition 1.

Definition 1. *A diagram $F$ consists of:*

- *a family of topological spaces $F(x)$, one for each real number $x$,*
- *and continuous transition functions $F_a^b : F(a) \to F(b)$, one for each pair of real numbers $a \leq b$,*

*such that*

- *$F_a^a(x) = x$ for all $a, x \in \mathbb{R}$,*
- *and $F_b^c(F_a^b(x)) = F_a^c(x)$ for all $x \in \mathbb{R}$ and $a \leq b \leq c$.*

The persistent homology of the object is the homology that is preserved by the transition functions. Each transition gives rise to a persistent homology group, defined as follows.

Definition 2. *Given a diagram of topological spaces $F$, the $n$th persistent homology group between $a$ and $b$ is defined as $H_n(\mathrm{im} F_a^b)$; and when $n$ is unambigous, is denoted by $H_a^b F$. (to do: first use of 'im'?)*

(to do: Insight: what does it mean to stay close? Why does this matter, or why is this a requirement?) Two diagrams that "always stay close" are said to be *interleaved*. This requirement (connection?) is formalized by Definition 3.

Definition 3. *Two diagrams $F, G$ are $\varepsilon$-interleaved if $F(x) \subseteq G(x+\varepsilon)$, $G(x) \subseteq F(x+\varepsilon)$ and the transitions are compatible, i.e. $G_{x+\varepsilon}^y(F(x)) = F(y+\varepsilon)$ and $F_{x+\varepsilon}^y(G(x)) = G(y+\varepsilon)$ for all $x, y \in \mathbb{R}$.*

We can interpret subsets of a metric space (with open balls drawn around each point), as diagrams. (to do: Insight to connect prev def'n with next one.)

Definition 4. *Let $M$ be a metric space. The sub-level diagram $f_\downarrow$ of a continuous function $f : M \to \mathbb{R}$ is defined as follows:*

- *for each $\varepsilon \in \mathbb{R}$, $f_\downarrow(\varepsilon) = \{x \in M \mid f(x) < \varepsilon\}$*
- *the transition functions ${f_\downarrow}_a^b$ are inclusion maps.*

*The sub-level diagram $S_\downarrow$ of the set $\mathfrak{M} \subseteq M$ is the sub-level diagram $f_\downarrow$ of the* *corresponding distance function*

$$f(x) = \inf_{m \in \mathfrak{M}} d(x, m)$$

.

We can now freely identify subsets of $M$ with their distance functions and their sub-level diagrams. (to do: Statement to connect sentence with Lemma 5.)

LEMMA 5. *If two functions $f$ and $g$ satisfy $|f(x) - g(x)| < \varepsilon$ for all $x$, then the sub-level diagrams $f_\downarrow$ and $g_\downarrow$ are $\varepsilon$-interleaved, and the persistent homology groups satisfy $H_{a-\varepsilon}^{b+\varepsilon} f_\downarrow \subseteq H_a^b g_\downarrow$.*

*Proof.* The transition functions are inclusion maps, so it suffices to prove that $f_\downarrow(k) \subseteq g_\downarrow(k+\varepsilon)$ and vice versa. For any $x \in f_\downarrow(k)$, we have that $f(x) < k$. Therefore,

$$g(x) < f(x) + \varepsilon < k + \varepsilon,$$

so $x \in g_\downarrow(k+\varepsilon)$. The other direction works identically.

Functoriality of homology yields $H_{a-\varepsilon}^{b+\varepsilon} f_\downarrow \subseteq H_a^b g_\downarrow$. □

(to do: Connect lemma with theorem.)

THEOREM 6. (COHEN-STEINER, EDELSBRUNNER, HARER) *Let $\mathfrak{M}$ be a closed subset of $\mathbb{R}^n$. Let $\mathfrak{p} \subseteq \mathfrak{M}$ be a finite set of points. If there exists $\ell < u \in \mathbb{R}$ such that*

1. *every point of $\mathfrak{M}$ is $\ell$-close to some point in $\mathfrak{p}$, and*
2. *the diagram $\mathfrak{M}_\downarrow$ is constant on the interval $(0, 4u)$*

*then the (singular) homology of $H\mathfrak{M}$ is the persistent homology of $H_\varepsilon^{3\varepsilon}\mathfrak{p}$ for any $\varepsilon \in (\ell, u)$.*

*Proof.* Choose any $\varepsilon$ such that $\ell < \varepsilon < u$. By the first condition and and Proposition 5, $\mathfrak{M}$ and $\mathfrak{p}$ are $\varepsilon$-interleaved. Since the diagram $\mathfrak{M}_\downarrow$ is constant on $(0, 4\epsilon) subseteq (0, 4u)$,

$$H\mathfrak{M} = H_0^{4\varepsilon}\mathfrak{M} \subseteq H_\varepsilon^{3\varepsilon}\mathfrak{p} \subseteq H_{2\varepsilon}^{2\varepsilon}\mathfrak{M} = H\mathfrak{M}$$

proving that $H\mathfrak{M} = H_\varepsilon^{3\varepsilon}\mathfrak{p}$. □

This connection between uniformly distributed sample points and their persistent homology provides the foundation for our topological tests. The foundation for the evaluation and interpretation of test results is provided by use of barcodes, described next.

**2.3. Barcodes.** Barcodes describe features that emerge when a homology is taken over a field $\mathbb{F}$, and when $\mathbb{F}$'s sub-level diagram changes at only finitely many points. The changes mark homological features that can be assigned a unique birth time, that is the value of the parameter $\varepsilon$ where the feature first appears. Correspondingly, features can be assigned a death time, i.e. the value of $\varepsilon$ above which the feature disappears; the death time may be infinite. The proof involves reinterpreting the time-parametrized diagram as a single graded module over $\mathbb{F}[x]$, where the scalar product with $x$ acts by "advancing time", and decomposes the module using the structure theory of principal ideal domains. A full development of the argument may be found in Zomorodian and Carlsson [15]

Returning to barcodes, the birth and death times of the homological features can be plotted on a *persistence diagram*. The persistence diagram is said to be the *barcode* that summarizes the homological features of the point set across all scales.

Barcodes fall outside of the standard families of statistical data types. As a consequence, direct statistical hypothesis testing on barcode distributions is inappropriate. However, statistical tools are appropriate on auxiliary distributions over the real numbers, that can be induced by passing the *distances* between barcodes, as follows:

Definition 7. *Let $A, B$ be barcodes, and consider two points $(a_1, a_2) \in A$ and $(b_1, b_2) \in B$. Let the distance between two such points be given by*

$$d_\infty (a, b) = \max \{ |a_1 - b_1| , |a_2 - b_2| \}$$

*. The bottleneck distance of the barcodes $A, B$ is defined as the quantity*

$$d_B (A, B) = \inf_\gamma \sup \{ d_\infty(a, \gamma(a)) \mid a \in A \}$$

*where $\gamma$ ranges over all bijections between $A, B$ considered as multisets.*

Bottleneck distance is a standard method for turning the set of barcodes into a metric space. Moreover, the bottleneck distance is equivalent to the interleaving distance used above, as stated by the isometry theorem of Lesnick [11]. (to do: Connect this statement to later use of Bette numbers.)

## 3. Multi-scale Equidistribution.

The goal of the Hypercube Homology Test is to verify that the observed distribution of barcodes obtained by taking $k$ samples of $t$ consecutive outputs of the PRNG approximates the expected distribution of barcodes for $k$-point uniform random samples from a $t$-dimensional hypercube.

Recall the null hypothesis: the generator produces a sequence of samples from independent identically distributed uniform random variables. It follows that $t$-tuples formed from consecutive output values must have uniform distribution in the unit hypercube $[0, 1]^k$.

Many interesting homological features tend to show up on very small scales. For example, the image of a linear congruential generator always consists of many connected components (the infamous parallel planes), but the separation between these components can be extremely small in good generators (e.g $10^{-5}$ for three dimensions in MINSTD).

Calculating the persistent homology of the unit hypercube is computationally intractable for the large sample sizes required to detect such small anomalies.

Fortunately, there is a trick to resolve this issue: one takes a large sample of the unit cube, and calculates the homology only for a small subcube of side length $\sigma$. At this point, generating enough samples becomes the new performance bottleneck. In particular, one might worry that the pool of true random numbers would be insufficient to empirically determine the expected distribution of barcodes. However, the homological features of the unit cube are identical to the homological properties of any of its subcubes, so the expected distribution derived for the unit cube remains valid for the subcubes as well.

The largest value $\sigma$ for which a generator systematically fails the Hypercube Homology Test can be used as an empirical "figure of merit" for comparing any two generators.

### 3.1. Description.

*Parameters:.* The dimension of the hypercube $d$, the size of the point samples $n$, the number of point samples $I$, the side length of the subcube $\sigma$.

*Algorithm:.*

1. Use the output of the generator to obtain $I$ sets of point samples (denoted $\mathfrak{p}_i$ for $i \in I$) from a cube of side length $\sigma$. Throw away any points that fall outside the subcube.
2. Calculate pairwise distances between the barcodes of $\mathfrak{p}_i$, yielding a distribution $D$ of distances over $\mathbb{R}_{>0}$.
3. Compare the distribution $D$ to the expected distribution of barcodes for uniform random samples, using the Kolmogorov-Smirnov goodness-of-fit test. Reject the null hypothesis if the $p$-value of the goodness-of-fit is sufficiently extreme ($< 0.001$).

**3.2. Remarks.** As in most empirical tests of random number generators, the distribution of the quantity of interest is approximated empirically. However, the asymptotic theory of random geometric complexes is already well-understood, so there is good reason to believe that explicit properties of these distributions will be determined in the future [7].

Hypercubes vacuously satisfy the second condition of Theorem 6, since the persistent homology diagrams of a cube are constant in any interval. For sufficiently large samples $\mathfrak{p}$ from the uniform distribution on the hypercube $\mathfrak{M}$, the expected largest distance between an arbitrary point on the hypercube and the nearest point in the sample goes to zero. Hence, the second condition of Theorem 6 is also satisfied with high probability. As such, Theorem 6 yields an exact interval of interest, as soon as one provides quantitative bounds for the first condition.

Unfortunately, the explicit intervals obtained from naive bounds are so big that they are unable to fail even the most ill-behaved generators. Obtaining better bounds is a non-trivial problem. Paradoxically, it may be easier to start with larger dimensions: the curse of dimensionality becomes a temporary blessing, as the distances start to approximate a normal distribution [3]. Needless to say, calculating the barcodes themselves is prohibitive in such high dimensions.

On the performance front, a few optimizations are used to make large sample tests computationally viable. For example, it is worthwhile to replace the bottleneck distance with the $L^2$ norm on the Betti sequences. The latter is easier to calculate than, but bounded above by a function of, the bottleneck distance and the number of barcodes, providing a tradeoff between statistical power and computational speed.

**3.3. Evaluation.** ANALYSIS OF TEST RESULTS HERE

TODO: Compare effect with discrepancy and explain how it is more general than Marsaglia's spectral test [**?**], since the tested generators need not belong to the Linear Congruential family.

**4. Linear Independence.** In this section, the techniques of persistent homology are used to create a powerful homological variant of Marsaglia's well-known Binary Matrix Rank Test. The resulting Matrix Rank Homology Test can be used to verify that there are no unexpected linear dependences between consecutive outputs of the sequence.

The main insight is that the norm $|A| = \operatorname{rank} A$ makes the set of $d \times d$ matrices into a normed vector space. The only non-trivial part is the triangle inequality: $\operatorname{rank}(A+B) \leq \operatorname{rank} A + \operatorname{rank} B$. If a vector belongs to $\operatorname{im}(A+B)$, then it has the form $(A+B)\,v = Av + Bv$ for some $v \in \mathbb{F}_2^d$. Thus, it can be written as the linear combination of a vector in $\operatorname{im} A$ and a vector in $\operatorname{im} B$, proving that

$$\operatorname{im}(A+B) \subseteq \operatorname{im} A + \operatorname{im} B$$

The rank is just the dimension of the image, so the triangle inequality holds.

The function $d\left(x,y\right) = \operatorname{rank}\left(x-y\right)$ is therefore a metric over the set of $d \times d$ matrices. Whenever there is a metric space, persistent homology can be computed. Proposition 8 counts the $n \times k$ binary matrices of a given rank $r$. Knowing that allows us to derive the exact distribution of the ranks of differences of random matrices and to make observations about the intervals of interest for the test.

### 4.1. Description.

*Parameters:.* The number of square matrices to generate $n$ per set, the number of sets to generate $i$, the dimension of each matrix $d$.

*Algorithm:.*

1. Use the output of the generator (considered not as a floating-point value but as a 64-bit integer) to obtain $i$ sets of $n$ matrices, each of dimension $d \times d$.
2. Calculate rank differences between the barcodes of the generated sets, yielding a distribution of distances over $\mathbb{R}_{>0}$.
3. Compare the distribution $D$ to the expected distribution of barcodes for uniform random samples, using the Kolmogorov-Smirnov goodness-of-fit test. Reject the null hypothesis if the $p$-value of the goodness-of-fit is sufficiently extreme.

**4.2. Remarks.** As pointed out above, the distribution of rank differences between matrices over the finite field $\mathbb{F}_2$ can be calculated explicitly.

THEOREM 8. *Let $N(k,r)$ denote the number of $d \times k$ matrices with rank $r$ for some $d \in \mathbb{N}$. Then the following recurrence relation holds:*

$$N(k+1, r+1) = (2^d - 2^r)N(k,r) + 2^{r+1}N(k, r+1)$$

*Proof.* There are two ways to construct a $d$-by-$(k+1)$ matrix of rank $r+1$:

1. Start from a $d$-by-$k$ matrix $M$ of rank $r$. Augment $M$ with a row $v \notin \operatorname{span} M$.
2. Start from a $d$-by-$k$ matrix $M$ of rank $r+1$. Augment $M$ with a row $v \in \operatorname{span} M$.

In a vector space over $\mathbb{F}_2$, we can form $2^n$ different linear combinations from $n$ lineraly independent vectors, so in the first case we can choose from $2^d - 2^r$ different vectors $v$, and in the second case we can choose from $2^{r+1}$ different vectors $v$. ∎

The probability that a uniformly random $d \times d$ matrix has rank $r$ is simply $\frac{N(d,r)}{2^{d \times d}}$. However, our main interest in the homological case is not the distribution of ranks, but the distribution of differences of ranks. Fortunately, we can use Theorem 8 as a stepping stone for that calculation.

THEOREM 9. *The probability that* $\operatorname{rank}(A-B) = r$*, where $A, B$ are $d$-by-$d$ matrices over $\mathbb{F}_2$, is also $\frac{N(d,r)}{2^{d \times d}}$.*

*Proof.* The space of matrices $\mathbb{F}_2^{d \times d}$ with matrix addition is a finite vector space, and so, a fortiori, a finite group. Consequently, every element occurs the same number of times in the addition table of $\mathbb{F}_2^{d \times d}$, and so the probability that $\operatorname{rank}(A+B) = r$ is just $\frac{N(d,r)}{2^{d \times d}}$. Finally, we observe that $A + B = A - B$ in $\mathbb{F}_2^{d \times d}$. ∎

The formula of Proposition 8 can be used to calculate $N(k,r)$ using e.g. dynamic programming. We get the following distribution of rank differences for $d = 64$.

Table 1
*Distribution of rank differences for all $64 \times 64$ matrices.*

| $r$ | 64 | 63 | 62 | $< 62$ |
|---|---|---|---|---|
| $N(k,r)/2^{r \times r}$ | 0.29 | 0.58 | 0.13 | $< 0.01$ |

Based on this distribution, one should expect the space to be fully disconnected for ranks below 62, and fully connected for rank 64, making $\varepsilon \in [61, 63]$ the only "interesting" parameter choices.

### 4.3. Evaluation.

TEST RESULTS HERE

## 5. Open and Future Directions.

The results of Sections 3 and 4 suggest that applying Topological Data Analysis can be used to create a wide variety of different tests for **RNG!**s (**RNG!**s).

In this section, we suggest a possible general approach for applying persistent homology in randomness testing.

## 6. Conclusion.

...

## Appendix A. An example appendix.

**Acknowledgments.** We would like to acknowledge the assistance of volunteers in putting together this example manuscript and supplement.

## REFERENCES

[1] L. E. Bassham, III, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks, N. A. Heckert, J. F. Dray, and S. Vo, *Sp 800-22 rev. 1a. a statistical test suite for random and pseudorandom number generators for cryptographic applications*, tech. report, Gaithersburg, MD, United States, 2010.

[2] H. Corrigan-Gibbs and S. Jana, *Recommendations for randomness in the operating system, or how to keep evil children out of your pool and other random facts*, in Workshop on Hot Topics in Operating Systems (HotOS), Kartause Ittingen, Switzerland, 2015, USENIX Association, https://www.usenix.org/conference/hotos15/workshop-program/presentation/corrigan-gibbs.

[3] D. Francois, V. Wertz, and M. Verleysen, *The concentration of fractional distances*, IEEE Transactions on Knowledge and Data Engineering, 19 (2007), pp. 873–886.

[4] R. Ghrist, *Barcodes: The persistent topology of data*, Bulletin of the American Mathematical Society, 45 (2008), pp. 61–75.

[5] P. Hellekalek, *Inversive pseudorandom number generators: Concepts, results and links*, in Proceedings of the 1995 Winter Simulation Conference, IEEE Press, 1995, pp. 255–262.

[6] P. Hellekalek, *Good random number generators are (not so) easy to find*, Math. Comput. Simul., 46 (1998), pp. 485–505, doi:10.1016/S0378-4754(98)00078-0, http://dx.doi.org/10.1016/S0378-4754(98)00078-0.

[7] M. Kahle, *Random geometric complexes*, Discrete Computational Geometry, 45 (2011), pp. 553–573.

[8] D. E. Knuth, *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1997.

[9] P. L'Ecuyer, *Testing random number generators*, in Proceedings of the 24th Conference on Winter Simulation, New York, NY, USA, 1992, ACM, pp. 305–313, doi:10.1145/167293.167354, http://doi.acm.org/10.1145/167293.167354.

[10] P. L'Ecuyer and R. Simard, *Testu01: A c library for empirical testing of random number generators*, ACM Transactions on Mathematical Software, 33 (2007), pp. 22:1–22:40.

[11] M. P. Lesnick, *Multidimensional interleavings and applications to topological inference*, PhD thesis, Stanford University, 2012.

[12] H. Niederreiter, *On a new class of pseudorandom numbers for simulation methods*, Journal of Computational and Applied Mathematics, 56 (1994), pp. 159 –

167, doi:http://dx.doi.org/10.1016/0377-0427(94)90385-9, http://www.sciencedirect.com/science/article/pii/0377042794903859.

[13] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, *Random numbers certified by bell's theorem*, Nature, 464 (2010), pp. 1021–1024, doi:10.1038/nature09008, http://dx.doi.org/10.1038/nature09008.

[14] J. Soto, *Statistical testing of random number generators*, in In: Proceedings of the 22nd National Information Systems Security Conference, 1999.

[15] A. Zomorodian and G. Carlsson, *Computing persistent homology*, Discrete Computational Geometry, 33 (2005), pp. 249–274.