Using Information Entropy to Measure Stationary Randomness of Quantum Random Sequences

Weizhong Yang, Yamin Luo, Zhefei Li and Jeffrey Zheng

Abstract Different statistical measurements can be used to determine stationary randomness for random sequences. This chapter proposes a testing scheme for random sequences using information entropy as measurements. Datasets are collected from University of Science & Technology of China (USTC), three quantum random sequences are selected for testing. Multiple results are created on three maps, entropy curves, and quantitative measurements of stationary randomness are compared. Three differences of Max-Min entropy variation ratios are bounded in $[0.08, 0.09]\%$ region. The whole structure has measurable stationary properties.

Keywords Variant maps · Quantum random sequence · Ordered measures Entropy · Stationary randomness

This work was supported by the Key Project on Electric Information and Next Generation IT Technology of Yunnan (2018ZI002), NSF of China (61362014), Yunnan Advanced Overseas Scholar Project.

W. Yang
Shanghai Key Laboratory of Intelligent Information Processing,
School of Computer Science, Fudan University, Shanghai, China

W. Yang
Key Laboratory of Quantum Information of Yunnan, School of Software,
Yunnan University, Kunming, China
e-mail: yangweizhong@126.com

Y. Luo · Z. Li · J. Zheng (✉)
Yunnan University, Kunming, China
e-mail: conjugatelogic@yahoo.com

Y. Luo
e-mail: 1047668416@qq.com

Z. Li
e-mail: 576167164@qq.com

1 Introduction

From a statistical viewpoint, various parameters of statistical process [2–4, 7] could be stationary invariant [6] under shift operations on random sequences. Using variant maps [8], it is a normal approach to transfer a long random sequence into 1D and 2D statistical distributions as three maps: 1DP, 1DQ, and 2DPQ [9]. For each map, it is easy to divide each number by the total number to transfer a counting number into a probability measure. By this way, three sets of probability measures can be generated. Applying information entropy function to summarize all pairs of probability parameters, one map corresponds an information entropy measurement determined by the distribution for stationary randomness.

2 Test Methodology

The test for a stationary randomness requires a sequence with length N . For the given input sequence, multiple segments M are divided from the sequence by a given length m , a 2-tuple pair of measures can be extracted from a 0-1 segment that are the number of 1 element and the number of 1 pattern in the segment. All paired measures are composed of a sequence of M pairs of measures as an ordered measuring set with M elements.

The pairs of the measuring sequence are directly separated as two independent measuring sequences to keep each parameter in the same order. A total of three sequences of distinct measures are constructed including two sequences on single measures and one sequence on 2-tuple measures.

Following this approach, two sets of single measuring sequences are sorted as two 1D numeric arrays as statistical histograms corresponding to 1D maps and the 2-tuple measuring sequence is sorted as a 2D integer array as statistic histograms being a 2D map. Under the controlling operations on the changes of shift displacement, multiple results of the three measuring sequences are transformed into 1D statistic histograms and 2D pseudo-color maps to show effective patterns from the generated sequence under various positions and conditions on a list of shift operations.

2.1 Dataset

2.1.1 USTC Resource

In the Key Laboratory of Quantum Information, USTC, CAS, and quantum random number sequences are generated [5]. This type of true random sequences supports advanced quantum communication devices of QKD systems [1].

Fig. 1 Methodology for information entropy testing stationary random sequences

More than 20 GB of quantum random number sequences are provided by USTC for random streams testing. Three sequences from eight sequences are selected from three stages (1 Initial, 2 Secure, and 4 Filtered). Each random sequence has a length of about 8 MB.

3 Method

3.1 Methodology

This method consists of five steps (Fig. 1): Input, Shifted Transformation (ST), Segment Measurement (SM), Combinatorial Projection (CP), and Output.

The input of the testing system is a selected 0-1 sequence and its output is composed of three maps, two in 1D and one in 2D for visual distributions, and three maximals to be processed by ST, SM, and CP.

3.2 Description of Steps

The testing system consists of three steps: {ST, SM CP}.

Input: X $N = m * M$ bit sequence; m segment length; M total segments; r shift length;

Output: Three maps {1DP, 1DQ, 2DPQ}; Three Maximals {1DP_x, 1DQ_x, 2DPQ_x}

Process: Shifting r position from X to be $Y = X(r)$ in ST. Making segment measuring sequences in SM and then projecting three measuring sequences as three maps and extracting three maximals in CP.

Let X, Y be 0-1 sequences with N elements, ST takes the sequence X as input, then shift r position on the whole sequence to be the shifted sequence $Y = X(r)$ (i.e., a cyclic shift right + or shift left -).

$$Y = X(r), Y[I] = X[I \pm r], I \pm r(\bmod N), \quad (1)$$

$$0 \leq I < N; X[I], Y[I] \in \{0, 1\}$$

SM takes the shifted vector as input and divides the vector into M segments. For the i th sub-vector $0 \leq i < M$ on the j th position $0 \leq j < m$, denoted as $Y_{i,j}$.

This sequence at the end of sub-vectors after the segmenting operation forms an $m * M$ matrix, m positions for the i th complete row vector in the sequence correspond to a pair of 2-tuple measures: (p_i, q_i) .

$$Y = \{Y_i\}_{i=0}^{M-1} \quad (2)$$

$$Y_i = \{Y_{i,0}, Y_{i,1}, \dots, Y_{i,j}, \dots, Y_{i,m-1}\} \quad (3)$$

$$0 \leq i < M, 0 \leq j < m$$

$$Y_i \rightarrow (p_i, q_i), 0 \leq i < M \quad (4)$$

$$\{Y_i\}_{i=0}^{M-1} \rightarrow \{(p_i, q_i)\}_{i=0}^{M-1} \quad (5)$$

The pair of 2-tuple measures (p_i, q_i) is determined by the following formula:

$$Y_{i,j} = Y[J] \in \{0, 1\}; J = i \times m + j, \quad (6)$$

$$0 \leq i < M, 0 \leq j < m, 0 \leq J < m \times M$$

$$p_i = \sum_{j=0}^{m-1} Y_{i,j}, Y_{i,j} \in \{0, 1\}, 0 \leq p_i \leq m; \quad (7)$$

$$q_i = \sum_{j=0}^{m-1} [(Y_{i,j-1}, Y_{i,j}) == (0, 1)], \quad (8)$$

$$j - 1(\bmod m), 0 \leq q_i \leq \lfloor m/2 \rfloor;$$

That is, $X = 0011010010$, $N = 10$, $M = 2$, $m = 5$; $(p_0 = 2, q_0 = 1)$; $(p_1 = 2, q_1 = 2)$.

The output from SM are M pairs of ordered 2-tuple measures $\{(p_i, q_i)\}_{i=0}^{M-1}$.

CP consists of Split and Projection steps. Split adapts the 2-tuple measuring sequence $\{(p_i, q_i)\}_{i=0}^{M-1}$, splitting it into two independent measuring sequences: $\{p_i\}_{i=0}^{M-1}$, $\{q_i\}_{i=0}^{M-1}$ to keep the original order invariant.

The Three measure sequences are $\{p_i\}_{i=0}^{M-1}$, $\{q_i\}_{i=0}^{M-1}$, $\{(p_i, q_i)\}_{i=0}^{M-1}$.

The Projection step turns the sequence into histograms: Project Array (PA), Color Map (CM), and Get Entropy (GE). For three measuring sequences, two types of 1D and 2D measures will be processed separately.

The PA processes measuring sequences to transform them into integer arrays and the CM will organize them on either normalized histograms (1D measures) or color maps (2D measures), respectively.

The 1D measures involve two measuring sequences: $\{p_i\}_{i=0}^{M-1}$, $\{q_i\}_{i=0}^{M-1}$. Let $P[m+1]$, $Q[\lfloor m/2 \rfloor + 1]$ and $NP[m+1]$, $NQ[\lfloor m/2 \rfloor + 1]$ be two 1D (integer, float) arrays to represent the corresponding elements.

The 1DP statistic histogram is generated from a sequence $\{p_i\}_{i=0}^{M-1}$, NP , P two arrays (floating point, integer) with $(m+1)$ elements. For the j th element $NP[j]$, $P[j]$, $0 \leq j \leq m$, and $1DP_e$ the entropy element, the output can be obtained by the following procedure:

Initialization: $\forall NP[j] = 0.0$,
 $P[j] = 0, 0 \leq j \leq m$;
 Calculation: $for(i = 0; i < M; i++)$
 $\{P[p_i]++;\}$
 Normalization: $for(j = 0; j \leq m; j++)$
 $\{NP[j] = P[j]/M;\}$
 Get Entropy: $1DP_e = -\sum_{i=0}^m NP[j] * \log_2(NP[j])$

In the 1DP map, the PA corresponds to Initialization and Calculation; the MA handles Normalization and the GE determines the entropy element of the map.

The 1DQ statistic histogram is generated from a sequence $\{q_i\}_{i=0}^{M-1}$, NQ , Q two arrays (floating point, integer) with $(\lfloor m/2 \rfloor + 1)$ elements; For the j th element $NQ[j]$, $Q[j]$, $0 \leq j \leq \lfloor m/2 \rfloor$, and $1DQ_e$ the entropy element, the output can be obtained from the following procedure:

Initialization: $\forall NQ[j] = 0.0$,
 $Q[j] = 0, 0 \leq j \leq \lfloor m/2 \rfloor$;
 Calculation: $for(i = 0; i < M; i++)$
 $\{Q[q_i]++;\}$
 Normalization: $for(j = 0; j \leq \lfloor m/2 \rfloor; j++)$
 $\{NQ[j] = Q[j]/M;\}$
 Get Entropy: $1DQ_e = -\sum_{j=0}^{\lfloor m/2 \rfloor} NQ[j] * \log_2(NQ[j])$

Using P , NP , Q , NQ arrays, it is possible to generate corresponding 1D statistical histograms as 1D maps.

In the 1DQ map, the PA corresponds to Initialization and Calculation; the MA handles Normalization and the GE identifies the entropy element of the map.

The 2D measures specially processes one measuring sequence: $\{(p_i, q_i)\}_{i=0}^{M-1}$. Let PQ , NPQ be two 2D (integer, float) arrays.

A 2DPQ statistic histogram is generated from a sequence $\{(p_i, q_i)\}_{i=0}^{M-1}$, PQ , NPQ 2D arrays with $(m+1) \times (\lfloor m/2 \rfloor + 1)$ elements. For the i, j th element $PQ[i, j]$, $NPQ[i, j]$, $0 \leq i \leq m$, $0 \leq j \leq \lfloor m/2 \rfloor$, and $2DPQ_e$ the entropy element, their values can be obtained by the following procedure:

Initialization: $\forall PQ[i, j] = 0,$
 $0 \leq i \leq m, 0 \leq j \leq \lfloor m/2 \rfloor;$
 Calculation: $for(i = 0; i < M; i++)$
 $\{PQ[p_i, q_i]++;\}$
 Pseudo-color: Matching proper color for
 $\forall PQ[i, j], 0 \leq i \leq m, 0 \leq j \leq \lfloor m/2 \rfloor$
 Normalization: $for(j = 0; j \leq m; j++)\{$
 $for(j = 0; j \leq \lfloor m/2 \rfloor; j++)$
 $\{NPQ[i, j] = PQ[i, j]/M;\}$
 Get Entropy: $1DPQ_e = - \sum_{j=0}^{\lfloor m/2 \rfloor} \sum_{i=0}^m NPQ[i, j] * \log_2(NPQ[i, j])$

In the 2DPQ map, the PA corresponds to Initialization and Calculation; the MA handles Pseudo-color, Normalization and the GE identifies the entropy element of the map.

Through the CP module, three measuring sequences are transformed into two 1D arrays and one 2D array with $(m + 1)$, $(\lfloor m/2 \rfloor + 1)$ and $(m + 1) \times (\lfloor m/2 \rfloor + 1)$ clusters.

The output of the testing system are three maps $\{1DP, 1DQ, 2DPQ\}$ and three entropies $\{1DP_e, 1DQ_e, 2DPQ_e\}$ as expected statistic distributions and representatives of the input 0-1 sequence, respectively.

4 Results

Three quantum random sequences are selected from USTC $\{1, 2, 4\}$ streams.

Typical results of testing stationary properties for three sequences in nine maps are shown in Fig. 2. Top part contains three 2D maps of global entropy curves on $r = 0 - 128$ condition. Three 2D maps of entropy curves for $r = 0 - 128$ are shown to illustrate refined properties in stationary random curves. Three sets of variant maps in $r = 0$ and their enlarged entropy curves on $r = 0 - 128$ are shown in three columns to illustrate corresponding 1DP, 1DQ, and 2DPQ maps for three sequences. Three larger maps of three global entropy curves are shown in Fig. 3.

For a G map, let G_e be an average entropy variation, ΔG_e be a region of entropy variations, and $G_e^R = \Delta G_e / G_e$ be an entropy variation ratio. Three entropy curves on three 2D maps are compared. Three entropy measurements and $\{\text{Max}, \text{Min}, \text{Max-Min}\}$ values for three sequences are listed in Table 1. Three variation ratios and their numeric quantities are listed in Table 2.

5 Result Analysis

Three 2D maps of global entropy curves show stronger stationary randomness under shift operations on $r = 0 - 128$. Three entropy curves on each map are three stable

Fig. 2 Three USTC random sequences: {1, 2, 4} on 2DPQ, 1DP, and 1DQ maps and $r = 0 - 128$ entropy curves

Fig. 3 Three enlarged 2D maps of global entropy curves for three USTC random sequences

Table 1 Comparisons on three measures for three USTC samples

	Q_e	ΔQ_e	Q_e^R
1:	2.4537	0.63%	0.2559%
2:	2.4638	0.83%	0.335%
4:	2.46	0.86%	0.3481%
Min:	2.4537	0.63%	0.2559%
Max:	2.4638	0.86%	0.3481%
Max–Min:	0.0101	0.23%	0.0922%
	P_e	ΔP_e	P_e^R
1:	2.9948	0.82%	0.2742%
2:	3.1502	0.84%	0.267%
4:	3.1472	0.61%	0.1937%
Min:	2.9948%	0.61%	0.1937%
Max:	3.1502%	0.84%	0.2742%
Max–Min:	0.1554%	0.23%	0.0805%
	PQ_e	ΔPQ_e	PQ_e^R
1:	5.4397	0.81%	0.1481%
2:	5.5998	1.14%	0.2035%
4:	5.5932	0.66%	0.1183%
Min:	5.4397	0.66%	0.1183%
Max:	5.5998	1.14%	0.2035%
Max–Min:	0.1601	0.48%	0.0852%

Table 2 $Q_e + P_e : PQ_e$ measures

No.	Q_e	Q_e	$Q_e + P_e$	PQ_e	$\Delta_e =$	$ Q_e + P_e - PQ_e $	PQ_e / Δ_e
1	2.4537	2.9948	5.4485	5.4397		0.0088	618
2	2.4638	3.1502	5.614	5.5998		0.0142	394
4	2.46	3.1472	5.6072	5.5932		0.014	400

horizontal lines. From a global viewpoint, there are significant differences compared with entropy curves between No. 1 (PQ and P) and No. 2 & 3 cases. Both No. 2 and 3 are in similar measures.

Nine variant maps in 2DPQ, 1DP, and 1DQ, three 2DPQ maps are 2D distributions and there are different symmetric distributions. Maximal elements in three maps show stronger vertical-oriented features. Three maps have a symmetry on left/right directions and have a broken symmetry on up/down directions. Pseudo-color pixels on three maps are shown in 3D shapes. Three 1DP maps have similar distributions in bell shapes to illustrate Poissonian distributions. Compared with three 1DP maps, three 1DQ maps have similar distributions and more narrow bell shapes to illustrate sub-Poissonian distributions.

However, nine enlarged entropy curves for each type have significantly different variations and distributions. Local curves are bounded in narrow regions with random variations.

It is difficult to tell detailed differences from entropy curves. Quantitative measurements in Table 1 are helpful to use numeric values in comparison. The difference of entropy variation ratios are on three sets, Q_e^R : [0.26, 0.35]%, P_e^R : [0.19, 0.27]%,

and PQ_e^R : [0.12, 0.20]%. Three Max-Min values of $\{Q_e^R, P_e^R, PQ_e^R\}$ are bounded in [0.08, 0.09]%. The whole structure illustrates measurable stationary properties. In Table 2, it is interesting to notice that $Q_e + P_e \sim PQ_e$.

All variation measurements are shown in distinct stationary randomness to be measured by entropy approaches.

6 Conclusion

Information entropy is a useful measurement to determine stationary randomness. Three quantum random sequences are used, distinct stationary randomness can be identified from both variant maps and numeric measurements. To explore various conditions of stationary properties, further investigations are required to explore theoretical boundaries on variant maps.

Acknowledgements Thanks to The Key project of Quantum Communication of Yunnan Province, National Science Foundation of China (61362014) and High-Level Overseas Professional Project of Yunnan Province for financial supports to this project. Thanks to the Key Laboratory of Quantum Information, USTC, and CAS for providing quantum random sequences.

References

1. W. Chen, Active phase compensation of quantum key distribution system. *Chin. Sci. Bul.* **53**(9), 1310–1314 (2008)
2. D.E. Knuth, in *The Art of Computer Programming*, vol. 4A, Combinatorial Algorithms Part 1 (Addison-Wesley, 2011)
3. NIST, in *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications* (NIST, Special Publication, 2010)
4. M.B. Priestley, in *Non-linear and Non-stationary Time Series Analysis* (Academic Press, 1988)
5. X.T. Song, Phase-coding self-testing Quantum random number generator. *Chin. Phys. Lett.* **32**(8), 080302–080310 (2015)
6. Stationary process, https://en.wikipedia.org/wiki/Stationary_process
7. W.Z. Yang, J. Zheng, Variant Pseudo-random number generator. *Hakin9 Extra Timing Attack* **06**(13), 28–31 (2012)
8. J. Zheng, C. Zheng, T.L. Kunii, Interactive Maps on Variant Phase Space, in *Emerging Application of Cellular Automata* (InTech Press, 2013), pp. 113–196
9. J. Zheng, C. Zheng, Stationary randomness of quantum cryptographic sequences on variant maps, in *Proceedings on ASONAM '17*, (ACM, 2017). ISBN 987-1-4503-4993-2/17/07 <https://doi.org/10.1145/3110025.3110151>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.