entropy

MDPI

Article

Complexity Reduction in Analyzing Independence between Statistical Randomness Tests Using Mutual Information

Jorge Augusto Karell-Albo [1], Carlos Miguel Legón-Pérez [1], Raisa Socorro-Llanes [2], Omar Rojas [3] and Guillermo Sosa-Gómez [3,*]

[1] Instituto de Criptografía, Facultad de Matemática y Computación, Universidad de la Habana, Habana 10400, Cuba; jorgekarellalbo@gmail.com (J.A.K.-A.); clegon58@gmail.com (C.M.L.-P.)
[2] Facultad de Ingeniería Informática, Universidad Tecnológica de la Habana José Antonio Echeverría (CUJAE), Habana 19390, Cuba; raisa@ceis.cujae.edu.cu
[3] Facultad de Ciencias Económicas y Empresariales, Universidad Panamericana, Álvaro del Portillo 49, Zapopan 45010, Jalisco, Mexico; orojas@up.edu.mx
* Correspondence: gsosag@up.edu.mx; Tel.: +52-3313682200

Abstract: The advantages of using mutual information to evaluate the correlation between randomness tests have recently been demonstrated. However, it has been pointed out that the high complexity of this method limits its application in batteries with a greater number of tests. The main objective of this work is to reduce the complexity of the method based on mutual information for analyzing the independence between the statistical tests of randomness. The achieved complexity reduction is estimated theoretically and verified experimentally. A variant of the original method is proposed by modifying the step in which the significant values of the mutual information are determined. The correlation between the NIST battery tests was studied, and it was concluded that the modifications to the method do not significantly affect the ability to detect correlations. Due to the efficiency of the newly proposed method, its use is recommended to analyze other batteries of tests.

Keywords: mutual information; complexity; PRNG; cryptography

Citation: Karell-Albo, J.A.; Legón-Pérez, C.M.; Socorro-Llanes, R.; Rojas, O.; Sosa-Gómez, G. Complexity Reduction in Analyzing Independence between Statistical Randomness Tests Using Mutual Information. *Entropy* **2023**, *25*, 1545. https://doi.org/10.3390/e25111545

Academic Editor: Haobin Shi

Received: 11 October 2023
Revised: 9 November 2023
Accepted: 13 November 2023
Published: 15 November 2023

1. Introduction

In cryptographic applications, sequences of random numbers are very important. One of the widest-spread applications is the generation of secret keys. These sequences are obtained using random number generators, which, depending on the source of randomness, can be classified as pseudo-random number generators (PRNGs) or truly random number generators (TRNGs). Some authors, such as [1], include a third group of generators called hybrid random number generators (HRNGs). The HRNG combines elements of the previous two generators.

According to [2], the best way to generate "unpredictable" random numbers is to use physical processes, such as atmospheric or thermal noise, cosmic radiation, and some other phenomena; however, number generation from physical processes is relatively inefficient since TRNGs are expensive devices both in terms of execution and applicability. For this reason, ref [3] states that most systems use PRNGs instead of TRNGs.

Weak RNGs have been identified as the cause of many security problems in computer systems. A concrete case is exemplified by MIFARE Classic, an RFID radio frequency identification) card manufactured by NXP Semiconductors that had a weak PRNG, making it possible to attack their channel of supposedly secure communications [4]. Furthermore, ref [5] described pre-play and man-in-the-middle attacks that demonstrated how, due to a weak RNG, a payment system could be abused. An explanation of how it was possible to attack Taiwan's national database of "Citizen Digital Certificates" was presented in [6], where it was due to poor implementation of the TRNG in the Office-certified Renesas AE45C1 v.01 smart card IC German Federal Information Security

(BSI) [7]. These examples highlight the need for further analysis of RNG design and implementation to mitigate the possibility of such attacks in the future.

Because PRNGs are based on deterministic algorithms, it is necessary to examine the output to confirm its suitability for cryptographic applications. This output is statistically analyzed by one or several randomness tests, and the results are evaluated to determine whether the generator is random or not. The statistic associated with each statistical test attempts to identify the presence or absence of a specific pattern. If the pattern is detected, then the lack of randomness in the sequence can be inferred. There are many statistical tests for randomness. The same test can even be applied multiple times with different combinations of parameters [8].

Although some authors [9] recommend using multiple statistical randomness tests to leverage each test's strengths, applying too many tests may lead to overestimating the properties of a PRNG. Thus, ref [10] suggest that two conceptually different tests may evaluate the same characteristic of randomness and produce correlated results. In addition, not only can the correlation between different randomness tests be studied, but the same test with different parameters can be correlated. Therefore, correlation analysis between randomness tests can also be used for parameter selection. In [11,12], the study of correlations between tests using the Pearson correlation coefficient (CCP) was proposed; however, this coefficient is not capable of detecting nonlinear correlations.

Results reported by [13] demonstrate the need for a greater variety of statistical tests. In their work, they studied eight RNGs, and, in some cases, generators showed difficulties in passing the tests of Alphabits, Rabbit, Small Crush, and Crush batteries, passing the tests of Dieharder and SP800-22. As [13] points out, future batteries of statistical tests should be thoroughly analyzed for possible correlations between their tests before publication. Likewise, a continuous review of the new randomness tests must be carried out to ensure their efficiency and statistical strength and integrate them into the batteries if they are not correlated with the existing tests or are superior to them.

In [12,14], Pearson's correlation coefficient was used to analyze correlations between tests; however, [2,11] focused on the proportion of regions where p-values are less than 0.01. Then, ref [15] analyzed the dependencies between nine NIST tests and found the same dependencies as [12]. Another approach proposed by [16] analyzed the difference between p-values corresponding to two different tests and detected correlations according to the distribution of that difference.

In [14], dependencies between some NIST tests were determined using the Pearson correlation coefficient, and patterns were found in the evolution of these dependencies according to specific factors, such as the length of the sequences analyzed by the tests. In [17], a novel method was proposed to detect the correlation between statistical randomness tests using mutual information for the first time. This measure was used to examine the tests present in the NIST SP 800-22 battery, detecting new correlations.

Different articles have verified the effectiveness of mutual information in studying the correlation between the randomness tests of different batteries. In [18], DieHarder, TufTest, and SmallCrush batteries were studied, and in [19], the FIPS 140-2 battery was analyzed using the method proposed in [17]. Mutual information can detect the correlations found by Pearson correlation and new nonlinear dependencies between tests. In [20], the complexity of the method proposed in [17] was estimated, and its high complexity was pointed out as a disadvantage, which prevents testing batteries with more tests or longer sequences. However, mutual information as a new measure of independence between randomness tests represents a significant advance and an important piece for building a definitive test battery [18].

The main objective of this work is to reduce the complexity of the method proposed by [17], which we will call from now on MIRT-1 (Mutual Information to analyze Statistical Randomness Test v1). The step in which the mutual information values corresponding to significant correlations are determined is modified to achieve this. The complexity reduction achieved with this modification is estimated theoretically and then confirmed

experimentally. The efficiency of the new proposed method (MIRT-2) will be determined by considering the complexity calculation. On the other hand, the efficacy will be measured according to the number of correlations that it can detect compared to the MIRT-1 method.

2. Preliminaries

2.1. Mutual Information

Mutual information (MI) is an important measure of statistical correlation. It stands out for its ability to detect the correlation between variables and possesses properties that make it an ideal measure of stochastic dependence [21–24]. Unlike the Pearson correlation coefficient (PCC), which only considers linear dependencies, or other correlation coefficients that only detect monotonic dependencies, MI considers all correlations regardless of their functional form.

The MI between two discrete random variables X and Y is defined as

$$MI(X,Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right), \tag{1}$$

where $p(x,y)$ is the joint probability function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y, respectively. MI can also be defined in terms of entropy as

$$MI(X,Y) = H(X) + H(Y) - H(X,Y), \tag{2}$$

where $H(X)$ and $H(Y)$ are the marginal entropy of the variables X and Y, respectively, and $H(X,Y)$ is the joint entropy of both variables. For the case of continuous random variables, the sums are replaced by integrals:

$$MI(X,Y) = \int_Y\int_X p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)dxdy. \tag{3}$$

More on the properties of entropy and mutual information for continuous variables and their relationship to the discrete case can be found in [25].

2.2. Estimating Mutual Information

When the probability distributions are unknown, it is not possible to calculate the exact value of $MI(X,Y)$, so it is necessary to calculate a sample estimator $\widehat{MI}(X,Y)$. Estimators of the mutual information $\widehat{MI}$ differ in estimating the probabilities of the marginal and joint densities. Some of the estimators proposed in the literature use discretization [26,27], kernels [28,29] or correlation integrals [30], k-nearest neighbors [31,32], B-splines [33], or Gram–Charlier polynomial expansion [34].

One of the simplest estimators is the maximum-likelihood (ML) estimator (plug-in). With this, entropy is estimated from the observed individual and joint frequencies. Since there are no assumptions about the data distribution, the ML estimator is considered non-parametric. Another of the estimators used is the James–Stein shrinkage estimator. The approach of this estimator is considered semi-parametric because it has characteristics of parametric and non-parametric methods [35]. A comprehensive comparison of estimators can be found in [35,36].

According to [37], the most common mutual information estimator is the naive equidistant binning estimator. This estimator considers each variable's domain partition in a finite number, n, of discrete intervals (equidistant partition). The number of intervals for each variable is the same, so the parameter to optimize is the number of intervals to discretize or, equivalently, the length of the interval. The binning process begins with selecting the support interval, (a,b). Usually, this interval is constructed from the smallest and largest values in the sample $\{x_1, x_2, \ldots, x_n\}$. Given the interval (a,b), a frequency histogram is constructed from the number of points in each interval. The intervals, B_k, are half-open, i.e., $B_k = [t_k, t_{k+1})$ for $k = -\infty, \ldots, -1, 0, 1, \ldots, \infty$, where the points t_k satisfy

that $t_k < t_{k+1}$. The width of each interval is denoted by $h_k = t_{k+1} - t_k$, and the number of elements in the interval is denoted by v_k. In this way, given the sample $x_1, x_2, \ldots, x_n$, we have

$$v_k = \sum_{i=1}^{n} \mathbb{1}_{x_i \in B_k}, \tag{4}$$

where $\mathbb{1}(\cdot)$ is the indicator function. It is clear that $v_k \geq 0$ and $\sum_k v_k = n$. Generally, the intervals are selected so they are of the same width, i.e., $h_k = h$, $\forall k$. For equal-width intervals, we define

$$h = \frac{b-a}{k}. \tag{5}$$

Many researchers have attempted to determine the optimal value, k, of the intervals, but these methods often rely on strong assumptions about the distribution of the data [38]. Depending on the distribution type and the analysis objective, different values of h can be selected. Experimentation is usually necessary to determine the optimal width. Table 1 presents some of the most used rules for selecting k.

Table 1. Rules for selecting the number k of intervals for the discretization.

Rules	k
Sturges [39]	$1 + \log_2(n)$
Cochran [40]	$\lfloor \sqrt{n/5} \rfloor$
Rice [41]	$\lfloor 2\sqrt[3]{n} \rfloor$
Cencov [42]	$\lfloor \sqrt[3]{n} \rfloor$
Bendat–Piersol [43]	$\lfloor 1.87(n-1)^{0.4} \rfloor$
Larson [44]	$1 + \lfloor 2.2 \log_2(n) \rfloor$
Velleman [45]	$\lfloor 2\sqrt{n} \rfloor$ if $n \leq 100$ $\lfloor 10 \log_{10}(n) \rfloor$ if $n > 100$
Doane [46]	$1 + \log_2(n) + \log_2\left(1 + \frac{\sqrt{b_1}}{\sigma\sqrt{b_1}}\right)$
Mosteller–Tukey [47]	$\lfloor \sqrt{n} \rfloor$
Terrell–Scott [48]	$\sqrt[3]{2n}$
Ishikawa [49]	$6 + \lfloor n/50 \rfloor$

In most cases, the number of bins to discretize is set to $k = 10$ [50]. This work will discretize it in $k = 10$ intervals because it is the same approach used in [17] to design the MIRT-1 method.

2.3. Distribution of Estimators

The ML estimator of mutual information $\widehat{MI}_{ML}(X,Y) = \widehat{H}_{ML}(X) + \widehat{H}_{ML}(Y) - \widehat{H}_{ML}(X,Y)$ has been extensively studied in the literature [51,52]. It is known from [51] that, under certain conditions, such as finite alphabet size and $MI > 0$, the following holds:

$$\frac{\sqrt{n}(\widehat{MI}_{ML} - MI)}{\widehat{\sigma}} \sim N(0,1). \tag{6}$$

On the other hand, in [52,53], it was stated that

$$2n\widehat{MI}_{ML} \sim \chi^2_{(I-1)(J-1)} \tag{7}$$

where I and J are the sizes of the alphabets of X and Y, respectively.

In [17], the ML, Miller–Madow, James–Stein, and Schurmann–Grassberger estimators were compared as the sample size increased. For the selection of the mutual information estimator, two pairs of variables were used, one of independent variables (see Figure 1) and the other of dependent variables (see Figure 2). It was concluded that, for more than 10,000 observations, the difference between the estimators is very small: 0.0074 for the pair of independent variables and 0.0065 for the dependent ones. However, the James–Stein

estimator (shrinkage) was selected since the mutual information took the value of 0 for the independent variables, even for very small sample sizes.

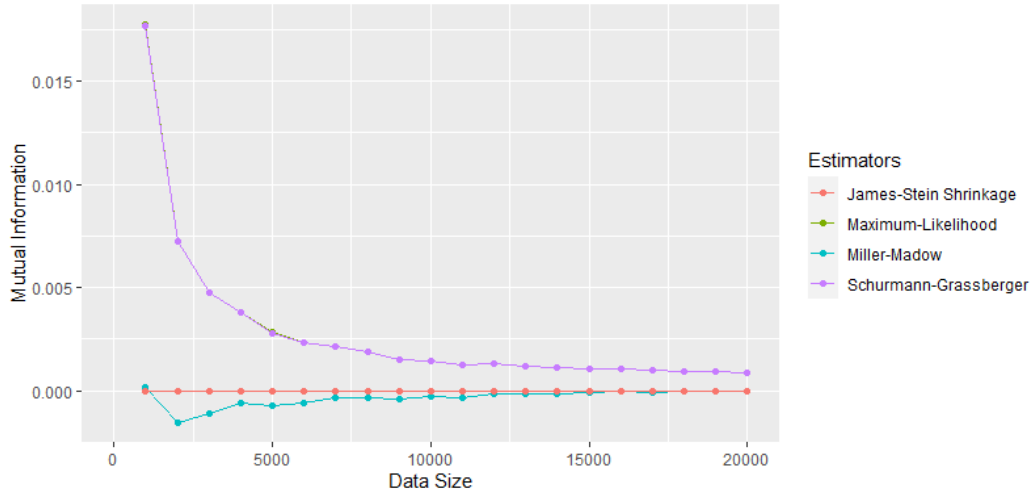

Figure 1. Comparison of mutual information estimators for a pair of independent tests [17].

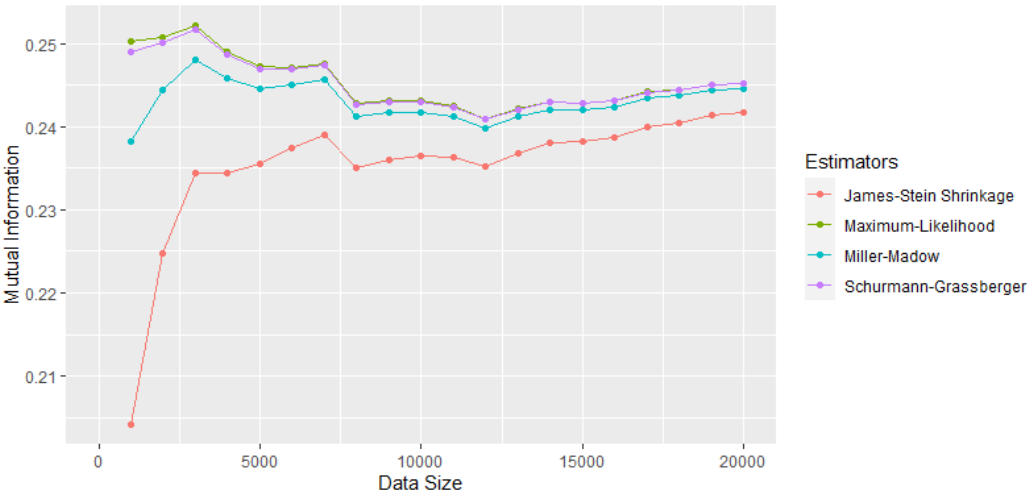

Figure 2. Comparison of mutual information estimators for a pair of correlated tests [17].

Although for this particular case, only these mutual information estimators are being compared, a more comprehensive study of the entropy estimators can be found in [36]. The entropy estimator James–Stein shrinkage used to calculate mutual information in [17] is defined as

$$\widehat{H}_{SH} = -\sum_{x \in K} \hat{p}_x^{SH} \log_2 \hat{p}_x^{SH} \tag{8}$$

where

$$\hat{p}_x^{SH} = \hat{\lambda} t_x + (1 - \hat{\lambda}) \hat{p}_x^{ML} \tag{9}$$

with

$$\hat{\lambda} = \frac{1 - \sum_{x=1}^{k} (\hat{p}_x^{ML})^2}{(n-1) \sum_{x=1}^{k} (t_x - \hat{p}_x^{ML})^2} \tag{10}$$

and

$$t_k = \frac{1}{k} \tag{11}$$

2.4. MIRT-1 Method

As the distribution of the shrinkage estimator of the mutual information is unknown, a permutation test was performed to decide which values were significantly greater than 0. A permutation test is a statistical method used to determine the significance of a particular statistic by comparing it to a null distribution generated by randomly permuting the data. In this case, the permutation test determines whether the mutual information between two variables significantly differs from what would be expected by chance.

The null hypothesis (H_0) is that there is no relationship between the two variables, and any observed mutual information is due to chance. The alternative hypothesis (H_1) is that there is a significant relationship between the two variables, and the observed mutual information is not due to chance. If the p-value is less than a predetermined significance level, then the null hypothesis can be rejected in favor of the alternative hypothesis, indicating that there is a significant relationship between the two variables. The permutation test is applied following these steps:

1. Let $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$ be continuous random variables.
2. Construct the permuted samples $\left(X, \pi_i(Y)\right)$, $\forall i = \overline{1,k}$ in such a way that the possible association between X and Y disappears, π_i being the permutation i of elements of Y, i.e.,
   - $\pi_i \in S_n$, $\forall i = \overline{1,k}$;
   - $\pi_i \neq \pi_j$, for $i \neq j$;
   - π_0 is the identity of S_n.
3. Estimate the MI of the allowed samples to obtain $\{Z_i\}_{i=0}^{k}$, where $Z_i = MI(X, \pi_i(Y))$.
4. The p-value associated with the test is calculated by

$$p\text{-value} = \frac{\sum_{j=1}^{k} \mathbb{I}_{\geq Z_0}(Z_i)}{k}, \tag{12}$$

   where $\mathbb{I}_{\geq Z_0}(Z_i)$ is the indicator function defined by

$$\mathbb{1}_{\geq Z_0}(x) = \begin{cases} 1 & \text{if} \quad x \geq Z_0, \\ 0 & \text{if} \quad x < Z_0. \end{cases} \tag{13}$$

5. If p-value $\geq \alpha$, then the null hypothesis is not-rejected.

In [17], these steps were performed for each pair of random statistical tests, computing $q = 10,000$ permutations in each case.

The MIRT-1 method proposed in [17] is described in the following steps:

1. Select PRNGs.
   - The selected generators must generate outputs that satisfy the randomness conditions.
2. Build the data samples using the selected generators.
   - Generate n sequences of random numbers of length L to be evaluated using the selected statistical randomness tests.
3. Evaluate each of the n sequences using the k statistical randomness tests to obtain the corresponding p-values for each T_i test (with $i = 1, \ldots, k$).
4. Compute the MI between sequences of p-values to detect the possible presence of correlations.
   - Estimate the MI between pairs (T_i, T_j) of sequences of p-values to detect the presence of correlation.
   - In the case of the MIRT-1 method, the estimator used is the shrinkage estimator.
   - For a better interpretation in [17], the MI values were normalized, i.e., $MI'(T_i, T_j) = \frac{IM(T_i, T_j)}{H(T_i)}$, where $H(T_i)$ represents the entropy of the variable T_i.

5. Determine the significant correlations to conclude the correlation between the tests using the permutation test. The MI values are grouped in a triangular matrix

$$M = \begin{pmatrix} H(T_1) & MI(T_1,T_2) & \dots & MI(T_1,T_k) \\ 0 & H(T_2) & \dots & MI(T_2,T_k) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & H(T_k) \end{pmatrix}_{k\times k}$$

where $MI(T_i, T_j)$ represents the MI between i and j. The diagonal of the matrix contains the values $H(T_i)$ that represent the entropy of the variable T_i.

The complexity of the MIRT-1 method is estimated in [20] using the following:

$$O(k^2 \cdot q(n + d^2)), \tag{14}$$

where k represents the number of statistical tests to be analyzed, q is the number of permutations to determine the significance of the correlations, n is the length of the sequences of p-values analyzed, and d is the number of intervals in the discretization process. In practical applications, it is of interest to increase k, and the parameter d is selected in such a way as to increase the effectiveness of detection. Therefore, the feasible parameters to reduce are q and n, but the reduction in n can affect the effectiveness of the estimator, so, in this work, we will focus on reducing the value q.

3. Reducing the Complexity of the Method Based on Mutual Information to Analyze the Independence between the Statistical Tests of Randomness

This section presents the main contribution of this work: significantly reducing the complexity of the MIRT-1 method.

3.1. Solution Idea

To reduce complexity, it is proposed to eliminate the permutation test, which poses a new problem. In Step 5 of the MIRT-1 method, deciding which values of the mutual information are significant is necessary to conclude the correlation between the tests. Given the continuous random variables X and Y, decide if the value $MI(X, Y)$ is significantly greater than 0, and thus conclude if there is some dependency between both variables. The hypothesis test asks the following:

$$\begin{cases} H_0: & MI(X,Y) = 0 \\ H_1: & MI(X,Y) > 0 \end{cases}, \tag{15}$$

where H_0 is the null hypothesis that states the independence between X and Y, and H_1 is the alternative hypothesis, where there would be some association between X and Y.

To determine the significant values without using the permutation test, we propose to replace the James–Stein (shrinkage)-type estimator $\widehat{MI}_{SH}$, whose distribution is unknown, with another estimator of mutual information whose distribution is known. The mean square error of the new estimator should be reasonably small for the selected sample size. In this paper, we will use a maximum-likelihood estimator transformation $\widehat{MI}_{ML}$ whose distribution is known [51,52]. It is known that the maximum-likelihood estimator of entropy is normally distributed [54] asymptotically with mean

$$E(\widehat{H}_{ML} - H) = -\frac{k-1}{2n} + \frac{1}{12n^2}\left(1 - \sum_{i=1}^{k}\frac{1}{p_i}\right) + O(n^{-3}), \tag{16}$$

and variance,

$$\sigma^2(\widehat{H}_{ML}) = \frac{1}{n}\left(\sum_{i=1}^{k} p_i \ln^2(p_i) - H^2\right) + \frac{k-1}{2n^2} + O(n^{-3}). \tag{17}$$

The mean can be calculated if the order n^2 is neglected. However, for the case of the variance, the terms of order n depend on the unknown probabilities. Therefore, the transformation of Equation (7) will be used since the mean and variance are known, and it is known that the normal distributes asymptotically when the number of degrees of freedom is greater than 30.

3.2. Selection of the Critical Value to Determine the Significant Correlations

To decide whether to reject the null hypothesis of independence between (T_i, T_j) in the new variant of the method, a priori knowledge of the distribution of the transformation $2n\widehat{MI}_{ML}(T_i, T_j)$ of estimator $\widehat{MI}_{ML}$ will be used. The transformation $2n\widehat{MI}_{ML}(T_i, T_j)$ distributes χ^2 (Equation (7)). For our case, $I = J = d$ is the number of discretization intervals. Then,

$$2n\widehat{MI}_{ML}(T_i, T_j) \sim \chi^2_{(d-1)^2}. \tag{18}$$

To guarantee that the χ^2 distribution of $2n\widehat{MI}_{ML}(T_i, T_j)$ approximates the normal distribution, the condition $(d-1)^2 > 30$ will be imposed on d, which implies that $d > 1 + \sqrt{30} = 6.48$, concluding that, for values of $d > 6.48$, we have

$$2n\widehat{MI}_{ML}(T_i, T_j) \sim N\Big((d-1)^2, \sqrt{2}(d-1)\Big). \tag{19}$$

For the particular case of study, if we discretize $d = 10 > 6.48$ intervals, we obtain $(d-1)^2 = (9)^2 = 81 \gg 30$, which guarantees that the theoretical distribution of $2n\widehat{MI}_{ML}(T_i, T_j)$ closely approximates the normal distribution (Equation (19)). Using this theoretical distribution, the critical value corresponding to the α prefix can be taken, and therefore it is unnecessary to use the permutation test.

For the sampling distribution of $2n\widehat{MI}_{ML}(T_i, T_j)$ to be close to its theoretical normal distribution, it is convenient that the length n of the sequence of p-values is big enough. So,

$$Z_{ML} = \frac{2n\widehat{MI}_{ML}(T_i, T_j) - (d-1)^2}{\sqrt{2}(d-1)} \sim N(0,1) \tag{20}$$

distributes the normal asymptotically when $(d-1)^2 \gg 30$. For an α significance level, the critical value for the right tail is $Z_{1-\alpha}$. Therefore, if the estimated values of Z_{ML} are greater than the critical value, it can be concluded that there is a correlation between T_i and T_j with probability α that they are independent. For example, for $\alpha = 0.01$, if $Z_{ML} > Z_{1-0.01} = 2.36$, it is concluded that the value Z_{ML} is significant.

3.3. Method Modification Proposal

From Step 1 to Step 3, the MIRT-1 method remains unchanged. In Step 4, in [17], the shrinkage estimator $\widehat{MI}_{SH}$ was used, while now it is proposed to use the ML estimator $\widehat{IM}_{ML}$. In Step 5, in [17], the values of $\widehat{MI}_{SH}$ were normalized. A permutation test was applied; on the other hand, in this new method used to determine the significant values of $\widehat{MI}_{ML}$, the transformation Z_{ML} is computed, and knowledge of the normal distribution of Z_{ML} will be used.

The application of the new MIRT-2 method from this modification is as follows (the steps that changed concerning MIRT-1 are indicated in bold):

1. Select PRNGs.
   - The selected generators must generate outputs that satisfy the randomness conditions.
2. Build the data samples using the selected generators.
   - Generate n sequences of random numbers of length L to be evaluated using the selected statistical tests of randomness.

3. Evaluate each of the n sequences using the k statistical tests of randomness to obtain the corresponding p-values for each T_i test (with $i = 1, \ldots, k$).
4. Compute the MI between sequences of p-values to detect the presence of correlations.
   - Calculate the MI between pairs (T_i, T_j) of sequences of p-values to detect the presence of correlation.
   - The MI estimator used is the ML $\widehat{MI}_{ML}$
5. Calculate Z_{ML} and compare it with the critical value associated with the default α value. If $\widehat{Z}_{ML} > Z_{1-\alpha}$, the null hypothesis of independence between the tests of randomness is rejected.

Table 2 summarizes the parameters chosen for the MIRT-1 and MIRT-2 methods.

Table 2. Parameters used for the MIRT-1 and MIRT-2 methods.

Method	MIRT-1	MIRT-2
Distribution of the p-values	U(0,1)	U(0,1)
MI estimator	shrinkage (SH)	maximum-likelihood (ML)
Discretization	k = 10	k = 10
Selection of significant values	Comparison with the critical value obtained by the permutation test.	Comparison with the critical value of the normal distribution for the α prefix.

The complexity of the MIRT-1 method was calculated in [20]

$$O(k^2 \cdot q(n + d^2)). \tag{21}$$

The modification proposed in this paper reduces it to

$$O(k^2(n + d^2)), \tag{22}$$

since not applying the permutations test is equivalent to taking $q = 1$. The reduction achieved is of the order

$$\frac{k^2 \cdot q(n + d^2)}{k^2(n + d^2)} = q, \tag{23}$$

so the modified method is expected to be about q-times faster. For example, in [17,20], $q = 10{,}000$ was used; for this case, the MIRT-2 method would have a complexity 10,000 times lower than MIRT-1.

The replacement of the shrinkage estimator $\widehat{MI}_{SH}$ of mutual information used in the MIRT-1 method with the maximum-likelihood estimator $\widehat{MI}_{ML}$ allows complexity to be reduced since the permutation test is eliminated and the known distribution of the new estimator is used to set the critical value and select the significant values. This modification could affect the ability to detect correlations if appropriate measures are not taken. To avoid this impact, it is proposed to apply the MIRT-2 method with a sample size equal to 10 000, and it is recommended not to reduce this value while this $\widehat{MI}_{ML}$ estimator is used. This recommendation is based on the results of Figure 2 of [17] since, for this sample size, the difference between the values of the estimators is very small and the ability to detect correlations is not affected.

4. Experimental Validation

In this section, the reduction in the complexity of MIRT-2 to MIRT-1 will be verified experimentally.

4.1. Experimental Check of Normality of Z_{ML}

4.1.1. Design of Experiment 1

For random sequences, the *p*-values generated by statistical randomness tests follow a uniform distribution with values between 0 and 1 [9]. For this reason, the experiments generated data that followed the same distribution. To study the normality of the transformation Z_{ML} and the pairs (X_i, Y_i) , $i = 1, \ldots, 1\,000$ of independent and identically distributed random variables, where $X, Y \sim U(0,1)$, $X = \{x_1, x_2, ..., x_{10,000}\}$ and $Y = \{y_1, y_2, ..., y_{10,000}\}$. Subsequently, the MI between the pairs was calculated using the ML estimator, and 1000 $\widehat{MI}_{ML}$ observations were obtained. The MI values were transformed using Equation (20) to obtain a sample of 1000 Z_{ML} values. The Anderson–Darling and Kolmogorov–Smirnov tests were applied to verify the normality of Z_{ML}.

4.1.2. Results

Table 3 shows *p*-values for the Anderson–Darling and Kolmogorov–Smirnov normality tests.

Table 3. Tests for normality for the values of Z_{ML}.

Normality Test		***p*-Value**
Anderson–Darling	A = 0.47518	0.2395
Kolmogorov–Smirnov	D = 0.019679	0.8334

For all cases, the *p*-value > 0.05; therefore, there is insufficient evidence to reject the null hypothesis of normality. In Figure 3, a histogram with the values of Z_{ML} and the estimate of the kernel density (KDE) of the sample data (black) and the curve of normal density (blue) can be observed. The blue line represents the theoretical normal probability density function, while the black curve represents a smoothed density estimate using a kernel density estimation technique.

In the Q-Q graph of Figure 4, it can be seen that the observed values are close to the expected ones. Through normality tests, it was possible to verify experimentally that the transformation Z_{ML} follows a normal distribution under the theoretical argumentation presented in Section 3.2. The red line in the Q-Q plot graph represents the theoretical quantiles of a standard normal distribution, and it serves as a reference line for comparing the distribution of the plotted data to the normal distribution.

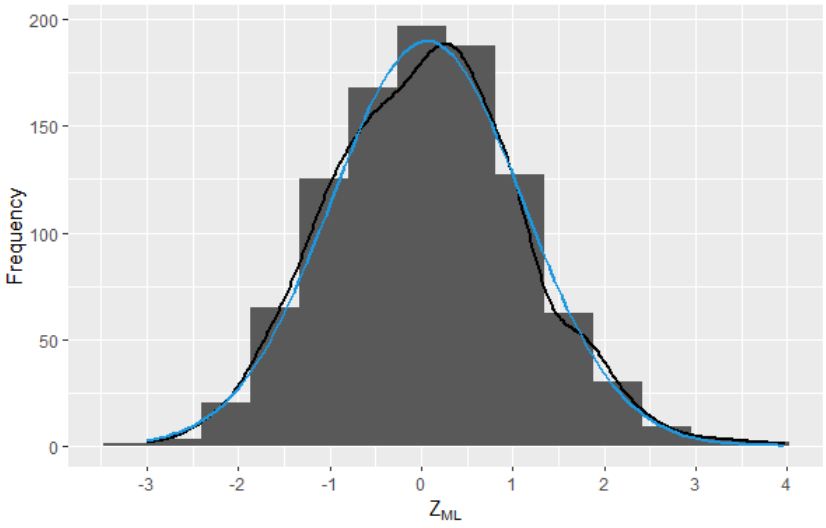

Figure 3. Sampling distribution of the 1000 observations of Z_{ML} obtained from pairs (X_i, Y_i).

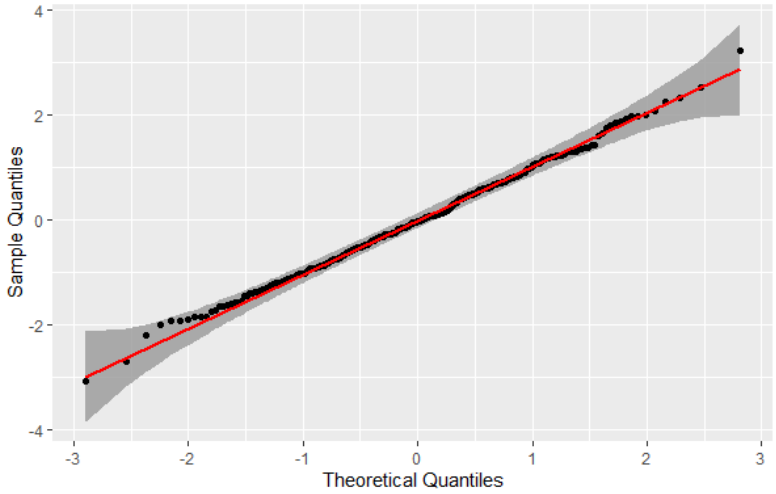

Figure 4. Quantile plot of the 1000 observations of Z_{ML} obtained from pairs (X_i, Y_i).

4.2. Analysis of the Effectiveness of the Proposed Variant

4.2.1. Design of Experiment 2

Based on the idea of the previous section, the MI between the NIST statistical randomness tests was calculated using the MIRT-2 method. For this, the ML estimator was used, and the symmetric matrix of dimension 17×17 was obtained with the values of the MI. It is important to highlight that for selecting the p-values, the approach proposed by [20] was followed. All sequences that did not comply with the requirement were discarded. There was a required number of cycles for the random excursions and random excursions variant tests, thus generating p-values equal to 0. Under the standard normal distribution and the independence assumption, a significance level $\alpha = 0.001$ corresponds to a critical value $CV = 3.090232$. Therefore, if $Z_{ML} > 3.09$, the null hypothesis of independence is rejected. Following Step 5 of the MIRT-2 method, those greater than the critical value were selected as significant values of the Z_{ML}.

4.2.2. Results

In Figure 5, the histogram is represented with the $k(k-1)/2 = 136$ observations of Z_{ML} for the values of the mutual information between the pairs of statistical tests of randomness. High values of Z_{ML} (greater than 300) are seen on the right, indicating possible correlations between test pairs. In Figures 5 and 6, the dotted line represents the critical value selected for the MI.

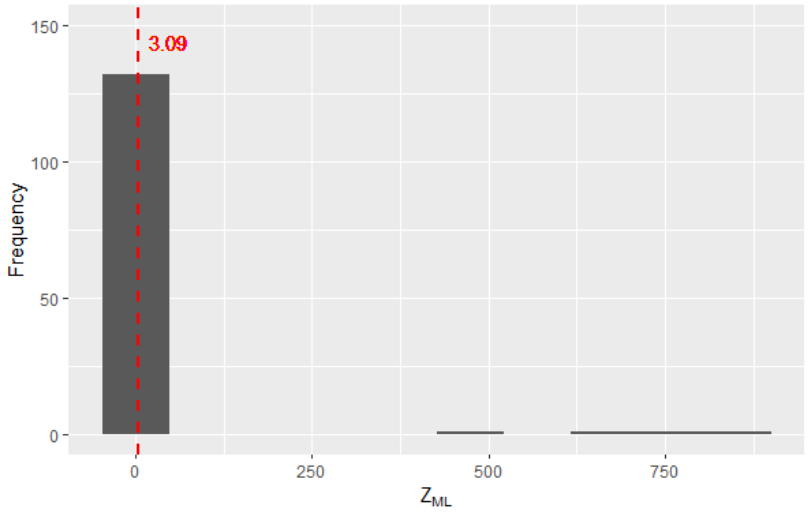

Figure 5. Distribution of the 136 observations of Z_{ML} for values between 0 and 1000.

In Figure 6, only the observations of Z_{ML} are represented for the values between 0 and 6. It can be noted that the values of Z_{ML} smaller than the critical value behave following a normal distribution, as expected under the independence hypothesis.

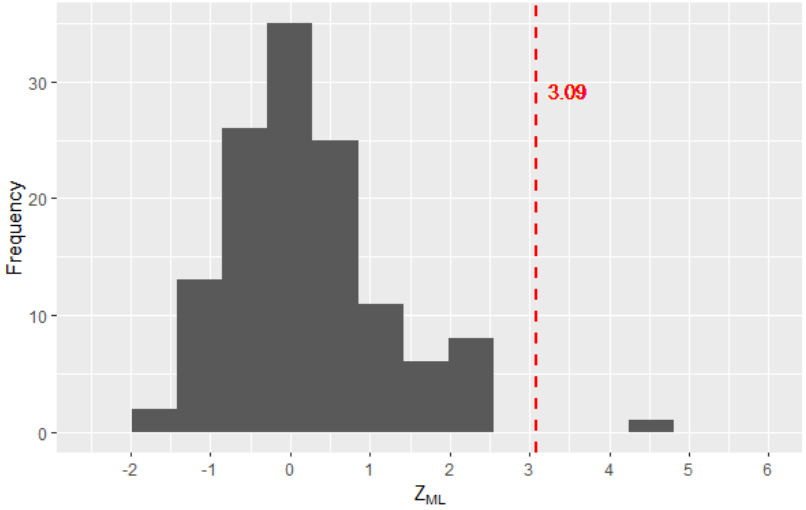

Figure 6. Distribution of Z_{ML} observations for values between 0 and 6.

Figure 7 shows the mutual information matrix with the significant values for the selected critical value. The values of Z_{ML} greater than $CV = 3.09$ are indicated in red.

Table 4 compares the correlations detected by the two variants. The paper presented by [17] did not consider the pre-processing of invalid sequences for the random excursions and random excursions variant tests. For this reason, the correlations detected with the MIRT-1 method in [20] and the new MIRT-2 variant proposed in this article will be compared.

The correlation between the CUSUM (b) and non-overlapping template tests was not detected by the new MIRT-2 method, indicating that the new variant for this case loses efficacy concerning MIRT-1. This may be due to substituting the shrinkage estimator for the ML. However, according to what was expressed in [20], this correlation is quite small, so its detection may be due to an error or not easy to detect.

Table 4. Comparison of the correlations detected by the MIRT-1 and MIRT-2 methods.

	MIRT-1	**MIRT-2**
App. Ent.	≈Serial 1	≈Serial 1
CUSUM (f)	≈CUSUM (b) ≈Frequency	≈CUSUM(b) ≈Frequency
CUSUM (b)	≈Frequency ≈Non-Overlapping	≈Frequency
Long. Run	≈Overlapping	≈Overlapping
Random Ex.	≈Random Ex. Variant	≈Random Ex. Variant
Serial 1	≈Serial 2	≈Serial 2

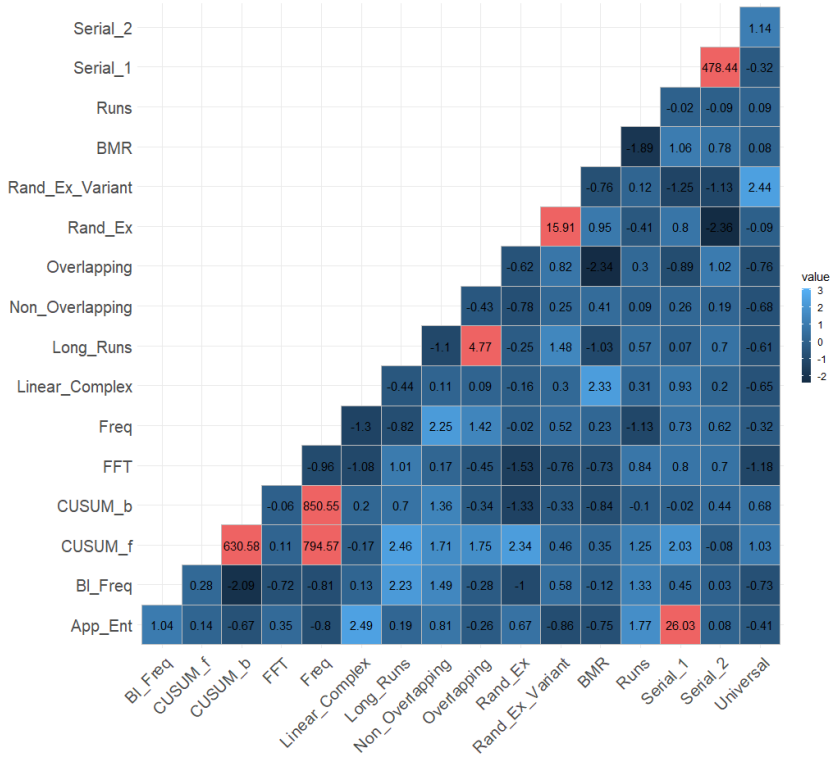

Figure 7. Mutual information matrix with the values of Z_{ML} between the pairs of statistical tests for randomness.

4.2.3. Design of Experiment 3

To verify the reduction in complexity of the new variant of the method concerning MIRT-1, the experiments presented in [17] were repeated. The methods were implemented in *R* using the entropy library in a computer running Windows 10 (64-bit) operating system, Intel Core i7-3770 3.40 GHz CPU, and 32 GB RAM. The MI with the two versions, MIRT-1 and MIRT-2, was calculated, and the times were compared. This experiment was repeated ten times. Results are shown in Table 5.

Table 5. Execution time in seconds of the methods for ten experiments.

Execution	MIRT-1	MIRT-2	Quotient
1	6249.36	0.33	18,937.45
2	6230.96	0.29	21,486.07
3	6250.03	0.3	20,833.43
4	6238.48	0.31	20,124.13
5	6240.17	0.41	15,219.93
6	6234.23	0.31	20,110.42
7	6236.86	0.29	21,506.41
8	6244.91	0.30	20,816.37
9	6241.02	0.30	20,803.40
10	6246.43	0.42	14,872.45
μ	6241.245	0.326	19,144.92
σ	40.713	0.002	17,415.26

As seen in Table 5, the times were reduced on average 19,444-fold approximately, where $19{,}444 > q = 10{,}000$. It can be seen that the reduction in time observed in practice is even greater than the expected theoretical reduction.

The times for detection of correlations between the 17 tests of the NIST battery with sequences of p-values of length 10,000 decreased from approximately 1.7 h to less than 1 s (see Figure 8).

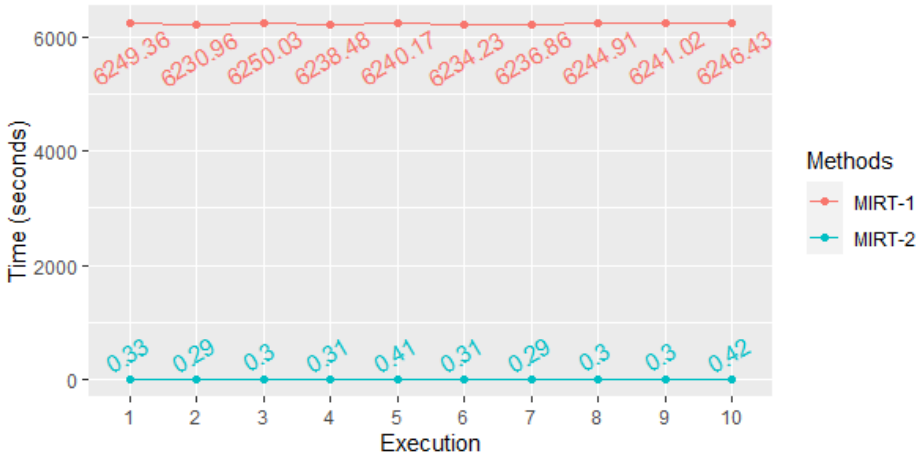

Figure 8. Representation of the execution time of the methods for ten experiments.

4.3. Analysis of the Consistency and Stability of the MIRT-2 Method

In this section, a detailed analysis of the consistency of the method MIRT-2 and MIRT-1 and the stability of MIRT-2 will be carried out.

4.3.1. Consistency

To analyze the consistency of the estimators $\widehat{MI}_{ML}$ and $\widehat{MI}_{SH}$ of mutual information, a Bland–Altman test was performed. The Bland–Altman test is a statistical method used to assess the agreement between two measurement methods. In this case, the experiment aimed to check the consistency between the estimators $\widehat{MI}_{ML}$ and $\widehat{MI}_{SH}$.

To perform the Bland–Altman test, the following steps were taken:

1. Generate 10,000 samples of independent and identically distributed random variables $U(0,1)$.
2. Calculate the mutual information using both $\widehat{MI}_{ML}$ and $\widehat{MI}_{SH}$ estimators for each sample.
3. Calculate the mean and difference between the two methods for each sample. To check the assumptions of normality of the differences, a test for normal distribution, such as the Shapiro–Wilk or Kolmogorov–Smirnov test, can be conducted for the hypothesis that the distribution of the observations in the sample is normal (Figure 9)) (if $p < 0.05$, then reject normality).
4. Calculate the mean difference and the LOAs (limits of agreement) (mean difference ± 1.96 times the standard deviation of the differences.)
5. Interpret the results. If the mean difference is close to zero and the limits of agreement are narrow, then the two methods are considered to be in good agreement.

If the mean difference is close to zero and the LOA is narrow, it suggests that the two methods are consistent and can be used interchangeably. On the other hand, if the mean difference is far from zero and/or the LOA is wide, it suggests that the two methods are inconsistent and cannot be used interchangeably.

In this case, the test was performed on 10,000 comparisons of the two estimators. The main measures are summarized in Table 6, resulting in a bias of 0.001768216 and a standard deviation of bias of 0.0002812867. The upper and lower LOAs were calculated to be 0.002319538 and 0.001216894, respectively.

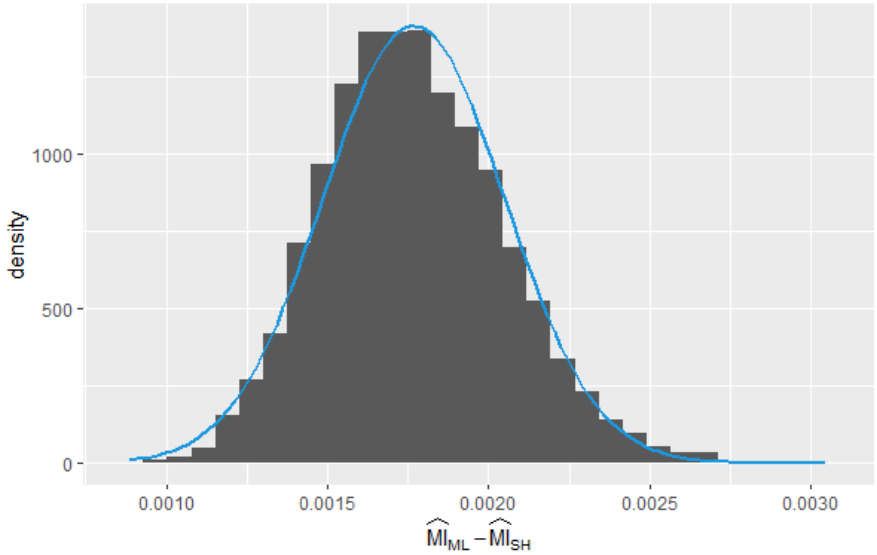

Figure 9. Distribution of the differences in the estimators $\widehat{MI}_{ML}$ and $\widehat{MI}_{SH}$.

Table 6. Bland–Altman test measurements.

Bias	0.001768216
The standard deviation of bias	0.0002812867
Upper LOA	0.002319538
Lower LOA	0.001216894
Mean of differences/means	202.4483

The bias is within the LOA, indicating that there is no significant difference between the two measures. However, the bias is not zero, indicating that there is some systematic difference between the two measures. The upper and lower LOAs are relatively narrow (Figure 10), indicating good agreement between the two measures.

Overall, the results of the Bland–Altman test suggest that the two measures being compared are in good agreement, with a small systematic difference between them.

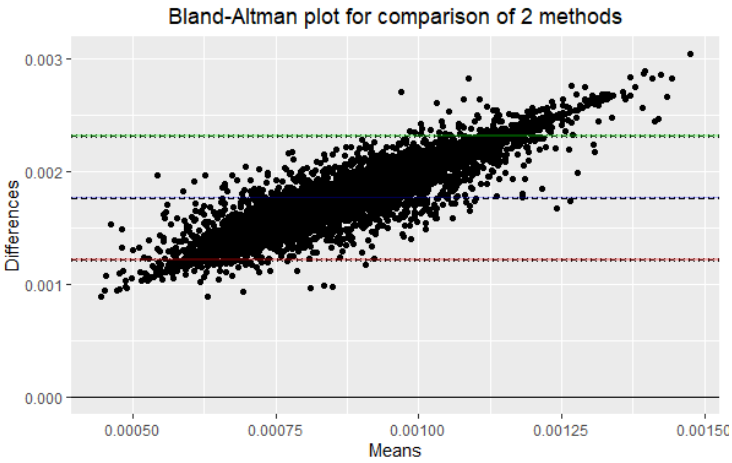

Figure 10. Bland–Altman plot for comparison of the estimators $\hat{M}I_{ML}$ and $\hat{M}I_{SH}$.

4.3.2. Stability

To assess the stability of this estimator, a stability test was carried out using a bootstrap test with 10,000 repetitions.

The hypothesis being tested is that the $\widehat{MI}_{ML}$ estimator is stable, producing consistent results when applied to different data samples. The null hypothesis is that the $\widehat{MI}_{ML}$ estimator is unstable, producing inconsistent results when applied to different data samples.

To perform the bootstrap test, multiple data samples were randomly selected with replacements from the original dataset. The $\widehat{MI}_{ML}$ estimator was applied to each sample, and the resulting mutual information values were recorded. This process was repeated 10,000 times to generate a distribution of mutual information values.

This was carried out for each dataset of each of the NIST statistical tests. The results showed that of the seven correlations detected with the original data, after carrying out the bootstrap test, six correlations were maintained (Figure 11). It can be concluded that, in general, the MIRT-2 method presents good stability. It would be necessary to study in depth the reason why the correlation between the overlapping and longest run tests did not remain stable.

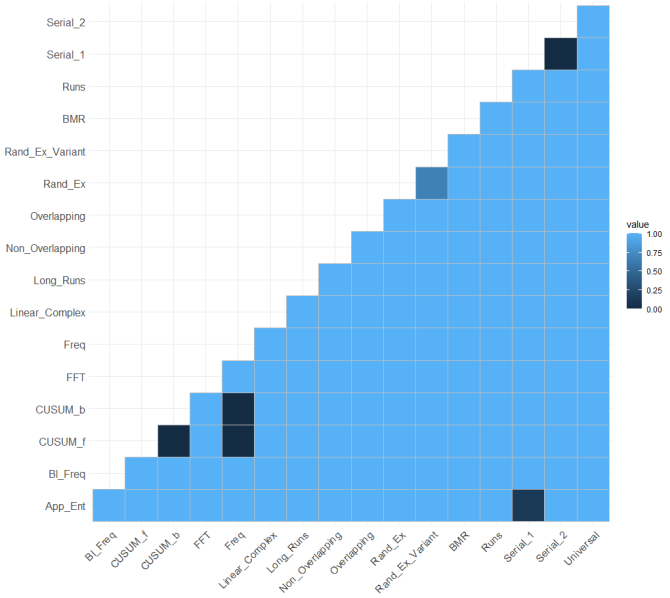

Figure 11. Matrix to illustrate the stability of the MI estimator $\widehat{MI}_{ML}$.

5. Conclusions

In this work, we reduced the complexity of the MIRT-1 method proposed by [17] q-fold by modifying the selection criteria for significant values. The complexity reduction was estimated theoretically and confirmed through experimentation. In addition, it was concluded that this modification does not significantly affect the method's effectiveness in detecting correlations. Therefore, it was proposed for this modification to be implemented in future applications of the MIRT-1 method to enhance its efficiency. As directions for future work, it is recommended to apply the MIRT-2 method to analyze the correlation in other batteries of statistical tests, such as the batteries analyzed in [18,19], with the MIRT-1 method. Some of the batteries that can be studied are ENT, FIPS 140-2, DieHarder, TufTests, and TestU01. On the other hand, it is proposed to continue reducing the complexity of the MIRT-2 method by reducing the value of n. Although it is important to note that reducing n may increase the mean square error of the $\widehat{MI}_{ML}$ estimator used in MIRT-2, this causes a decrease in the method's effectiveness. The challenge is to reduce n and complexity without losing effectiveness.

Author Contributions: Conceptualization, J.A.K.-A. and C.M.L.-P.; methodology, C.M.L.-P.; formal analysis, R.S.-L. and J.A.K.-A.; investigation, J.A.K.-A., R.S.-L. and G.S.-G.; writing—original draft, J.A.K.-A.; R.S.-L. and G.S.-G.; writing—review and editing, O.R. and G.S.-G.; project administration, O.R. and G.S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Skliar, O.; Monge, R.E.; Medina, V.; Gapper, S.; Oviedo, G. A Hybrid Random Number Generator(HRNG). *Rev. De Matemática Teoría Apl.* **2011**, *18*, 265. [CrossRef]
2. Turan, M.S.; Doğanaksoy, A.; Boztaş, S. On Independence and Sensitivity of Statistical Randomness Tests. In Proceedings of the Sequences and Their Applications-SETA 2008, Lexington, KY, USA, 14–18 September 2008; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5203, pp. 18–29.
3. Koçak, O. A Unified Evaluation of Statistical Randomness Tests and Experimental Analysis of Their Relations. Ph.D. Thesis, Middle East Technical University, Üniversiteler Mahallesi, Dumlupınar Bulvarı No:1 06800, Ankara, Turkey, 2016.
4. Garcia, F.D.; de Koning Gans, G.; Muijrers, R.; van Rossum, P.; Verdult, R.; Schreur, R.W.; Jacobs, B. Dismantling MIFARE Classic. In Proceedings of the Computer Security—ESORICS 2008, Torremolinos, Spain, 6–8 October 2008; Lecture Notes in Computer Science; Jajodia, S., Lopez, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 97–114. [CrossRef]
5. Kasper, T.; Silbermann, M.; Paar, C. All You Can Eat or Breaking a Real-World Contactless Payment System. In Proceedings of the International Conference on Financial Cryptography and Data Security, Tenerife, Spain, 25–28 January 2010; Sion, R., Ed.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6052, pp. 343–350. [CrossRef]
6. Bernstein, D.J.; Chang, Y.A.; Cheng, C.M.; Chou, L.P.; Heninger, N.; Lange, T.; van Someren, N. Factoring RSA Keys from Certified Smart Cards: Coppersmith in the Wild. In Proceedings of the Advances in Cryptology—ASIACRYPT, Bengaluru, India, 1–5 December 2013; Lecture Notes in Computer Science; Sako, K., Sarkar, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 341–360. [CrossRef]
7. Bundesamt für Sicherheit in der Informationstechnik. *Certification Report BSI-DSZ-CC-0212-2004 for Renesas AE45C1 (HD65145C1) Smartcard Integrated Circuit Version 01*; Technical Report; Federal Office for Information Security: Bonn, Germany, 2004.
8. Dunn, W.L.; Shultis, J.K. Pseudorandom Number Generators. In *Exploring Monte Carlo Methods*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 47–68.
9. Rukhin, A. Statistical Testing of Randomness: New and Old Procedures. In *Randomness through Computation*; World Scientific: Singapore, 2011.
10. Hernandez-Castro, J.; Barrero, D.F. Evolutionary Generation and Degeneration of Randomness to Assess the Indepedence of the Ent Test Battery. In Proceedings of the 2017 IEEE Congress on Evolutionary Computation (CEC), Donostia, Spain, 5–8 June 2017; pp. 1420–1427. [CrossRef]
11. Doğanaksoy, A.; Ege, B.; Muş, K. Extended Results for Independence and Sensitivity of NIST Randomness Tests. In Proceedings of the Information Security and Cryptography Conference, Istanbul, Turkey, 25–27 December 2008.
12. Doğanaksoy, A.; Sulak, F.; Uğuz, M.; Şeker, O.; Akcengiz, Z. Mutual Correlation of NIST Statistical Randomness Tests and Comparison of Their Sensitivities on Transformed Sequences. *Turk. J. Elec. Eng. Comp. Sci.* **2017**, *25*, 655–665. [CrossRef]
13. Hurley-Smith, D.; Hernandez-Castro, J. Great Expectations: A Critique of Current Approaches to Random Number Generation Testing & Certification. In Proceedings of the Security Standardisation Research, Darmstadt, Germany, 26–27 November 2018; Lecture Notes in Computer Science; Cremers, C., Lehmann, A., Eds.; Springer: Cham, Switzerland, 2018; Volume 11322, pp. 143–163. [CrossRef]
14. Burciu, P.; Simion, E. A systematic approach of NIST statistical tests dependencies. *J. Electr. Eng. Electron. Control. Comput. Sci.* **2019**, *5*, 1–6.
15. Sulak, F.; Uğuz, M.; Koçak, O.; Doğanaksoy, A. On the Independence of Statistical Randomness Tests Included in the NIST Test Suite. *Turk. J. Elec. Eng. Comp. Sci.* **2017**, *25*, 3673–3683. [CrossRef]
16. Fan, L.; Chen, H.; Gao, S. A General Method to Evaluate the Correlation of Randomness Tests. In Proceedings of the International Workshop on Information Security Applications, Jeju Island, Republic of Korea, 19–21 August 2013; pp. 52–62.
17. Karell-Albo, J.A.; Legón-Pérez, C.M.; Madarro-Capó, E.J.; Rojas, O.; Sosa-Gómez, G. Measuring independence between statistical randomness tests by mutual information. *Entropy* **2020**, *22*, 741. [CrossRef] [PubMed]
18. Luengo, E.A.; Cerna, M.B.L.; Villalba, L.J.G.; Hernandez-Castro, J. A New Approach to Analyze the Independence of Statistical Tests of Randomness. *Appl. Math. Comput.* **2022**, *426*, 127116. [CrossRef]
19. Luengo, E.A.; Cerna, M.B.L.; Villalba, L.J.G.; Hernandez-Castro, J.; Hurley-Smith, D. Critical Analysis of Hypothesis Tests in Federal Information Processing Standard (140-2). *Entropy Int. Interdiscip. J. Entropy Inf. Stud.* **2022**, *24*, 613.
20. Cerna, M.B.L. Nuevas Técnicas Computacionales Para La Estimación de La Independencia de Los Tests de Aleatoriedad. 2021. Available online: https://docta.ucm.es/entities/publication/9c972153-b581-456f-b2a3-3bb7c3c256c9 (accessed on 1 October 2023)
21. Thomas, J.A.; Cover, T. *Elements of Information Theory*; John Wiley & Sons, Inc.: New York, NY, USA, 1991; Volume 6, pp. 187–202.
22. Darbellay, G.A. An Estimator of the Mutual Information Based on a Criterion for Conditional Independence. *Comput. Stat. Data Anal.* **1999**, *32*, 1–17. [CrossRef]
23. Joe, H. Relative Entropy Measures of Multivariate Dependence. *J. Am. Stat. Assoc.* **1989**, *84*, 157–164. [CrossRef]
24. Renyi, A. On the Foundations of Information Theory. *Rev. Int. Stat. Inst.* **1965**, *33*, 1–14. [CrossRef]
25. Michalowicz, J.V.; Nichols, J.M.; Bucholtz, F. *Handbook of Differential Entropy*; CRC: Boca Raton, FL, USA, 2013.
26. Fraser, A.M.; Swinney, H.L. Independent Coordinates for Strange Attractors from Mutual Information. *Phys. Rev. A* **1986**, *33*, 1134–1140. [CrossRef] [PubMed]

27. Darbellay, G.; Vajda, I. Estimation of the Information by an Adaptive Partitioning of the Observation Space. *IEEE Trans. Inf. Theory* **1999**, *45*, 1315–1321. [CrossRef]
28. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, USA, 1986; Volume 26.
29. Moon, Y.I.; Rajagopalan, B.; Lall, U. Estimation of Mutual Information Using Kernel Density Estimators. *Phys. Rev. E* **1995**, *52*, 2318–2321. [CrossRef] [PubMed]
30. Diks, C.; Manzan, S. Tests for Serial Independence and Linearity Based on Correlation Integrals. *Stud. Nonlinear Dyn. Econom.* **2002**, *6*. [CrossRef]
31. Paninski, L. Estimation of Entropy and Mutual Information. *Neural Comput.* **2003**, *15*, 1191–1253. [CrossRef]
32. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating Mutual Information. *Phys. Rev. E Stat. Physics Plasmas Fluids Relat. Interdiscip. Top.* **2004**, *69*, 066138. [CrossRef]
33. Daub, C.O.; Steuer, R.; Selbig, J.; Kloska, S. Estimating Mutual Information Using B-spline Functions—An Improved Similarity Measure for Analysing Gene Expression Data. *BMC Bioinform.* **2004**, *5*, 118. [CrossRef]
34. Blinnikov, S.; Moessner, R. Expansions for Nearly Gaussian Distributions. *Astron. Astrophys. Suppl. Ser.* **1998**, *130*, 193–205. [CrossRef]
35. Yavuz, Z.K.; Aydin, N.; Altay, G. Comprehensive review of association estimators for the inference of gene networks. *Turk. J. Electr. Eng. Comput. Sci.* **2016**, *24*, 695–718.
36. Contreras Rodríguez, L.; Madarro-Capó, E.J.; Legón-Pérez, C.M.; Rojas, O.; Sosa-Gómez, G.S.G. Selecting an Effective Entropy Estimator for Short Sequences of Bits and Bytes with Maximum Entropy. *Entropy* **2021**, *23*, 561. [CrossRef]
37. Papana, A.; Kugiumtzis, D. Evaluation of Mutual Information Estimators for Time Series. *Int. J. Bifurc. Chaos* **2009**, *19*, 4197–4215. [CrossRef]
38. Sulewski, P. Equal-Bin-Width Histogram versus Equal-Bin-Count Histogram. *J. Appl. Stat.* **2021**, *48*, 2092–2111. [CrossRef] [PubMed]
39. Sturges, H.A. The Choice of a Class Interval. *J. Am. Stat. Assoc.* **1926**, *21*, 65–66. [CrossRef]
40. Cochran, W.G. Some Methods for Strengthening the Common χ^2 Tests. *Biom. J. Int. Biom. Soc.* **1954**, *10*, 417–451.
41. Lane, D. Online Statistics Education: A Multimedia Course of Study. Available online: https://onlinestatbook.com/ (accessed on 13 May 2023).
42. Cencov, N.N. Estimation of an Unknown Distribution Density from Observations. *Soviet Math.* **1962**, *3*, 1559–1566.
43. Bendat, J.S.; Piersol, A.G. *Measurement and Analysis of Random Data*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 1966.
44. Larson, H.J. *Statistics: An Introduction*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1975.
45. Velleman, P. Interactive Computing for Exploratory Data Analysis i: Display Algorithms. *Proc. Stat. Comput. Sect.* **1976**, 142–147.
46. Doane, D.P. Aesthetic Frequency Classifications. *Am. Stat.* **1976**, *30*, 181–183.
47. Mosteller, F.; Tukey, J.W. *Addison-Wesley Series in Behavioral Science: Quantitative Methods, Reading, Mass*; Addison-Wesley: Boston, MA, USA, 1977.
48. Terrell, G.R.; Scott, D.W. Oversmoothed Nonparametric Density Estimates. *J. Am. Stat. Assoc.* **1985**, *80*, 209–214. [CrossRef]
49. Ishikawa, K. *Guide to Quality Control, UNIPUB/Kraus International*; White Plains: New York, NY, USA, 1986.
50. Boulle, M. Optimal Bin Number for Equal Frequency Discretizations in Supervized Learning. *Intell. Data Anal.* **2005**, *9*, 175–188. [CrossRef]
51. Zhang, Z. *Statistical Implications of Turing's Formula*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2016. [CrossRef]
52. Wilks, S.S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Ann. Math. Stat.* **1938**, *9*, 60–62. [CrossRef]
53. Menezes, A.J.; van Oorschot, P.C.; Vanstone, S.A. *Handbook of Applied Cryptography*; CRC Press: Boca Raton, FL, USA, 1997.
54. Madarro-Capó, E.J.; Legón-Pérez, C.M.; Rojas, O.; Sosa-Gómez, G. Information theory based evaluation of the RC4 stream cipher outputs. *Entropy* **2021**, *23*, 896. [CrossRef] [PubMed]