Revisiting the DFT Test in the NIST SP 800-22 Randomness Test Suite

Hiroki Okada[a] and Kazuhide Fukushima
KDDI Research, Inc., Fujimino-shi, 356-8502 Japan

Keywords: Randomness, RNG, NIST SP 800-22, Discrete Fourier Transformation.

Abstract: The National Institute of Standards and Technology (NIST) released SP 800-22, which is a test suite for evaluating pseudorandom number generators for cryptographic applications. The discrete Fourier transform (DFT) test, which is one of the tests in NIST SP 800-22, was constructed to detect some periodic features of input sequences. There was a crucial problem in the construction of the DFT test: its reference distribution of the test statistic was not derived mathematically; instead, it was numerically estimated. Thus, the DFT test was constructed under the assumption that the pseudorandom number generator (PRNG) used for the estimation generated "truly" random numbers, which is a circular reasoning. Recently, Iwasaki (Iwasaki, 2020) performed a novel analysis to theoretically derive the correct reference distribution (without numerical estimation). However, Iwasaki's analysis relied on some heuristic assumptions.

In this paper, we present theoretical evidence for one of the assumptions. Let $x_0, \cdots, x_{n-1}$ be an n-bit input sequence. Its Fourier coefficients are defined as $F_0, \ldots, F_{n-1}$. Iwasaki assumed that $\sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 = n^2/2$. We use a quantitative analysis to show that this holds when n is sufficiently large. We also verify that our analysis is sufficiently accurate with numerical experiments.

1 INTRODUCTION

Random numbers are used in many applications, such as cryptography and numerical simulations. However, it is not easy to generate "truly" random sequences. Pseudorandom number generators (PRNGs) generate sequences by iterating a recurrence relation; therefore, the sequences are produced deterministically and are not "truly" random. The binary "truly" random sequence is defined as the sequence of random variables that have a probability of exactly $\frac{1}{2}$ of being "0" or "1" and are mutually independent: We can write an n-bit "truly" random sequence as $\varepsilon_0, \ldots, \varepsilon_n \overset{\text{iid}}{\sim} \mathcal{U}(\{0,1\})$.

NIST SP 800-22 (Rukhin et al., 2010; Bassham III et al., 2010) is a well-known statistical test suite for evaluating pseudorandom number generators for cryptographic applications. This test suite consists of 15 tests, and every test is a hypothesis test, where the hypothesis is that the input sequence is truly random. If this hypothesis is not rejected in any of the tests, it is concluded that the input sequences are random.

The discrete Fourier transform (DFT) test in NIST SP 800-22 is of interest to us. This test was constructed to detect periodic features in an input sequence. It performs discrete Fourier transformation on input sequences and constructs the test statistic from the Fourier coefficients.

Kim *et al.* (Kim et al., 2003; Kim et al., 2004) reported that the DFT test in the original NIST SP 800-22 (Rukhin et al., 2010) has a crucial theoretical problem. They reported that the reference distribution of the test statistic of the DFT test was erroneously derived. Kim *et al.* numerically estimated the distribution of the test statistic with sequences generated with a more accurate PRNG and proposed a new DFT test with an estimated distribution. Hamano (Hamano, 2005) also performed an analysis on the distribution of the Fourier coefficients in the original DFT test and made the DFT test problems clearer; however, the theoretical distribution of the test statistic was not derived. In 2005, in response to these reports, NIST revised the DFT test according to the report of Kim *et al.* and published NIST SP 800-22 version 1.7. The DFT test has not been subsequently revised. Pareschi *et al.* (Pareschi et al., 2012) reviewed the DFT test included in NIST SP 800-22 version 1.7, and they reported a more accurate numerical estimation on the reference distribution of the DFT test than that given by Kim *et al.* (Kim et al., 2003; Kim et al., 2004). Okada and Umeno (Okada and Umeno, 2017) proposed another test based on discrete Fourier

[a] https://orcid.org/0000-0002-5687-620X

Okada, H. and Fukushima, K.
Revisiting the DFT Test in the NIST SP 800-22 Randomness Test Suite.
DOI: 10.5220/0011626300003405
In *Proceedings of the 9th International Conference on Information Systems Security and Privacy (ICISSP 2023)*, pages 366-372
ISBN: 978-989-758-624-8; ISSN: 2184-4356

transformation that can avoid the problem, but they failed to theoretically derive the reference distribution of the original test statistic.

Iwasaki (Iwasaki, 2020) finally solved this long-standing open problem of the DFT test with a novel analysis on the joint probability density function of the (square of the absolute value of) Fourier coefficients. However, Iwasaki's analysis relied on some heuristic assumptions.

In the following subsections, we describe the details of the procedure of the DFT test (Sect. 1.1) and the problem of the DFT test (Sect. 1.2).

Then, we clarify our contribution in Sect. 1.3.

1.1 The DFT Test

We describe the details of the procedure of the original DFT test ($\mathrm{DFTT}_{\mathrm{original}}$) from 2001 (Rukhin et al., 2010), which was released before the revision in 2005 (Bassham III et al., 2010). The focus of this test is the peak heights in the discrete Fourier transformation of the input sequence. The purpose of this test is to check whether the input sequence periodic features indicate a deviation from the assumption of randomness.

1. Throughout this paper, let n be an even integer. The input sequence is an n-bit sequence $\varepsilon_0, \cdots, \varepsilon_{n-1} \in \{0,1\}$. The null hypothesis of this test is that
$$\varepsilon_0, \cdots, \varepsilon_{n-1} \overset{\text{iid}}{\sim} \mathcal{U}(\{0,1\}). \tag{1}$$
2. Convert the input sequence to $x_0, \cdots, x_{n-1}$, where
$$x_i = 2\varepsilon_i - 1 \ \ (i \in \{0, \ldots, n-1\}).$$
3. Apply a discrete Fourier transform (DFT) to $x_0, \cdots, x_{n-1}$ to produce Fourier coefficients $\{F_j\}_{j=0}^{n-1}$. The Fourier coefficient F_j and its real and imaginary parts $c_j(X)$ and $s_j(X)$, respectively, are defined as follows:
$$F_j := \sum_{k=0}^{n-1} x_k \exp\left(i\frac{2\pi kj}{n}\right). \tag{2}$$
4. Compute $\{|F_j|\}_{j=0}^{\frac{n}{2}-1}$. Note that $\{|F_j|\}_{j=\frac{n}{2}}^{n-1}$ are not of concern because $|F_j| = |\overline{F_{n-j}}|$ holds.
5. Set a threshold value $T_{0.95} = \sqrt{3n}$ such that 95% of $\{|F_j|\}_{j=0}^{\frac{n}{2}-1}$ are $< T_{0.95}$, assuming that Eq. (1) holds.
According to NIST SP800-22, $\frac{2}{n}|F_j|^2$ is considered to follow χ_2^2, and $T_{0.95}$ is defined by the following equation.
$$\begin{aligned} P(|F_j| < T_{0.95}) &= P\left(\frac{2}{n}|F_j|^2 < \frac{2}{n}T_{0.95}^2\right) \\ &= \int_0^{\frac{2}{n}T_{0.95}^2} \frac{1}{2}e^{-\frac{y}{2}}dy = 1 - e^{-\frac{T_{0.95}^2}{n}} \\ &:= 0.95 \\ \therefore T_{0.95} &= \sqrt{-n\ln(0.05)} \simeq \sqrt{3n} \end{aligned}$$
As several researchers (Kim et al., 2004; Hamano, 2005) reported, it is obvious that $T_{0.95}$ should be set as $T_{0.95} := \sqrt{-n\ln(0.05)}$ without approximation ($T_{0.95} := \sqrt{3n}$). Thus, $T_{0.95} := \sqrt{-n\ln(0.05)}$ in the revised version of the DFT test (Bassham III et al., 2010).
6. Count
$$N_1 = \#\left\{|F_j| \;\middle|\; |F_j| < T_{0.95}, 0 \leq j \leq \frac{n}{2} - 1\right\}.$$
If $\{|F_j|\}_{j=0}^{\frac{n}{2}-1}$ are mutually independent, then under the assumption of Eq. (1), N_1 can be considered to follow $\mathcal{B}(\frac{n}{2}, 0.95)$, where $\mathcal{B}$ is the binomial distribution.
Since $\mathcal{B}(n,p)$ can be approximated as the normal distribution $\mathcal{N}(np, np(1-p))$ when n is sufficiently large, we can approximate
$$N_1 \sim \mathcal{N}\left(0.95\frac{n}{2}, (0.95)(0.05)\frac{n}{2}\right)$$
under the assumption of Eq. (1).
7. Compute a test statistic
$$d = \frac{N_1 - 0.95\frac{n}{2}}{\sqrt{(0.95)(0.05)\frac{n}{2}}}.$$
The test statistic d follows $\mathcal{N}(0,1)$ when n is sufficiently large, under the assumption of Eq. (1).
8. Compute the *P-value*; $p = \operatorname{erfc}\left(\frac{|d|}{\sqrt{2}}\right)$.
If $p < \alpha$, where α is the significance level of the DFT test, then it is concluded that the sequence is not random. NIST recommends $\alpha = 0.01$ (Bassham III et al., 2010). If $p \geq \alpha$, conclude that the sequence is random.

Perform steps 1 to 7 for m sample sequences and compute m *P-value*s $\{p_1, p_2, \ldots, p_m\}$. Then, we perform second-level tests I and II to test the proportion of sequences passing the tests and the uniformity of the distribution of the *P-value*s $\{p_1, p_2, \ldots, p_m\}$. See (Bassham III et al., 2010) for the details.

1.2 The Problem of the DFT Test

Kim *et al.* (Kim et al., 2004) and Hamano (Hamano, 2005) reported that

N_1 does not follow $\mathcal{N}\left(0.95\frac{n}{2}, (0.95)(0.05)\frac{n}{2}\right)$,

and as a consequence, the test statistic $d := \frac{N_1 - 0.95\frac{n}{2}}{\sqrt{(0.95)(0.05)\frac{n}{2}}}$ does not follow $\mathcal{N}(0,1)$. Furthermore, Kim *et al.* estimated that

$$N_1 \sim \mathcal{N}\left(0.95\frac{n}{2}, (0.95)(0.05)\frac{n}{4}\right) \tag{3}$$

and $d := \frac{N_1 - 0.95\frac{n}{2}}{\sqrt{(0.95)(0.05)\frac{n}{4}}} \sim \mathcal{N}(0,1)$ using a secure hash generator (G-SHA1) (Bassham III et al., 2010) as a PRNG. According to this result, $\mathrm{DFTT}_{\mathrm{original}}$ was revised in (Bassham III et al., 2010). The present DFT test, denoted as $\mathrm{DFTT}_{\mathrm{present}}$, has not been revised.

Furthermore, Pareschi *et al.* reported that the numerical estimation is *not* sufficiently accurate; they numerically estimated that

$$N_1 \sim \mathcal{N}\left(0.95\frac{n}{2}, (0.95)(0.05)\frac{n}{3.8}\right), \tag{4}$$

and $d := \frac{N_1 - 0.95\frac{n}{2}}{\sqrt{(0.95)(0.05)\frac{n}{3.8}}} \sim \mathcal{N}(0,1)$. $\mathrm{DFTT}_{\mathrm{present}}$ and $\mathrm{DFTT}_{\mathrm{pareschi}}$ are constructed based on numerical estimation using PRNGs. However, the randomness of PRNGs are the target that should be evaluated with a randomness test. Thus, these tests cannot be used unless the reference distribution is mathematically derived.

The crucial problem here is that the reference distribution of N_1 (or the test statistic d), Eqs. (3) and (4), are derived by the numerical estimation with some PRNG. The DFT test is constructed under the assumption that the PRNG that is used for the estimation generates truly random numbers, which is circular reasoning.

Iwasaki (Iwasaki, 2020) finally solved this problem. He derived the reference distribution of N_1 (and d) theoretically (without the numerical estimation with some PRNG), which is given as follows

$$N_1 \sim \mathcal{N}\left(0.95\frac{n}{2}, (0.95)(0.05)\frac{n}{3.79}\right),$$

and $d := \frac{N_1 - 0.95\frac{n}{2}}{\sqrt{(0.95)(0.05)\frac{n}{3.79}}} \sim \mathcal{N}(0,1)$. This was realized by a novel analysis of the joint probability density function of the $|F_j|$s.

However, the analysis by (Iwasaki, 2020) was based on several heuristic assumptions:

- Assumption 1: The value of $\mathrm{V}(N_1)$ can be analyzed in a sufficiently correct manner even if we consider $\frac{2}{n}|F_0|^2 \sim \chi_2^2$ (χ-squared distribution with 2 degrees of freedom, see Definition 2.2) for sufficiently large n.
  Note that $\frac{2}{n}|F_0|^2 \sim \chi_1^2$ correctly.
- Assumption 2: $\sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 = n^2/2$ holds.
  Note that by Parseval's theorem (see Lemma 2.6), $\sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 = \frac{1}{2}(n^2 + |F_0|^2 - |F_{\frac{n}{2}}|^2)$, correctly.
- Assumption 3 (Iwasaki, 2020, Assumption 3.1): Let $\mathbf{y} := (|F_0|^2, \ldots, |F_{\frac{n}{2}} - 1|^2)$; then, $\mathbf{y}$ uniformly distributes over the set

  $$\{\mathbf{y} \in \mathbb{R}^{\frac{n}{2}-1} \mid y_i \geq 0, |\mathbf{y}|^2 = n^2/2\}.$$

  This assumption is used on the premise that Assumption 2 holds.

1.3 Our Contribution

In this paper, we show that Assumption 2 above holds when n is sufficiently large. We rephrase Assumption 2 as follows

- Assumption 2': $\lim_{n\to\infty} \sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 - \frac{n^2}{2} = 0$,

and we give the rigorous proof of Assumption 2'.

As previously mentioned, we have

$$\sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 = \frac{1}{2}(n^2 + |F_0|^2 - |F_{\frac{n}{2}}|^2), \tag{5}$$

by Parseval's theorem. We analyze the distribution of the term $z := \frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2)$ and show that it follows $\mathcal{N}(0,4)$ when n is sufficiently large. Specifically, we analyze the characteristic function of z, denoted by $\phi_z(t)$, which satisfies that $\lim_{n\to\infty} \phi_z(t) = \exp(-8t^2)$ and coincides with the characteristic function of $\mathcal{N}(0,4)$. Furthermore, we perform an experiment and confirm that the empirical distribution of z is close to $\mathcal{N}(0,4)$.

By the definition of z, we can rewrite Eq. (5) as

$$\sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 = \frac{n^2}{2}(1 + \frac{1}{n} \cdot z).$$

As we prove that $z \sim \mathcal{N}(0,4)$, we have $z = O(1)$ with overwhelming probability. Thus, we have $\sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 = \frac{n^2}{2}(1 + O(\frac{1}{n}))$, and we conclude that $\lim_{n\to\infty} \sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 - \frac{n^2}{2} = 0$, which proves Assumption 2'.

2 PRELIMINARIES

Vectors are in column form and are written using bold lowercase letters, e.g., $\mathbf{x}$. The i-th component of $\mathbf{x}$ will be denoted by x_i. For any $s \in \mathbb{N}$, the set of the first s nonnegative integers is denoted $[s] = \{0, 1, \ldots, s-1\}$.

For any set X, $\mathcal{U}(X)$ denotes the uniform distribution over the set X. For a random variable (or distribution) X, we denote the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) by $f_X(\cdot)$ and $F_X(\cdot)$, respectively. We say that X and Y are (statistically) independent if $f_{XY}(x,y) = f_X(x)f_Y(y)$, where $f_{XY}(x,y)$ denotes the joint probability function of X and Y. We denote $X_1,\cdots,X_n \overset{\text{iid}}{\sim} \mathcal{D}$ if the random variables $X_1,\cdots,X_n$ are independent and identically distributed (i.i.d.) according to the distribution $\mathcal{D}$. We denote *the normal distribution* with mean μ and variance σ^2 by $\mathcal{N}(\mu,\sigma^2)$.

For clarity, we describe the definitions of the characteristic function, and χ-squared distribution.

Definition 2.1 (Characteristic Function). *If X is a random variable over $\mathbb{R}$, then for $0 < t \in \mathbb{R}$, the characteristic function of X is defined as*

$$\varphi_X(t) = \mathrm{E}\left[\exp(itX)\right].$$

Definition 2.2 (χ-squared Distribution χ_p^2). *Let p be a degree of freedom. Let $X_1,\cdots,X_p \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, then the χ-squared distribution χ_p^2 is defined as $\sum_{i=1}^{p} X_i^2$. The p.d.f. and c.d.f. of χ_p^2 are*

$$f_{\chi_p^2}(x) = \frac{1}{2^{p/2}\Gamma(p/2)} x^{p/2-1}\exp(-x/2),$$
$$F_{\chi_p^2}(x) = \frac{\gamma(\frac{p}{2},\frac{x}{2})}{\Gamma(p/2)},$$

respectively. Specifically,

$$f_{\chi_2^2}(x) = \frac{1}{2}\exp(-\frac{x}{2}), F_{\chi_2^2}(x) = \frac{\gamma(\frac{p}{2},\frac{x}{2})}{\Gamma(p/2)}$$

For clarity, we describe Parseval's theorem and give a proof to it in Lemma 2.6. For this proof, we use the n-th root of unity and some useful characteristics of it.

Definition 2.3 (n-th Root of Unity). *For any $n \in \mathbb{N}$ and $k, j \in \mathbb{Z}$, we define*

$$\omega_j := \exp\left(i\frac{2\pi j}{n}\right), \text{ and}$$
$$\omega_{k,j} := \omega_j^k = \exp\left(i\frac{2\pi k j}{n}\right),$$

both of which are an n-th root of unity.

Note that $\omega_{k,j}\overline{\omega}_{k,j} = 1$, and thus, $\overline{\omega}_{k,j} = \omega_{k,j}^{-1} = \omega_{k,-j} = \omega_{-k,j}$ holds for any $k, j \in \mathbb{Z}$.

Fact 2.4. *For any $n \in \mathbb{N}$ and $j \in \mathbb{Z}$,*

$$\sum_{k=0}^{n-1}\omega_{k,j} = \sum_{k=0}^{n-1}\exp\left(i\frac{2\pi k j}{n}\right) = \begin{cases} n & (j=0), \\ 0 & (j \neq 0). \end{cases}$$

Proof. Trivially, $\sum_{k=0}^{n-1}\omega_{k,0} = n$ and $\omega_j^n = \exp(2\pi i \cdot jn) = 1$ hold. Thus, we have

$$\begin{aligned} 0 &= \omega_j^n - 1 \\ &= (\omega_j - 1)(\omega_j^{n-1} + \cdots + \omega_j + 1) \\ &= (\omega_j - 1)\sum_{k=0}^{n-1}\omega_j^k \end{aligned}$$

Hence, when $j \neq 0$, i.e., when $\omega_j \neq 1$, we have $\sum_{k=0}^{n-1}\omega_j^k = \sum_{k=0}^{n-1}\omega_{k,j} = 0$. □

Corollary 2.5. $\forall j_1, j_2 \in \mathbb{Z}$ *such that* $j_1 + j_2 \neq 0$, $\sum_{k=0}^{n-1}\omega_{k,j_1}\omega_{k,j_2} = 0$.

Proof. Since $\omega_{k,j_1}\omega_{k,j_2} = \omega_{k,j_1+j_2}$, the corollary follows from Fact 2.4. □

Finally, we state Parseval's theorem and give a proof to it:

Lemma 2.6 (Parseval's Theorem). *For any $n \in \mathbb{N}$ and $j \in \mathbb{Z}$, let $x_0,\cdots,x_{n-1}$ be an n-bit input sequence, and let its Fourier coefficients be defined as $F_0,\ldots,F_{n-1}$, i.e., $F_j := \sum_{k=0}^{n-1}x_k\omega_{k,j}$ for $j \in [n]$. Then, we have the following:*

$$\sum_{j=0}^{n-1}|F_j|^2 = n\sum_{k=0}^{n-1}x_k^2.$$

When n is even:

$$\sum_{j=0}^{\frac{n}{2}-1}|F_j|^2 = \frac{1}{2}(n\sum_{k=0}^{n-1}x_k^2 + |F_0|^2 - |F_{\frac{n}{2}}|^2).$$

When n is odd:

$$\sum_{j=0}^{\frac{n}{2}-1}|F_j|^2 = \frac{1}{2}(n\sum_{k=0}^{n-1}x_k^2 + |F_0|^2).$$

Proof. For any $n \in \mathbb{N}$ and $j \in \mathbb{Z}$, we have

$$\begin{aligned} \sum_{j=0}^{n-1}|F_j|^2 &= \sum_{j=0}^{n-1}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}x_{k_1}x_{k_2}\omega_{k_1,j}\overline{\omega}_{k_2,j} \\ &= \sum_{j=0}^{n-1}\left(\sum_{k=0}^{n-1}x_k^2 + \sum_{k_1\neq k_2}x_{k_1}x_{k_2}\omega_{k_1-k_2,j}\right) \\ &= \sum_{j=0}^{n-1}\sum_{k=0}^{n-1}x_k^2 + \sum_{k_1\neq k_2}x_{k_1}x_{k_2}\sum_{j=0}^{n-1}\omega_{k_1-k_2,j} \\ &= n\sum_{k=0}^{n-1}x_k^2 \quad (\because \text{Corollary 2.5}) \end{aligned}$$

Thus, when n is even, we have

$$
\begin{aligned}
n\sum_{k=0}^{n-1} x_k^2 &= \sum_{j=0}^{n-1} |F_j|^2 \\
&= \sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 + \sum_{j=\frac{n}{2}}^{n-1} |F_j|^2 \\
&= |F_0|^2 + |F_{\frac{n}{2}}|^2 + \sum_{j=1}^{\frac{n}{2}-1} |F_j|^2 + \sum_{j=\frac{n}{2}+1}^{n-1} |F_j|^2
\end{aligned}
$$

$$
\therefore \sum_{j=1}^{\frac{n}{2}-1} |F_j|^2 = \frac{1}{2}(n\sum_{k=0}^{n-1} x_k^2 - |F_0|^2 - |F_{\frac{n}{2}}|^2)
$$

Similarly, when n is odd, we have

$$
\begin{aligned}
n\sum_{k=0}^{n-1} x_k^2 &= \sum_{j=0}^{n-1} |F_j|^2 \\
&= \sum_{j=0}^{\frac{n}{2}-1} |F_j|^2 + \sum_{j=\frac{n}{2}+1}^{n-1} |F_j|^2 \\
&= |F_0|^2 + \sum_{j=1}^{\frac{n}{2}-1} |F_j|^2 + \sum_{j=\frac{n}{2}+1}^{n-1} |F_j|^2
\end{aligned}
$$

$$
\therefore \sum_{j=1}^{\frac{n}{2}-1} |F_j|^2 = \frac{1}{2}(n\sum_{k=0}^{n-1} x_k^2 - |F_0|^2) \qquad \square
$$

3 OUR ANALYSIS

Our goal of this section is to give a proof of Assumption 2' stated in Sect. 1.3, which can be obtained as Corollary 3.4. For the proof, we show in Theorem 3.3 that $\frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2) \sim \mathcal{N}(0,4)$ when n is sufficiently large.

3.1 Building Blocks

We show some useful facts related to the random variable $x \sim \mathcal{U}(\{-1,1\})$. These facts are used for the proof of Theorem 3.3.

We first show that for $x_1, x_2, x_3 \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$, $X := x_1x_2$ and $Y := x_1x_3$ are mutually independent. For general independent random variables x_1, x_2, x_3, $X := x_1x_2$ and $Y := x_1x_3$ are not necessarily mutually independent since both are composed of the common random variable x_1. Interestingly, X and Y are mutually independent when $x_1, x_2, x_3 \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$.

Fact 3.1. *Let* $x_1, x_2, x_3 \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$ *and* $X = x_1x_2, Y = x_1x_3$, *then* $X, Y \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$.

Proof. We can show that $f_{XY}(x,y) = f_X(x)f_Y(y)$, i.e., X and Y are mutually independent, as follows:

$$
f_{XY}(x,y) = P[x_1x_2 = x, x_1x_3 = y]
$$

$$
= \begin{cases}
\frac{1}{4} & (x,y) = (1,1) \\
 & (\because (x_1,x_2,x_3) = (1,1,1), (-1,-1,-1)) \\
\frac{1}{4} & (x,y) = (1,-1) \\
 & (\because (x_1,x_2,x_3) = (1,1,-1), (-1,-1,1)) \\
\frac{1}{4} & (x,y) = (-1,1) \\
 & (\because (x_1,x_2,x_3) = (1,-1,1), (-1,1,-1)) \\
\frac{1}{4} & (x,y) = (-1,-1) \\
 & (\because (x_1,x_2,x_3) = (1,-1,-1), (-1,1,1))
\end{cases}
$$

$$
f_X(x) = P[x_1x_2 = x]
= \begin{cases}
\frac{1}{2} & x = 1 \ (\because (x_1,x_2) = (1,1), (-1,-1)) \\
\frac{1}{2} & x = -1 \ (\because (x_1,x_2) = (1,-1), (-1,1))
\end{cases}
$$

$$
f_Y(y) = P[x_2x_3 = y]
= \begin{cases}
\frac{1}{2} & y = 1 \ (\because (x_2,x_3) = (1,1), (-1,-1)) \\
\frac{1}{2} & y = -1 \ (\because (x_2,x_3) = (1,-1), (-1,1))
\end{cases}
$$

$$
f_X(x)f_Y(y) = P[x_1x_2 = x]P[x_1x_3 = y]
= \begin{cases}
\frac{1}{4} & (x,y) = (1,1) \\
\frac{1}{4} & (x,y) = (1,-1) \\
\frac{1}{4} & (x,y) = (-1,1) \\
\frac{1}{4} & (x,y) = (-1,-1)
\end{cases} \qquad \square
$$

Fact 3.2. *For* $x \leftarrow \mathcal{U}(\{-1,1\})$ *and constant* C, *we have*

$$
\mathrm{E}[\exp(iC\cdot x)] = \frac{\exp(-iC) + \exp(iC)}{2} = \cos C
$$

3.2 Proof of Assumption 2'

We analyze the distribution of $\frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2)$, and then we give a proof of Assumption 2', which was stated in Sect. 1.3.

Let us define $y_0 := \frac{2}{n}|F_0|^2$ and $y_{\frac{n}{2}} := \frac{2}{n}|F_{\frac{n}{2}}|^2$. Then, we have

$$
y_0 = \frac{2}{n}\left(\sum_{k=0}^{n-1} x_k\right)^2
$$

$$
y_{\frac{n}{2}} = \frac{2}{n}\left(\sum_{k=0}^{n-1} x_k(-1)^k\right)^2
$$

Since $x_0, \ldots, x_{n-1} \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$, we have $\mathrm{E}[x_k] = 0, \mathrm{V}(x_k) = \mathrm{E}[x_k^2] = 1$ for any $k \in [n]$. Thus, by the central limit theorem, we have

$$\left(\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}x_k\right) \overset{n\to\infty}{\sim} \mathcal{N}(0,1), \text{ and}$$

$$\frac{1}{n}|F_0|^2 = \left(\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}x_k\right)^2 \overset{n\to\infty}{\sim} \chi_1^1,$$

where $X \overset{n\to\infty}{\sim} \mathcal{D}$ means that the random variable X follows the distribution $\mathcal{D}$ when $n\to\infty$. Additionally, note that $\frac{1}{n}|F_{\frac{2}{n}}|^2 = \left(\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}x_k(-1)^k\right)^2$ also holds since $x_k(-1)^k$s for $k\in[n]$ are i.i.d according to $\mathcal{U}(\{-1,1\})$. However, it is not trivial to analyze the distribution of $y_0 - y_{\frac{n}{2}} = \frac{2}{n}(|F_0|^2|-F_{\frac{n}{2}}|^2)$ since y_0 and $y_{\frac{n}{2}}$ ($\frac{2}{n}(|F_0|^2|$ and $|F_{\frac{n}{2}}|^2$) are not necessarily mutually independent.

3.2.1 The Distribution of $\frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2)$

We analyze the asymptotic distribution of $z := y_0 - y_{\frac{n}{2}} := \frac{2}{n}(|F_0|^2|-F_{\frac{n}{2}}|^2)$ as follows:

Theorem 3.3. *Let* $x_0,\cdots,x_{n-1} \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$*,* $F_j := \sum_{k=0}^{n-1}x_k\omega_{k,j}$ *for* $j\in[n]$*, and* $z := \frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2)$*. Then, we have* $\lim_{n\to\infty}\phi_z(t) = \exp(-8t^2)$*, i.e., z follows* $\mathcal{N}(0,4)$ *when n is sufficiently large.*

Proof. By a routine calculation, we have

$$\begin{aligned}
z &:= y_0 - y_{\frac{n}{2}}\\
&= \frac{2}{n}\left(\left(\sum_{k=0}^{n-1}x_k\right)^2 - \left(\sum_{k=0}^{n-1}x_k(-1)^k\right)^2\right)\\
&= \frac{2}{n}\left(\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}x_{k_1}x_{k_2} - \sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}x_{k_1}x_{k_2}(-1)^{k_1+k_2}\right)\\
&= \frac{2}{n}\left(\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}(1-(-1)^{k_1+k_2})x_{k_1}x_{k_2}\right)\\
&= \frac{2}{n}\left(\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\delta_{k_1+k_2}x_{k_1}x_{k_2}\right),
\end{aligned}$$

where we define $\delta_k := 1-(-1)^k$, which satisfies

$$\delta_k = \begin{cases} 2 & k \text{ is odd,}\\ 0 & k \text{ is even.}\end{cases}$$

Then, we calculate the characteristic function of z as follows:

$$\begin{aligned}
&\phi_z(t)\\
&= \mathrm{E}\left[\exp\left(it\frac{2}{n}\left(\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\delta_{k_1+k_2}x_{k_1}x_{k_2}\right)\right)\right]\\
&= \mathrm{E}\left[\exp\left(it\frac{4}{n}\sum_{k_1=1}^{n-1}\sum_{k_2=0}^{k_1-1}\delta_{k_1+k_2}x_{k_1}x_{k_2}\right)\right]\\
&= \prod_{k_1=1}^{n-1}\prod_{k_2=0}^{k_1-1}\mathrm{E}\left[\exp\left(it\frac{4}{n}\delta_{k_1+k_2}x_{k_1}x_{k_2}\right)\right] (\because \text{Fact 3.1})\\
&= \prod_{k_1=1}^{n-1}\prod_{k_2=0}^{k_1-1}\cos\left(t\frac{4}{n}\delta_{k_1+k_2}\right). (\because \text{Fact 3.2})
\end{aligned}$$

By using Taylor series expansion, we obtain

$$\begin{aligned}
&\ln\phi_{y_0}(t)\\
&= \sum_{k_1=1}^{n-1}\sum_{k_2=0}^{k_1-1}\ln\cos\left(t\frac{4}{n}\delta_{k_1+k_2}\right)\\
&= \sum_{k_1=1}^{n-1}\sum_{k_2=0}^{k_1-1}\left(-\frac{8\delta_{k_1+k_2}^2t^2}{n^2} - \frac{64\delta_{k_1+k_2}^4t^4}{3n^4} + O(1/n^5)\right)\\
&= -8t^2 + O(1/n^2),
\end{aligned}$$

where we use the following fact:

$$\begin{aligned}
\sum_{k_1=1}^{n-1}\sum_{k_2=0}^{k_1-1}\delta_{k_1+k_2}^2 &= \frac{1}{2}\left(\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} - \sum_{k_1=k_2=0}^{n-1}\right)\delta_{k_1+k_2}^2\\
&= \frac{1}{2}\left(\frac{n^2}{2}\cdot 4 - 0\right)\\
&= n^2.
\end{aligned}$$

Therefore, we have

$$\lim_{n\to\infty}\phi_{y_0}(t) = \exp(-8t^2)$$

□

3.2.2 Proof of Assumption 2'

As stated in Sect. 1.3, a proof of Assumption 2' can be obtained as a corollary of Theorem 3.3

Corollary 3.4 (Proof of Assumption 2'). *Let* $x_0,\cdots,x_{n-1} \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$*,* $F_j := \sum_{k=0}^{n-1}x_k\omega_{k,j}$ *for* $j\in[n]$ *and* $z := \frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2)$*. Then, we have*

$$\lim_{n\to\infty}\sum_{j=0}^{\frac{n}{2}-1}|F_j|^2 - \frac{n^2}{2} = 0.$$

Proof. By Parseval's theorem (Lemma 2.6), we have

$$\begin{aligned}
\sum_{j=0}^{\frac{n}{2}-1}|F_j|^2 &= \frac{1}{2}(n^2 + |F_0|^2 - |F_{\frac{n}{2}}|^2)\\
&= \frac{n^2}{2}(1 + \frac{1}{n}\cdot z),
\end{aligned}$$

where $z := \frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2)$. By Theorem 3.3, $z \sim \mathcal{N}(0,4)$ holds when $n \to \infty$. Thus, we have $z = O(1)$ when $n \to \infty$, and the corollary follows. $\square$

3.2.3 Experimental Verification

We showed in Theorem 3.3 that $z := \frac{2}{n}(|F_0|^2 - |F_{\frac{n}{2}}|^2)$ distributes according to $\mathcal{N}(0,4)$ when n is sufficiently large. We now empirically verify how accurately z distributes according to $\mathcal{N}(0,4)$ when we set $n = 100000$. We generate 5000 sets of input sequences $x_0, \ldots, x_{n-1} \overset{\text{iid}}{\sim} \mathcal{U}(\{-1,1\})$ by the default PRNG in R (Comprehensive R Archive Network, 2022), and then calculate 5000 samples of z. Fig. 1 shows the empirical c.d.f. of the samples of z and the theoretical c.d.f. of $\mathcal{N}(0,4)$. We can observe that they match well, although not perfectly. It is sufficient to conclude that $z = O(1)$, which is required for the proof of Corollary 3.4.

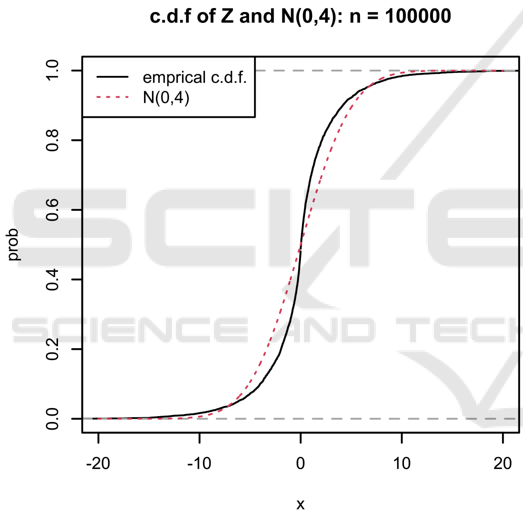

Figure 1: Experimental verification of Theorem 3.3.

4 CONCLUSION

Iwasaki (Iwasaki, 2020) proposed a novel analysis to solve the long-standing problem of the DFT test under the 3 heuristic assumptions described in Sect. 1.2. In this paper, we showed that Assumption 2, which is also required for Assumption 3, holds when n is sufficiently large. The rest of the heuristic assumptions remain unproved, and they remain future work.

REFERENCES

Bassham III, L. E., Rukhin, A. L., Soto, J., Nechvatal, J. R., Smid, M. E., Barker, E. B., Leigh, S. D., Levenson, M., Vangel, M., Banks, D. L., et al. (2010). Sp 800-22 rev. 1a. a statistical test suite for random and pseudorandom number generators for cryptographic applications.

Comprehensive R Archive Network (2022). R: The r project for statistical computing. https://www.r-project.org/.

Hamano, K. (2005). The distribution of the spectrum for the discrete fourier transform test included in sp800-22. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 88(1):67–73.

Iwasaki, A. (2020). Deriving the variance of the discrete fourier transform test using parseval's theorem. *IEEE Transactions on Information Theory*, 66(2):1164–1170.

Kim, S.-J., Umeno, K., and Hasegawa, A. (2003). On the nist statistical test suite for randomness. *TECHNICAL REPORT OF IEICE*, 103(499):21–27.

Kim, S.-J., Umeno, K., and Hasegawa, A. (2004). Corrections of the nist statistical test suite for randomness.

Okada, H. and Umeno, K. (2017). Randomness evaluation with the discrete fourier transform test based on exact analysis of the reference distribution. *IEEE Transactions on Information Forensics and Security*, 12(5):1218–1226.

Pareschi, F., Rovatti, R., and Setti, G. (2012). On statistical tests for randomness included in the nist sp800-22 test suite and based on the binomial distribution. *IEEE Transactions on Information Forensics and Security*, 7(2):491–505.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., et al. (2010). Nist special publication 800-22: a statistical test suite for the validation of random number generators and pseudo random number generators for cryptographic applications.