擬似乱数生成系の検定方法に関する調査報告書

NIST SP800-22 の離散フーリエ変換検定について

廣瀬勝一
京都大学情報学研究科

2005 年 1 月

1 まえがき

与えられた系列が良い乱数列であるかどうかを判定するための手法として，従来より，その系列の統計的性質を利用した検定法が多数提案されている．したがって，統計的手法による乱数検定法に関する文献やそれらを利用した検定ツールも多く存在するが，どの検定法が採用あるいは推奨されているかはそれぞれ異なっており，明確な判定基準はない．

このような状況において，2002 年度,「擬似乱数検証ツールの調査開発」が行われ，その調査報告書が情報処理振興事業協会から発行されている [2]．この調査報告では，まず，これまでに提案されている統計的手法に基づく乱数検定法が分類・整理され，さらに，乱数検定のための必要最小限の検定法の組が選出されている．この検定法の組はミニマムセットと呼ばれている．なお，この調査開発は，NIST Special Publication 800-22 (SP800-22)[1] と DIEHARD [9] の二つの乱数検定ツールで採用されている検定法を主な対象として行われた．

上で述べた「擬似乱数検証ツールの調査開発」では，検定法の分類・整理，ミニマムセットの導出にあたり，検定法の解析および計算機実験が行われ，それによって，幾つかの問題点が明らかにされた．この内で最も大きな二つの問題点は，NIST SP800-22 の離散フーリエ変換検定および Lempel-Ziv 圧縮検定の不具合である．

これを受けて，昨年度には，CRYPTREC プロジェクトにおいて，これら二つの検定法に関する調査が行われ，その報告書が公開されている [3, 6].

本報告書は，昨年度に引き続き，離散フーリエ変換検定について，今年度に学会等で報告された離散フーリエ変換検定に関する解析結果 [7, 8] と，それに関連して筆者が行った解析結果とをまとめたものである．

本報告書の構成は以下の通りである．まず，2 章では，NIST SP800-22 の離散フーリエ変換検定法と，それについて NIST が示した理論的根拠を述べると共に，本検定法について明らかとなっている問題点を記す．3 章では，2 章で示された離散フーリエ変換の問題点に関して行った調査結果を述べる．4 章は本報告書の結論である．

2 NIST SP800-22 の離散フーリエ変換検定

本章では，NIST SP800-22 の離散フーリエ変換検定のアルゴリズムと理論的根拠を述べると共に，本検定に関して報告されている問題点を述べる．なお，以下では簡単のため，本検定を DFT 検定と呼ぶ.

2.1 DFT 検定のアルゴリズム

入力系列 $x = (x_0, x_1, \ldots, x_{n-1})$ を 0 と 1 からなる系列とする．n は系列の長さである．以下では簡単のため，特にことわらない限り，n は偶数であると仮定する．

1. 0 と 1 からなる長さ n の入力系列 $x = (x_0, x_1, \ldots, x_{n-1})$ を，次のように -1 と 1 からなる系列 $X = (X_0, X_1, \ldots, X_{n-1})$ へ変換する．

$$X_i = 2x_i - 1 \quad (1 \le i \le n)$$

2. 系列 X を離散フーリエ変換し，複素数の列 $S = (S_0, S_1, \ldots, S_{n-1})$ を得る．

3. S' を S の最初の $n/2$ 個からなる部分列とし，$M = modulus(S') = |S'|$ を求める．すなわち，$S' = (S_0, S_1, \ldots, S_{\frac{n}{2}-1})$ に対し，$i = 0, 1, \ldots, n/2 - 1$ について，$|S_i|$ を求める．

4. $T = \sqrt{3n}$ とする．なお，これは，$|S_i| \leq T$ である確率が 0.95 となるような値である．

5. $N_0 = 0.95(n/2)$ とする．N_0 は，X が真の乱数列であるときに，$|S_0|, |S_1|, \ldots, |S_{\frac{n}{2}-1}|$ のうちで T を越えない $|S_i|$ の個数の理論値である．

6. $|S_0|, |S_1|, \ldots, |S_{\frac{n}{2}-1}|$ の中で実際に T を越えなかった $|S_i|$ の個数 N_1 を求める．

7. $d = \dfrac{N_1 - N_0}{\sqrt{(0.95)(0.05)(n/2)}}$ を求める．

8. X が真の乱数列であるとき，d は標準正規分布に従うので，この分布から p-value $= erfc(|d|/\sqrt{2})$ を計算する．

入力系列は，p-value ≥ 0.01 のとき，良い乱数列であると判定され，p-value < 0.01 のとき，良い乱数列でないと判定される．

p-value は，真の乱数生成器が，検定対象として入力された系列よりもランダムでない系列を生成する確率と解釈できる．今の場合，関数 $erfc$ は以下の通りである．

$$erfc(z) = \int_z^\infty \frac{2}{\sqrt{\pi}} \mathrm{e}^{-x^2} \mathrm{d}x$$

2.2 DFT 検定の理論的根拠

$\{-1, 1\}$ 系列 $X = (X_0, X_1, \ldots, X_{n-1})$ の離散フーリエ変換 $S = (S_0, S_1, \ldots, S_{n-1})$ は以下のように定義される．

$$S_j = \sum_{k=0}^{n-1} X_k \cos \frac{2\pi kj}{n} - \mathrm{i} \sum_{k=0}^{n-1} X_k \sin \frac{2\pi kj}{n}$$

ここで，$\mathrm{i} = \sqrt{-1}$ である．

簡単のために，

$$\begin{aligned} c_j(X) &= \sum_{k=0}^{n-1} X_k \cos \frac{2\pi kj}{n} \\ s_j(X) &= \sum_{k=0}^{n-1} X_k \sin \frac{2\pi kj}{n} \end{aligned}$$

と表記すると，

$$|S_j(X)|^2 = (c_j(X))^2 + (s_j(X))^2$$

である．ここで，X が真の乱数列であれば，$c_j(X)$，$s_j(X)$ の分布は共に，n が大きくなるにつれて，平均 $\mu = 0$，分散 $\sigma^2 = n/2$ の正規分布 $N(0, n/2)$ に近づくことが証明できる．なお，$c_j(X)$ について，平均，分散は付録 A のように計算できる．$s_j(X)$ についても同様に計算できる．

DFT 検定の理論的根拠は，以下の定理である．

定理 1 [5] 正規分布 $N(\mu, \sigma^2)$ に従う母集団から，大きさ n の標本 $z_1, z_2, \ldots, z_n$ をランダムに選び，

$$y = \frac{1}{\sigma^2}\sum_{i=1}^{n}(z_i - \mu)^2$$

を作ると，これは自由度 n の χ^2 分布に従う． □

この定理に基づいて，DFT 検定では，

$$\frac{(c_j(X))^2}{\sigma^2} + \frac{(s_j(X))^2}{\sigma^2} = \frac{|S_j(X)|^2}{\sigma^2} = \frac{2}{n}|S_j(X)|^2$$

が自由度 2 の χ^2 分布に従うことが仮定されている．この分布の確率密度関数は以下の通りであり，平均は 2，分散は 4 である．

$$T_2(y) = \begin{cases} \frac{1}{2}\mathrm{e}^{-\frac{y}{2}} & y > 0 \text{ のとき} \\ 0 & y \leq 0 \text{ のとき} \end{cases}$$

$|S_j(X)|^2/\sigma^2 = (2/n)|S_j(X)|$ の平均，分散はそれぞれ，2，$4 - 8/n$ となる（付録 B）．

DFT 検定のステップ 3 で，$(S_0, S_1, \ldots, S_{\frac{n}{2}-1})$ のみが考慮されているのは，$1 \leq i \leq n-1$ について，S_{n-i} が S_i の共役であり，$|S_{n-i}| = |S_i|$ となるためである．なお，S_0, $S_{\frac{n}{2}}$ はそれぞれ以下の通りである．

$$S_0 = \sum_{k=0}^{n-1} X_k,\ S_{\frac{n}{2}} = \sum_{k=0}^{n-1}(-1)^k X_k$$

ステップ 4 の T は，T より大きな $|S_j|$ が 5%であるように定められているので，$y = |S_j|^2/\sigma^2$ とすれば，

$$\int_{T^2/\sigma^2}^{\infty} \frac{1}{2}\mathrm{e}^{-\frac{y}{2}}\mathrm{d}y = \mathrm{e}^{-\frac{T^2}{2\sigma^2}} = 0.05$$

が成立する．したがって，

$$\begin{aligned} -\frac{T^2}{2\sigma^2} &= \ln 0.05 \\ T^2 &= -2\sigma^2 \ln 0.05 = (-\ln 0.05)n \end{aligned}$$

であり，$-\ln 0.05 = 2.9957322735\cdots$ であることから，

$$T \approx \sqrt{3\,n}$$

とされている．

ステップ 5 から 8 では，$|S_0|, |S_1|, \ldots, |S_{\frac{n}{2}-1}|$ が独立であると仮定され，N_1 が，独立な試行の回数を $n/2$，1 回の試行における生起確率を 0. 95 としたときの二項分布に従うと考えられている．このとき，N_1 の平均は $N_0 = 0.95\,n/2$，分散は $(0.95)(0.05)(n/2)$ であり，d が標準正規分布 $N(0, 1)$ に従うと考えられている．

2.3 DFT 検定の問題点

金，梅野，長谷川 [4] は，DFT 検定に関する以下の二つの問題点を指摘している．

1. **ステップ 3 で，$-\ln 0.05$ の近似値として 3 が使われているが，これらの間の誤差が無視できない程度に大きいこと．**

2. **ステップ 5 から 7 で，N_1 は生起確率 0.95 の二項分布に従うと考えられているが，計算機実験では，分散が想定される値の 1/2 に近い値となること．**

問題点 1 については，文献 [4] で，金らが，G using SHA-1 を利用して得られる長さ $n = 10^6$ の 300,000 個の擬似乱数列について，T を $\sqrt{3\,n}$ とした場合と $\sqrt{2.995732274\,n}$ とした場合のそれぞれについて N_1 の分布を調べ，両者の相違を示している．

実際，$\mathrm{e}^{-3} \approx 0.0497871$ であり，$n = 10^6$ のとき，$T = \sqrt{3\,n}$ を用いた場合の N_1 の平均値は

$$(1 - \mathrm{e}^{-3})\frac{n}{2} \approx 475106$$

になると見積もることができる．文献 [4] でもこの見積もりに合致した結果が示されている．以上のことから，この問題点は，単に近似の精度が十分ではなかったことと結論付けて良いと考えられる．

問題点 2 については，文献 [4] で，金らが，G using SHA-1 を利用して得られる擬似乱数列について，$T = \sqrt{2.995732274\,n}$ とした場合の N_1 の分布を調べ，分散が $(0.95)(0.05)(n/2)/2$ に近い値となることを確認している．このように，この問題点については，N_1 の分散値を $(0.95)(0.05)(n/2)$ とすることは明らかな誤りであることが確認されたが，実験によって得られた $(0.95)(0.05)(n/2)/2$ が理論的に正しい分散値であるかどうかについては明らかになっていない．

次章では，この問題点に関する調査結果を述べる．

3 調査結果

前章の定理 1 と DFT 検定の理論的根拠との間には，以下の不一致が見られる．

1. DFT 検定では，$c_j(X)$ と $s_j(X)$ が独立ではないこと．

2. 定理 1 が保証するのは，j を固定して，X がランダムに選択された場合に，$S_j(X)$ が自由度 2 の χ^2 分布に従うということである．一方，DFT 検定では，X を固定して，$S_0(X), S_1(X), \ldots, S_{\frac{n}{2}-1}(X)$ の分布を検査している．なお，これら $S_0(X), S_1(X), \ldots, S_{\frac{n}{2}-1}(X)$ は独立ではない．

これらの不一致が，DFT 検定で検査される d の分散の値が理論値と大きく異なる原因であると考えられる．以下ではこれらそれぞれの不一致について調査結果を述べる．

1 について　$c_j(X)$ と $s_j(X)$ の共分散は 0 となる（付録 C 参照）ので，相関係数も 0 であることが分かるが，これは $c_j(X)$ と $s_j(X)$ とが独立であることを意味するものではない．そこで以下では，c_j と s_j が独立の場合との相違について検討する．

長さ n の独立な二つの $\{-1,1\}$ 系列 $X=(X_0,X_1,\ldots,X_{n-1})$，$Y=(Y_0,Y_1,\ldots,Y_{n-1})$ について，

$$\begin{aligned} c_j(X) &= \sum_{k=0}^{n-1} X_k \cos\frac{2\pi kj}{n} \\ s_j(Y) &= \sum_{k=0}^{n-1} Y_k \sin\frac{2\pi kj}{n} \end{aligned}$$

とし，

$$|S_j(X,Y)|^2 = (c_j(X))^2 + (s_j(Y))^2$$

とする．このとき，$(2/n)|S_j(X,Y)|^2$ の平均，分散はそれぞれ，2，$4-6/n$ である（付録 D 参照）．

付録 B に示した通り，$(2/n)|S_j(X)|^2$ の平均と分散はそれぞれ，2，$4-8/n$ である．したがって，$(2/n)|S_j(X)|^2$ と $(2/n)|S_j(X,Y)|^2$ の平均は等しい．一方，これらの分散は漸近的には同じ値に収束するものの，$(2/n)|S_j(X)|^2$ の分散は $(2/n)|S_j(X,Y)|^2$ の分散よりも小さい．

本調査では，DFT 検定において $S_j(X,Y)$ を用いた場合について，この検定で得られる N_1 の標本の分散値を計算し，$S_j(X)$ を用いた場合と比較した．

2.1 節で示した DFT 検定のアルゴリズムのステップ 4 と 6 について，これを一般化した以下のアルゴリズムを実行し，各 $p=0.05,\ 0.10,\ 0.15,\ 0.20,\ \ldots,\ 0.90,\ 0.95$ について N_1 の標本の分散値を計算した．本実験では，各 p について，X, Y それぞれ，長さ $n=4096$ の 2000 個の系列を用いた．なお，これらの擬似乱数系列の生成には Matematica の `Random` 関数を用いた．

4. $T=\sqrt{(-\ln(1-p))n}$ とする．なお，これは，$|S_i|\le T$ である確率が p となるような値である．

6. $|S_0|,|S_1|,\ldots,|S_{\frac{n}{2}-1}|$ の中で実際に T を越えなかった $|S_i|$ の個数 N_1 を求める．

図 1 に，$S_j(X,Y)$ について得られた N_1 の標本の分散値（$N_1(X,Y)$）を示す．この図には，比較のため，$S_j(X)$ について得られた N_1 の標本の分散値（$N_1(X)$）と，DFT 検定の根拠とされた理論による N_1 の分散値（theory）も示されている．この値は $p(1-p)(n/2)=2048p(1-p)$ である．図 1 から直ちに分かるように，$N_1(X,Y)$ は $N_1(X)$ より大きくなるものの，依然として理論による分散よりも小さな値をとる．また，$N_1(X)$ と同様に p が 0.5 より小さい値のときに最大値をとると考えられるなど，理論による分散とは異なる振舞を示している．

図 2 に，$S_j(X,Y)$ について得られた N_1 の標本の分散値と DFT 検定の根拠とされた理論による N_1 の分散値の比（$N_1(X,Y)$）を示す．この図には，比較のため，$S_j(X)$ について得られた N_1 の標本の分散値と理論による N_1 の分散値の比（$N_1(X)$）も示されている．これらは図 1 に示した結果から直ちに得られる結果である．

$S_j(X)$ について得られた N_1 の標本の分散値に関しては，既に，[7, 8] で，長さ 10^6 の擬似乱数系列を用いて精密な検討がなされている．ここで行った実験では，長さ 4096 の擬似乱数系列を用いたが，$S_j(X)$ について得られた N_1 の標本の分散値の振舞いは [7, 8] の実験とほぼ同じであり，十分妥当性のある結果と考えられる．

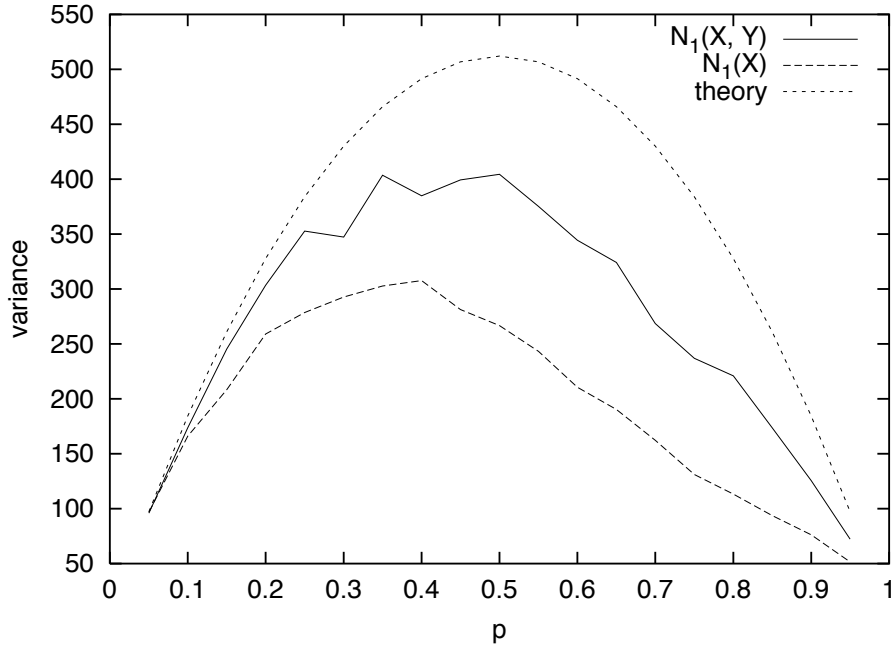

図 1: N_1 の標本の分散．$N_1(X,Y)$，$N_1(X)$ はそれぞれ，$S_j(X,Y)$，$S_j(X)$ によって得られた N_1 の標本の分散値を表す．theory は DFT 検定の根拠とされた理論に基づく N_1 の分散値 $p(1-p)(n/2)$ を表す．X，Y 各々について，長さ $n = 4096$ の系列が 2000 個用いられている．

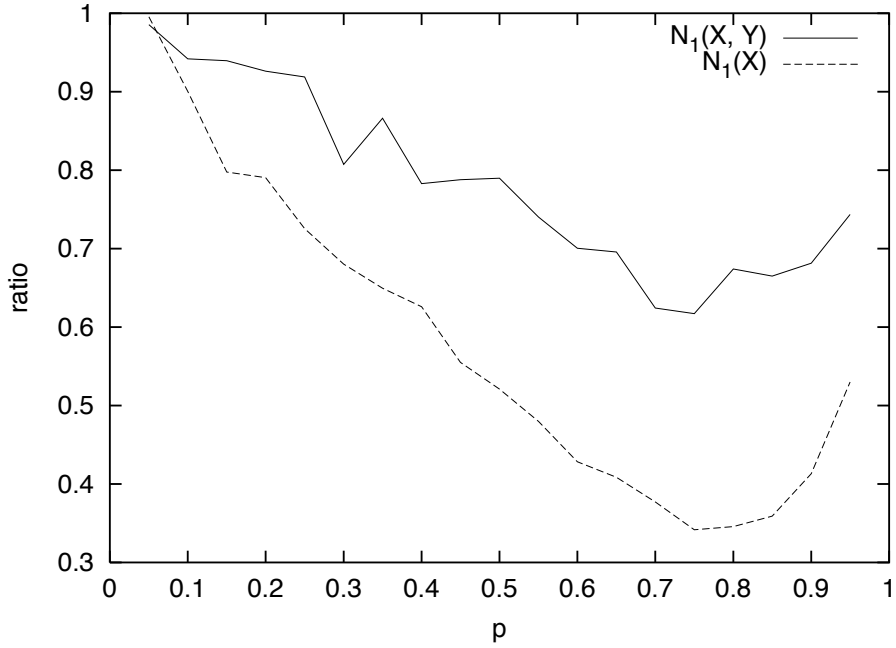

図 2: N_1 の標本の分散値と DFT 検定の根拠とされた理論による分散値 $p(1-p)(n/2)$ との比．$N_1(X,Y)$，$N_1(X)$ はそれぞれ，$S_j(X,Y)$，$S_j(X)$ によって得られた N_1 の標本の分散値に関する結果を示す．X，Y 各々について，長さ $n = 4096$ の系列が 2000 個用いられている．

2 について　パーセバルの定理より

$$\sum_{j=0}^{n-1}|S_j(X)|^2 = n\sum_{k=0}^{n-1}|X_k|^2 = n^2$$

が成立するので，$S_0(X), S_1(X), \ldots, S_{\frac{n}{2}-1}(X)$ が独立でないことは明らかである．濱野 [8] は，以下の分布について検討し，この従属性を定量的に評価することを試みている．

- $\sqrt{\dfrac{2}{n}}\displaystyle\sum_{j=1}^{n/2-1} c_j(X)$ の分布
- $\sqrt{\dfrac{2}{n}}\displaystyle\sum_{j=1}^{n/2-1} s_j(X)$ の分布
- $\sqrt{\dfrac{2}{n}}\displaystyle\sum_{j=1}^{n/2-1} |S_j(X)|$ の分布
- $\sqrt{\dfrac{2}{n}}\displaystyle\sum_{j=1}^{n/2-1} |S_j(X)|^2$ の分布

更に詳しく述べると，濱野はこれらについて，n が素数であると仮定して検討を行っている．このとき，

$$\begin{aligned}
\sum_{j=1}^{(n-1)/2} c_j(X) &= \sum_{j=1}^{(n-1)/2}\left(\sum_{k=0}^{n-1} X_k\cos\frac{2\pi kj}{n}\right) = \sum_{k=0}^{n-1} X_k \sum_{j=1}^{(n-1)/2}\cos\frac{2\pi kj}{n} \\
&= \frac{n-1}{2}X_0 + \sum_{k=1}^{n-1} X_k \sum_{j=1}^{(n-1)/2}\cos\frac{2\pi kj}{n} \\
&= \frac{n-1}{2}X_0 - \frac{1}{2}\sum_{k=1}^{n-1} X_k
\end{aligned}$$

より，

$$\sum_{j=1}^{(n-1)/2}(c_j(X) - X_0) = -\frac{1}{2}\sum_{k=1}^{n-1} X_k$$

が成立する．このことから，先に示した確率変数の分布については，$S_j(X)$，$c_j(X)$，$s_j(X)$ の代わりに以下について検討を行っている．

$$\begin{aligned}
S_j(X) - X_0 &= (c_j(X) - X_0) + \mathrm{i}\, s_j(X) \\
c_j(X) - X_0 &= \sum_{k=1}^{n-1} X_k\cos\frac{2\pi kj}{n} \\
s_j(X) &= \sum_{k=1}^{n-1} X_k\sin\frac{2\pi kj}{n}\left(= \sum_{k=0}^{n-1} X_k\sin\frac{2\pi kj}{n}\right)
\end{aligned}$$

以上の検討にもかかわらず，DFT 検定おいて，統計量 d がどのような分布に従うかは，これまでのところ解明されていない．

山本と金子 [7] は，$S_0(X), S_1(X), \ldots, S_{\frac{n}{2}-1}(X)$ について，長さ $(n/2)/M$ の M 個の部分列 $S_{\frac{n}{2}\frac{\ell}{M}}, \ldots, S_{\frac{n}{2}\frac{\ell+1}{M}-1}$ （$\ell = 0, 1, \ldots, M-1$）を考え，それぞれを個別の乱数系列を用いて計算して，DFT 検定における d の標本の分散値を確認している．

本調査ではこれに倣い，2.1 節で示した DFT 検定のアルゴリズムのステップ 4 と 6 について，これを一般化した以下のアルゴリズムを実行し，各 $p = 0.05, 0.10, 0.15, 0.20, \ldots, 0.90, 0.95$ について N_1 の標本の分散値を計算した．本実験では，各 p について，長さ $n = 32768$ の 5000 個の系列を用いた．なお，これらの擬似乱数系列の生成には Matematica の `Random` 関数を用いた．

4. $T = \sqrt{(-\ln(1-p))n}$ とする．

6. $|S_0|, |S_1|, \ldots, |S_{\frac{n}{2}\alpha-1}|$ の中で T を越えなかった $|S_i|$ の個数 N_1 を求める．

$\alpha = 1, 1/2, 1/4$ について，この実験の結果を図 3 に示す．この図から分かるように，利用する $S_0(X), S_1(X), \ldots, S_{\frac{n}{2}-1}(X)$ の部分列の長さを小さくするほど，標本の分散値が，DFT 検定の根拠とされた理論による分散値 $(\alpha\, n/2)p(1-p)$ に近付くことが分かる．但し，p が 0.7 から 0.8 の辺りで標本の分散値が $(\alpha\, n/2)p(1-p)$ に対して最小となるなど，いずれの α の値についても，同様の特徴を示すことが分かる．

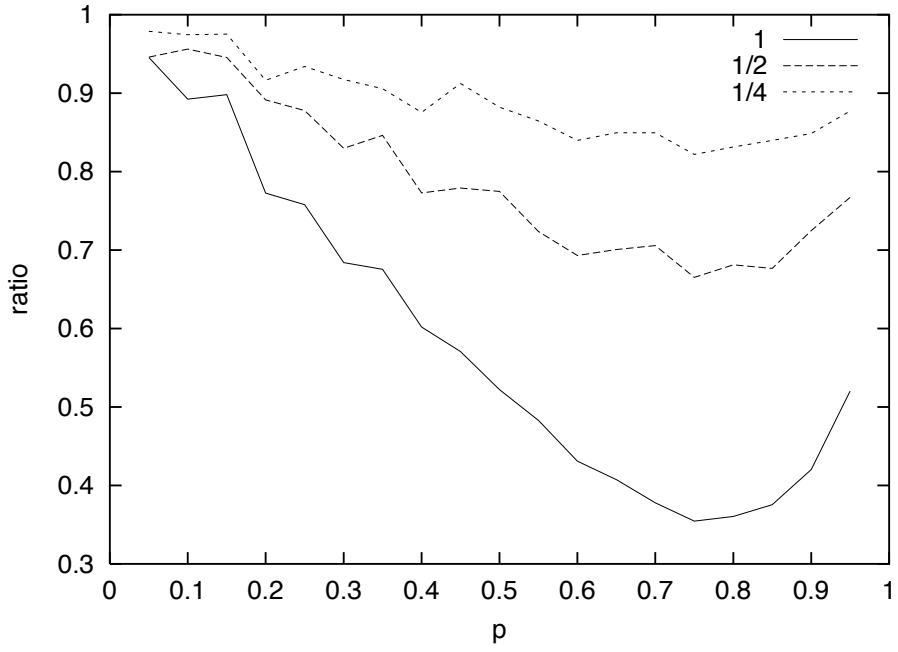

図 3: N_1 の標本の分散値と DFT 検定で用いられる分散の理論値 $(\alpha\, n/2)p(1-p)$ との比．右上の 1，1/2，1/4 は α の値を表す．長さ $n = 32768$ の系列が 5000 個用いられている．

4 むすび

本調査結果から明らかなように，DFT 検定とその根拠とされていた理論との間には明確な不一致が確認できる．この不一致は，DFT 検定で計算される統計量 d の分布が，理論の示す標準正規分布 $N(0,1)$ とは大きく異なる分布を示すということである．なお，計算機実験によると，d の分布は $N(0,0.5)$ に近い分布となることが確認できるが，これを裏付ける理論は構築されていない．

以上より，NIST SP800-22 の DFT 検定は，乱数列の検定法として不適切である．更に，DFT 検定をどのように改良すれば適切な検定法となるかは，現在のところ未解決問題である．

参考文献

[1] NIST, Special Publication 800-22, A statistical test suite for random and pseudorandom number generators for cryptographic applications, 2001.

[2] 電子政府情報セキュリティ技術開発事業「擬似乱数検証ツールの調査開発」調査報告書, 情報処理振興事業協会セキュリティセンター, 2003 年 2 月. http://www.ipa.go.jp/security/fy14/crypto/.

[3] 金子敏信, 擬似乱数生成系の検定方法に関する調査報告書　－ Lempel-Ziv 圧縮検定について－, http://cryptrec.nict.go.jp/PDF/wat-rep2003/rep_ID0206.pdf. http://www.ipa.go.jp/security/enc/CRYPTREC/fy15/documents/rep_ID0206.pdf.

[4] 金成主, 梅野健, 長谷川晃朗, NIST のランダム性評価テストについて, 信学技法, ISEC2003-87, pp. 21–27, 2003.

[5] 小針 [日見] 宏, 確率・統計入門, 岩波書店, 1973.

[6] 廣瀬勝一, 擬似乱数生成系の検定方法に関する調査　調査報告書, http://cryptrec.nict.go.jp/PDF/wat-rep2003/rep_ID0207.pdf. http://www.ipa.go.jp/security/enc/CRYPTREC/fy15/documents/rep_ID0207.pdf.

[7] 山本尚史, 金子敏信, NIST SP800-22 の DFT 検定に関する一考察, 信学技法, ISEC2004-50, pp. 61–64, 2004.

[8] K. Hamano, The distribution of the spectrum for the discrete Fourier transform test included in SP800-22, IEICE Trans. Fundamentals, vol. E88-A, no. 1, pp. 67–73, 2005.

[9] G. Marsaglia, DIEHARD, http://stat.fsu.edu/~geo/diehard.html, http://stat.fsu.edu/pub/diehard.html.

A　$c_j(X)$，$s_j(X)$ の平均と分散

$c_j(X)$ の平均は,

$$\frac{1}{2^n}\sum_X c_j(X) = \frac{1}{2^n}\sum_X\sum_{k=0}^{n-1} X_k \cos\frac{2\pi kj}{n} = 0$$

$c_j(X)$ の分散は,

$$\begin{aligned}
\frac{1}{2^n}\sum_X (c_j(X))^2 &= \frac{1}{2^n}\sum_X \left(\sum_{k=0}^{n-1} X_k \cos\frac{2\pi kj}{n}\right)^2 \\
&= \frac{1}{2^n}\sum_X \left(\sum_{k=0}^{n-1} {X_k}^2 \left(\cos\frac{2\pi kj}{n}\right)^2 + \sum_{\ell_1\neq\ell_2} X_{\ell}X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n}\right) \\
&= \frac{1}{2^n}\sum_X\sum_{k=0}^{n-1}\left(\cos\frac{2\pi kj}{n}\right)^2 = \sum_{k=0}^{n-1}\left(\cos\frac{2\pi kj}{n}\right)^2 = \sum_{k=0}^{n-1}\frac{1}{2}\left(1+\cos\frac{4\pi kj}{n}\right) \\
&= \frac{n}{2}
\end{aligned}$$

$s_j(X)$ の平均と分散も同様に計算できる.

B　$(2/n)|S_j(X)|^2$ の平均，分散

$$\begin{aligned}
|S_j(X)|^2 &= (c_j(X))^2 + (s_j(X))^2 \\
&= n + \sum_{\ell_1\neq\ell_2} X_{\ell_1}X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n} + \sum_{\ell_1\neq\ell_2} X_{\ell_1}X_{\ell_2}\sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n} \\
&= n + \sum_{\ell_1\neq\ell_2} X_{\ell_1}X_{\ell_2}\left(\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n} + \sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}\right) \\
&= n + \sum_{\ell_1\neq\ell_2} X_{\ell_1}X_{\ell_2}\cos\frac{2\pi(\ell_1-\ell_2)j}{n} \\
&= n + 2\sum_{\ell_1<\ell_2} X_{\ell_1}X_{\ell_2}\cos\frac{2\pi(\ell_1-\ell_2)j}{n}
\end{aligned}$$

である．したがって，$(2/n)|S_j(X)|^2$ の平均は

$$\frac{1}{2^n}\sum_{X\in\{-1,1\}^n}\frac{2}{n}|S_j(X)|^2 = \frac{1}{2^n}\sum_X\frac{2}{n}\left(n + 2\sum_{\ell_1<\ell_2} X_{\ell_1}X_{\ell_2}\cos\frac{2\pi(\ell_1-\ell_2)j}{n}\right) = 2$$

$(2/n)|S_j(X)|^2$ の分散は

$$\begin{aligned}
&\frac{1}{2^n}\sum_{X\in\{-1,1\}^n}\left(\frac{2}{n}|S_j(X)|^2-2\right)^2=\frac{16}{n^2}\frac{1}{2^n}\sum_X\left(\sum_{\ell_1<\ell_2}X_{\ell_1}X_{\ell_2}\cos\frac{2\pi(\ell_1-\ell_2)j}{n}\right)^2\\
&=\frac{16}{n^2}\frac{1}{2^n}\sum_X\sum_{\ell_1<\ell_2}(X_{\ell_1}X_{\ell_2})^2\left(\cos\frac{2\pi(\ell_1-\ell_2)j}{n}\right)^2=\frac{16}{n^2}\frac{1}{2^n}\sum_X\sum_{\ell_1<\ell_2}\left(\cos\frac{2\pi(\ell_1-\ell_2)j}{n}\right)^2\\
&=\frac{16}{n^2}\sum_{0\le\ell_1<\ell_2\le n-1}\left(\cos\frac{2\pi(\ell_1-\ell_2)j}{n}\right)^2=\frac{8}{n^2}\sum_{0\le\ell_1<\ell_2\le n-1}\left(1+\cos\frac{4\pi(\ell_1-\ell_2)j}{n}\right)\\
&=\frac{8}{n^2}\left(\binom{n}{2}+\sum_{0\le\ell_1<\ell_2\le n-1}\cos\frac{4\pi(\ell_1-\ell_2)j}{n}\right)\\
&=\frac{8}{n^2}\left(\frac{n(n-1)}{2}-\frac{n}{2}\right)=4-\frac{8}{n}
\end{aligned}$$

となる.

C $c_j(X)$ と $s_j(X)$ の共分散

$c_j(X)$ と $s_j(X)$ の平均は共に 0 であるから，$c_j(X)$ と $s_j(X)$ の共分散は以下のように計算できる.

$$\begin{aligned}
&\frac{1}{2^n}\sum_X c_j(X)\,s_j(X)=\frac{1}{2^n}\sum_X\left(\sum_{k=0}^{n-1}X_k\cos\frac{2\pi kj}{n}\right)\left(\sum_{k=0}^{n-1}X_k\sin\frac{2\pi kj}{n}\right)\\
&=\frac{1}{2^n}\sum_X\left(\sum_{k=0}^{n-1}{X_k}^2\cos\frac{2\pi kj}{n}\sin\frac{2\pi kj}{n}+\sum_{\ell_1\neq\ell_2}X_\ell X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}\right)\\
&=\sum_{k=0}^{n-1}\cos\frac{2\pi kj}{n}\sin\frac{2\pi kj}{n}\\
&=0
\end{aligned}$$

D $(2/n)|S_j(X,Y)|^2$ の平均，分散

$$\begin{aligned}
|S_j(X,Y)|^2&=(c_j(X))^2+(s_j(Y))^2\\
&=n+\sum_{\ell_1\neq\ell_2}X_{\ell_1}X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n}+\sum_{\ell_1\neq\ell_2}Y_{\ell_1}Y_{\ell_2}\sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}
\end{aligned}$$

である．$(2/n)|S_j(X,Y)|^2$ の平均は,

$$
\begin{aligned}
&\frac{1}{2^{2n}}\sum_{X,Y}\frac{2}{n}|S_j(X,Y)|^2 \\
&= \frac{1}{2^{2n}}\sum_{X,Y}\frac{2}{n}\left(n+\sum_{\ell_1\neq\ell_2}X_{\ell_1}X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n}+\sum_{\ell_1\neq\ell_2}Y_{\ell_1}Y_{\ell_2}\sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}\right) \\
&= 2
\end{aligned}
$$

$(2/n)|S_j(X,Y)|^2$ の分散は,

$$
\begin{aligned}
&\frac{1}{2^{2n}}\sum_{X,Y}\left(\frac{2}{n}|S_j(X,Y)|^2-2\right)^2 \\
&= \frac{4}{n^2}\frac{1}{2^{2n}}\sum_{X,Y}\left(\sum_{\ell_1\neq\ell_2}X_{\ell_1}X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n}+\sum_{\ell_1\neq\ell_2}Y_{\ell_1}Y_{\ell_2}\sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}\right)^2 \\
&= \frac{4}{n^2}\frac{1}{2^n}\sum_{X}\left(\sum_{\ell_1\neq\ell_2}X_{\ell_1}X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n}\right)^2+\frac{4}{n^2}\frac{1}{2^n}\sum_{Y}\left(\sum_{\ell_1\neq\ell_2}Y_{\ell_1}Y_{\ell_2}\sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}\right)^2 \\
&= \frac{16}{n^2}\frac{1}{2^n}\sum_{X}\left(\sum_{\ell_1<\ell_2}X_{\ell_1}X_{\ell_2}\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n}\right)^2+\frac{16}{n^2}\frac{1}{2^n}\sum_{Y}\left(\sum_{\ell_1<\ell_2}Y_{\ell_1}Y_{\ell_2}\sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}\right)^2 \\
&= \frac{16}{n^2}\sum_{\ell_1<\ell_2}\left(\left(\cos\frac{2\pi\ell_1 j}{n}\cos\frac{2\pi\ell_2 j}{n}\right)^2+\left(\sin\frac{2\pi\ell_1 j}{n}\sin\frac{2\pi\ell_2 j}{n}\right)^2\right) \\
&= \frac{16}{n^2}\sum_{\ell_1<\ell_2}\left(\frac{1}{4}\left(1+\cos\frac{4\pi\ell_1 j}{n}\right)\left(1+\cos\frac{4\pi\ell_2 j}{n}\right)+\frac{1}{4}\left(1-\cos\frac{4\pi\ell_1 j}{n}\right)\left(1-\cos\frac{4\pi\ell_2 j}{n}\right)\right) \\
&= \frac{8}{n^2}\sum_{\ell_1<\ell_2}\left(1+\cos\frac{4\pi\ell_1 j}{n}\cos\frac{4\pi\ell_2 j}{n}\right) \\
&= 4-\frac{4}{n}+\frac{8}{n^2}\sum_{\ell_1<\ell_2}\cos\frac{4\pi\ell_1 j}{n}\cos\frac{4\pi\ell_2 j}{n} \\
&= 4-\frac{4}{n}+\frac{4}{n^2}\sum_{\ell_1\neq\ell_2}\cos\frac{4\pi\ell_1 j}{n}\cos\frac{4\pi\ell_2 j}{n} \\
&= 4-\frac{4}{n}+\frac{4}{n^2}\left(\sum_{\ell_1=0}^{n-1}\sum_{\ell_2=0}^{n-1}\cos\frac{4\pi\ell_1 j}{n}\cos\frac{4\pi\ell_2 j}{n}-\sum_{\ell=0}^{n-1}\left(\cos\frac{4\pi\ell j}{n}\right)^2\right) \\
&= 4-\frac{4}{n}+\frac{4}{n^2}\left(\sum_{\ell_1=0}^{n-1}\left(\cos\frac{4\pi\ell_1 j}{n}\right)\sum_{\ell_2=0}^{n-1}\cos\frac{4\pi\ell_2 j}{n}-\frac{1}{2}\sum_{\ell=0}^{n-1}\left(1-\cos\frac{8\pi\ell j}{n}\right)\right) \\
&= 4-\frac{4}{n}-\frac{2}{n} \\
&= 4-\frac{6}{n}
\end{aligned}
$$