L-020

線形複雑度のプロファイルを用いた乱数検定に関する一考察

A Study on Randomness Test Using Linear Complexity Profile

芝山 直喜* 金子 敏信† 半谷 精一郎*
Naoki Shibayama Toshinobu Kaneko Seiichiro Hangai

1 はじめに

様々な乱数検定がある中で，米国商務省標準技術局（以下,「NIST」という.）の乱数検定 [1] が暗号研究者の間で特に利用されている．NIST の乱数検定には線形複雑度検定が含まれており，線形複雑度（以下,「LC」という.）を用いた乱数検定は重要な評価項目の一つである．2009 年に濱野らは LC のプロファイルを用いた乱数検定 [2]（以下,「LC プロファイル検定」という.）を提案した．NIST の LC 検定は 2 進系列をブロックに分割し，各ブロックの最終ビットでの LC の値を検定に利用するのに対し，LC プロファイル検定は最終ビットでの LC が直線 $y = \frac{1}{2}x$ 上に存在するブロックの LC のプロファイルと直線 $y = \frac{1}{2}x$ とで囲まれた三角形の面積の和を検定に利用する．

本稿では，LC プロファイル検定の妥当性評価結果について報告する．結果として，LC プロファイル検定の 3 シグマ法における棄却率を考慮しない場合，検定合格率の経験分布が理論分布に適合しないことがわかった．このため，LC プロファイル検定を精密に実施するためには，3 シグマ法を排除する，または，3 シグマ法における棄却率を考慮した検定合格率を適用する必要がある．

2 LC プロファイル検定

LC プロファイル検定は，ランダムな 2 進系列から計算された LC のプロファイルは図 1 のように直線 $y = \frac{1}{2}x$ の近くを不規則に増加するという性質に着目しており，最終ビットでの LC が直線 $y = \frac{1}{2}x$ 上に存在するブロックの LC のプロファイルと直線 $y = \frac{1}{2}x$ とで囲まれた三角形の面積の和を利用する．LC プロファイル検定は特定ブロック全体の LC に関する統計量を検定に使用するため，部分的にランダムではない系列に対し，NIST の LC 検定よりも精密な結果が得られる．ここでは，LC プロファイル検定の手順について示す．なお，n は系列長，M はブロック長であり，それぞれ文献 [2] で使用している値（n=10^6, M=500）とする．また，α は有意水準であり，α=0.01 とする．

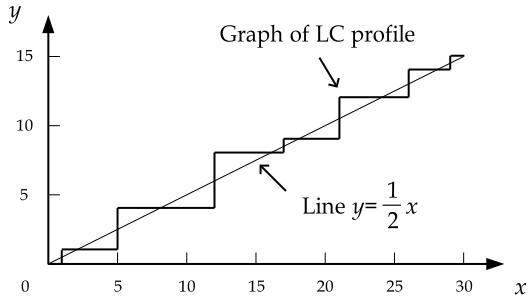

図 1: LC のプロファイルと直線 $y = \frac{1}{2}x$

検定手順

S1 2 進系列 $\varepsilon = \varepsilon_0\varepsilon_1 \cdots \varepsilon_{n-1}$ を N 個の M ビット長ブロック $B_j(1 \le j \le N)$ に分割する．ここで，$n = MN$ である．

S2 ブロック B_jの 2 進系列 $\varepsilon_{M(j-1)}\varepsilon_{M(j-1)+1} \cdots \varepsilon_{Mj-1}$ において，$\varepsilon_{M(j-1)}\varepsilon_{M(j-1)+1} \cdots \varepsilon_{M(j-1)+k}$ $(0 \le k \le M-1)$ の LC; $L_{k+1}^{(j)}$を求め，LC のプロファイルと直線 $y = \frac{1}{2}x$ とで囲まれた三角形の面積の和 $A_M^{(j)} = \sum_{k=1}^{M} \left| \frac{k}{2} - L_k^{(j)} \right|$ を計算する．

S3 集合 $\mathcal{A} = \{A_M^{(j)} | L_M^{(j)} = \frac{M}{2}\}$ とその要素数 $N' = \#\mathcal{A}$ を求める．

S4 $|\frac{N'}{N} - \frac{1}{2}| > \frac{3}{2\sqrt{N}}$ならば，3 シグマ法より帰無仮説 H_0を棄却する．

S5 $N'\phi \le 5$ のとき，S1 に戻り，$N'\phi > 5$ となるまで N を増やす．

S6 $p = \dfrac{\#\{A_M^{(j)} | A_M^{(j)} \in \mathcal{A}, A_M^{(j)} > T\}}{N'}$を求める．ただし，$T = 441$ である．

S7 $z = \dfrac{p - \phi}{\sqrt{\frac{\phi(1-\phi)}{N'}}}$を計算する．ただし,$\phi = 0.049611$ である．

S8 P 値 $= \mathrm{erfc}\left(\frac{|z|}{\sqrt{2}}\right)$ を計算する．なお, $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} \mathrm{d}u$ である．

S9 P 値 $< \alpha$ならば帰無仮説 H_0を棄却する．

本稿では，NIST の連検定 [1]（以下，単に「連検定」という.）に準じ，S4 の 3 シグマ法により帰無仮説 H_0 が棄却されたとき，P 値=0 とし S9 へ進むものとする．

* 東京理科大学大学院工学研究科電気工学専攻，Department of Electrical Engineering, Faculty of Engineering, Tokyo University of Science.
† 東京理科大学大学院理工学研究科電気工学専攻，Department of Electrical Engineering, Faculty of Science and Technology, Tokyo University of Science.

LC プロファイル検定の結果として出力される P 値は，NIST の乱数検定と同様の扱いが可能である [2]．すなわち，n ビット $\mathcal{N}$ 系列を検定し，$\mathcal{N}$ 個の P 値から検定合格率及び χ^2 統計量を算出し，それらの値によって乱数生成器の統計的評価を行う．ただし，$\mathcal{N}$=10^3 である．なお，ここでは検定合格率及び χ^2 統計量についての説明は省略する．

3 検定の妥当性評価

LC プロファイル検定における検定合格率の経験分布が分布関数に適合するかどうかを調べる [8]．検定合格率 r_α を (1) 式により定義する．

$$r_\alpha = \frac{\#\{\text{P 値} : \text{P 値} \geq \alpha\}}{\mathcal{N}} \tag{1}$$

次に，規格化された合格率 ξ を (2) 式により定義する．

$$\xi = \frac{r_\alpha - (1-\alpha)}{\sqrt{\frac{\alpha(1-\alpha)}{\mathcal{N}}}} \tag{2}$$

確率 $1-\alpha$ で 1，確率 α で 0 となる確率変数 X_i を考える．X_i が互いに独立ならば，確率変数 $Y = \sum_{i=1}^{\mathcal{N}} X_i$ は B($\mathcal{N}, \alpha$) に従う．したがって，ξ の経験分布は $\frac{\frac{Y}{\mathcal{N}} - (1-\alpha)}{\sqrt{\frac{\alpha(1-\alpha)}{\mathcal{N}}}}$ の分布関数に近いことが期待される．

$\mathcal{N}$ 回試行して 1 つの r_α を求めることを 1 セットの実験とする．良質な乱数を生成することが知られている DES の OFB モード [4] から生成した系列に対して LC プロファイル検定を適用し，10^3 セットの実験から得られた ξ の経験分布と二項分布の規格化された分布関数（以下,「理論分布」という.）を図 2 に示す．図 2 の両分布の適合度を表 1 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 = 853.80 となった．これは，自由度 6 の χ^2 分布の上側 5% が 12.59 であるので，853.80 $\geq$ 12.59 より両分布が同一分布であるとする帰無仮説が有意水準 5% で棄却される．また，検定合格率の平均値は 0.987317 であり，その期待値 $1-\alpha$=0.99 から乖離が生じている．したがって，LC プロファイル検定における検定合格率の平均値が期待値より低いため，乱数系列のランダム性が確率 α よりも多く棄却されることがわかった．さらに，DES の代わりに AES を使用し，ξ の経験分布と理論分布を図 3 に示す．図 3 の両分布を表 2 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量=718.59 ($\geq$ 12.59) となり，両分布が同一であるとする帰無仮説が有意水準 5% で棄却され，DES と同様の結果が得られた．

4 実際の検定合格率

これまでに，LC プロファイル検定における検定合格率は期待値に従わないことがわかった．ここでは，その理由について説明し，LC プロファイル検定における検定合格率の理論値を導出する．なお，以降では DES の OFB モードより生成した系列のみを使用する．

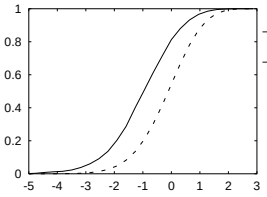

図 2: ξ の経験分布 (実線) と理論分布 (破線)（DES の場合）

表 1: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty, -2)$	0.026	0.136
$(-2, -1)$	0.108	0.264
$(-1, 0)$	0.282	0.315
0	0.126	0.098
$(0, 1)$	0.328	0.153
$(1, 2)$	0.119	0.032
$(2, \infty)$	0.010	0.002
合計	1.000	1.000

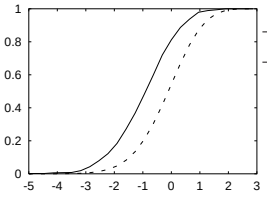

図 3: ξ の経験分布 (実線) と理論分布 (破線)（AES の場合）

表 2: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty, -2)$	0.026	0.122
$(-2, -1)$	0.108	0.246
$(-1, 0)$	0.282	0.352
0	0.126	0.092
$(0, 1)$	0.328	0.167
$(1, 2)$	0.119	0.021
$(2, \infty)$	0.010	0.000
合計	1.000	1.000

4.1 3 シグマ法の排除

LC プロファイル検定における検定合格率が期待値に従わない原因として，S4 の 3 シグマ法による帰無仮説 H_0 の棄却が考えられる．そのため，本節では 3 シグマ法を排除した LC プロファイル検定における ξ の経験分布が理論分布に適合するかを調べた．

ξ の経験分布と理論分布を図 4 に示す．図 4 の両分布を表 3 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 = 6.70 ($<$ 12.59) となり，両分布が同一であるとする帰無仮説が有意水準 5% で棄却されない．よって，LC プロファイル検定における検定合格率が期待値に従わない要因が 3 シグマ法であることが推測される．

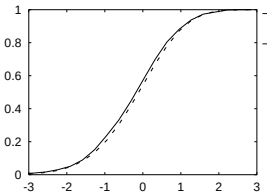

図 4: ξ の経験分布 (実線) と理論分布 (破線)

表 3: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty, -2)$	0.026	0.031
$(-2, -1)$	0.108	0.120
$(-1, 0)$	0.282	0.293
0	0.126	0.127
$(0, 1)$	0.328	0.310
$(1, 2)$	0.119	0.105
$(2, \infty)$	0.010	0.014
合計	1.000	1.000

4.2 3 シグマ法を考慮した検定合格率

LC プロファイル検定の 3 シグマ法における棄却率を考慮した検定合格率の理論値を導出し，LC プロファイル検定における ξ の経験分布がこの理論分布に従うかを調べる．

はじめに，3 シグマ法が適切に構成されていることを確認する．10^3 個の $\frac{N'}{N}$ を求め，規格化された $\frac{N'}{N}$ の経験分布が二項分布 $\mathrm{B}(10^3, 0.5)$ の規格化された理論分布に適合するかを調べた．規格化された $\frac{N'}{N}$ の経験分布と二項分布 $\mathrm{B}(10^3, 0.5)$ の規格化された理論分布を図 5 に示す．図 5 の両分布を表 4 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 $= 11.18\,(< 12.59)$ となった．よって，両分布が同一であるとする帰無仮説を有意水準 5% で棄却されないため，3 シグマ法が適切に構成されていることがわかった．

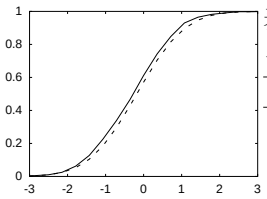

図 5: 規格化された $\frac{N'}{N}$ の経験分布 (実線) と二項分布 $\mathrm{B}(10^3, 0.5)$ の規格化された理論分布 (破線)

表 4: 規格化された $\frac{N'}{N}$ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty, -2)$	0.023	0.015
$(-2, -1)$	0.163	0.133
$(-1, 0)$	0.324	0.318
0	0.025	0.022
$(0, 1)$	0.324	0.350
$(1, 2)$	0.140	0.141
$(2, \infty)$	0.023	0.021
合計	1.000	1.000

$\frac{N'}{N}$ は近似的に $\mathrm{N}(\frac{1}{2}, (\frac{1}{2\sqrt{N}})^2)$ に従うため，3 シグマ法における合格率の期待値は約 0.997 である．さらに，3 シグマ法を排除した LC プロファイル検定の検定合格率の期待値は $1-\alpha = 0.99$ であるから，LC プロファイル検定の検定合格率の理論値は $0.997 \times 0.99 \approx 0.987$ となり，3 節で得られた LC プロファイル検定における検定合格率の平均値にほぼ一致する．これより，$\alpha = 1-0.987 = 0.013$ としたとき，ξ の経験分布は理論分布に従うものと考えられる．

$\alpha = 0.013$ とし，ξ の経験分布が理論分布に適合するかを調べた．ξ の経験分布と理論分布を図 6 に示す．図 6 の両分布を表 5 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 $= 9.57\,(< 12.59)$ となり，両分布が同一であるとする帰無仮説を有意水準 5% で棄却されない．以上より，3 シグマ法により LC プロファイル検定における検定合格率は期待値から 0.003 程度低下することがわかった．

5 LC プロファイル検定の適切な事前判定

連検定は，4 シグマ法を用いて事前に系列が頻度検定に合格していること（以下,「事前判定」という.）を確認し，これに合格した系列に対して連の検定を適用する．ここでは，その妥当性評価結果を示す．また，事前判定に 3 シグマ法を用いた連検定の妥当性評価結果について示し，LC プロファイル検定の LC に関する適切な事前判定について述べる．

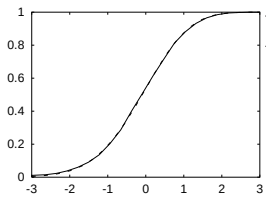

図 6: ξ の経験分布 (実線) と理論分布 (破線)

表 5: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty, -2)$	0.024	0.023
$(-2, -1)$	0.139	0.113
$(-1, 0)$	0.264	0.264
0	0.111	0.106
$(0, 1)$	0.298	0.307
$(1, 2)$	0.154	0.172
$(2, \infty)$	0.010	0.015
合計	1.000	1.000

5.1 連検定における事前判定

連検定の手順について説明する．2 進系列 $\varepsilon = \varepsilon_0 \varepsilon_1 \cdots \varepsilon_{n-1}$ の 1 の割合 $\pi = \frac{1}{n}\sum_{i=0}^{n-1} \varepsilon_i$ を計算し，$|\pi - \frac{1}{2}| \geq \frac{2}{\sqrt{n}}$（4 シグマ法）のとき，2 進系列における 0 及び 1 の割合に偏りがあると判定し，帰無仮説 H_0 を棄却する．このとき，連の検定を適用せず，P 値 $= 0$ とする．$|\pi - \frac{1}{2}| < \frac{2}{\sqrt{n}}$ のとき，$V_n(\mathrm{obs}) = \sum_{k=1}^{n-1} r(k) + 1$ を求め，P 値 $= \mathrm{erfc}(\frac{|V_n(\mathrm{obs}) - 2n\pi(1-\pi)|}{2\sqrt{2n}\pi(1-\pi)})$ を計算し，P 値 $< \alpha$ ならば帰無仮説 H_0 を棄却する．ここで，$\varepsilon_{k-1} = \varepsilon_k$ のとき $r(k) = 0$，$\varepsilon_{k-1} \neq \varepsilon_k$ のとき $r(k) = 1$ である．

検定の妥当性評価

ξ の経験分布が理論分布に適合するかを調べた．ξ の経験分布と理論分布を図 7 に示す．図 7 の両分布を表 6 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 $= 10.83\,(< 12.59)$ となり，両分布が同一であるとする帰無仮説を有意水準 5% で棄却することはできない．

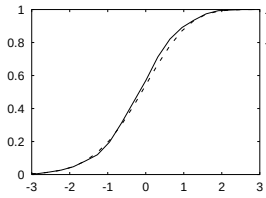

図 7: ξ の経験分布 (実線) と理論分布 (破線)

表 6: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty, -2)$	0.026	0.027
$(-2, -1)$	0.108	0.093
$(-1, 0)$	0.282	0.321
0	0.126	0.128
$(0, 1)$	0.328	0.322
$(1, 2)$	0.119	0.102
$(2, \infty)$	0.010	0.007
合計	1.000	1.000

π は近似的に $\mathrm{N}(\frac{1}{2}, (\frac{1}{2\sqrt{n}})^2)$ に従うため，4 シグマ法における棄却率は約 0.00006 である．よって，$\mathcal{N}$=10^3 のとき，その確率が検定合格率に与える影響は非常に小さい．

3 シグマ法を用いた連検定の妥当性評価

連検定の 4 シグマ法を 3 シグマ法に変更し，その ξ の経験分布が理論分布に適合するかを調べる．すなわち，値 $\frac{2}{\sqrt{n}}$ を $\frac{3}{2\sqrt{n}}$ に変更することによって，検定合格率が期待値から 0.003 程度低下するため，LC プロファイル検定と同様の結果が得られることを確認する．

ξ の経験分布と理論分布を図 8 に示す．図 8 の両分布を表 7 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 $= 868.79\,(\geq 12.59)$ となった．よって，両分布が同一

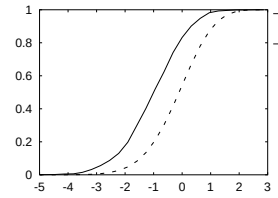

図 8: ξ の経験分布 (実線) と理論分布 (破線)

表 7: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty,-2)$	0.026	0.130
$(-2,-1)$	0.108	0.272
$(-1,0)$	0.282	0.339
0	0.126	0.091
$(0,1)$	0.328	0.150
$(1,2)$	0.119	0.017
$(2,\infty)$	0.010	0.001
合計	1.000	1.000

であるとする帰無仮説は有意水準 5% で棄却される．

次に，α=0.013 とし，その検定合格率の経験分布が理論分布に適合するかを調べた．ξ の経験分布と理論分布を図 9 に示す．図 9 の両分布を表 8 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 $= 10.41\,(< 12.59)$ となり，両分布が同一であるとする帰無仮説を有意水準 5% で棄却することはできない．

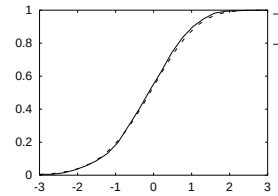

図 9: ξ の経験分布 (実線) と理論分布 (破線)

表 8: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty,-2)$	0.024	0.015
$(-2,-1)$	0.139	0.115
$(-1,0)$	0.264	0.272
0	0.111	0.115
$(0,1)$	0.298	0.315
$(1,2)$	0.154	0.161
$(2,\infty)$	0.010	0.007
合計	1.000	1.000

以上より，LC プロファイル検定における検定合格率の低下を導く要因が 3 シグマ法であること，また 3 シグマ法における棄却率を考慮した場合，ξ の経験分布が理論分布に適合することが再確認された．

5.2 4 シグマ法を用いた LC プロファイル検定の妥当性評価

本節では，LC プロファイル検定の 3 シグマ法を連検定と同じ 4 シグマに変更，すなわち，値 $\frac{3}{2\sqrt{N}}$ を $\frac{2}{\sqrt{N}}$ に変更し，その ξ の経験分布が理論分布に適合するかを調べる．

ξ の経験分布と理論分布を図 10 に示す．図 10 の両分布を表 9 に基づき χ^2 適合度検定を行ったところ，χ^2 統計量 $= 9.58\,(< 12.59)$ となり，両分布が同一であるとする帰無仮説を有意水準 5% で棄却することはできない．したがって，LC プロファイル検定の 3 シグマ法を 4 シグマ法に変更することによって，$\mathcal{N}$=10^3 のとき，4 シグマ法における棄却率は非常に小さく，その検定合格率は期待値に従うとみなせるため，検定が適切に構成されていることがわかった．

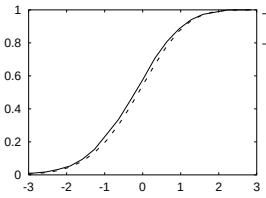

図 10: ξ の経験分布 (実線) と理論分布 (破線)

表 9: ξ の区間分けされた確率値

区間	理論確率	観測値
$(-\infty,-2)$	0.026	0.033
$(-2,-1)$	0.108	0.121
$(-1,0)$	0.282	0.300
0	0.126	0.122
$(0,1)$	0.328	0.308
$(1,2)$	0.119	0.102
$(2,\infty)$	0.010	0.014
合計	1.000	1.000

6 まとめ

本稿では，LC プロファイル検定の妥当性評価結果について報告した．LC プロファイル検定の 3 シグマ法における棄却率を考慮しない場合，検定合格率の経験分布が理論分布に適合しないことがわかった．また，連検定の 4 シグマ法を 3 シグマ法に変更したときも同様の結果が得られることがわかった．よって，LC プロファイル検定を精密に実施するためには，3 シグマ法を排除する，または，3 シグマ法における棄却率を考慮した検定合格率を適用する必要がある．さらに，従来通り LC プロファイル検定に LC に関する事前判定を適用するならば，$\mathcal{N}$=10^3 のとき，3 シグマ法を 4 シグマ法に変更する必要がある．このとき，検定合格率には本質的な問題が生じないことがわかった．ただし，$\mathcal{N}$ を 10^5 程度以上に大きくしたとき，4 シグマ法における棄却率により，検定合格率は期待値から乖離が生じると考えられるため，その棄却率を考慮しなければならない．

参考文献

[1] A. Rukhin, “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Apprications”, NIST Special Publication 800-22.

[2] K.Hamano, F.Sato, H.Yamamoto, “A New Randomness Test Based on Linear Complexity Profile”, IEICE Trans. on Fundamentals, vol.E92-A, no.1, pp.166-172, 2009.

[3] K.Hamano, “Errata for A New Randomness Test Based on Linear Complexity Profile”, http://www.it.k.u-tokyo.ac.jp/〜hamano/errata.pdf, 2009.

[4] R. R. Jueneman, “Analysis of Certain Aspects of Output Feedback Mode”, CRYPTO 1982, pp.99-127.

[5] 金子研究室, “疑似乱数生成系の検定方法に関する調査報告書”, http://www.cryptrec.go.jp/estimation/rep_ID0211.pdf, 2004.

[6] 奥富秀俊, 金田学, 山口健二, 中村勝洋, “NIST 乱数検定を用いた乱数性の評価に関する考察”, SCIS2006-1E2-3, 2006.

[7] 濱野健二, “NIST の乱数検定に含まれる最長連検定の修正”, ISEC2007-3, 2007.

[8] 濱野健二, “全自己相関に基づく新しい乱数検定法”, 信学技法, ISEC2008-4, 2008.