NIST 及び DIEHARD テストによる

RPG１００乱数評価

FDK（株）RPG 推進室

2003/12/16

導入 RPG100 から生成される乱数を、2 つの有名なテストを用いて評価します。一方は米国機関 NIST により公開されている資料に基づくテストで、他方は Marsaglia 博士により提供されているテストです。乱数を一度テストしただけでは、それが常にテストを満足する乱数性を持っていることを確認できないことから、ここでは 1Gbit の乱数列を用いてそれぞれのテストを複数回行います。そして、これらテストの合格率の妥当性と P-Value の一様性を調べることで乱数性の評価をします。

このテストは、電圧３.３Ｖ、温度２５℃、乱数生成クロック２５０ＫＨｚの条件で、RPG100 サンプル一つから取得した乱数について行われています。

１．ＮＩＳＴ　Special Publication 800-22　乱数テスト

NIST Special Publication 800-22（以下 NIST SP800-22 と略す）は、NIST（米国商務省標準技術研究所）により発行された資料で、そのテストプログラムも入手することができます[1]。

テストに用いられる乱数データ量は、指定された範囲内でユーザーが決めることが出来ますが、本資料では、NIST SP800-22 を一回行うのに用いる乱数量を 1 Mbit としています。評価はテスト１０００回で行います。

1-1. テストの種類とパラメータの設定

NIST SP800-22 では１６種類の乱数テストが設けられており、それぞれのテストの有意水準は１％となっています。これらのテストのうち幾つかはパラメータの設定が必要となっており、ここでの設定を下記テーブルに示します。また、情報処理振興事業協会（IPA）で採択された資料には、たくさんある乱数テストの中からミニマムセットを決めたものがあります[2]。以下、このミニマムセットに該当するテストには＊マークを付けます。

テストは１６種類ありますが、そのうちの **Lempel-Ziv Compression** と **Discrete Fourier Transform (Spectral) Test** の統計量は NIST で期待した分布とずれていることが判り、P-Value は一様とならず、テストの有意水準も設定値である１％となっていません（[3],及び[2]の p40 参照）。したがって、本資料ではこれらのテストを除外し、１４種類のテストをもって乱数性の評価をします。

乱数テスト名　（＊はミニマムセット）	テストパラメータ
Frequency (Monobit) Test	－
＊**Frequency Test within a Block**	m=20000
Runs Test	－
＊**Test for the Longest Run of Ones in a Block**	M=10000
Binary Matrix Rank Test	－

~~**Discrete Fourier Transform (Spectral) Test**~~	－
Non-overlapping Template Matching Test	m=9　,　B=000000001
Overlapping Template Matching Test	m=9
Maurer's "Universal Statistical" Test	L=7　,　Q=1280
~~**Lempel-Ziv Compression Test**~~	－
＊**Linear Complexity Test**	M=500
＊**Serial Test**	m=5　,　$\nabla\Psi^2$
Approximate Entropy Test	m=5
＊**Cumulative Sums (Cusum) Test**	Forward
Random Excursions Test	X= +3 , - 3
Random Excursions Variant Test	X= +3 , - 3

本資料では､NISTテストを１０００回行って得られたP-Valueの値を用いて乱数性を評価します。ただし、 **Random Excursions Test** および **Random Excursions Variant Test** ではランダムウォークのサイクル数がJ<５００であった場合にテストが中止されます。そのため１０００回のNISTテストを行っても、これらのテストが行われるのはそれより少ない回数なります（テストが中止される確率は理論的に３８％となっています）。テストの標本数を１０００に統一するため、これら2つのテストではX= +3 のP-Value５００個とX= - 3のP-Value５００個を合わせて１０００個の標本とします。

1-2. 各テストの合格率評価

NIST SP800-22では、各々のテストについて有意水準を１％としています。これは理想的な乱数をテストした場合に合格と判断される確率が９９％（不合格１％）であるということです。ここではテストを１０００回行い、RPG１００の乱数がこれら有意水準１％のテストに合格する割合の妥当性について評価します。

標本数をn、テストに合格する確率をp、合格した回数をxとします。合格数xは試行回数n、確率pの二項分布に従いますが、n が大きいときは近似的に期待値 $m = np$ 、標準偏差$\sigma = \sqrt{np(1-p)}$ の正規分布とみなすことができます。ここで、$z \equiv (x-m)/\sigma$と置くと、この変数 z は期待値ゼロ、標準偏差１の標準正規分布 N(0,1)に従います。合格数xを標準化変数zをつかって書くと、$x = m + z\sigma$ となり、両辺をnで割ると、

$$p' \equiv x/n = p + z\sqrt{p(1-p)/n} \qquad \cdots(1)$$

となります。式から明らかですが、p'は測定による合格率となっています。NISTでは、$-3 \leqq z \leqq 3$となる範囲にp'が入った時に『合格率は妥当』、もしくは『テストされた乱数の質は適正』と判断することが推奨されており、これは有意水準が約０.２７％のテストとなっています。各テストに合格する確率は p=0.99であり、標本数はn=1000なので、乱数の質が適正であると判断されるp'の範囲は、

$$p' = 0.99 \pm 3\times\sqrt{0.99\times 0.01/1000} = 0.99 \pm 0.0094392 \qquad \cdots(2)$$

で与えられます。結果は次の通りです。

標本数 n=1000　採択域 0.980561≦p'≦0.999439

乱数テスト名　(＊はミニマムセット)	合格率 p'	結果
Frequency (Monobit) Test	0.994	SUCCESS
＊Frequency Test within a Block	0.983	SUCCESS
Runs Test	0.992	SUCCESS
＊Test for the Longest Run of Ones in a Block	0.989	SUCCESS
Binary Matrix Rank Test	0.989	SUCCESS
Non-overlapping Template Matching Test	0.993	SUCCESS
Overlapping Template Matching Test	0.988	SUCCESS
Maurer's "Universal Statistical" Test	0.991	SUCCESS
＊Linear Complexity Test	0.991	SUCCESS
＊Serial Test	0.989	SUCCESS
Approximate Entropy Test	0.991	SUCCESS
＊Cumulative Sums (Cusum) Test	0.985	SUCCESS
Random Excursions Test	0.990	SUCCESS
Random Excursions Variant Test	0.995	SUCCESS

1-3. 各テストの P-Value 一様性評価

乱数が理想的であれば、各テストのアウトプットである P-Value は[０,１)の間で一様に出現することが期待されます。ここでは１０００回のテストで得られた P-Value を等確率１/１０となるような１０区間に分け、それぞれの度数について χ^2 検定を行います（一様分布に対する適合度の検定)。この場合、χ^2 分布の自由度は９となります。ここで、F_i を i 番目の区間に入った P-Value の個数とすると、χ^2 統計量は、次式で与えられます。

$$\chi^2 = \sum_{i=1}^{10} \frac{(F_i - n/10)^2}{n/10} \qquad \cdots (3)$$

NIST では、有意水準０．０１％での判断を推奨しており、対応する χ^2 値の範囲は $\chi^2 \le 33.72$ です。

採択域　$\chi^2 \le 33.72$

乱数テスト名　(＊はミニマムセット)	χ^2 値	結果
Frequency (Monobit) Test	5.20	SUCCESS
＊Frequency Test within a Block	10.51	SUCCESS
Runs Test	2.64	SUCCESS
＊Test for the Longest Run of Ones in a Block	10.33	SUCCESS
Binary Matrix Rank Test	17.84	SUCCESS
Non-overlapping Template Matching Test	5.18	SUCCESS
Overlapping Template Matching Test	9.40	SUCCESS

Maurer's "Universal Statistical" Test	20.53	SUCCESS
＊Linear Complexity Test	13.87	SUCCESS
＊Serial Test	9.978	SUCCESS
Approximate Entropy Test	14.02	SUCCESS
＊Cumulative Sums (Cusum) Test	4.67	SUCCESS
Random Excursions Test	5.64	SUCCESS
Random Excursions Variant Test	6.18	SUCCESS

1-4. テスト全体の合格率評価

1-2 では、テスト毎に合格数を集計して、乱数の良し悪しを判断しましたが、ここではテスト全体での合格率について評価します。方法は同じですが、標本数が n=14000 となります。乱数の質が適正であると判断される p'の範囲は式（1）より、

$$p' = 0.99 \pm 3 \times \sqrt{0.99 \times 0.01 / 14000} = 0.99 \pm 0.0025228 \quad \cdots (4)$$

となります。

標本数 n=14000　採択域 $0.987477 \leqq p' \leqq 0.992523$

乱数テスト名	合格率 p'	結果
NIST SP800-22	0.990	SUCCESS

1-5. テスト全体の P-Value 一様性評価

1-3 では、テスト毎に P-Value 一様性を調べましたが、ここではテスト全体の P-Value についてその一様性を調べます。統計量は式（3）で与えられますが、その際の標本数は n=14000 となります。

採択域　$\chi^2 \leq 33.72$

乱数テスト名	χ^2値	結果
NIST SP800-22	12.85	SUCCESS

２．DIEHARD　乱数テスト

フロリダ州立大学統計学部の Marsaglia 博士による乱数テストです[4]。１８種類の統計テストで構成されており、乱数が理想的であれば P-Value が[０,１)の間に一様に分布するとしていますが、NIST 800-22 の様に明確な判定基準は設けられていません。テストに必要な乱数データ量は８０Mbit で、本資料ではこれを１２回行って評価を行います。

2-1. テストの種類と P-Value の数

DIEHARD では１８種類の乱数テストが設けられておりますが、各テストで出力される P-Value は複数あり、その数もテスト毎に異なります。テスト種と P-Value の個数は以下の通りです。DIEHARD から出力される P-Value は２２０個あり、テストは１２回行われていますので、全 P-Value 数は２６４０個となります。

乱数テスト名　(＊はミニマムセット)	P-Value の数
＊The Birthday Spacings Test	１０
＊The Overlapping 5-Permutation Test	2
The Binary Rank Test for 31x31 Matrices	1
The Binary Rank Test for 32x32 Matrices	1
The Binary Rank Test for 6x8 Matrices	２６
＊The Bitstream Test	２０
＊The Overlapping-Pairs-Sparse-Occupancy Test	２３
＊The Overlapping-Quadruples-Sparse-Occupancy Test	２８
The DNA Test	３１
＊Count-The-1's Test on a Stream of Bytes	2
＊Count-The-1's Test for Specific Bytes	２５
The Parking Lot Test	１１
The Minimum Distance Test	1
The 3D-Spheres Test	２１
The Squeeze Test	1
The Overlapping Sums Test	１１
The Runs Test	4
The Craps Test	2

2-3. テスト合格率評価

各テストの有意水準をNISTと同じく１％とし､テスト合格率について全P-Valueを用いて評価します。標本数は n=2640 となります。判断基準となる合格率 p'の範囲は式（１）より、

$$p' = 0.99 \pm 3 \times \sqrt{0.99 \times 0.01 / 2640} = 0.99 \pm 0.0058094 \quad \cdots (5)$$

となります。

標本数 n=2640　採択域 $0.984191 \leqq p' \leqq 0.995809$

乱数テスト名	合格率 p'	結果
DIEHARD	0.989	SUCCESS

次に、0.025≦P-Value≦0.975 を合格とした場合についても評価します。この場合の有意水準は５％（合格率 p=９５％）なので、判断基準となる合格率 p'の範囲は式（１）より、

$$p' = 0.95 \pm 3 \times \sqrt{0.95 \times 0.05 / 2640} = 0.95 \pm 0.0127252 \quad \cdots (6)$$

となります。

標本数 n=2640　採択域 $0.937275 \leqq p' \leqq 0.962725$

乱数テスト名	合格率 p'	結果
DIEHARD	0.950	SUCCESS

2-4. P-Value 一様性評価

得られた P-Value について一様性を調べます。統計量は式（３）で与えられますが、その際の標本数は n=2640 となります。

採択域　$\chi^2 \leq 33.72$

乱数テスト名	χ^2値	結果
DIEHARD	4.57	SUCCESS

３．評価まとめ

NIST 800-22 および DIEHARD テストを複数回行い、その合格率と P-Value の一様性についてテストを行いました。合格率のテストは有意水準０．２７％、P-Value の一様性のテストは有意水準０．０１％で評価し、RPG１００の乱数はこれらのテストに不合格となるものはありませんでした。RPG１００の乱数は十分な乱数性をもっていると考えられます。

参考文献

[1] NIST, Special Publication 800-22, “A STASTICAL TEST SUITE FOR RANDOM AND PSEUDO-RANDOM NUMBER GENERATORS FOR CRYPTOGRAPHIC APPLICATIONS”, 2001.
(http://csrc.nist.gov/rng/)

[2] IPA,調査報告書,”疑似乱数検証ツールの調査開発”, 2003.
(http://www.ipa.go.jp/security/fy14/crypto/pseudo_rundum/rundum_inve.pdf)

[3] 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”, 電子情報通信学会信学技報, Vol.103, No.449, pp21-27, 2003.

[4] G. Marsaglia, “DIEHARD”. (http://stat.fsu.edu/~geo/diehard.html)