擬似乱数の検定

2000MM110 大池 拓真
指導教員　伏見 正則

1　はじめに

現在，”シミュレーション”と聞いて何も思い浮かばないという人はいないと思われる．それほど，この言葉はポピュラーなものになってきた．そういったシミュレーションにおけるランダムな要素や PC 等のセキュリティー等を決定する方法のひとつとして乱数が使われている．しかし，乱数を何十万個も必要とするようなシミュレーションやパスワードにサイコロを使って行なうことは現実的でない．そこで電子計算機を使って乱数らしきもの（擬似乱数）を多量に発生させる方法が考えられた．近年使われているほとんどのプログラム言語には，それぞれのプログラム言語によって，一様乱数列を発生させるプログラムが初めから組み込まれている．しかし，これらの一様乱数列の一様性を検証した文献は少ない．そこで，本研究では身近なプログラムである $VisualBasic$ によって求められる擬似乱数が本当に一様性を持つかどうかを検証していく．$VisualBasic$ は $Excel$ に VBA（$VisualBasicforApplications$）という形で標準装備している．今回は，その VBA を使用し検証を行なった．

2　本研究の目的と研究方法

擬似乱数は計算機で数列を作る以上，次に出る数はプログラムによって完全に決まってしまうという点において，真の乱数と擬似乱数とは大きく違うが，数の頻度分布の”一様性”だけが満たされていれば十分であると考えられる．そこで乱数列の局所的な性質を調べるために実際に一様乱数列を発生させ，統計的な手法を使用し，解析を行なった．今回は $Lehmer$ が述べているように，極めて多数提案されている統計的検定法のうちから利用目的に応じていくつかの検定法を選び，検定を行なうことで，擬似乱数が一様性を持つかを検証していく．なお本研究では低次元（3 次元程度）までの検定をカイ 2 乗検定 [3][6] を用いた度数検定 [1] で検証し，さらにその結果を $Kolmogorov$-$Smirnov$ 検定（以下 K-S 検定と略す.）[1][2] を用いて検定を積み重ねた．また高次元（4 次元以上）における検定では度数検定で行なうと膨大な時間がかかってしまうので衝突検定 [1] を用いて検証した。

3　試行結果

以下に行なった 3 つの検定の試行結果を記載する．

3.1　度数検定

ここでは，乱数列の文字の出現頻度を求め，その度数分布が一様になっているかどうかをカイ 2 乗検定を用いた 1 次元度数検定，2 次元度数検定，3 次元度数検定を用いて検定する．検定を行なう際に，複雑な計算や，単純ではあるが大量の計算を必要とする為に，ほとんどの $Office$ ツールに含まれている $Excel$ に付属されている $VisualBasicforApplications$（以下 VBA）を使用し，プログラムし，結果を $Excel$ シート上に表記する方法をとった．なお，乱数の発生には $randomize$ 環境を使用して行なった実験と，使用せずに行なった実験の 2 種類が存在するが，今回は $randomize$ 環境を使用せずに行なった実験の結果を記載する．またすべての場合において $np_i = 100$ となるように擬似乱数の発生個数を決定してある．以下に実行結果を示す．表 1 はカイ 2 乗検定の片側 95% で検定し，棄却されない個数を数えたものである．

表 1　度数検定で棄却されなかった個数

度数検定	1 次元	2 次元	3 次元
棄却されない個数	92 個	100 個	100 個

3.2　K-S 検定

1 - 3 次元度数検定のそれぞれで 100 回実験を繰り返し，得た統計量を使用し，K-S 検定を行なった．本検定は表計算ソフト $SPSSBase10.0J$(以下 $SPSS$ と略す．) を使用し算出した．本来なら経験分布関数とカイ 2 乗分布の分布関数を比べ検定を進めていくのだが，検定する統計量の個数が 100 個と大きいので中心極限定理 [4] により，標準正規分布の分布関数と経験分布関数を比べることで両側検定を行なった．以下に結果を記載する．

表 2　K-S 検定における検定結果

次元	1	2	3
N	100	100	100
漸近有意確率	0.600	0.537	0.340

3.3　衝突検定

高次元の度数検定は算出時間が非常にかかる為に効率的とはいえない．実際，4 次元度数検定は時間がかかりすぎて，算出することが出来なかった．そこで高次元の検定を行なう為に衝突検定を使用し，検定を行なった．乱数列の k 個の数値の組み合わせを k 次元座標上の点をみなし，セルに配置していく．セルの個数は m=d^k とする．ただし，d は 1 次元の分割数である．また，配置する点の個数を n とする．既に点が入っているところに入ったら，衝突として衝突回数を数える．そして数えた

衝突回数を検定統計量にあてはめて両側 95% で検定する．今回は統計量 $Time$(本論文内では 100) 個のうち棄却されない統計量が何個あるかを調べた．なお，衝突検定の検定統計量は，統計表が存在しないため，[1] に記載されている式から算出し，検定を行った．

表 3　$m = 2^{20} = 4^{10} = 16^5$,$n = 2^{14}$ の衝突検定の結果

$m =$	2^{20}	4^{10}	16^5
棄却されなかった個数	87	88	95

表 4　$m = 2^{24} = 64^5 = 256^3 = 4096^2$, $n = 2^{18}$ の衝突検定の結果

$m =$	2^{24}	64^5	256^3	4096^2
棄却されなかった個数	40	0	0	0

表 5　$m = 2^{24} = 64^4 = 256^3 = 4096^2$, $n = 2^{18}$ の衝突検定の結果

$m =$	2^{24}	64^4	256^3	4096^2
棄却されなかった個数	0	0	0	0

4　考察

度数検定で得た低次元（3 次元程度）における擬似乱数は表 1 を見ると片側 95% で検定した場合，棄却されない個数がどれもきわめて高く，低次元における度数検定は，カイ 2 乗分布に従っていると考えられる．また度数検定で求めた統計量 100 個を使用し，$Kolmogorov$-$Smirnov$ 検定を使用し，検定を積み重ねた結果が記載されている表 2 を見ても，1 – 3 次元での漸近有意確率がすべて 0.05 を上回っていることよりカイ 2 乗分布に従っているといえる．しかし，衝突検定の結果表 3 を見てみるといずれもかなり高い値を出しているが，棄却されなかった個数が 95 個以上だったのは $m = 16^5$ の場合だけである．m と n の数値を増加して検定した結果を表した表 4 を見ると $m = 2^{24}$ がかろうじて 40 個という値を観測しているがその他はどれも 0 である．そこから表 5 のように，さらに数値を増加させ $m = 2^{24}$,$n = 2^{20}$ で行なった実験にいたっては棄却されない値が 1 つもないという結果を得た．このことから基本的には単位正方形あたりの分割するセルの個数が増えるが，次元を減らしたほうがより高い精度を持っている，すなわち高次元での擬似乱数を使用する場合は精度に不安が残ると考えられる．

しかし，表 2 を見てみると表 1 とはまったく逆の現象，すなわち単位正方形あたりの分割するセルの個数を増やし，次元を減らしたにもかかわらず棄却されない値は減ってしまった．これは m, n の値を増やすことにより多次元疎結晶構造がより明確に現れた為と考えられる．

以上の考察をまとめると，$Excel$ の擬似乱数は一様性の面に於いて，低次元（3 次元程度まで）で使用する分にはあまり問題はないが高次元（4 次元位以上）では精度に不安があると考えられる．

5　おわりに

いくつかの統計的検定すべてに合格したら結果を保証するというような検定の組み合わせはなく，むしろ，不良な数列を排除するための手段と考えたほうが無難である．また，統計的な乱数検定は，真にランダムな系列の持つべき種々の性質の一部を保証するものであり，各種検定に合格したからといって検定対象である乱数列の性質の十分性を示しているわけではないことに注意する必要がある．したがって，その検定結果で分かることは，その数列が非常に悪いか，あまり望ましくないか，望ましくないとは言えないかという程度なので，検定は良い数列を選び出す方法と考えず，悪い数列を排除するための方法だと考えておいたほうが無難である．そのことを踏まえた上で今回行なった実験の最終的な結論は，$Excel$ に含まれている擬似乱数の精度は次元が変化すると一様性の精度も変化することもあり，あまり望ましくないと考えられる．

今後の課題として，衝突検定の結果を数多く算出することで，擬似乱数の多次元化による一様性の精度を確かめる必要があると考えられる．また，今回はいろいろな問題があり実行出来なかったのだが，擬似乱数を発生させる上での合同法のパラメータが分かればスペクトル検定を行なうことがでる．今後はそこに重点をおき研究を進めることにより，より良い擬似乱数を発生させたい．

6　謝辞

本研究を進めるにあたり，伏見正則教授には，筆者の不勉強からくる至らない点を根気よく何度もご指導して頂いた．また，伏見研究室に在籍するすべての学生が笑顔で力を貸してくれた事は本当に心の支えとなった．この場を借りて深く感謝の意を表したい．

参考文献

[1] 伏見正則：乱数（UP 応用数学選書），東京大学出版会 (1989).
[2] 森戸晋, 逆瀬川浩孝：システムシミュレーション（経営工学ライブラリー 5），朝倉書店 (2000).
[3] Donald Ervin Knuth 著，渋谷政昭（訳）：準数値算法/乱数, サイエンス社 (1981).
[4] 白旗慎吾：統計解析入門, 共立出版株式会社 (1992).
[5] 土屋和人：Excel VBA（Perfect Master 44），秀和システム (2002).
[6] 日本工業標準調査会：乱数発生及びランダム化の手順（JIS Z 9031），日本規格協会 (2001).