置換回数の分布による乱数の検定について

一　村　　稔

最近，乱数の利用はシミュレーションをはじめ広範囲になり，かつ発生個数もコンピュータの発達にともない，数億を用いることも困難なことではなくなった．乱数の発生にはある程度数学的にその性質がわかっている合同法が主として用いられているが，その他乱数発生器などによる物理的な発生法も用いられる．これらの方法で発生した数列がはたして乱数の性質である一様性と，無規則性の２つの性質を満足しているかどうかについては別に検定する必要がある．

ここでは乱数の無規則性の検定を考え，特に多量に発生した乱数にコンピュータの比較演算の高速性を利用する方法として，置換回数 Rank を用いるものについて考察する．

無規則性の検定には現在，系列相関検定，組合せ（Poker）検定，連（Run）の検定，ギャップ（Gap）検定等が考えられている，しかしながら，多量に発生した乱数を高速に検定するには不適切なところが多い．そこで置換回数を用いることによりコンピュータの比較演算のみを利用し，できるだけ情報を失わないで，発生と同時に高速で検定を行なうことが必要となってくる．もちろん与えられた数列の無規則性を調べるためには，いろいろな検定を併用する必要があり，検定相関の関連性についても吟味する必要があるが，この報告では２，３の悪い性質を検出するための与えられた数列における置換回数，また最大値の位置，最小値の位置の分布を利用するノンパラメトリックな検定について述べる．

置換回数の分布は K. N. Chandler[1) によってはじめられ，分布は多項式近似によって求められている．その後，F. G. Foster と A. Stuart[2) により，置換回数の和，分散について研究されたが，分布の正確な値は求められていない．

方法と解析

(*a*)　置換回数を用いる方法

分布関数 F(x) をもつように発生させた N 個の乱数を

$$x_1,\ x_2, \cdots\cdots,\ x_N$$

とし，はじめから n 個ずつの m 個のブロック（簡単のために nm＝N とする）に分割し次の仮説を検定する．

仮説：x_1，x_2，……，x_N は F(x) からの独立な確率変数の実現値

各々のブロック x_1，x_2，……，x_N において

$$u_i = \begin{cases} 1 & (x_i > x_1,\ x_2,\ \cdots\cdots,\ x_{i-1}) \\ 0 & (\text{その他}) \quad i=1,\ 2, \cdots,\ n \end{cases}$$

$$l_i = \begin{cases} 1 & (x_i < x_1,\ x_2,\ \cdots\cdots,\ x_{i-1}) \\ 0 & (\text{その他}) \quad i=1,\ 2, \cdots,\ n \end{cases}$$

（ただし $u_1=l_1=0$ と定義する）

$$t=\sum_{i=i}^{n} u_i, \quad s=\sum_{i=1}^{n} l_i$$

なる統計量をつくる．t は n 個の乱数において，最大値を得るまでの置換回数であり，s は最小値を得るまでの置換回数である．

$n\to\infty$ の場合の t, s の極限分布については A. Renyi[3] によって仮説のもとで正規分布に従うことが示された．

さて，仮説を検定する場合に n がどれくらい大きければ正規近似を用いてよいかということが当然問題となる．n が小さいとき t, s の確率分布が正規分布から大きくはずれるために，検定に用いることができない．

そこで t, s の正確な確率分布を求める．t, s の正確な分布は

$$\mathrm{P}_n(t)=\frac{1}{n}\sum_{d_1=t+1}^{n}\sum_{d_2=t}^{d_1-1}\cdots\sum_{d_t=2}^{d_{t-1}-1}\frac{1}{d_1-1}\cdot\frac{1}{d_2-1}\cdots\frac{1}{d_t-1}$$

となることが知られている．[3]

この正確な値を計算するために次のような漸化式を用いる．

$$S(n,\ k)=\sum_{d_1=k+1}^{n}\sum_{d_2=k}^{d_1-1}\cdots\sum_{d_k=2}^{d_{k-1}-1}\frac{1}{d_1-1}\cdot\frac{1}{d_2-1}\cdots\frac{1}{d_k-1}$$

$$S(n,\ k)=\frac{1}{n-1}S(n-1,\ k-1)+S(n-1,\ k)$$

計算結果について過去に文献が見つからないので，ここにその一部を記載し，正規分布 $P_n^*(t)$ への近似度合を同時に示す．これは２ブロック以上 m ブロックまでにおいても同様に考えられる．

註）$N=1,000,000$ のとき，$n=1,000$, $m=1,000$ としても $n=100$, $m=10,000$ としてもよく，n, m, の値は自由にとることができる．数億におよぶ乱数列においては，$N=1,000,000$ として検定し，また次の 1,000,000について検定し，順次やっていくこともできる．

Fig.　1 (a)

t, s ＼ n	4	6	8
0	0.250000	0.166666	0.125000
1	0.458333	0.380555	0.324107
2	0.250000	0.312500	0.325694
3	0.416666	0.118055	0.1678820
4	——	0.020833	0.048611
5	——	0.001388	0.007986
6	——	——	0.000694
7	——	——	0.000025

Fig,　1 (b)

t, s ＼ n	10	12	14
0	0.100000	0.083333	0.071429
1	0.282897	0.251656	0.227152
2	0.323165	0.315068	0.405084
3	0.199427	0.219745	0.233014
4	0.074218	0.096024	0.114222
5	0.017436	0.027849	0.038268
6	0.002604	0.005506	0.009073
7	0.000240	0.000746	0.001549
8	0.000012	0.000068	0.000191
9	——	0.000004	0.000017
10	——	——	0.000001

Fig. 2 (a)

t, s \ n	16	18	20
0	0.062500	0.555555	0.050000
1	0.207389	0.191086	0.177387
2	0.294694	0.284520	0.274820
3	0.241698	0.247301	0.250777
4	0.129372	0.142025	0.152651
5	0.048257	0.057637	0.066353
6	0.013039	0.017217	0.021481
7	0.002611	0.003885	0.005324
8	0.000392	0.000673	0.001029
9	0.000044	0.000090	0.000157
10	0.000004	0.000009	0.000019
11	——	——	0.000002
12	——	——	——

Fig. 2 (b)

t, s \ n	25	30	35
0	0.040000	0.033333	0.028571
1	0.151038	0.132055	0.117663
2	0.253075	0.234728	0.219196
3	0.253788	0.252404	0.248963
4	0.172617	0.186092	0.195384
5	0.085339	0.100852	0.113599
6	0.032036	0.041995	0.051150
7	0.009406	0.013852	0.018405
8	0.002206	0.003700	0.005410
9	0.000420	0.000813	0.001321
10	0.000065	0.000149	0.000272
11	0.000008	0.000023	0.000048
12	——	0.000003	0.000007
13	——	——	——

Fig. 2 (c)

t, s \ n	40	60	100
0	0.025000	0.016666	0.010000
1	0.106339	0.077720	0.051774
2	0.205913	0.167645	0.125852
3	0.244558	0.225085	0.192986
4	0.201868	0.213011	0.211204
5	0.124169	0.152301	0.176717
6	0.059512	0.086021	0.118151
7	0.022925	0.039592	0.065101
8	0.0072611	0.015195	0.030245
9	0.001924	0.004949	0.012057
10	0.000432	0.0013873	0.004183
11	0.000083	0.000338	0.001277
12	0.000014	0.000073	0.008347
13	0.000002	0.000014	0.000084
14	——	0.000002	0.000018
15	——	——	0.000004

Fig, 3 n=20 の場合

t	Pn(t)	Pn*(t)	Pn(t)−Pn*(t)
0	0.0500000	0.1108923	−0.0608923
1	0.1773870	0.2318904	−0.0545034
2	0.2748198	0.3171203	−0.0423005
3	0.2507771	0.2306437	+0.0201334
4	0.1526509	0.0891470	+0.0635039
5	0.0663527	0.0182730	+0.0480797
6	0.0214809	0.0019798	+0.0195011
7	0.0053244	0.0001131	+0.0052113
8	0.0010292	0.0000034	+0.0010258
9	0.0001570	0.0000000	+0.0001570

このようにして求めた表（置換回数の正確な確率分布）と，発生した乱数の m ブロックにおける置換回数の実現値とから適合度検定ができる．この表から収束の速度は意外と遅いことがわかる．$n\leqq300$ においては，検定に $P_n(t)$ を用いることをすすめたい．

（b）各ブロックにおける最大値，最小値の利用し各ブロックにおいて仮説のもとでは，n 個の乱数列の最大値 x_i，最小値 x_j の位置（番号）は，1～n の値を一様にとり得るので，

$$P(i\,|\max x_i)=\frac{1}{n},\quad P(j\,|\min\ x_j)=\frac{1}{n}$$

Fig, 4　n=50 の場合

t	Pn(t)	Pn*(t)	Pn(t)−Pn*(t)
0	0.0200000	0.0422972	−0.0222972
1	0.0895841	0.1140426	−0.0244585
2	0.1843855	0.2278091	−0.0434233
3	0.2347958	0.2787588	−0.0439630
4	0.2095354	0.2089937	+0.0005417
5	0.1404912	0.0959733	+0.0445179
6	0.0740035	0.0269733	+0.0470302
7	0.0315923	0.0046338	+0.0269585
8	0.0111839	0.0004858	+0.0106981
9	0.0033413	0.0000311	+0.0033102
10	0.0008542	0.0000013	+0.0008529

Fig, 5　n=150 の場合

t	Pn(t)	Pn*(t)	Pn(t)−Pn*(t)
0	0.0066666	0.0132788	−0.0066122
1	0.0372301	0.0430707	−0.0058406
2	0.0984952	0.1135925	−0.0150973
3	0.1656897	0.2035528	−0.0378631
4	0.2003684	0.2479102	−0.0475418
5	0.1865065	0.2052314	−0.0187249
6	0.1396132	0.1154741	+0.0241391
7	0.0866611	0.0441456	+0.0425155
8	0.0456269	0.0114620	+0.0341649
9	0.0207346	0.0020196	+0.0187150
10	0.0082466	0.0002414	+0.0080052
11	0.0029031	0.0000196	+0.0028835
12	0.0009131	0.0000011	+0.0009120
13	0.0002587	0.0000000	+0.0002587

となる．そこで（a）で述べた検定と同時に一様性の検定を行なうこともできる．また各ブロックで最大値が最小値よりも前に出れば＋，反対なら－とすると，＋－の２つの記号から連の検定も可能である．

この検定の特徴

(1)　コンピュータの比較演算のみを使用するために非常に高速でできる．また連の検定を除いては発生する乱数を store することを必要とせず，発生させながら同時に検定できる．

(2)　部分的には増減をくり返しながら，全体として増加しているような数列については連の検定ではとらえることができないが，この検定法では検証できる．

Fig. 6　n=300 の場合

t	Pn(t)	Pn*(t)	Pt(t)−Pn*(t)
0	0.0033333	0.0064001	−0.0030668
1	0.0209311	0.0225056	−0.0015745
2	0.0629807	0.0670449	−0.0040642
3	0.1217076	0.1420273	−0.0203197
4	0.1706015	0.2139984	−0.0433969
5	0.1855954	0.2293756	−0.0437802
6	0.1636337	0.1749010	−0.0112673
7	0.1205080	0.0948646	+0.0256434
8	0.0758033	0.0365925	+0.0392108
9	0.0414338	0.0100352	+0.0313986
10	0.0199502	0.0019556	+0.0179946
11	0.0085565	0.0002707	+0.0082858
12	0.0032993	0.0000267	+0.0032726
13	0.0011526	0.0000019	+0.0071507
14	0.0003673	0.0000000	+0.0003673
15	0.0001074	0.0000000	+0.0001074

最後に本研究を遂行するにあたり助言をいただいた，岡山大学教養部，脇本和昌助教授に感謝の意を表します．

文　献

1. Chandler, K. N. (1952) "The distribution and frequency of record values", J. R. S. B. 14 220—228
2. Foster, F. G. & Stuart, A. (1954) "Distribution-free tests in time-series based on the Breaking of records", J. R. S. S. B. 16 1—23
3. Rényi, A. (1962), "Théorie des éléments saillants dûne suite d'observations", Calloqium, Aarhus, August 1—10, 1962