石田式乱数作成機のランダム性について

立教大学理学部数学教室　西　平　美　恵　子

(1957 年 12 月受付)

On the Test of Randomness on the Isida's Random Number Generator

Mieko NISIHIRA

We find the randomness of a series of the random numbers, which we obtained from the *Isida's random number generator*, by frequency test of each number from 0 to 9, by frequency test of each number from 00 to 99, by poker test, by gap test and by sequential test.

See, M. Isida and H. Ikeda, Random number generator. *Ann. Inst. Statist. Math.*, Vol. Ⅷ No. 2.

St. Paul's University.

最近，統計数理研究所において新たに乱数作成機がつくられた．これは，放射能のカウント数を計測する方法によつているものである[1),2)].

ここでは，この乱数作成機によつてつくられた数列のランダム性について述べることにしよう．

そのために，この乱数作成機から得られた 6000 個の数字について，以下のような 5 種類の検定をおこなつた*．1° 頻度検定，2° 継次検定，3° ポーカー検定，4° ギャップ検定，5° 系列相関による検定．

1° 頻度検定　1 桁の数字について，0 から 9 までの 10 通りの各数字の出現する回数が，等確率 $\left(\frac{1}{10}\right)$ になつているかどうかを調べること．

この検定をするために，6000 個の数字を始めから順に 300 個ずつの 20 組に分けて，各組ごとに各数字の頻度をとり，これを χ^2 検定によつて調べた．

その結果は第 1 表に示す通りである．これから分るように，95% の信頼度で，等確率という仮説が棄却されたものは，20 組のうち 1 組（第 9 組）で，ちようど 100 回のうち 5 回はずれる割合となつている．又，20 組全体をひとまとめにした 6000 個についての頻度をとり χ^2 検定をすると，これも 95% の信頼度で，等確率という仮説は棄却されない．

2° 継次検定　相次ぐ 2 つの数字を組にして，2 桁の数を考えると，00, 01,……,99 の 100 通りあるが，これらの数字の出現回数が等確率 $\left(\frac{1}{100}\right)$ になつているかどうかを調べること．

6000 個の数字全体の相次ぐ 2 つずつを組合わせて 2 桁の数と考え，この 3000 個の 2 桁の数について頻度をとると第 2 表のようになる．この表について χ^2 検定をする場合自由度は 99 となる．自由度 ν が $\nu \geq 30$ の時には，$\sqrt{2\chi^2}$ は漸近的に $N(\sqrt{2\nu-1}, 1)$ に従う．この表においては，

$\chi^2 = 108.133$　　したがつて，

$\sqrt{2\chi^2} = 14.706, \quad \sqrt{2\nu-1} = 14.036$

* 統計数値表[3)] の乱数表の検定は，このうちの 1°～4° についておこなわれている．

第 1 表 頻 度 検 定

組番号＼数字	0	1	2	3	4	5	6	7	8	9	計	χ^2
1	33	27	41	31	27	32	29	27	32	21	300	8, 267
2	31	26	27	32	35	28	22	38	33	28	300	6, 667
3	39	37	30	25	21	29	31	33	28	27	300	8, 667
4	26	36	30	31	36	30	35	23	28	25	300	6, 400
5	28	28	35	35	25	34	29	25	33	28	300	4, 600
6	28	37	33	29	26	30	46	27	25	19	300	16, 333
7	36	24	36	24	27	30	31	28	25	39	300	8, 800
8	31	29	19	27	20	39	34	40	28	33	300	14, 733
9	39	35	32	32	30	37	30	16	33	16	300	**18, 800**
10	28	32	30	29	28	28	29	32	41	23	300	6, 400
11	33	27	33	33	26	26	27	25	34	36	300	5, 133
12	38	30	29	34	27	25	34	29	25	29	300	5.267
13	29	33	32	34	24	26	36	34	20	32	300	7, 933
14	31	26	34	29	29	25	32	31	33	30	300	2, 467
15	26	26	38	26	26	26	31	33	27	41	300	9, 467
16	34	29	29	31	27	39	30	30	29	22	300	5, 800
17	33	36	28	15	25	27	32	35	37	32	300	13, 000
18	22	36	30	27	36	40	22	36	27	24	300	13, 000
19	36	21	22	30	30	25	32	31	46	27	300	15, 867
20	29	23	37	31	44	32	23	33	25	23	300	14, 400
(1〜20) 全 体	630	598	625	585	569	608	615	606	609	555	6, 000	8, 577

$d.f.=9 \quad P_r\{\chi^2>16.919\}=0.05$

第 2 表 継 次 検 定

10 位＼1 位	0	1	2	3	4	5	6	7	8	9	計
0	36	32	33	26	34	37	34	38	27	33	330
1	24	47	30	19	29	23	28	29	22	26	277
2	28	38	35	37	38	39	25	25	29	28	322
3	31	35	38	27	35	25	23	34	28	32	308
4	31	26	28	28	18	26	26	30	35	22	270
5	30	24	22	28	26	34	33	34	29	23	283
6	34	38	30	29	23	35	33	34	21	25	302
7	31	35	25	28	26	35	33	29	36	27	305
8	28	31	30	31	39	40	45	31	25	29	329
9	27	15	32	24	31	31	33	17	28	36	274
計	300	321	303	277	299	325	313	301	280	281	3,000

となり，95％の信頼度で等確率 $\left(\frac{1}{100}\right)$ という仮説は棄却されない．

3° ポーカー検定 例えば，相次ぐ5つの数字のパターンを考えると，このパターンは *aaaaa, aaaab, aaabb, aaabc, aabbc, aabcd, abcde* の7通りがある．これに対しては，次表に示すような確率分布が計算されている．これを帰無仮説として検定すること．

6000 個の数字全体を前半と後半の2つに分けて頻度をとると第3表のようになる．前半も後半

パターン	確率分布
$a\ a\ a\ a\ a$	0.0001
$a\ a\ a\ a\ b$	0.0045
$a\ a\ a\ b\ b$	0.0090
$a\ a\ a\ b\ c$	0.0720
$a\ a\ b\ b\ c$	0.1080
$a\ a\ b\ c\ d$	0.5040
$a\ b\ c\ d\ e$	0.3024
計	1.0000

第 3 表　ポーカー検定

パターン	実現頻度		理論頻度	実現頻度	理論頻度
	前半	後半		全体	
$a\ a\ a\ a\ a$ / $a\ a\ a\ a\ b$ / $a\ a\ a\ b\ b$	10	7	8	17	16
$a\ a\ a\ b\ c$	41	40	43	81	86
$a\ a\ b\ b\ c$	65	69	65	134	130
$a\ a\ b\ c\ d$	301	295	303	596	605
$a\ b\ c\ d\ e$	183	189	181	372	363
計	600	600	600	1200	1200
χ^2	0.628	1.145	—	0.834	

$d.f.=4\quad P_r\{\chi^2>9.488\}=0.05$

も χ^2 検定において，95 %の信頼度で棄却されない．又，6000 個の数字全体についても，95 %の信頼度でポーカー帰無仮説は棄てられない．

4° ギャップ検定　例えば 0 に着目して，0 がどれだけの間隔（ギャップ）をおいて現われるかを調べ，これについてギャップ帰無仮説の成否を検定すること．ただしこの仮説は 0 から 9 までの数字が一様に且つ独立にあらわれるとすると，ギャップに関して右表に示すような確率分布となることを意味する．

ギャップの長さ	確率分布	
0	0.1000	0.2710
1	0.0900	
2	0.0810	
3	0.0729	0.1975
4	0.0656	
5	0.0590	
6	0.0531	0.1439
7	0.0478	
8	0.0430	
9	0.0387	0.1050
10	0.0349	
11	0.0314	
12	0.0282	0.2826
13	0.0254	
14	0.0229	
15	0.0206	
16～20	0.0759	
21～25	0.0448	
26以上	0.0648	
計	1.0000	

ここでは 6000 個の数字を順に 4 組 (A, B, C, D) に等分して，まず，各組ごとに検定をし，つぎに全体をひとまとめにして検定する．いま，数字 0 に着目して，相次ぐ 0 の間隔を調べたところ，第 4 表のようになつた．

ここでは，A 組の最後に現われる 0 と B 組の最初に現われる 0 との間のギャップは，A 組のギャップとして数えた．したがつて，'全体' の一番初めに現われる 0 の前の数字と，一番あとに現われる 0 の後の数字はギャップとしては数えられない．

この表から分る様に，ギャップ帰無仮説に対して χ^2 検定を行つた結果，95 %の信頼度でこの仮説は棄却されない．

5° 系列相関による検定　出てくる数がすぐその前の数に影響されていないか，又，いくつか前の数に従属していないか，即ち数列の順序が独立であるかどうかを検定すること．この検定のためには系列相関係数を用いる．

いま，ズレが h なる系列相関係数を R_h とすると，

$$R_h=\sum_{i=1}^{n} x_i x_{i+h}$$

ここで $i+h>n$ に対しては $x_{i+h}=x_{i+h-n}$ とする．このとき $E(R_h)$ と $D^2(R_h)$ の値は h に関係しないから，それぞれ $E(R), D^2(R)$ と書くことにする．いま，n が大きくなれば $\dfrac{R_h-E(R)}{D(R)}$

第4表 ギャップ検定

ギャップの長さ＼組	A 実際	A 理論	B 実際	B 理論	C 実際	C 理論	D 実際	D 理論	全体 実際	全体 理論
0～2	39	43	35	44	47	43	35	42	156	170
3～5	33	31	36	32	30	31	33	30	132	124
6～8	21	23	31	23	17	23	29	22	98	91
9～11	19	16	21	17	19	16	13	16	72	66
12 以上	45	44	39	46	44	44	43	43	171	178
計	157	157	162	162	157	157	153	153	629	629
χ^2	1.261		7.130		2.532		4.257		3.027	

$d.f.=4 \quad P_r\{\chi^2>9.488\}=0.05$

第5表 系列相関による検定

h ＼ 組番号	1	2	3	4	5	6	7	8	9	10
1	5661	6024	5480	5866	6103	5404	6239	6533	4751	6136
2	5273	6108	5542	5612	5791	5365	5990	6854	4512	6058
3	5418	6253	5435	5554	6107	5386	6119	6893	4727	6136
4	5504	6233	5535	5783	6140	5515	6099	6948	4650	6479
5	5498	6468	5676	5733	6029	5574	6166	6580	5007	6148
6	5199	6257	5154	5502	5818	5343	6138	6674	5135	6286
7	5473	6137	5405	5372	5802	5543	6150	6594	4879	6185
8	5502	6131	5461	5836	6046	5299	6137	6942	4704	6114
9	5289	6111	5763	5702	5970	5571	6054	6935	4786	6185
10	5289	6196	5442	5727	5926	5718	6161	6610	4840	5866
11	5426	6430	5385	5417	6189	**5846**	6086	6900	4996	6177
12	5605	6210	5726	5389	6246	5515	6348	6969	4780	6051
13	5544	6211	5529	3936	6038	5428	6113	6517	4907	6431
14	5728	6197	5475	5490	6062	5474	6139	6992	4870	6080
15	5520	6441	5777	5548	6151	5296	6112	6612	4922	6313
16	5380	6074	5376	5481	6097	5298	6435	7107	4831	6175
17	5351	6215	5617	5692	5860	5469	6138	6827	4816	6130
18	5597	6426	5253	5902	5814	5510	6035	6956	5048	6145
19	5338	6269	5550	5778	5877	5311	6540	6775	4705	6265
20	5029	6356	5471	5610	5782	5532	6158	7007	4727	6295
$E(R)-3D(R)$	5026	5823	5037	5253	5542	5065	5688	6358	4390	5769
$E(R)+3D(R)$	5846	6673	5953	6050	6377	5843	6606	7220	5194	6618

は漸近的にガウス分布 $N(0,1)$ に従うことが示されている[4].

ここでは 6000 個の数字全体のうちの前半の 3000 個を 300 個ずつの 10 組に分けて検定した．各組について，1つズラシ，2つズラシ，……，20 ズラシの系列相関係数 20 個を求めた*．その結果は第5表の通りである．n が 300 で相当大きいから，ガウス分布を使つて検定すれば，信頼度を 99.7% としたときの信頼区間からはずれるものは第6組の $h=11$ だけである．

* この計算は統計数理研究所のリレー計算機で行つた．

以上5種類の検定の結果，この乱数作成機による数列はランダム性をもつものと考えられる．

なお，この報告作製にあたつて示された，統計数理研究所の林知己夫氏，石田正次氏の御指導に対して厚く御礼申し上げる．

参考文献

1) 石田正次，放射能のランダム性について．統計数理研究所彙報，4巻2号，1956．
2) Masatugu Isida, and Hiroji Ikeda, Random number generator. *Ann. Inst. Statist. Math.*, Vol. VIII, No. 2, 1956.
3) 統計科学研究会編，新編統計数値表，河出書房，1952．
4) 水野坦 他，統計数値表の使い方――ノンパラメトリック検定――，朝倉書店，1954．