擬似乱数の評価方法の検証

理学専攻　情報科学コース　高橋絢那

1　はじめに

擬似乱数とは, 一見乱数列のように見えるが, 実際には計算機による確定的な計算により, 求められている数列である.

良い擬似乱数の条件には, 周期が長いこと・高速に生成できること・統計的検定に耐えられること, があげられる.

本研究では擬似乱数の周期と, 検定に使用するビット長の関係性を見つけ, 各擬似乱数の周期ごとに最適な評価方法を提案することを目的としている.

2　言葉の定義

ランダムウォーク　原点を出発点として「1 ならば $(1,1)$ 方向に 1 歩進み, 0 ならば $(1,-1)$ 方向に 1 歩進む」というルールで生成する折れ線グラフ.

$X_1,\cdots,X_n$ は独立な確率変数で, $P_r(X_i=1)=P_r(X_i=0)=\frac{1}{2}$ を満たすものとするとき, $S_0=0, S_i=\sum_{j=1}^{i}X_j$ で定義される確率変数列 $S_0,S_1,\cdots,S_n$ が n 歩のランダムウォークである.

正の区域の滞在時間　$y\geq 0$ の上半平面を歩いた歩数を正の区域の滞在時間という.

ランダムウォークの長さ n が偶数であれば, 正の区域の滞在時間と負の区域の滞在時間はともに偶数になる.

n 歩のランダムウォークにおいて, 正の区域の滞在時間が k 時間となる確率を $P_{k,n}$ と表す. n が偶数の時の $P_{k,n}$ の実現確率は以下の式で表される.

$$P_{k,n}=u_{2k}\cdot u_{2n-2k}\quad ,u_{2k}=\begin{pmatrix}2k\\k\end{pmatrix}\cdot\frac{1}{2^{2k}}\qquad(1)$$

χ^2 検定　観測度数と期待度数から計算して求めた χ^2 値を, χ^2 分布を用いて評価する検定.

χ^2 値は以下の式で求める.

$$\chi^2=\sum_{i=1}^{m}\frac{(O_i-E_i)^2}{E_i}\qquad(2)$$

O_i : 観測度数, E_i: 期待度数

m : グループ数

3　擬似乱数生成法

3.1　rand

C 言語の 70〜90 年代の標準擬似乱数.
線形合同法であり, 以下の式で定義されている.

$$x_{n+1}:=ax_n+c \bmod M$$

$$a=1103515245,\ c=12345, M=2^{31}$$

周期は $M=2^{31}$.

31 桁はすべて, 桁ごとに周期が異なる.

3.2　CST(Combined Small Twister)

2010 年, 山形大学の西村先生によって提案された高速で周期の短い乱数. 高速で生成できる, 周期 $2^{64}-1$ の擬似乱数 x_i と周期 $2^{89}-1$ の擬似乱数 y_i を各ビットごとに排他的論理和で足し合わせたもの. 周期は 2^{153}.

$$x_{i+2}=LROT((x_{i+1}>>7)\oplus x_{i+1}\ ,9)\oplus((x_i>>12)\oplus x_i)A$$
$$y_{i+3}=LROT((y_{i+2}>>9)\oplus y_{i+2}\ ,17)\oplus((z>>12)\oplus z)B$$
$$z=U(y_i,25)|L(y_{i+1},7)\qquad(i=0,1,...)$$

$$A=\begin{pmatrix}0&1&&&\\&&1&&\\&&&\ddots&\\&&&&1\\a_{32}&a_{31}&&\cdots&a_1\end{pmatrix}\qquad B=\begin{pmatrix}0&1&&&\\&&1&&\\&&&\ddots&\\&&&&1\\b_{32}&b_{31}&&\cdots&b_1\end{pmatrix}$$

ただし, $>>$ は右方向のビットシフト, U は上位桁を, L は下位桁をとる.
CST : $x_i\oplus y_i$.

4　検定方法

1. rand で生成した擬似乱数 (31 桁) について, 各桁ごとに様々な歩数のランダムウォークを複数本生成する.
2. ランダムウォークの正の滞在時間を測定し, 正の滞在時間ごとにランダムウォークを 3 つのグループに分ける (本検定ではグループ数 $m=3$ とする. 3 つのグループはほぼ同確率でランダムウォークが分けられるように設定).
3. 自由度 2 の χ^2 検定を行い, 乱数性を評価する. 有意水準は 0.05 とし, 棄却域は $\chi^2\geq 5.991$ である.

以上の検定方法について, 棄却された擬似乱数の周期と使用ビット長の関係性を見つける.

5　検定結果

5.1　歩数を固定, 本数を変化

歩数を 1000 歩に固定し, 本数のみを変化させた時の χ^2 値を表 1 に示す (棄却箇所はグレー背景) :

表 1: rand:1000 歩, 本数を変えたときの χ^2 値

	6 本	8 本	10 本	20 本	30 本	40 本	50 本	100 本	1000 本	2000 本
2^1bit	2.995	3.993	4.992	9.984	14.975	19.967	24.959	49.918	499.182	998.364
2^1bit	1.000293	1.00497	1.403	2.512	3.815	5.0249	6.308	12.562	125.621	251.243
2^1bit	1.000293	1.748	2.596	6.101	6.194	8.740	1.305	23.400	0.149	436.991
2^1bit	1.00680	1.761	1.405	2.512	3.822	5.181	6.314	12.562	125.634	251.243
2^1bit	0.991	1.738	0.794	1.312	3.822	3.960	6.314	11.115	124.125	251.243
2^1bit	2.979	1.740	1.385	0.688	1.799	0.194	0.524	2.023	22.719	46.725
2^1bit	4.00332	7.0107	7.409	13.322	16.827	21.995	23.122	32.106	279.224	547.310
2^1bit	1.000293	0.247	1.403	6.421	4.209	2.444	5.0773	8.197	67.6515	128.665
2^1bit	3.0124	1.761	3.816	3.609	3.200	3.335	5.282	11.776	108.254	215.0772
2^1bit	1.000293	0.247	0.194	2.770	1.373	3.598	1.450	6.296	109.624	224.897
2^{10}bit	0.992	0.248	0.796	0.700	1.386	2.132	0.501	0.490	10.314	18.653
2^{11}bit	6.967	3.227	3.775	3.0840	4.202	6.0293	6.0445	10.260	10.295	22.740
2^1bit	0.000230	0.988	3.174	2.469	0.782	3.757	3.970	2.247	1.869	2.219
2^{15}bit	3.0124	3.262	3.811	6.432	2.616	1.400	1.234	4.784	5.967	10.363
2^{15}bit	4.0075	1.751	0.204	3.126	1.414	2.467	3.892	8.597	16.367	38.683
2^{16}bit	1.00828	0.248	0.194	0.105	0.803	0.928	0.149	1.314	1.404	3.758
2^{16}bit	0.000023	0.247	0.794	0.405	0.209	0.355	0.501	0.357	2.215	1.393
2^{17}bit	0.992	0.247	0.202	1.316	0.803	0.0535	0.0456	0.127	0.0491	2.288
2^{17}bit	1.000293	1.748	2.585	1.897	2.607	2.610	3.667	0.972	0.876	0.645
2^{20}bit	4.0154	3.266	3.820	3.127	2.422	3.062	2.446	2.967	3.977	8.552
2^{20}bit	3.980	3.227	3.774	6.697	1.407	1.425	1.507	0.896	4.550	2.829
2^{21}bit	3.0124	3.266	3.216	0.700	2.616	1.566	1.234	0.237	12.469	3.254
2^{22}bit	1.00689	0.247	0.194	0.0969	0.191	0.635	0.628	1.898	0.271	2.340
2^{17}bit	1.00251	3.262	3.911	4.917	3.209	1.400	1.131	0.559	2.902	0.0897
2^{25}bit	1.00828	3.266	3.820	1.620	2.430	1.276	0.294	0.257	0.149	0.183
2^{27}bit	1.00680	1.761	0.810	1.305	1.799	0.951	0.276	0.0913	1.144	6.567
2^{26}bit	7.0146	3.262	1.403	1.305	0.799	0.643	1.219	3.106	1.112	0.941
2^{27}bit	7.0212	6.272	3.811	0.690	0.209	0.648	0.632	0.540	3.605	1.0348
2^{28}bit	3.0124	4.769	7.427	0.105	1.413	1.859	1.131	0.748	0.590	0.659
2^{29}bit	1.00689	3.262	3.816	0.712	0.000115	0.0524	2.692	9.787	1.696	4.114
2^{30}bit	1.00251	1.00746	3.216	6.432	15.0621	6.985	5.348	1.493	5.172	0.740

表 1 より, 歩数を固定すると本数が増えるほどに棄却しやすくなるということがわかった.

5.2　本数を固定, 歩数を変化

本数を 1000 本に固定し, 歩数のみを変化させた時の χ^2 値を表 2 に示す (棄却箇所はグレー背景)：

表 2: rand:1000 本, 歩数を変えたときの χ^2 値

	6 歩	8 歩	10 歩	20 歩	30 歩	40 歩	50 歩	100 歩	1000 歩	2000 歩
2^0bit	300.000	496.103	668.934	659.370	432.0054	484.817	519.783	525.974	499.182	499.589
2^1bit	33.333	216.132	90.917	58.355	195.0652	141.276	113.654	107.497	125.621	125.310
2^2bit	33.333	226.521	279.388	246.030	37.929	214.537	377.616	369.606	0.149	218.622
2^3bit	0.0165	28.934	71.495	245.507	36.950	141.438	128.730	215.750	125.634	100.637
2^4bit	8.467	1.0789	15.450	80.563	92.492	101.718	0.0876	106.034	124.125	123.821
2^5bit	2.323	10.373	28.141	12.290	45.678	85.403	6.331	104.696	22.719	22.598
2^6bit	0.00213	0.0007	3.723	1.728	50.992	95.345	107.323	56.334	279.224	278.745
2^7bit	0.207	0.262	1.201	24.064	17.412	1.521	0.384	3.545	67.0515	66.832
2^8bit	0.0873	0.370	0.271	2.741	9.461	6.395	17.528	3.322	108.254	116.477
2^9bit	0.299	0.0121	0.156	0.648	0.0790	0.0890	2.630	0.217	109.624	110.742
2^{10}bit	0.141	0.513	0.885	3.882	1.349	0.383	1.220	1.940	10.314	101.239
2^{11}bit	0.341	0.310	0.0164	0.612	1.00608	1.642	0.400	11.403	10.295	53.752
2^{12}bit	0.0873	1.435	2.520	0.186	0.192	0.0308	2.950	1.260	1.869	11.554
2^{13}bit	0.209	0.130	0.0859	0.410	1.488	4.276	2.465	6.630	5.967	34.927
2^{14}bit	0.399	1.776	0.829	0.0487	0.731	0.988	1.743	0.926	16.367	32.761
2^{15}bit	0.106	0.0649	0.213	3.189	1.327	0.394	0.474	0.0229	1.404	7.997
2^{16}bit	0.0841	0.232	0.721	0.409	1.469	0.998	0.784	3.078	2.215	1.437
2^{17}bit	1.267	0.125	0.0675	2.758	0.331	0.388	0.625	3.139	0.0491	2.848
2^{18}bit	0.064	0.0787	0.0904	0.483	2.417	3.304	3.421	1.447	0.876	0.364
2^{19}bit	0.0585	0.418	0.0550	0.606	2.183	2.137	2.398	2.406	3.977	1.828
2^{20}bit	0.789	0.769	0.720	0.646	3.622	0.569	0.0793	2.547	4.550	4.937
2^{21}bit	2.153	1.701	1.655	0.407	0.496	0.569	0.675	1.710	12.469	4.285
2^{22}bit	0.349	2.485	3.411	4.085	0.927	1.698	4.352	1.0888	0.271	0.747
2^{23}bit	0.123	0.404	0.117	0.0266	2.837	2.206	1.382	2.307	2.902	0.309
2^{24}bit	7.594	10.795	6.838	2.907	1.270	2.771	0.971	0.002	0.149	0.0481
2^{25}bit	3.963	7.364	5.941	0.240	1.850	1.171	0.283	1.501	1.144	0.279
2^{26}bit	0.0521	0.176	0.446	1.880	3.481	4.710	1.811	0.0769	1.112	0.465
2^{27}bit	0.371	1.0978	0.549	2.920	3.624	2.690	2.875	4.411	3.605	4.927
2^{28}bit	0.104	0.0225	0.443	2.684	0.595	0.0834	0.751	0.635	0.590	1.220
2^{29}bit	0.335	0.0558	0.223	1.145	0.751	0.000996	0.275	4.559	1.696	7.436
2^{30}bit	0.485	0.366	0.572	0.826	0.480	1.481	0.861	0.314	5.172	5.265

表 2 より, 本数を固定すると歩数が増えるほどに棄却しやすくなるということがわかった.

5.3　歩数と本数を様々に変化

歩数×本数を一定に保ち, 様々に変化させて検定を行なった. 歩数×本数 $= 10^6$ のときの各桁ごとの χ^2 値を表 3 に示す (棄却箇所はグレー背景)：

表 3: rand:歩数×本数 $= 10^6$ の χ^2 値

	10 歩 100000 本	20 歩 50000 本	40 歩 25000 本	50 歩 20000 本	100 歩 10000 本	200 歩 5000 本	400 歩 2500 本	1000 歩 1000 本	2000 歩 500 本
2^0bit	66893.406	32968.511	12120.430	10395.658	5259.737	2480.508	1244.982	499.182	249.764
2^1bit	9091.727	2917.740	3531.906	2273.075	1074.968	640.676	316.397	125.621	62.655
2^2bit	27938.839	12301.493	5363.423	7552.318	3696.056	1087.780	546.83	0.1496	110.823
2^3bit	7164.313	12301.493	3532.062	2580.835	2150.184	640.710	250.785	125.634	49.781
2^4bit	1570.897	4134.079	2529.086	0.981	1074.968	639.111	314.881	124.125	61.167
2^5bit	2961.768	651.970	2169.006	141.353	1073.557	182.949	61.007	22.719	11.298
2^6bit	372.731	70.337	2245.291	2045.306	527.149	644.329	679.420	279.224	141.675
2^7bit	76.461	1228.557	22.991	4.331	27.204	18.560	33.991	67.0515	33.967
2^8bit	18.514	141.892	184.716	380.684	35.0998	65.740	256.145	108.254	59.728
2^9bit	16.018	32.820	4.869	49.254	2.0422	300.895	65.117	109.624	54.378
2^{10}bit	57.055	124.955	2.256	5.233	12.903	8.0797	23.320	10.314	50.517
2^{11}bit	1.0625	68.982	2.252	1.695	37.140	8.0797	98.756	10.295	41.110
2^{12}bit	0.416	7.330	14.808	2.160	0.377	42.128	37.665	1.869	7.571
2^{13}bit	0.708	0.653	4.626	0.921	0.313	7.127	0.620	5.967	22.391
2^{14}bit	0.615	1.867	10.675	8.606	1.763	4.339	3.198	16.367	22.227
2^{15}bit	0.110	0.997	0.914	2.651	1.750	0.490	8.441	1.404	5.143
2^{16}bit	0.183	0.625	2.105	0.469	2.852	2.774	0.590	2.215	0.823
2^{17}bit	0.531	0.310	0.418	1.050	0.808	1.133	0.429	0.0491	1.557
2^{18}bit	3.070	0.264	0.0168	0.050	1.189	2.152	0.429	0.876	3.940
2^{19}bit	0.394	0.199	2.813	1.838	0.822	0.844	1.223	3.977	1.024
2^{20}bit	1.168	2.059	2.688	2.470	0.822	0.0332	0.904	4.550	1.693
2^{21}bit	3.022	1.468	1.537	1.346	2.153	4.500	0.324	12.469	1.00914
2^{22}bit	2.889	0.122	1.484	0.783	1.0854	0.128	1.411	0.271	0.356
2^{23}bit	0.385	1.072	0.646	0.331	5.158	5.763	0.00884	2.902	1.623
2^{24}bit	0.553	0.847	4.157	1.900	0.171	1.741	0.246	0.149	0.366
2^{25}bit	0.101	3.008	4.143	1.025	0.611	2.484	0.700	1.144	0.723
2^{26}bit	0.638	1.705	0.851	1.869	0.869	3.586	0.298	1.112	0.0351
2^{27}bit	0.0876	0.825	0.185	1.184	0.355	1.0345	0.866	3.605	0.564
2^{28}bit	0.743	0.312	1.418	0.108	1.320	6.968	0.106	0.590	3.718
2^{29}bit	0.728	0.413	2.023	0.144	0.227	3.0955	0.411	1.696	9.193
2^{30}bit	0.235	1.723	2.001	0.342	0.745	0.490	0.177	5.172	4.983

表 3 から, 歩数×本数を一定にしてバランスを変えても, 棄却される桁数にはあまり差がなく, 使用ビット長（歩数×本数）が検定結果に影響しているということがわかった.

同様に, 歩数×本数$= 10^5, 10^4, 10^3$ の場合も検定を行なった.

5.4　周期と使用ビット長の関係性

5.1, 5.2, 5.3 の結果より, 擬似乱数の周期と検定に使用したビット長（歩数×本数）の 2 視点から棄却された箇所を表したものを図 1 に示す：

なお, 本研究では同じ検定を二度ずつ行なった. 一度目は最初の n 本を用い, 二度目は同じ初期値の次の n 本を用いた. 図 1 では, 二度のうち一度でも棄却された箇所はプロットしている. また, 各検定において飛び地は考慮せずに棄却された最上位桁付近で直線を引いている. この直線の傾きは, 乱数性を表していると考えられる.

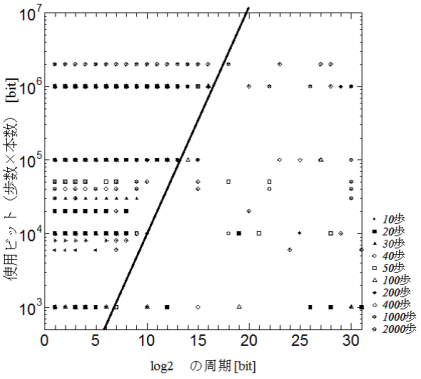

図 1: 周期と使用ビット長の 2 視点から見る, 棄却された箇所

図 1 より, 使用ビットが増えるに連れて, 周期の長い擬似乱数も棄却できるようになっていることがわかる.

10^{11} 程度の長さを検定に用いれば, 周期 2^{31} の擬似乱数が評価できると期待できる.

5.5　本検定を CST へ適用

CST へ本検定を適用したときの χ^2 値を表 4 に示す (棄却箇所はグレー背景)：

表 4: CST:100 歩, 本数を変えたときの χ^2 値

	50 本	100 本	200 本	500 本	1000 本
2^0 桁	0.006	0.140	0.947	0.066	0.072
2^8 桁	0.149	0.017	0.119	1.929	2.337
2^{16} 桁	2.016	4.175	4.585	4.159	3.526
2^{24} 桁	0.904	0.105	0.134	1.206	2.406
2^{30} 桁	0.437	0.418	0.817	0.023	0.825
2^{31} 桁	0.444	0.667	1.425	4.512	4.819

表 4 を見ると, rand よりも棄却される桁が少なく, 乱数性を表す直線の傾きが大きくなることを予想していたが, 予想以上に CST の乱数性が高く, 使用ビット長によらず, CST が棄却されることはなかった.

よって, 本検定方法においては, rand により生成された擬似乱数よりも CST により生成された擬似乱数の方が乱数性が高いと言える.

6　まとめと今後の課題

ランダムウォークを用いて擬似乱数を評価する本検定では, 検定に使用するビット長を増やすにつれて, 周期の長い擬似乱数も評価できるようになることがわかった.

本検定の結果より, 10^{11} 程度の長さを検定に用いれば, 周期 2^{31} の擬似乱数が評価できると期待できる.

今後は, 本検定において他の擬似乱数生成法について, また, 別の評価方法についても, 周期と使用ビット長の関係性を明確にすることで, 各擬似乱数の周期ごとに最適な評価方法を提案したい.

参考文献

[1] 津野義道：ランダム・ウォーク 乱れに潜む不思議な現象, 牧野書店, 2002

[2] 草間武：統計学, サイエンス社, 1975

[3] 東京大学教養学部統計学教室 編：統計学入門, 東京大学出版会, 1991