デタラメさの効用と、１＋１＝０の世界

松本 眞（東京大学数理科学研究科）

2012年1月18日　東京大学学術俯瞰講義

第壱話　デタラメさの効用：意外なところで
　ランダムネスのお世話に
第弐話　デタラメさを生み出すのは意外に難しい：
　（デタラメ禅問答）
第参話　近現代数学によるデタラメさの生成
　（メルセンヌ・ツイスター）

第弐話「デタラメさを生み出すのは意外に難しい」(天之章)

デタラメさを生みだすむずかしさ：

さいころ：

- 偏っているかもしれない
- おそい。計算機シミュレーションでは何兆個も乱数を使う
  計算機に取り込むのも大変
- 途中でさいころの角が欠けたら？
  物理的なランダムネス生成法は、故障の危険がある。

乱数表

人力でランダムネスを作りだす難しさから乱数表が誕生

第壱回で述べた「世論調査」のように、巨大な母集団からランダムにサンプリングを行うには乱数は不可欠である。

特に第二次産業革命（１９世紀終わり）ごろから、品質管理の目的で、「大量に生産された製品からサンプルをランダムに選び出し、検査を行う」といったことが行われるようになった。

ランダムネスの生成には、サイコロや、
「数を記入した大量のカードを帽子の中にいれて、人間が無作為に抽出する」
といった方法が実際に使われた。

L.H.C Tippettは、多くの紙片に数を書き、箱から無作為に抽出するという実験を繰り返したところ、得られた乱数列には大きな偏りがあることを発見した。偏りがなくなるまで紙片を混ぜるのは大変であり、そのような努力を何度も払う手間を考えれば、乱数を記入した表を出版することに意義があると判断した。

⇒世界初の**乱数表の出版**(Tippett, 1927年)

Tippett: 数理統計学の始祖 Karl Pearson（カイ自乗検定の考案者）の弟子

乱数表：0から9までの数字をデタラメ（一様独立）に選んで並べ、表として印刷したもの（せいぜい百万個くらい）。乱数表を読み始める始点は、鉛筆を投げて決める。

１９８０年代ごろまで、ランダムサンプリングなどに広く用いられてきた。

欠点の一つ：「予測不可能な数列」を作るのには向かない。

例えば、年末ジャンボ宝くじの組・番号は７ケタある。７ケタの数は$10^7 = １０００$万種類ある。長さ１０万の乱数表から始点を変えるだけでは、１０万通りの数しか作られない。よって、たからくじの当選番号を決めるのには乱数表は適切ではない。

同様に、電子文書にパスワードを掛ける際にも、乱数表の使用は適切ではない（１０万通りをコンピュータで調べるのは難しくない。）

二つ目の欠点：　計算機の発展により、乱数を何億、何兆個も消費するような応用が生まれた。乱数を電子的な表として記録しておくとすると、量が膨大すぎる。

物理乱数発生器　電気回路の熱雑音や放射性物質の崩壊と言った、物理的に存在するデタラメさを計測して乱数を発生する
実際に使われて来たし、販売もされている
高価、比較的低速（測定時間）、故障の危険性、

少し脱線:**デタラメさの保証の難しさ**（哲学的に）
トランプ52枚のカードの並べ方は$52! = 8.065\cdots \times 10^{67}$
1分程度トランプのカードを切り混ぜることで、これだけの並べ方を全て均等に作り出せるのか？

参考：宇宙の素粒子の数 10^{80} 個くらい（カール・セーガン）

疑似乱数の登場

von Neumannは1940年ごろコンピュータの演算のみを用いて＝漸化式によって、「乱数のように見える数列」を発生して核反応シミュレーションに利用し、大きな成功を収めた。

ある漸化式

$$x_n = f(x_{n-1})$$

と初期値x_0に対し、

$$x_1 = f(x_0), x_2 = f(x_1), x_3 = f(x_2), \ldots$$

によって数列を発生し、それを乱数列（デタラメな数の列）であるかのように利用する。

　このように、漸化式によって「乱数のように見える数列」を発生する方法を**疑似乱数発生法**という。

平方採中法von Neumann, 1940 ごろ？発表は1949

例10進4ケタの場合

2012

$\rightarrow 2012^2 = 04048144 \rightarrow 0481$

$\rightarrow 0481^2 = 00231361 \rightarrow 2313$

$\rightarrow 2313^2 = 05349969 \rightarrow 3499$

$\rightarrow 3499^2 = 12243001 \rightarrow 2430$

$\rightarrow 2430^2 = 05904900 \rightarrow 9049$

$\rightarrow 9049^2 = 81884401 \rightarrow 8844$

$\cdots$

- 異なる擬似乱数列を得るには、初期値を変える。
- 初期値（短い）をランダムに選べば、長い擬似乱数列が得られる。

平方採中法にはいろいろな問題がある。
例
$3792 \to 3792^2 = 14379264$　　周期1
$0540 \to 0540^2 = 00291600 \to 2916^2 = 08503056 \to 5030^2 = 25300900$
$\to 3009^2 = 09054081$　　周期4

平均的な周期は（4ケタの場合）100前後。
von Neumannはもっと大きな桁数を用い、かつ周期や分布を監視するプログラムを併用した。

実際にもっとも広く使われた擬似乱数は、

- Lehmerの**線形合同法**(1960〜90後半)... 循環小数の一般化
- Mersenne Twister('96〜現在)などの$\mathbb{F}_2$生成法

→次回のテーマ＝人之章（不可能ならごまかしてでもやる）

疑似乱数の特徴

- 高速で安価（漸化式をうまく選べば高速。センサーで取り込む物理乱数に比べて安価。）
- 再現性がある（漸化式と初期値を公開すれば、誰でも同じ疑似乱数列を生成できる）。
- →ストリーム暗号（前回参照）に使える。
  漸化式を公開し、短い初期値のみを秘匿・交換して、生成された長い擬似乱数でメッセージを攪乱。
- 宝くじの当選番号を決めるのには不向き。
  （物理的乱数生成法のような公平性があるのか不明）

哲学的・実用的問題(禅問答)
このような数列を乱数として用いることの正当性は？
いまだに良く分からない。それは、定義の不在による。
「デタラメな数列とは何か。定義せよ。」

von Neumann:「演算で乱数を作りだすのは一種の犯罪」

実用上も問題：定義がないので、未だに新しい疑似乱数
発生法が提唱されては、欠陥が見つかる

ランダムネスを定義する３つのアプローチ

- 情報エントロピーによるアプローチ（確率論的）
- 計算可能性によるアプローチ（計算機での生成不可能性）
- 計算量によるアプローチ（計算時間による困難性）

それぞれ成功と失敗の側面を持つ。
(絵に描いたもち)

現在、広く用いられている疑似乱数は「妥協の産物」だが
「近現代数学をうまく利用」して実用的で高性能である
（次回のテーマ＝人之章＝「無理ならごまかしてでもやる」）

情報エントロピー:不確定さの量(Shannon 1948)

確率1/2で表(0)裏(1)が投げるたびに独立に決まるコインがある。このとき、コインを一回投げて得られる結果の
　「不確定さの量」
を**1ビットのエントロピー**という。

より一般に、ある事象Xが$1, 2, \ldots, m$のm通りの値をとり、それぞれの確率が$p_1, p_2, \ldots, p_m$であるとき、この事象の持つエントロピーは

$$\sum_{i=1}^{m}(p_i \log_2(1/p_i))$$

で定義される。

例：上のようなコインを１個投げる場合、結果は0,1の２通りがある。この確率現象の持つエントロピーは

$$\sum_{i=1}^{2} \frac{1}{2} \log_2(2) = \log_2(2) = 1$$

で、1bitのエントロピーを持つ。

コインを同時に２個投げる場合、00,01,10,11の４通りがありこの確率現象の持つエントロピーは

$$\sum_{i=1}^{4} \frac{1}{4} \log_2(4) = \log_2(4) = 2$$

で、2bitのエントロピーを持つ。

例：サイコロの場合

$$\sum_{i=1}^{6} \frac{1}{6} \log_2(6) = \log_2(6) = 2.5849\cdots$$

したがって、サイコロを一回振って得られるデタラメさの量は、コインを2.58回投げて得られるデタラメさの量とほぼ等しい。

→サイコロ１００回と、コイン２５８回とがほぼ等しい。

$$6^{100} = 6.53 \times 10^{77}, 2^{258} = 4.63 \times 10^{77}$$

コイン三つあれば、サイコロ一個の代用ができる。
コインを3個なげると000,001,010,011,100,101,110,111の8通りがデタラメに得られる。二進数だと思うと0から7。このうち、0と7が出たら飛ばすことにすればよい。

事実(Shannonの観察)
確率的要素のない操作によっては、エントロピーを増やすことはできない。

例：ランダムに選ばれる初期値から擬似乱数列を発生する場合は、初期値の選択のときに取りこまれるエントロピー以外にエントロピーは増やせない。

注：働いている電子部品は雑音を出しエントロピーを発生させつづけている。デジタルコンピュータは、このような雑音によって結果が左右されないよう、信号を0-1の二値に分けて、常に決定的な動作をするように非常な努力を持って設計されている（部品のもつエントロピーが計算結果に全く反映されない)。

エントロピーによる結論
　確定的に動作する計算機で、エントロピーの意味で
　デタラメさを生み出すことはできない。

不可能の証明：やろうとしてもできないことがある(天之章)

注：エントロピーを取り込むことはできる。実際、
擬似乱数発生法の初期値を決める際には、計算機の時刻、
ディスクへのアクセスタイミングやキーボードタッチの
タイミングなどからデタラメさを取り込むことがある。

脱線：
脳はエントロピーを生み出せるのでしょうか。
きまぐれ。

禅問答：デタラメさを定義せよ

与えられた一つの有限数列が乱数列であるかどうか、定義することができるか？

エントロピー概念は、確率的現象に対して定義される。一つの固定した数列が「デタラメかどうか」は考慮できない。

デタラメな６ケタの数字には、000000も142857も123456も、全て同じ確率$1/10^6$であらわれる。

与えられた数列に対し「乱数列であるかいなか」を定義することは、不可能に思える。この困難さを解決（？）したのがコロモゴロフ-チャイティンである。

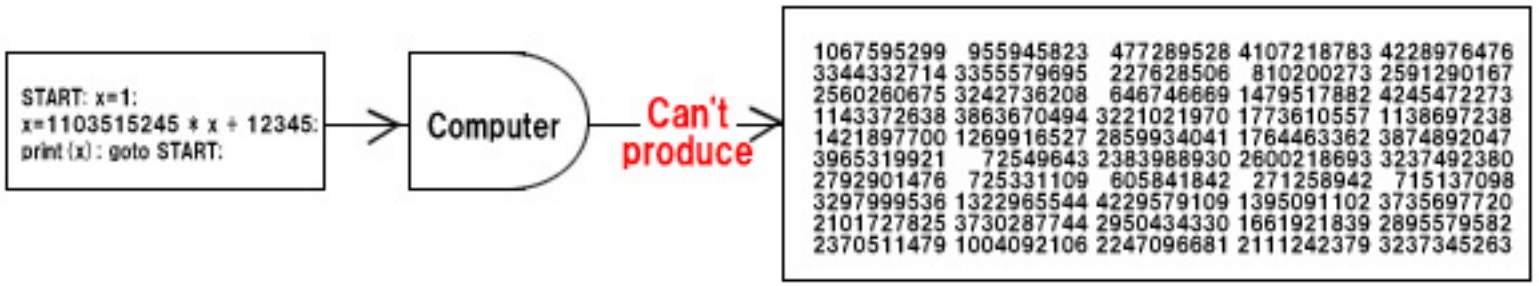
Kolmogorov-Chaitin 複雑性による乱数の定義

定義(Kolmogorov-Chaitin 60年代)

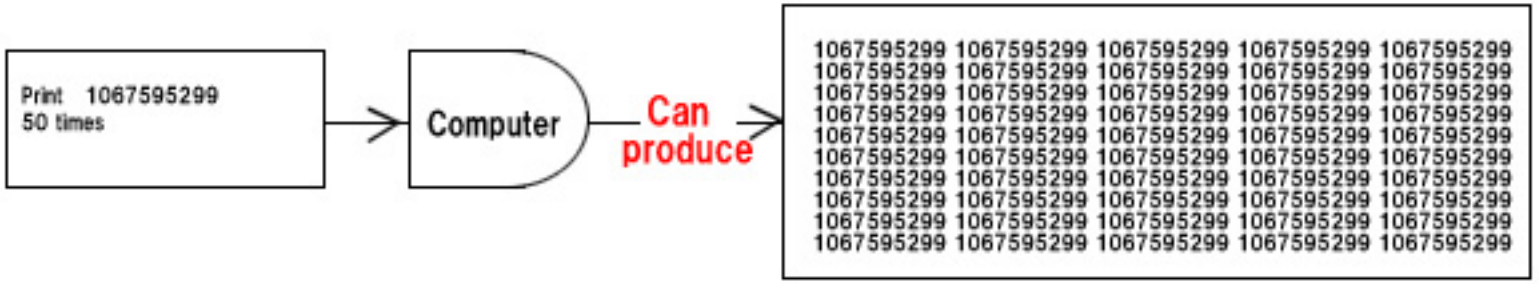
- 有限数列に対し、その数列を生成するのに必要な最小のプログラムの長さを、数列の Kolmogorov-Chaitin 複雑さという。
- 有限数列に対し、その数列を生成するのに必要なプログラムの長さが、数列の長さより短くできないとき、その数列を Kolmogorov の意味で「乱数である」という。

比ゆ的にいえば、Kolmogorov の意味での乱数とは
「全部記録しておく以上に効率的な生成方法が存在しない数列」パスワードで言えば「丸暗記以外に良い覚え方のない文字列」となる。
注：確率論を用いていない。計算可能理論を用いる。

より短いプログラム　　では発生のできない数列



短いプログラムで長い数列が発生できる例



Kolmogorovの意味での乱数列はたくさん（高い密度で）存在していることが証明でき、そのような数列に対する数学的研究はすすんでいる。

しかし、効率的乱数発生の立場からすると、
「Kolmogorov-Chaitinによる乱数の定義」は今のところ
「絵に描いたもち」（どんなに美しくても、使えない）。

なぜなら、「計算機で効率的に発生できる数列は、全て乱数でない」ということになってしまうから。
それどころか

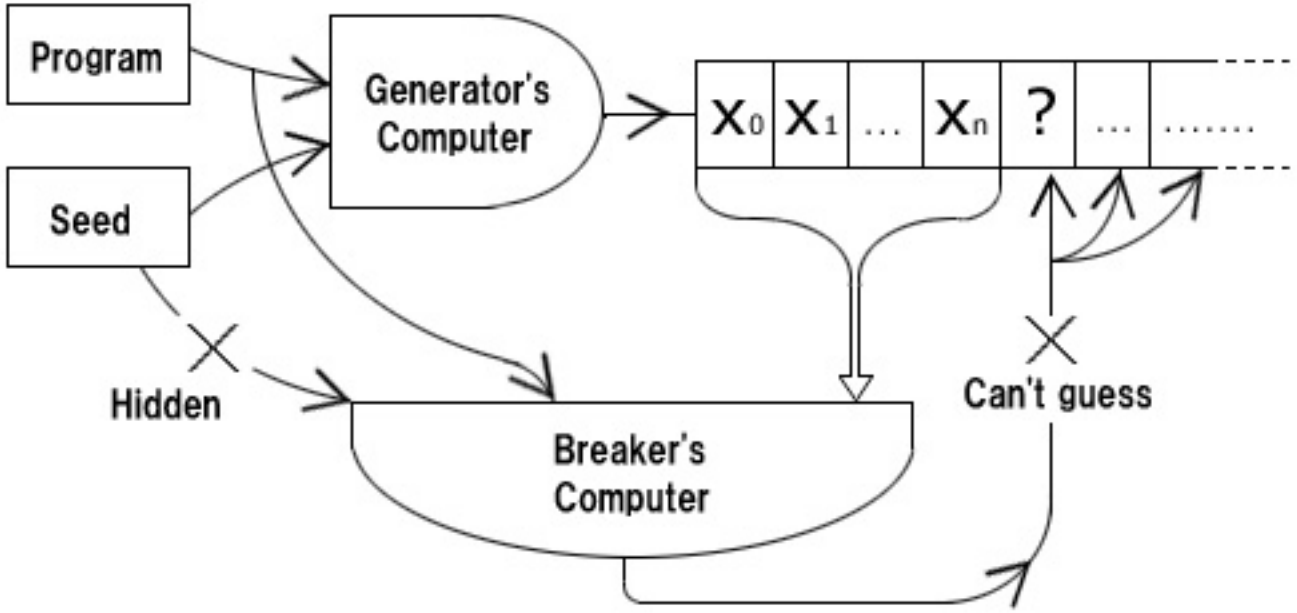
定理数列に対し、そのKolmogorov複雑性を計算するアルゴリズムは存在しない。

Gödelの不完全性定理、チューリング機械の停止判定不能性などと同様の「定義はできるけれども計算はできない」例

（不可能性の証明：天之章）

ベリーのパラドックス
「２０字以内では定義できない最小の自然数」

計算量理論による乱数の定義（Blum-Blum-Shub ’86）
数列であって、シードが与えられれば多項式時間で計算できるが、シードを与えられなければ数列の一部をいくら見ても、他の部分を推測することが多項式時間ではできない。



計算量理論による乱数の定義隠された鍵を持っていれば効率的に生成できる数列であるが、鍵を持っていない人にとってはどんなコンピュータを使っても、その数列に関する一切の予測ができず、真の乱数と区別することが、現実的な計算時間ではできない。

このような数列発生法は、ある種の予想が正しければ作れる。例えば整数の素因数分解問題が「多項式時間クラスP」に属さないことが証明できればよい。

注：もし、このような数列発生法があれば、
100万ドル懸賞金付きクレイ数学未解決７大問題の一つ、
$P \neq NP$ が解けたことになる。
もし、予想に反して$P = NP$だったとすると、このような乱数は存在しないし、ほとんどの電子的暗号は危険になる。

まとめ

エントロピー：不確定さの量を確率論で定義。
推測不能な数列を得るには、十分なエントロピーが必要。
計算機では作り出せない。一つ一つの数列は扱えない。

Kolmogorov-Chaitin 複雑性：確率論を使わずに「乱数」を定義した。一つ一つの数列に対して定義できる。
計算できない。
効率的に発生可能な数列は、この意味では乱数ではない。

計算時間による定義：秘密鍵と計算時間を用いて、乱数を定義した。
鍵があれば効率的に数列が作られ、鍵がなければ現実的な時間ではなんらの規則性も見つけれられない。
そのような発生法の候補はあるが、発生には時間がかかる。

以上、第弐話：天之章
デタラメさを生みだすのは意外に難しい：
ていうか、不可能？

次回予告：第参話 近現代数学によるデタラメさの生成
（人之章：妥協の産物）