確 率 と 乱 数

杉 田 洋

(大阪大学大学院理学研究科)

「確率」は日常的に使われている言葉ですし，高校の数学で習いますので，皆さん，それぞれいろんな印象をお持ちだと思います．一方,「乱数」という言葉を聞かれたことがあるでしょうか．デタラメな数列のことです．高校の数学の教科書 (数学 C) には「乱数表」が載っていますが，習わなかった人も多いでしょう．今日は両者についてできるだけきちんとした話をしようと思います．

1 確率論の目的

確率を計算する目的はいろいろあります．そもそも確率を考えるようになったきっかけは，どうやら賭け事にあったようです．たとえばサイコロを使うゲームをするとき，どういう手を選ぶのが有利か，とか，途中で中止しなければならなくなった賭けにおいて賭け金を公平に配分するにはどうすればよいか，等々．現在のゲーム理論や数理ファイナンスの分野はまさにこの延長線上にあります．それから，不確かな現象を予測するのに確率が計算されます．降水確率とか，破産確率とか，リスク管理に役立ちますね．しかし確率論の第一の目的は，やはりランダムな現象の解析でしょう．それも個々のランダムな現象を調べるというよりは，様々なランダムな現象に共通して見られる普遍的性質──「ランダム性」とでも呼びましょうか──を詳しく調べることこそが確率論の第一の目的です．

では，ランダム性の解析のためには私たちは何をすればよいのでしょうか．その答えは「確率」という言葉の源に見出すことができます．

「確率」はもちろん日本語です．英語の probability の訳です．probability は probable の名詞形，だから「probable なこと」を意味します．probable というのは「起こりそうな」という意味ですが，私たち日本人にとっては probable より副詞形の probably の方が親しみがありますね．probably を私は「多分(たぶん)」と訳すと習いました．皆さんはどうでしたか.「多分」を表す副詞は他にもあります．maybe，perhaps，possibly，likely，等々．これらの副詞を確実性の低いものから並べると，possibly<maybe ≈ perhaps<likely<probably となるんだそうです．probably は最も確実性の高い「多分」なのですね．あるアメリカの子供向けの (英英) 辞典[注1]には

$$\text{probably} = \text{almost surely}$$

---

[*] 日本数学会年会市民講演会 (2013 年 3 月 24 日)

[注1] *Macmillan Dictionary for Children*, Robert B. Costello (Ed.), Simon & Schuster, 2001.

とあります.[注2]だから probably の日本語訳として「十中八九」と書いてある英和辞典[注3]もあります．確率論を英語でいうと probability theory ですから，これを直訳すれば

十中八九確実なことに関する理論

ということになります．数学の確率論のことを何も知らない英語圏の人が probability theory という言葉を聞くと，きっとそんなふうな印象を持つのではないでしょうか．ちなみに中国語では確率のことを「概率 (gàilǜ)」といいます.「概」は「ほとんど，おおよそ」の意味です.

じつは面白いことに，確率論の第一の目的「ランダム性の解析」のためにはまさに「十中八九確実なこと」を詳しく調べることが重要なのです．そう，確率論は本当に「十中八九確実なことに関する理論」ということができるのです．それはどうしてか，説明していきましょう.

2　乱数とは

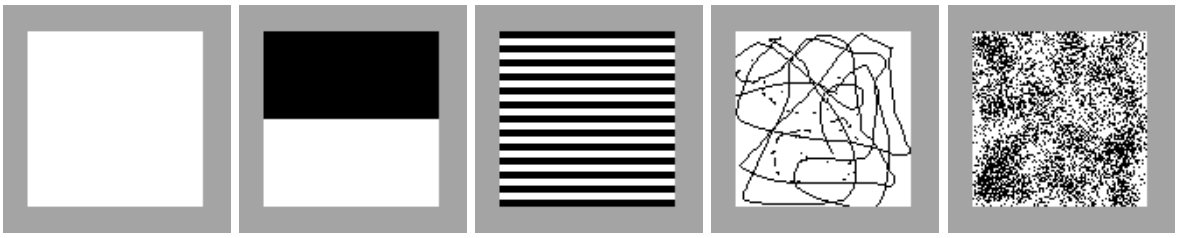
そもそも「ランダムである」とはどういうことでしょうか．この問いに答えるために考え出されたのが「乱数」という概念です．突き詰めれば，確率論の第一の目的は「乱数の性質を詳しく調べること」ということになります.

これからお話しする乱数の定義は，1960 年代にコルモゴロフ，チャイティン，ソロモノフらがそれぞれ独立に与えたものです．それを現代風にアレンジしてご紹介します.

2.1　情報の圧縮とランダム性

「ランダムである」とはどういうことか，を考えるために，その正反対の「ランダムでない」, すなわち「規則的である」とはどういうことか，を考えてみましょう.

図 1: 画像の圧縮



583 バイト　589 バイト　666 バイト　1853 バイト　2967 バイト

[注2]これは確率論研究者にとってはちょっとショックですね！
[注3]ランダムハウス英和大辞典，小学館，1973 年.

データの規則性を利用してコンピュータを有効に活用する技術があります．情報の圧縮です．[注4]画像の圧縮の例を見てみましょう．図 1 の 5 個の画像[注5]はそれぞれ 100×100 画素でファイルサイズはすべて同じ 40054 バイトです．これらは左から右へ順に規則正しいものからランダムなものへと並んでいます．各画像の下に書かれたバイト数は ZIP 形式で圧縮した場合のファイルサイズです．規則正しいほど圧縮の効率が高いことが分かります．

圧縮されたデータをもう一度圧縮するとどうなるでしょうか．やってみると，もうそれ以上，圧縮されなくなります．[注6] 一度圧縮されてしまうと，もう規則性が取り尽くされているので圧縮できなくなるのでしょう．だから「圧縮できないデータは規則性がない」といってよいと思われます．規則性がない，とはデタラメである，ランダムである，と言い換えることができます．

圧縮方法は ZIP のほかにも tar や gz などいろいろあります．それらだけでなく，考え得るすべての圧縮方法で圧縮されないようなものをランダムと呼びます．たとえば円周率 π は，現在，数兆桁が計算されています．それを印刷しますと数億ページになりますが，その数字の並びはランダムではありません．実際，π を数兆桁計算するプログラムはせいぜい数十ページ程度の短いものです．つまり π の数億ページにおよぶ数字の並びは，数十ページのプログラムというとても小さな情報に圧縮可能なのです．

2.2 デジタル符号化と乱数

森羅万象を数学で扱うことができるわけではありませんので，自ずと制限がつくでしょうが，できるだけ多くの物事の規則性，裏を返せばランダム性を測りたい．ランダムな数字，ランダムな文章，ランダムな画像，ランダムな音声，ランダムな映像...，そういうものを調べたい．幸いなことに，数字も，文章も，画像も，音声も，映像も，それらすべてコンピュータで扱うことができますよね．それは「デジタル符号化」によって，すべての情報はコンピュータでは有限個の 0 と 1 からなる列── $\{0,1\}$-列と呼びましょう──として記録され，複写され，配信されるからです (図 2)．だから，物事のランダム性を調べたければ本質的に $\{0,1\}$-列のランダム性について調べればよいのです．

コルモゴロフらは乱数を次のように定義しました．

定義 1　長い $\{0,1\}$-列で，どのような方法でもほとんど圧縮することができないものをランダムな $\{0,1\}$-列，すなわち**乱数**という．

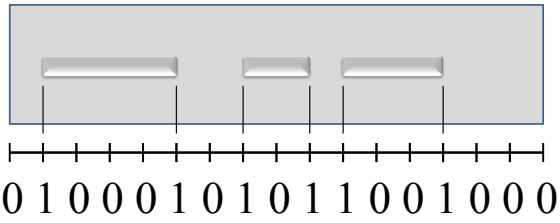
乱数の定義はいささか曖昧にならざるを得ません．乱数である $\{0,1\}$-列と乱数でない $\{0,1\}$-列の間に明確な境界線を引くことができないからです．また，ここでは情報の圧縮という日頃馴染みのある概念を用いて乱数を定義しましたが，ありとあらゆる情報の圧縮法を考えて厳密に議論するにはじつは計算論という分野の知識が必要になります．

[注4] ここでは圧縮されたデータから元のデータが完全に復元できる「可逆圧縮」だけを扱います．

[注5] bitmap 形式です．

[注6] それどころか実際には逆に少し増えてしまうようです．

図 2: CD や DVD におけるデジタル符号化



平らな部分 (ランド) とくぼんだ部分 (ピット) の境界部分で 1 をそれ以外の部分で 0 を記録する.

乱数のお話を初めて聞く皆さんには定義 1 はピンとこないかも知れません．文字式を習ったばかりの中学一年生には「$x^2+y^2=1$ は円を表す」といってもピンとこないでしょう．それと同じです．でも高校三年生になったら $x^2+y^2=1$ ほど真ん丸なものはこの世に存在しないことがよく分かるようになります．定義がピンとくるまでにはそれなりの経験が必要なのですね.

3　硬貨投げはランダムか？

硬貨を投げて，表が出たら 1，裏が出たら 0，を次々に記録してできる $\{0,1\}$-列はランダムだといえるでしょうか？ ランダムに決まってるって？ 本当ですか？ それを確かめるには，もちろん圧縮してみればよいのです．圧縮できれば規則的であり，圧縮できなければランダムです.

圧縮できるかどうか，を問うのですから，あまり短い $\{0,1\}$-列では意味がありません．長い $\{0,1\}$-列，たとえば長さ 10000 の $\{0,1\}$-列について考えることにしましょう．いま，10000 回の硬貨投げの結果が，たまたま

$$\overbrace{0,0,\ldots,0}^{5000},\overbrace{1,1,\ldots,1}^{5000} \tag{1}$$

となった，としましょう．さすがに，これはランダムとはいえないでしょうね.「圧縮できる」からです．(1) という表現自身が長さ 10000 の $\{0,1\}$-列を圧縮したものではありませんか．意地悪な例ですね．つまり硬貨投げの結果であっても，ランダムでないこともあるのです．しかし，このように 10000 回の硬貨投げの結果がランダムでない $\{0,1\}$-列，すなわち圧縮可能な $\{0,1\}$-列になることは滅多(めった)に起こりません．そのことを示してみましょう.

その前に，硬貨投げについて確認します．10000 回の硬貨投げの結果，長さ 10000 の $\{0,1\}$-列が得られますが，それは全部で 2^{10000} 通りあり，そのどれもが同じ確率 $1/2^{10000}$ で実現されます．高校数学の表現ですと,「どの出方も同様に確からしい」ということです．だから，それらのうちどれがランダムでどれがランダムでないか，は確率では分かりません.

本題に戻ります．10000 回の硬貨投げの結果がランダムでない $\{0,1\}$-列になることは滅多にないことを示そうとしているところでしたね．

長さ 10000 の $\{0,1\}$-列のうち，たとえば長さ 9990 の $\{0,1\}$-列に圧縮できるものはどれくらいあるでしょうか．圧縮された結果はすべて異なっていなければなりませんから，その個数は長さ 9990 の $\{0,1\}$-列の総数 2^{9990} を越えません．同様に圧縮された結果が長さ 9989 の $\{0,1\}$-列になるものの個数は 2^{9989} を越えません．この議論を繰り返すと，圧縮された結果が長さ 9990 以下の $\{0,1\}$-列になるものの個数は高々

$$2^{9990} + 2^{9989} + \cdots + 2^2 + 2^1 \tag{2}$$

であることが分かります．さあ，この足し算の答えはいくらでしょう.「等比数列の和の公式」を知っている人には易しいですが，知らなくても大丈夫です．(2) の和の値を x としますと

$$\begin{aligned} 2x &= 2 \times (2^{9990} + 2^{9989} + \cdots + 2^2 + 2^1) \\ &= 2^{9991} + 2^{9990} + \cdots + 2^3 + 2^2 \\ &= 2^{9991} + x - 2^1, \end{aligned}$$

両辺から x を引けば答え

$$x = 2^{9991} - 2$$

が得られます．

さて，長さ 10000 の $\{0,1\}$-列は全部で 2^{10000} 個あるのですから，そのうち長さ 9990 以下の $\{0,1\}$-列に圧縮できるものの割合は全体の高々

$$\frac{2^{9991} - 2}{2^{10000}} < \frac{2^{9991}}{2^{10000}} = \frac{1}{2^9} = \frac{1}{512}$$

です．逆に，長さ 10000 の $\{0,1\}$-列で長さ 9990 以下の $\{0,1\}$-列に圧縮できないものの割合は 511/512 より大きいです．つまり;

命題 1　乱数は長い $\{0,1\}$-列の圧倒的多数を占める．

10000 回の硬貨投げの表裏の出方は 2^{10000} 個あり，そのすべてが同様に確からしく実現されます．ですから，命題 1 を言い換えると，10000 回の硬貨投げにおいて少なくとも確率 511/512 以上で，長さ 9990 以下に圧縮されない列が実現されるのです．つまり;

命題 2　硬貨投げによって得られる長い $\{0,1\}$-列はきわめて高い確率でほとんど圧縮できない列，すなわち乱数である．

なお，乱数の定義 1，命題 1，および命題 2 の曖昧さは有限の長さの $\{0,1\}$-列を扱っているからで，無限乱数列については曖昧さなしに記述できることが知られています．

4　極限定理

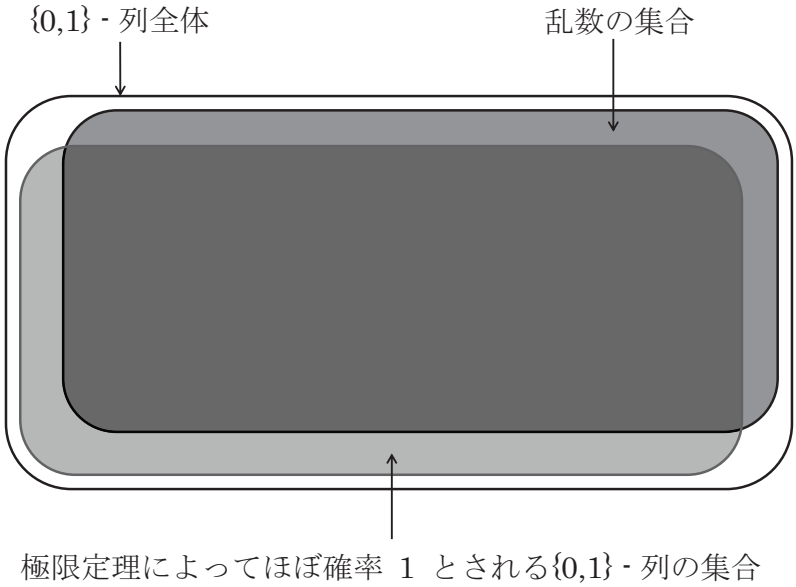
4.1　ランダム性を調べる方法

確率論の第一の目的はランダム性の解析でしたね．ランダム性を調べるには乱数の性質を調べればよいわけです．では乱数の性質を調べるにはどうしたらよいのでしょうか？

答えは意外にも簡単です．長い $\{0,1\}$-列の圧倒的多数は乱数でした．だから，長い $\{0,1\}$-列の圧倒的多数の持つ性質を調べればよいのです．硬貨投げの言葉でいえば，確率が 1 に近い事象を調べればよい，ということになります．思い出して下さい，英語 probability theory の直訳は「十中八九確実なことに関する理論」でしたよ！

確率が 1 に近い事象の性質は，ほとんどそのまま乱数の性質といってかまいません．一方，確率論でさかんに研究されている**極限定理**と総称される定理の多くは確率が 1 に近い事象の性質を記述します．ですから，それらの定理が記述する性質はほぼ乱数の性質である，といってよいでしょう．

図 3: 乱数と極限定理 (概念図)



確率が 1 に近い事象の性質を記述する極限定理は，ランダム性の解析という純粋数学の目的ばかりでなく，確率論を現実の問題に応用する目的のためにも，とてもよく研究されています．ランダムな状況であっても，ほとんど確実に起こることが予測できるから大変役立つのです．極限定理は，理論においても応用においても，確率論で最も重要な研究課題ということができます．

4.2　ベルヌーイの定理——大数の法則

今日は極限定理を一つだけ紹介します．ヤコブ ベルヌーイによって最初に発見された**大数(たいすう)の法則**です．

硬貨投げは表と裏がそれぞれ同じ確率 1/2 で出る，といっても，表と裏がいつも同じ回数だけ出る，ということではありません．試みに私が 1 円玉を 100 回投げて表が出たら 1 を，裏が出たら 0 を記録したところ，次のようになりました．

1110110101 1011101101 0100000011 0110101001 0101000100
0101111101 1010000000 1010100011 0100011001 1101111101

これら 100 個の数字の中で 1 の個数，すなわち 100 回の硬貨投げで表が出た回数，は 51 です．同じ実験を繰り返すたびに表裏の出方は変化するでしょう．でも，多くの場合，表の出る回数は 50 から大きく外れることはありません．硬貨を 100 回投げて表が k 回出る場合の数は

$$ {}_{100}\mathrm{C}_k \;=\; \frac{100!}{(100-k)!\,k!}, \quad k = 0, 1, \ldots, 100 $$

です．だから表の出る回数が 50 ± 15 以内，すなわち 35 から 65 である場合の数は

$$ \sum_{k=35}^{65} {}_{100}\mathrm{C}_k \;=\; 126538159389309434317301668 2006 $$

です．硬貨を 100 回投げるときの表裏の出方の総数は

$$ 2^{100} \;=\; 126765060022822940149670320 5376 $$

で，それらはすべて同等に確からしく起こりますから，硬貨を 100 回投げて表の出る回数が 35 から 65 である確率は

$$ \frac{\displaystyle\sum_{k=35}^{65} {}_{100}\mathrm{C}_k}{2^{100}} \;=\; \frac{126538159389309434317301668 2006}{126765060022822940149670320 5376} \;=\; 0.9982100696 $$

となります．すなわち確率 99.8% で表の出る回数は 50 ± 15 以内に入るわけです．

投げる回数を 10 倍にしてみましょう; 硬貨を 1000 回投げるときに表の出る回数が 500 ± 150 以内，すなわち 350 から 650 の間に入る確率を計算してみると

$$ \frac{\displaystyle\sum_{k=350}^{650} {}_{1000}\mathrm{C}_k}{2^{1000}} \;=\; 0.99999999999999999999991357036 93 $$

となります．これはきわめて 1 に近い確率です．区間の幅をもっと狭くして，表の出る回数が 500 ± 50 以内，すなわち 450 から 550 の間に入る確率を求めると

$$ \frac{\displaystyle\sum_{k=450}^{550} {}_{1000}\mathrm{C}_k}{2^{1000}} \;=\; 0.99860825840557791377767485889 6 $$

で約 99.9% ですね．さらに頑張ってみましょう．硬貨を 10000 回投げて表の出る回数が 5000 ± 150 以内，すなわち 4850 から 5150 の間に入る確率は

$$\frac{\displaystyle\sum_{k=4850}^{5150} {}_{10000}\mathrm{C}_k}{2^{10000}} = 0.99869463104666066665210123664458 \tag{3}$$

で，これも約 99.9% です．

大変面倒な計算をわざわざ正確に計算してみせたのは，皆さんのお持ちのパーソナルコンピュータ (PC) の計算能力を知って貰いたかったからです．最も大変な (3) の計算さえも，最近の PC は瞬時に答えを出します．[注7]オイラーもガウスも計算ばかりやっていました．そのような偉人たちがやっていた計算はこのように PC を使えば瞬時にできてしまいます．皆さんも PC を使って数学の計算をどしどしやりましょう．

さて，先ほどの計算結果を表(ひょう)にしてみました．

投げる回数	表の出る回数の区間	区間の幅 / 投げる回数	確率
100	50 ± 15	0.30	99.8%
1000	500 ± 50	0.10	99.9%
10000	5000 ± 150	0.03	99.9%

この表を見ると，各行の「確率」の欄はほぼ同じなのに,「区間の幅 / 投げる回数」の欄の数字は回数が増えるごとに小さくなっていきます．投げる回数が 10000 の場合は表が出る回数のとり得る範囲は 0 から 10000 まであるのに，実際にはその範囲のわずか 3% に過ぎない 5000 ± 150 以内に入る確率が 99.9% となります．次の定理が成り立ちます．

定理 1　どんなに小さな $\varepsilon > 0$ に対しても，硬貨を投げる回数 n を大きくすると，表の出る相対度数が区間 $(1/2) \pm \varepsilon$ に入る確率は 1 に収束する．

長さ 10000 の $\{0,1\}$-列の圧倒的多数が乱数だったことを思い出して下さい．一方，$\{0,1\}$-列の 99.9%——これも圧倒的多数といってよいでしょう——は 1 の個数が 5000 ± 150 の範囲にあります．ということは，長さ 10000 の乱数は 1 の個数が 5000 ± 150 の範囲にあると考えてほぼ間違いないでしょう．じつは詳しく調べてみますと，どんなに小さな $\varepsilon > 0$ に対しても，十分大きな n をとれば，長さ n の乱数において 1 の現れる相対度数は区間 $(1/2) \pm \varepsilon$ に入ることが証明されます．

ベルヌーイは表の出る確率が 1/2 とは限らない一般の硬貨投げの場合をも含む次の定理を証明しました．

定理 2　(ベルヌーイの定理——大数の法則，1713 年) 表の出る確率が p，裏の出る確率が $1-p$ の硬貨を投げ続ける．このとき，どんなに小さな $\varepsilon > 0$ に対しても，投げる回数 n を大きくすると，表の出る相対度数が区間 $p \pm \varepsilon$ に入る確率は 1 に収束する．

[注7]私は MacBook Pro 2.7GHz Intel Core i7 + *Mathematica* 9 で計算しました．

定量的には次の**チェビシェフの不等式**が知られています; 投げる回数が n のとき，表の出る相対度数が区間 $p \pm \varepsilon$ に入る確率は $1 - (1/4n\varepsilon^2)$ 以上である．式で書けば

$$P\left(\left|\frac{\text{表の出る回数}}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4n\varepsilon^2}. \tag{4}$$

大数の法則は今までに様々な確率変数の列について示されてきました．そして現代の確率論においても新たな大数の法則が次々と発見されています．

4.3 経験的確率

ベルヌーイの時代,「確率」はまだ数学において市民権を得た概念ではありませんでした．そもそも様々な事象の確率をどのように定めるか，が問題です．硬貨投げの場合，その刻印が表裏の出方に影響するのはわずかだろうから，表裏の出方はほぼ同等に確からしいとして，それぞれ確率 1/2 である，としてもかまわないでしょう．しかし，たとえば画鋲(がびょう)を投げるとき，針が上を向く確率はどのように考えたらよいでしょうか.「針は上を向くか下を向くか，二つに一つだから，それぞれ確率 1/2 である」というのでは乱暴過ぎます．

このような場合の確率の定め方は，おそらく賭博師たちがよく知っていたことでしょう．それは何回も画鋲を投げてみて，針が上を向く回数の相対度数を観測することです．投げる回数をどんどん大きくしていくと針が上を向く回数の相対度数がある一定の値に近づいていくことが観測されます．その値を「針が上を向く確率」と定めればよいのです．これを「経験的確率」あるいは「統計的確率」と呼びます．ベルヌーイの定理は，この経験的確率によって事象の確率を定めることが確かに可能であることを理論的に示したものということができます．このことによってベルヌーイの定理は「確率」が数学における市民権を獲得するのに大いに貢献しました．

5 モンテカルロ法

偶然によって人生が左右されたり，ランダム性は不規則で予測不能でいろんな意味で厄介です．でも役立つことだってたくさんあります．ここではベルヌーイの定理の応用としてコンピュータを用いたモンテカルロ法を紹介します．

5.1 例題

次の例題を考えてみましょう．

例題 硬貨投げを 100 回行うとき，表が続けて 6 回以上出る確率 p を求めよ．

経験的確率のアイデアを用いて例題の確率 p を求めます.「硬貨投げを 100 回行う」という試行を 10^6 回行って，そのうち表が続けて 6 回以上出る回数を S とします.「硬貨投

げを 100 回行う」という試行を「画鋲を投げる」という試行に,「表が続けて 6 回以上出る」という事象を「針が上を向く」という事象に，それぞれなぞらえて考えます．すると S は「針が上を向く確率が p の画鋲を 10^6 回投げるとき針が上を向く回数」ということになります．ですから，ベルヌーイの定理によって，n を大きくするとき，$S/10^6$ が区間 $p \pm \varepsilon$ に入る確率は 1 に収束します．チェビシェフの不等式は

$$P\left(\left|\frac{S}{10^6} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4 \cdot 10^6 \varepsilon^2},$$

とくに $\varepsilon = 1/200$ とすれば

$$P\left(\left|\frac{S}{10^6} - p\right| < \frac{1}{200}\right) \geq \frac{99}{100} \tag{5}$$

となります.

いま S は長さ $100 \times 10^6 = 10^8$ の $\{0,1\}$-列の関数であって，硬貨投げは全部で 10^8 回行わなければなりません．これを人間が行うにはあまりに手間と時間が掛かるので，コンピュータに仮想的にやらせて，$S/10^6$ の値を計算します．このとき (5) によれば，その $S/10^6$ の値は確率 99% 以上で誤差 $1/200$ 未満の p の近似値となるはずです．これが**モンテカルロ法**です.

5.2 乱数を選ぶことの不可能性

ただし，不等式 (5) に意味を持たせるためには長さ 10^8 のすべての $\{0,1\}$-列 ω を同様に確からしく選ばなければなりません．そのためには ω を主として圧倒的多数を占める乱数から選ぶべきでしょう.[注8]

ところが万能と思われるコンピュータでも乱数を選ぶことはできません．なぜでしょうか？ じつはコンピュータプログラムで長さ 10^8 の乱数 ω を得るには，どうしても長さがほぼ 10^8 の $\{0,1\}$-列を入力しなければなりません．なぜなら，もしそれより短い入力 ω' で得られるなら，ω は短い ω' に圧縮されることになり，乱数の定義に反するからです．しかもその長い入力は乱数でなくてはなりません．もしその長い入力が短い $\{0,1\}$-列に圧縮されるなら ω 自身もその短い $\{0,1\}$-列に圧縮されることになるからです．だから，コンピュータでは乱数を選ぶことはできない，というか，そもそも乱数を選ぶためにコンピュータを用いることは意味がないのです.

乱数は圧縮できない $\{0,1\}$-列でした．それは長い $\{0,1\}$-列の圧倒的多数を占めます．硬貨投げの結果できる長い $\{0,1\}$-列はきわめて高い確率で乱数です．同時に乱数はコンピュータを用いても選び出すことができません．そして，じつは $\{0,1\}$-列がどのくらいまで圧縮できるか，を計算する手続きが存在しないため，乱数が与えられても「それが乱数であること」を確かめることができません．——いかがでしょう，皆さん,「ランダムであるもの」の数学モデルとして乱数が適切なものである，という実感が湧いてきませんか.

[注8] これがモンテカルロ法に乱数が必要だといわれる理由です.

5.3 疑似乱数

実際に行われているモンテカルロ法のほとんどは，乱数の代わりコンピュータによって生成される疑似乱数が用いられています．疑似乱数を生成するために用いられるプログラム，数学的には関数

$$g : \{0,1\}^l \to \{0,1\}^n, \quad l < n$$

を**疑似乱数生成器**といいます．ここで $\{0,1\}^l$，$\{0,1\}^n$ はそれぞれ長さ l と n の $\{0,1\}$-列全体の集合を表します．l は十分小さくて，私たちは各 $\omega' \in \{0,1\}^l$ を同様に確からしく選ぶことができるとします．そのように選ばれた ω' を入力としてコンピュータは l よりずっと長い $\{0,1\}$-列 $g(\omega') \in \{0,1\}^n$ を生成します．$g(\omega')$ を**疑似乱数**，ω' をその**種**と呼びます．

疑似乱数 $g(\omega')$ は長さ n の乱数では決してありません．なぜなら——繰り返しになりますが——それは短い ω' に圧縮可能だからです．だから $\omega' \in \{0,1\}^l$ をどのように選んでも，$\omega = g(\omega') \in \{0,1\}^n$ を同様に確からしく選ぶことはできません．このことから，疑似乱数は妥協の産物であり，疑似乱数を用いたモンテカルロ法は数学的に正当化できない，と考えられていました．

しかし，じつはモンテカルロ法に乱数が絶対に必要というわけではないのです．例題の場合はベルヌーイの定理を利用するのでした．乱数でないけれども

$$\left| \frac{S(\omega)}{10^6} - p \right| < \frac{1}{200}$$

を満たす $\{0,1\}$-列 ω が——わずかですが——存在します．そのような ω ならばコンピュータで生成できる可能性があります．それが疑似乱数です．

実際，例題の場合には，うまく疑似乱数生成器 $g : \{0,1\}^{238} \to \{0,1\}^{10^8}$ を定義して，各 $\omega' \in \{0,1\}^{238}$ を同様に確からしく選ぶとき，(5) と同様の不等式

$$P\left(\left| \frac{S(g(\omega'))}{10^6} - p \right| < \frac{1}{200} \right) \geq \frac{99}{100}$$

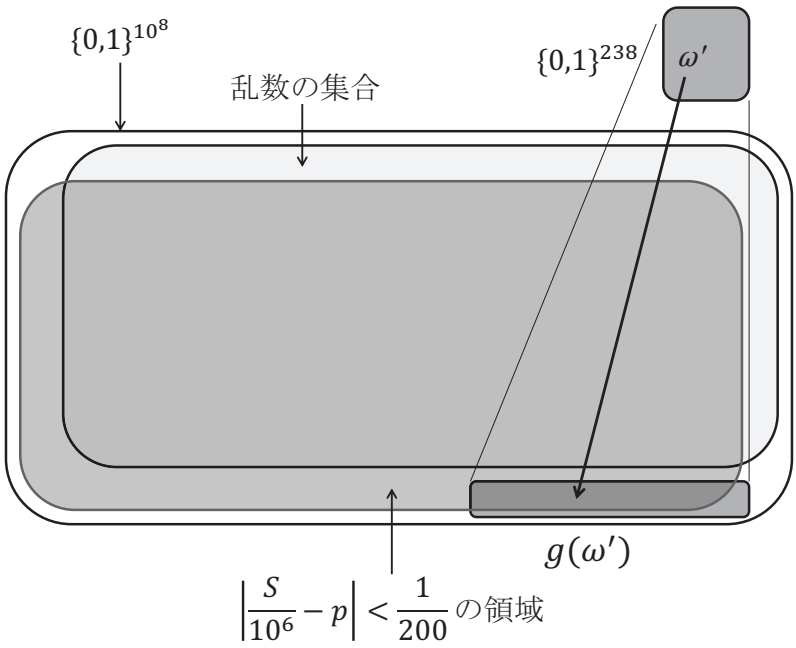
が成り立つようにできるのです (ランダム-ワイル-サンプリング (RWS 法)，図 4)．$\omega' \in \{0,1\}^{238}$ なら私たちは自分の意思でどれでも選ぶことができますね．また本物の硬貨を 238 回投げて，その結果として ω' を得れば，人の意思によらない無作為なサンプリングも可能です．だから例題の場合は長大な乱数は必要なく，g の生成する疑似乱数によって高い確率で p の値の精度のよい近似値が求まります．

例題は硬貨投げに関する確率の問題でしたが，現実の問題は硬貨投げよりずっと複雑です．しかし，じつはどのような問題でも硬貨投げの問題に還元して考えることが可能なので，疑似乱数を $\{0,1\}$-列に限って考えてもかまいません．

参考．詳しく知りたい人のために，きちんと定式化しておきます.[注9] S は $\{0,1\}^{10^8}$ で定

[注9] 詳しくは H. Sugita，*Monte Carlo method, random number, and pseudorandom number*, MSJ Memoirs vol.25，Chapter 2 を参照して下さい．

図 4: RWS 法による疑似乱数生成器 $g : \{0,1\}^{238} \to \{0,1\}^{10^8}$ のはたらき (概念図)



義された関数でした．詳しく書くと次のようになります;

$$S(\omega) \;=\; \sum_{k=1}^{10^6} X(\omega_{100(k-1)+1}, \ldots, \omega_{100k}), \quad \omega = (\omega_1, \ldots, \omega_{10^8}) \in \{0,1\}^{10^8},$$

ただし X は各 $(\xi_1, \ldots, \xi_{100}) \in \{0,1\}^{100}$ に対して

$$X(\xi_1, \ldots, \xi_{100}) \;=\; \begin{cases} 1 & (\xi_1, \ldots, \xi_{100}\text{の中に 1 が続けて 6 回以上現れる場合}), \\ 0 & (\text{それ以外の場合}), \end{cases}$$

によって定まる関数です．

では RWS 法による疑似乱数生成器 $g : \{0,1\}^{238} \to \{0,1\}^{10^8}$ の定義を述べます．以下 $m = 100$，$N = 10^6$，$j = \lfloor \log_2 N \rfloor = 19$ とします．ですから $238 = 2m + 2j$，$10^8 = Nm$ であり，$g : \{0,1\}^{2m+2j} \to \{0,1\}^{Nm}$ ということになります．

まず，種 $\omega' = (\omega'_1, \ldots, \omega'_{2m+2j}) \in \{0,1\}^{2m+2j}$ に対して

$$x \;=\; \sum_{i=1}^{m+j} 2^{-i} \omega'_i, \qquad \alpha \;=\; \sum_{i=1}^{m+j} 2^{-i} \omega'_{m+j+i}$$

とおきます．次に

$$Z_k \;=\; (d_1(x + k\alpha), \ldots, d_m(x + k\alpha)) \in \{0,1\}^m, \quad k = 1, \ldots, N$$

とします．ここに $d_i(t)$ は実数 $t \geq 0$ の 2 進小数展開における小数第 i 桁の数 (0 または 1) です．各 Z_k は ω' の関数であることに注意します．そして最後に

$$g(\omega') \;=\; (Z_1, \ldots, Z_N) \in \{0,1\}^{Nm}$$

と定義します.

例題の場合ですと，たとえば $\omega' \in \{0,1\}^{238}$ が

$$
\begin{array}{l}
1110110101\ 1011101101\ 0100000011\ 0110101001\ 0101000100 \\
0101111101\ 1010000000\ 1010100011\ 0100011001\ 1101111101 \\
1101010011\ 1111001001\ 1000001110\ 1110001000\ 0011010111 \\
0010000010\ 0100010001\ 0101011011\ 1101011100\ 0100100111 \\
0000000110\ 1010001100\ 1011100100\ 10111111
\end{array}
$$

のとき

$$S(g(\omega')) \ = \ 546177$$

となり，求める p の近似値として 0.546177 を得ます.[注10]

6　ひとこと...

タイトルは「ひとこと」ですが，せっかくの機会ですので，二言，三言，話させて下さい.

6.1　応用数学とは

応用数学というと，既存の数学の理論を様々な応用分野で役立てることを想像する人が多いと思います．たとえば，いま見てきた「モンテカルロ法」がそうですね．しかし，それは狭い意味の応用数学です．前半の確率と乱数のお話でお分かりのように，応用分野から新しい数学の芽を見つけ，それを一つの数学として育てることも応用数学といえるでしょう.

そのような意味の応用数学はとても難しい分野です．定義のないところで考え抜く力が必要です．自分の考えていることが何なのか，自分でもその正体が分からない，ましてや他人に説明することなどできない，そういう状況で耐えなければならない．いや，耐えるというより，そのような状況をも，なお「面白い」と感じることのできる，そんな感受性こそ必要です.

私は高校生の頃から「確率って代数や幾何，微分積分と違ってモヤモヤ感がある」とずっと思ってきました．そういうモヤモヤ感こそが私にとって確率論を研究する原動力であったように思います．その中でも乱数やモンテカルロ法はモヤモヤ感が最も強烈でした．コルモゴロフもそうだったのかもしれません.

6.2　不可能から可能へ——奇想天外 vs 荒唐無稽

私は数学は「不可能を可能にする学問」だと思います．正確にいうと「今まで不可能だと思われていたことでもやり方によっては可能だということがある．数学は，まずそうい

[注10] この計算も PC であっという間です.

うことを見出し，そしてそれが可能である，ということを証明する，そういう学問」だということです．不可能に見えることは大抵やはり不可能です．しかし稀だけど，実際に不可能と思われていたことが奇想天外な方法によって可能になることがあるのです．

一昔前のドラマの一場面[注11]です．飛行機の整備工場で働く青年が試験で失敗した失意の恋人を励まします．セスナを操縦しながら「飛行機はなぜ飛ぶのか，知ってるか」と尋ねます．恋人は「翼の上と下を流れる気流の速度が違うので...」と答えようとしますが，青年はそれを遮って「鉄でできた機械に空を飛ばせることが無理だって，思わなかった人がいたからだよ」と言います．だから「無理って言うな，諦めるな」と続けます．

飛行機の場合は「鳥」という実際に飛ぶ生き物がいます．ですから「鉄でできた機械を飛ばす」ことはそれほど奇想天外ではないかも知れません．しかし数学の場合はそんなお手本がないことが多いのです．だから本来の意味で奇想天外なのです.「2 乗して負になる数は本当に存在しないのか．あってもよいのではないか」と考えた人がいたから複素数が誕生しました.「掛け算をより易しい足し算で実行できないか」というアイデアから対数が誕生しました．確率論もそうして生み出された数学の代表的な例だと思います．デタラメ，偶然，不規則，予測不能，...，そういうものを数学として，つまり幾何学や代数学と同じように，厳密に捉えようと考えた人たちがいたわけです．

数学者はいつも奇想天外なことを考え続けています．それが仕事だからです．そのためにまず数学者はいつも「それは本当に正しいのか」と疑うことから始めます．私は「疑似乱数を用いたモンテカルロ法は数学的に正当化されない」という常識を疑うところから乱数と疑似乱数の研究を始めました．数学者が物分かりの良い常識者ばかりになってしまう様な世の中では未来はありません．

ただし数学における奇想天外とは完全に裏付けのあるものでなければなりません．荒唐無稽ではないのです．歌舞伎役者の坂東玉三郎が芸について養父の守田勘弥に言われ続けていたことがあるそうです[注12]; 「型破りってえのは型を持っている人間の言うことなんだ．型も何もないヤツラがやれば，いいかい，それは形なしっていうんだよ.」数学も同じです．型，すなわちしっかりした学力の基盤があってこそ，奇想天外なアイデアが実を結ぶのです．型を持つということは本来の意味で自由を獲得することです．私は受験勉強のような訓練は型を身につけるのに非常に役立つと考えています．

今日のお話は，はじめに確率の英語 probability の語源について話しました．そして最後に「数学」の語源について皆さんに宿題を出して終わりたいと思います．英語の mathematics はギリシャ語の $\mu\alpha\theta\eta\mu\alpha\tau\iota\kappa\acute{o}\varsigma$ に起源を持ちます．その原意は現在の日本で用いられているような意味ではありませんでした．興味のある人はぜひ調べてみて下さい．

本日はご静聴ありがとうございました．

(すぎた ひろし)
http://www.math.osaka-u.ac.jp/~sugita/mcm.html

注11「アテンションプリーズ」第 8 話，2006 年，フジテレビ，上戸彩，錦戸亮ほか.
注12朝日ジャーナル 1991 年 6 月 7 日号，坂東玉三郎,「風の四季 21」.