

1億人の考えをたった1000人の意見から推測できる

現在の政権は、どれくらいの人に支持されているのだろうか？ この疑問に答えるために、1億人をこえる国民に意見を聞くのはあまりにも非効率だ。実は、たった1000人にアンケート調査を行えば、この疑問に確率を使って答えることができる。アンケートで政権支持率を調べ、「支持率70%」という回答が出たとしよう。すると、「真の支持率は

67～73%の範囲にある可能性が95%である」と推定できるのだ（このような表現になる理由は82～87ページで解説）。なぜ、1億人のうちたった1000人に聞くだけで、全体の意見を推測できるのだろうか。それは、鍋のスープの味見をするのに、スプーン一杯で十分であるのと同じ理由である。スープがよくまざっているなら、スプーン一杯の味のバラン

スと、鍋全体の味のバランスは同じであるはずだ。同じように、国民と、男女比や年齢の比など、あらゆる要素の割合が同じである「一握りの回答者の集団」を選びだすことができれば、その集団の意見を聞くことで国民の意見を推測できるのである。

ではどうすれば、国民と同じ構成の、1000人の回答者集団を選びだすことができるだろうか。実は、男女差や年齢差などを考慮する必要はない。ただランダムに選べば、国民とほぼ同じ構成の1000人を選びだせるのである。

ただし、この「ランダムに選ぶこと」が実はむずかしい。ランダムに選ぶことは、国民のどの人も同じ確率で選ばれ

る方法で選ぶということを意味する。たとえば、平日の朝、東京駅でただ目についた人を選ぶ方法では、回答する人々は国民の構成から偏った集団になってしまうだろう。『ランダムに選ぶ』ことは、単に『あてずっぽうに選ぶ』こととはまったくちがうのだ。

たとえば左ページのイラストのように、国民に番号をふり、「0」から「9」の10個の面をもつサイコロで番号をつくらせて選びだす方法なら、回答者に選ばれる確率は、国民のどれと同じなので、ランダムな選び方だといえる。実際の新聞社のアンケートでも、下のイラストのように、できるだけランダムに回答者を選ぶ方法がとられている。

1億人から1000人をランダムに選ぶしくみ

国民のどの人も選ばれる可能性がある方法で回答者を選ぶには、国民に番号を振り、その番号の各けたを、0～9の数字をもつ10面サイコロで決めればよい（左ページで解説）。ただし、調査を行う民間企業などがある権者の連絡先を自由に知ることにはできないので、この方法を用いることはできない。そこで新聞社の世論調査では、電話番号を用いて回答者を選びだしている（右ページで解説）。

100,000,000人

00000000

34728810 34728811

99999998 99999999

1. 国民に番号を振る

「00000000」～「99999999」の8けたの番号を振る。



34728810

57726231

99328116

1,000人

2. 10面サイコロを振って8けたの数字をつくる

10面サイコロを8回振って、「3」「4」「7」「2」「8」「8」「1」「0」と出たら、「34728810」番の人を選ぶ。これを1000回くりかえして1000人を選びだす。



1. 電話番号の最初の6けた（局番）から1万個の番号をランダムに選ぶ

電話番号のはじめの6けたは「局番」といい、地域ごとに割りふられている。現在、有効な局番は約2万4000個存在する。その中から1万個の局番をランダムに選ぶことで、全国の1万地域を選びだす。たとえば、「03-5352」は、東京都23区などに使われる局番の一例だ。

2. 電話番号の下4けたの数字をランダムに選び、電話番号をつくる

電話番号の下4けたは、「0000」から「9999」の1万通りの番号がありうる。そのうち一つの数がランダムに選ばれ、1で選んだ局番と組み合わせて電話番号がつけられる。たとえば「6054」が選ばれ、1で選んだ局番と合わせて電話番号「03-5352-6054」ができる。ちなみにこの番号は、ニュートンを刊行している株式会社ニュートンプレスの営業部の番号である。

3. 電話をかけ、アンケートを行う

2でつくった1万個の電話番号のうち、実際に使われている番号は1600個程度であることが知られている。この約1600件の番号に電話をかけ、留守の場合や調査への協力を断られる場合を除いて、約1000件の協力が得られることを目標に調査を行う。

電話をかける調査員

「無差別抽出」と「有効回答率」が信頼性の要

世論調査は、たった1000人にアンケートをとるだけで全人口の意見を推定できる。ただしその際、「ランダムに選ばれた人々に聞く」という条件をできる限り厳密に守る必要がある。この条件を守らないとどうなるのだろうか？

回答者200万人以上！大規模調査の失敗

1936年、アメリカの雑誌「リテラリー・ダイジェスト」は、次期大統領を予測するため、大規模なアンケート調査を実施した。雑誌の登録者や、電話や自動車をもつ人など、1000万人もの人々にハガキを出し、「共和党の候補ランドン氏と民主党の候補ルーズベルト氏、どちらに投票するか」を問い

たのだ。雑誌社は、237万人の回答を元に「ランドン氏が勝利する」と予想した。

ところが大統領選挙の結果は、民主党のルーズベルト候補の勝利だった。母集団（全国民）と、標本（アンケートの回答者）の意見が食いちがってしまったのだ。その理由は、アンケートの対象になった「電話や自動車をもつ人」が当時は富裕層に偏っていたためだった。回答者をランダムに選ばなかったために、ルーズベルト候補を支持する庶民の意見を見落としてしまったのだ。

しかも、1000万人にハガキを出したのに対し、回答数は237万人と少ない。このことも、結果の偏りにつながった可

能性がある。

回答者数よりも、回答者の割合が重要

電話が普及している現代の日本では、72～73ページで紹介した電話を使った調査で、ランダムな選別ができると考えられている。しかし、それでも偏りは生じてしまう。

下は、毎日新聞社と産経新聞社が、政権の支持率を知るため、同じ方法で世論調査を行った結果である。どちらの調査でも支持率が低下しつづけていることがわかる。ただし、両社の調査結果に数パーセントの差があり、産経新聞社の調査結果の方がつねに低い。このような差が偶然生じる可能性は低い。この差は、どちらの新聞社からの電話かによって、調査に応じる人々の構成がかわるために生じたと考えられる。

このような媒体ごとのずれをできるだけ小さくするには、無作為抽出で選ばれた回答者のうち、できるだけ多くの人

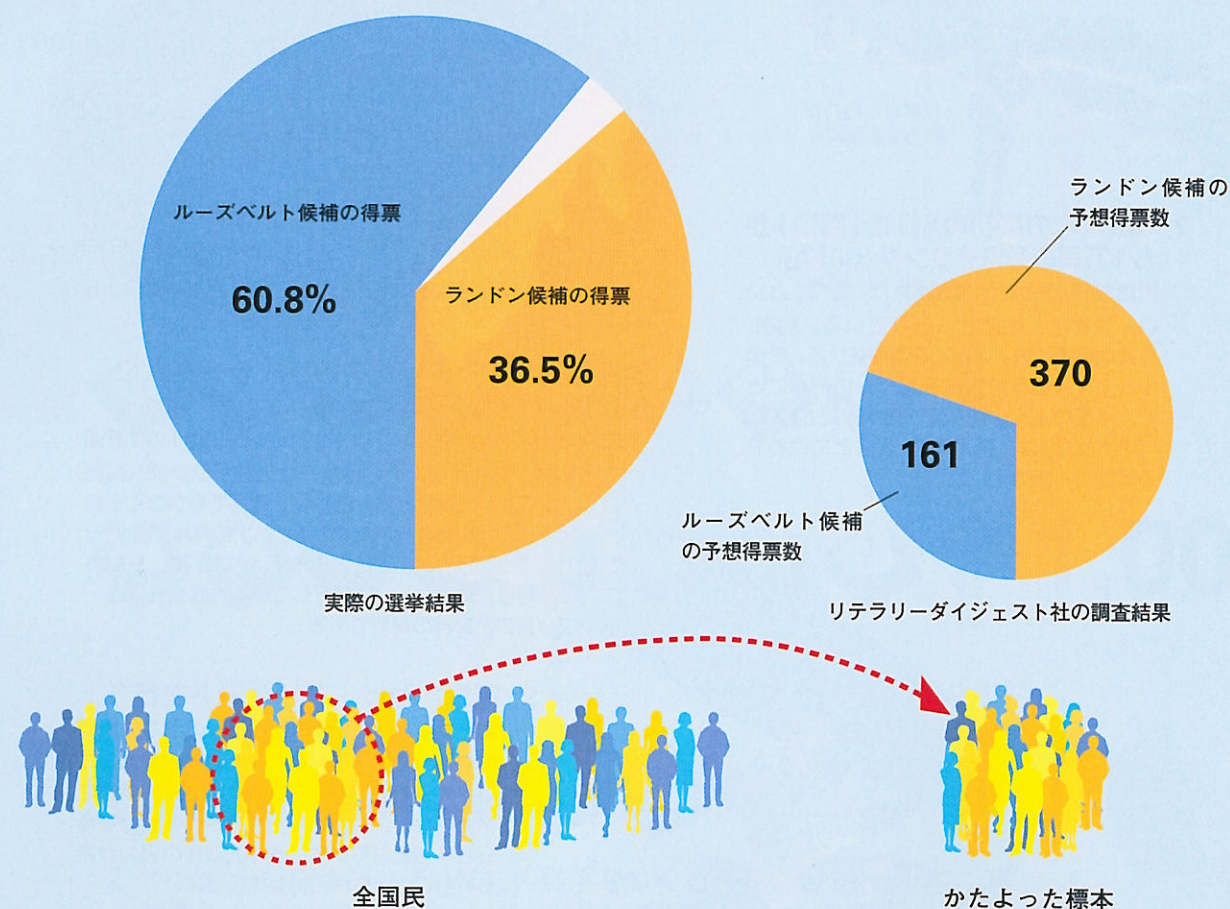
が回答することが重要である。「調査の対象に選ばれた人（標本数）」に占める、「調査に協力して有効な回答を行った人（有効回答数）」の割合を「有効回答率」という。この有効回答率が高いほど、調査を拒否した人が少ないため、回答者が特定の層にかたよっている危険も少ないと見積もれる。一つの目安として、有効回答率は60%を上まわることが望ましいといわれている※。

極端に有効回答率が低いと、たとえ標本をランダムに選んでも、そこからさまざまな結論を得ることができてしまう。ランダム抽出であることに加え、有効回答率が高いことも重要な条件なのだ。

※近年では、世論調査の有効回答率が全体的に低下する傾向にある。その一因として、「プライバシー意識の高まり」があげられる。個人情報の流出をきらって、調査を拒否する例が増加している可能性があるのだ。過去に行われた世論調査と現在の世論調査を比較するときなどには、こうした事情も頭に入れておくことが参考になるだろう。

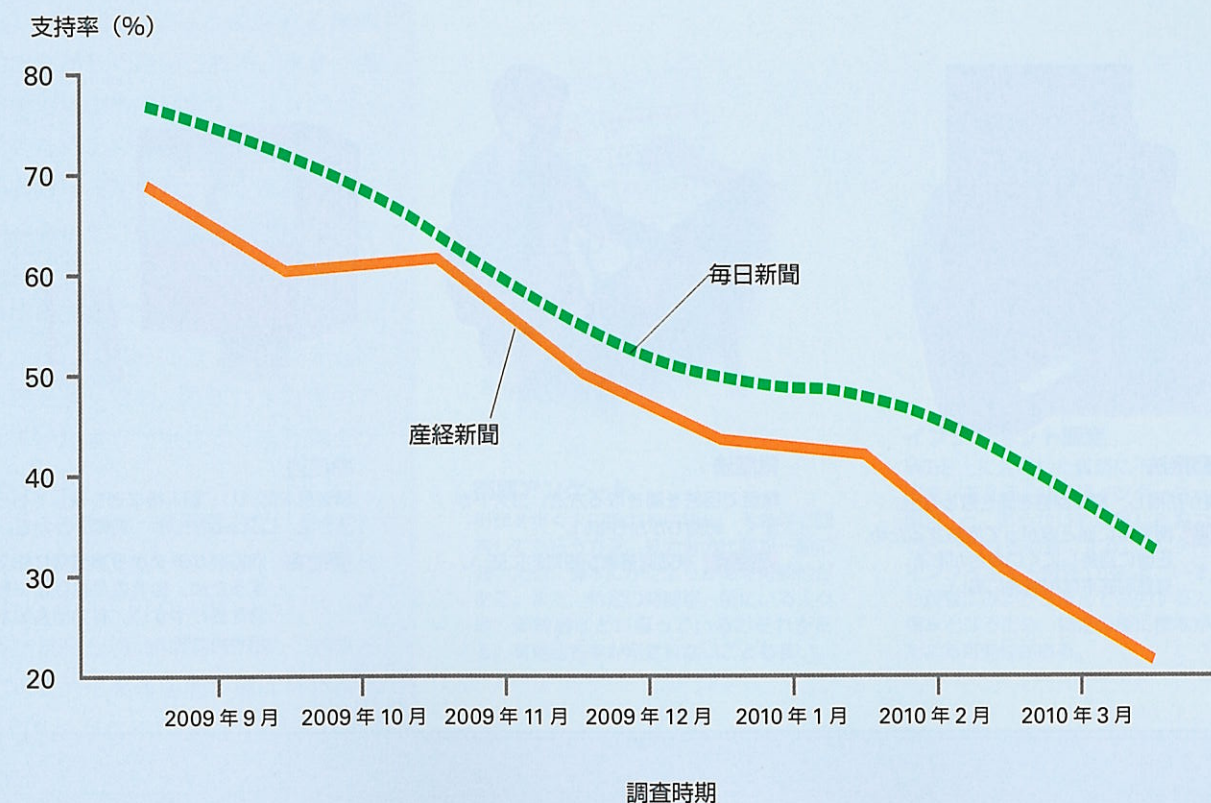
リテラリー・ダイジェスト社の調査結果と実際の選挙結果

リテラリー・ダイジェスト社の調査では共和党のランドン候補が優勢だったが、実際の選挙では、民主党のルーズベルト候補が勝利した。



産経新聞社と毎日新聞社の内閣支持率調査結果

同時期に行われた調査にもかかわらず、内閣支持率の調査結果はつねに産経新聞社の方が低かった。グラフは『実践する統計学』（藪友良著、東洋経済新報社）を元に作成した。



世論調査を見るときは 調査方式に注意しよう

「地球環境問題、9割以上が関心」、
「サマータイム制、過半数が賛成」、
「内閣支持率が最低を更新」……。

テレビのニュースや新聞の見出しは、「世論調査」であふれている。テレビ局や新聞社などの報道機関は、定期的に世論調査を行い、「国民の声」として伝えている。このほか、内閣府や自治体などの行政機関が、住民の声を把握するために世論調査を行うこともある。

内閣府大臣官房政府広報室が取りまとめた『世論調査年鑑』によれば、国内で近年行われている世論調査の数は、実に年間1000をこえるという。

世論調査は、「世論」という目に見えないものを、具体的に説得力のある「数字」へと置きかえる。だがそこに、“世論調査の落とし穴”がひそんでいる可能性がある。世論調査の結果を目にしたときには、結果だけではなく、どんな

方法で行われたかなど、調査方法に注目する必要があるのだ。

さまざまな調査方式の長短

同じ時期に実施した内閣支持率などの世論調査の結果が、新聞やテレビ局によってことなることがある。その理由には、調査方式のちがいや、80ページ以降にのべる「標本誤差」など、さまざまな要因が考えられる。

世論調査の調査方式には注意が必要

世論調査で行われる各方式の留意点をまとめた。ここにあげた「訪問面接法」「電話法」「郵送法」のほかに、調査員が訪問して調査票を対象者にあずけ、後日回収する「訪問留置き法」などがある。



訪問面接法

調査員が訪問し、直接回答を聞き取る方法。

留意点 調査員に面と向かって回答するため、正直に回答しにくい場合がある。有効回答率は相対的に高い。



電話法

電話で回答を聞き取る方法。コストが安く、時間もかからない。

留意点 有効回答率が相対的に低い。



郵便法

調査票を郵送し、記入後に送り返してもらう方法。コストは安いですが、時間がかかる。

留意点 回答者がみずから調査票に記入するため、他者の意見などに影響を受けやすい。有効回答率も相対的に低い。

現在行われている世論調査の方法には、訪問面接法、電話法、郵便法などがあり、それぞれ長所と短所がある(左ページのイラスト参照)。

調査員が訪問して回答を聞き取る「訪問面接法」では、全住民の名簿である「住民基本台帳」から、乱数表などを使って無作為的に(ランダムに)標本を選び出す方法などがとられる。住民基本台帳が原則非公開となった2006年以降は、調査対象地域の地図上の点をランダムに選んで訪問する方法なども用いられつつある。

また、新聞社などが行う電話世論調査では、「ランダム・ディジット・ダイヤリング(RDD)法」とよばれる無作為抽出法が多く用いられる。まず、調査対象地域の市内局番に、コンピュータがでたらめにつくりだした4けたの加入者番号を足して、固定電話の番号をつくる。それが一般の世帯につながる番号であった場合に、調査対象とするのがRDD法である。

しかし、これらの方法でランダムに回答者を選んでも、さまざまな理由で標本にかたよりが生じてしまう場合があるという。

たとえば近年、携帯電話だけをもち、固定電話をもたない世帯が、若い世代を中心にふえている。RDD法では、こうした世帯が標本に選ばれず、結果として相対的に高齢層が選ばれやすくなる可能性がある。

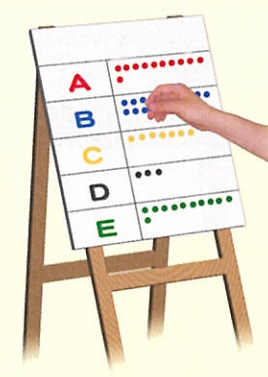
世論調査ではない調査

テレビ報道などでは、街頭アンケートを行って結果をグラフにしたり、道行く人にインタビューし、コメントを紹介したりする場合がある。これらも調査の一種ではあるが、世論調査とはちがいで、その調査上の多数派が全国民の多数派だという保証はない。一意見として考え、世論調査とは区別して考えるべきだろう。

調査の方法にはそれぞれ一長一短があり、100点満点といえるものはない。調査方法を確認し、その特徴を頭に入れた上、調査結果を解釈するのが望ましいだろう。

世論調査とは区別すべきもの

「街頭アンケート」や「インターネット調査」は、無作為に選んだ標本が対象ではないため、結果の解釈には注意が必要である。



街頭アンケート

街頭を歩く人に協力を依頼し、その場で回答してもらう方法。協力者の比率が一般に低いため、標本にかたよりの可能性がある。また、特定の時間帯、街にいる人々は、年齢層などが偏っているおそれがある。有効回答率が示されないことも多い。



インターネット調査

WEB上に公開した質問に不特定の参加者が答える「公開型インターネット調査」と、あらかじめ確保した協力者(モニター)を対象にする「モニター型インターネット調査」がある。いずれも調査にみずから進んで協力する人が標本となるため、特定の層に標本がかたよりの可能性がある。

不合格品の割合を調べたい。 サンプルは何個必要？

あなたは缶詰工場の工場長です。あなたは工場の経営者から、在庫の缶詰を調査して、品質基準を満たさない不合格品の割合を報告するように指示されました。

この割合を、いっさいの誤差なく調べる唯一の方法は、在庫の缶詰をすべて開けて調べる「全数調査」です。しかし、それをしてしまったら、売りのための商品がなくなってしまいます。すると経営者は、「調査に必要な最小限の個数だけ開けてよい」といいました。はたして、何個の缶詰を開けるべきでしょうか。

全体（母集団）から一部のサンプル（標本）を抜き出す調査を「標本調査」といいます。標本調査は、全体ではなく、あくまで一部の標本を調べたもので、どうしても誤差が生じます。

標本調査で調べるサンプルの量を「サンプル・サイズ」といいます。そして、サンプル・サイズをふやせばふやすほど（全数調査に近づくほど）、誤差はゼロに近づいていくのです。したがって、「どの程度までの誤差を許容するか」が決まれば、必要なサンプル・サイズを決めることができます。右ページの下の方では、その具体的な計算方法を紹介しています。

誤差を10分の1にするには？

さて、調査結果を報告したあなたは、経営者から「誤差が大きすぎる。標本の誤差を10分の1にせよ」といわれてしまいました。

実は、これは簡単なことではありません。もし、誤差を10分の1にしたければ、サンプル・サイズを100倍にふやさなければならないからです。誤差は、サンプル・サイズの平方根に反比例するのです。実際には、調査にかかるコストなども考慮して、現実的なサンプル・サイズを設定することになります。

標本調査とは

工場で作られた缶詰から一部の缶詰を抜き出して、不合格品の割合を調べるイメージをえがきました。缶詰の品質検査のように、調査によって商品としての価値を失ってしまう検査（破壊検査）や、母集団の数がきわめて多い場合など、全数調査が困難な場合には標本調査が適しています。



母集団

標本をランダムに
抜き出す

標本

母集団の特徴を
標本から推定する

サンプル・サイズの決め方

標本の抜き出し方にはかたよりが生じる可能性があるため、標本調査には必ず誤差がつきまといます。この誤差を「標本誤差」といいます。標本調査によって得られた割合（標本の缶詰に含まれる不合格品の割合）が p だったとき、母集団の割合（すべての缶詰に含まれる不合格品の割合）は、「 $p \pm$ 標本誤差」と幅をもたせて推定します。

許容できる標本誤差を2%として、必要なサンプル・サイズを求めましょう。 p は未知ですが、前回の調査結果などがあればその値とし、なければ仮に0.5とします。前回の調査結果が5%だったとして、 $p = 0.05$ とすれば、必要なサンプル・サイズは次の計算で求められます。

$$\text{サンプル・サイズ} = \left(\frac{1.96 \times \sqrt{0.05 \times (1-0.05)}}{0.02} \right)^2 = 456.19$$

こうして、456個の缶詰を開ければよいことがわかりました。

必要なサンプル・サイズの計算式

$$\text{サンプル・サイズ} = \left(\frac{1.96 \times \sqrt{p(1-p)}}{\text{標本誤差}} \right)^2$$

信頼係数95%の場合。信頼係数とは、標本調査が母集団の姿を正しくいいてる確率のことです。

未成年の飲酒率を正直に答えさせるには？

「10代で飲酒したことがある人がどれくらいいるのか」を調査したいとしよう。しかし、正直に答えてもらえず、不正確な調査になってしまうおそれがある。どうすれば、答えにくい質問に対して正直に答えてもらえるだろうか。

未成年飲酒をしたことがある人が、アンケートで「未成年飲酒をしたことはない」とうそをつくのは、自分が未成年飲

酒したことを知られたくないからだ。一方質問者は、未成年飲酒の経験者の割合を知りたいだけで、それがだれかまで知りたいわけではない。

そこで、「回答のランダム化」という方法を使うと、だれが未成年飲酒の経験者かを秘密にしたまま、割合だけを調べることができる。

質問のしかた次第で、正直に答えるかどうかかわる

「未成年飲酒をしたことがあるか」を調べたいが、正直に答えてもらえないかもしれない(左ページ)。このような場合、「回答のランダム化」という方法を使うと、答えにくい質問にも、正直に答えてもらいやすくなる(右ページ)。

「未成年飲酒したことがありますか？」と聞くと……



未成年飲酒をしたことがある人は、「未成年飲酒をしたことがない」とうそをつく可能性がある。そのため、調査結果が示す未成年の飲酒率は、実際の未成年の飲酒率よりも低くなるおそれがある。

「コイン投げ」で回答者のプライバシーを守る

まず、回答者たちに、質問者には見えないようにして、それぞれコイン投げをしてもらう。そして、質問者は回答者全員に次のように告げる。

「出たコインが『表』だった人は『はい』と答えて下さい。出たコインが『裏』だった人は『未成年飲酒をしたことがありますか』という質問に『はい』か『いいえ』で答えて下さい」。

この場合、「はい」と答えた人たちが、コインが表だったために「はい」と答えたのか、それとも未成年飲酒をしたことがあるから「はい」と答えたのか、質問者からは区別がつかない。このことを理解している回答者は、未成年飲酒をしたことがある場合に、正直に「はい」と答えやすくなると期

待できる。この方法では、回答の意味がランダムに変わることから、「回答のランダム化」とよばれている。

質問者は、回答を集計したあと、次のようにして未成年飲酒の経験者の割合をみちびける。今、300人の回答者の内、200人が「はい」と答えたとしよう。

コイン投げの結果、表が出る確率は2分の1なので、300人の内150人はコインが表だったために「はい」と答えたと推定できる。そこで、「はい」という回答から150人分を引けば、「未成年飲酒をしたことがあるか」という質問に答えた人の割合がわかる。この質問に「はい」と答えた人は50人、「いいえ」と答えた人は100人なので、約33%が未成年飲酒をしたことがあると推定できる。

「コイン投げをして、コインが表だった人は『はい』と言ってください。また、コインが裏だった人のうち、未成年飲酒したことがある人も『はい』と言ってください」と聞くと……



質問者から見て、「はい」と答えた人たちの内、だれが未成年飲酒をしたことがあるのかはわからない。そのため、未成年飲酒をしたことがある回答者は「正直に答えても大丈夫だ」と考え、素直に答える可能性が高くなる。

ランダムと乱数の 奇妙な世界

「究極の不規則さ」をめざして

ランダム（でたらめ）に数字が並んだものを「乱数」といいます。ランダムに数字を並べることは、いい加減にやればできそうです。しかし、「真にランダム」な数字の並びをつくるのは、実は一筋縄ではいかない問題なのです。コンピューターゲームやAI（人工知能）の開発、インターネット上の情報セキュリティなど、現在社会のさまざまな場面で役立つ乱数の奥深い世界をのぞいてみましょう。

協力

田村義保

情報・システム研究機構 統計数理研究所 特任教授・名誉教授

サイコロと乱数

サイコロが出す目は、あらかじめ予測できません。イラストの背景に並ぶ数字の列が乱数です。乱数は、サイコロなどの「乱数発生器」を使うことでつくられます。

サイコロが生みだす乱数が、ゲームや AI 開発に必要

数学の教科書の巻末に、0 から 9 までの数字がでたらめに並んだ表が載っていることがあります。この表は「乱数表」とよばれ、世論調査などの対象者をかたよりなく選ぶ「無作為抽出」などに利用されます。

乱数とは、「次の数を決める規則がいっさい存在しない数」のことです。いいかえれば、「次に何が出るかまったくわからない数」が乱数だといえます。

円周率の値はサイコロで求めることができる

正 20 面体のサイコロに 0 から 9 までの数字を等しく割りあてたものを乱数表といいます。色分けされた 3 個の乱数表を用意して、たとえば赤の目を小数第 1 位、黄の目を第 2 位、青の目を第 3 位とすれば、0.000 から 0.999 までの 3 桁の乱数ができます。右では、この 3 桁の乱数を使って、円周率 π の値を求める方法を紹介しています。これは、最も単純なモンテカルロ法の応用例です。

サイコロは、最も身近な「乱数発生器」

乱数を生みだす装置を「乱数発生器」とよびます。すごろくなどのゲームに使う正 6 面体（立方体）のサイコロは、1 から 6 までの乱数を生みだす乱数発生器です。もしサイコロの目がランダムでなく、次に出る目があらかじめわかったとしたら、すごろくは実につまらないゲームになってしまうでしょう。

正 20 面体の各面に、0 から 9 までの数字を 2 回

ステップ 1

3 個の乱数表を 2 回振り、3 桁の乱数を 2 個つくります。

1 0 3 → 0.103
7 8 2 → 0.782

ステップ 2

2 個の乱数を xy 座標とする点を置きます。

$(x, y) = (0.103, 0.782)$

ステップ 3

これを何回もくりかえして、多数の点を置きます。

ステップ 4

多数の点は、1 辺の長さが 1 の正方形の上に置かれることになります。この正方形の上に、半径が 1 の扇形をえがきます。正方形の面積は 1 です。半径が 1 の円の面積は、半径 $\times$ 半径 $\times \pi = 1 \times 1 \times \pi = \pi$ であり、その 4 分の 1 の扇形の面積は $\pi/4$ です。つまり、正方形に占める扇形の面積の割合は $\pi/4$ ということになります。

ステップ 5

乱数表を使って上の手順で正方形の上に点を置くと、その点が扇形の中に入る確率は $\pi/4$ になります。そのため、点の数をふやしていき、扇形の中に入った点の割合を実際に数えて求めてやれば、その値は $\pi/4$ に近づいていくことになります。ここから π の値を求めることができます。点の数をふやすほど、 π は本来の値である 3.14... に近づいていきます。

ずつ割りあてれば、0 から 9 までの乱数を生みだす乱数発生器になります。これは「乱数表」とよばれ、これを振りつづければ数学の教科書に載っているような乱数表をつくることができます。

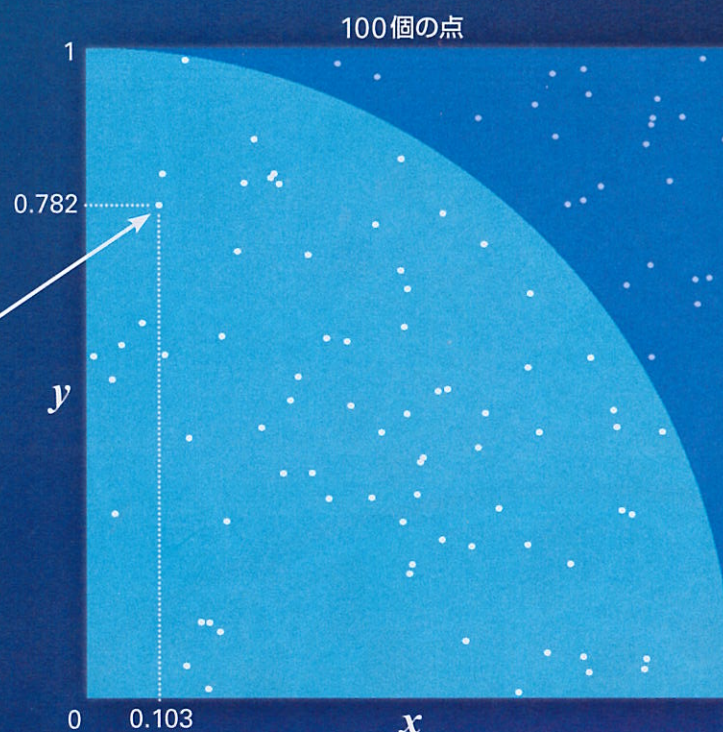
ビンゴゲームや福引きに使う抽選器や、ギャンブルで使うルーレットなども、乱数発生器の例です。コンピューターゲームにも、敵キャラクターの動きなどが単調にならないよう、乱数が使われています。

乱数は、科学研究や AI 開発も支えている

乱数が役立つのは、ゲームだけではなく。

科学研究や人工知能 (AI) の開発でも乱数が使われています。その重要な例が、乱数を使って数値計算やシミュレーションを行う「モンテカルロ法」です。トッププロ棋士を倒した囲碁 AI「AlphaGO」も、モンテカルロ法を使って膨大な数の対局を経験・学習し、強くなったのです。

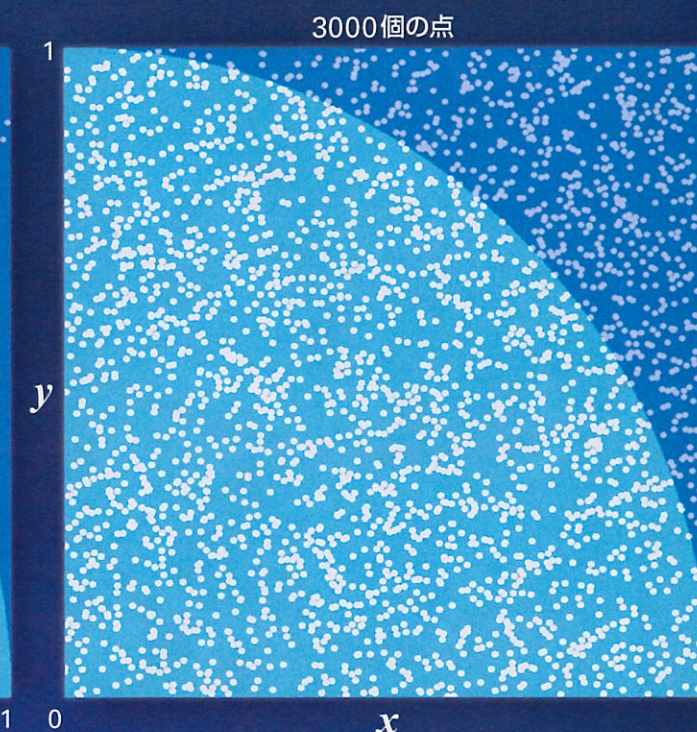
下の図では、モンテカルロ法を使って円周率 π を求める方法を紹介しています。雑貨店などで購入できる正 20 面体のサイコロを用意して、ぜひ自分で円周率を求めてみましょう。



扇形の中の点は 74 個

$$\frac{\text{扇形の面積}}{\text{正方形の面積}} = \frac{\pi/4}{1} \div \frac{74}{100}$$

$$\text{式を変形して } \pi \div \frac{74}{100} \times 4 = 2.96$$



扇形の中の点は 2430 個

$$\frac{\text{扇形の面積}}{\text{正方形の面積}} = \frac{\pi/4}{1} \div \frac{2430}{3000}$$

$$\text{式を変形して } \pi \div \frac{2430}{3000} \times 4 = 3.24$$

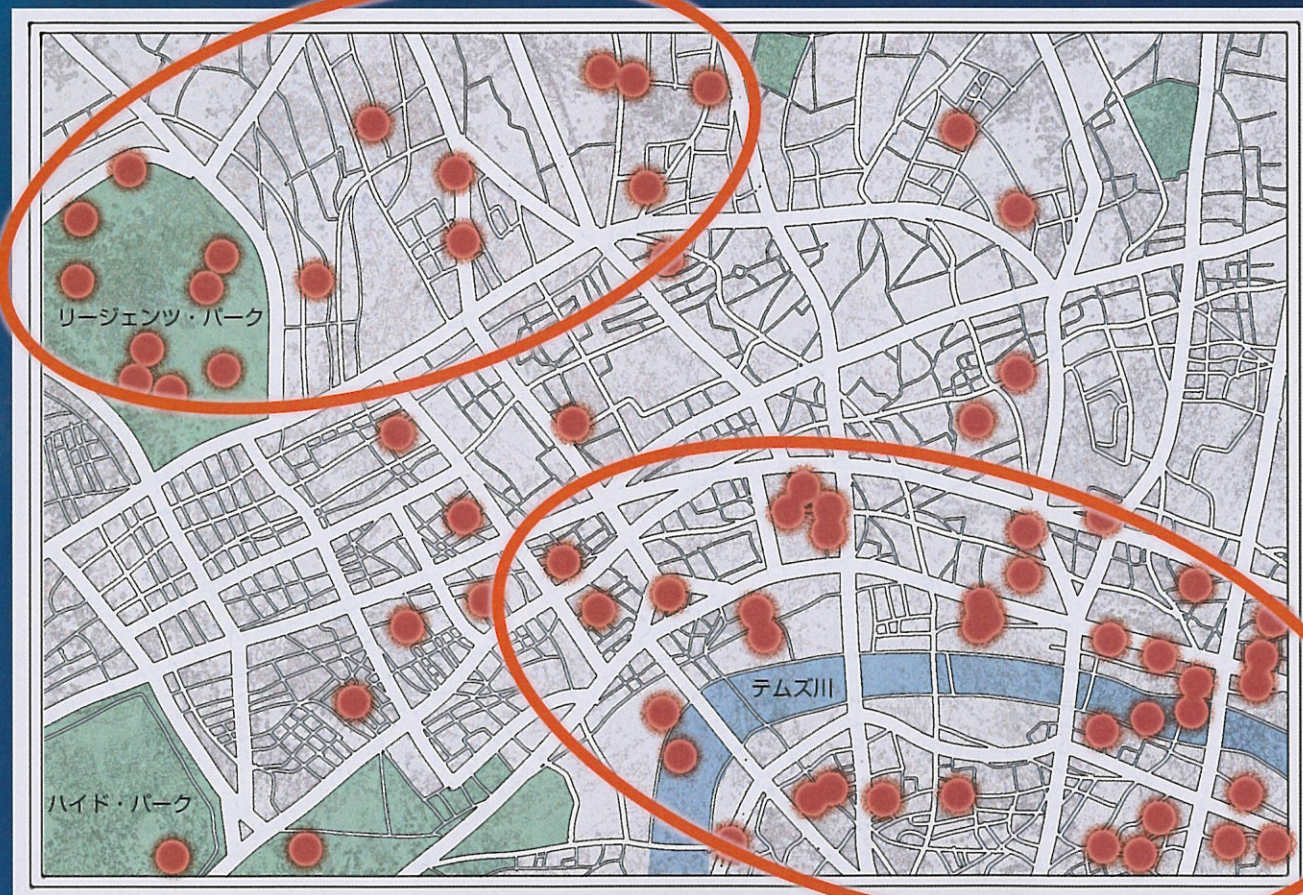


人は、ランダムから“意味のあるパターン”を見つけがち

サイコロが5回連続で「1」を出しました。これを見て「1が出やすいように細工されているのでは？」と疑う人がいたとしても無理はありません。心理学では、実際はランダムであるにもかかわらず、たまたま同じ出来事がつづけておきると、それをランダムではないと錯覚してしまう傾向があることが知られています。これは「クラスター錯覚」とよべます。

クラスター錯覚の例として、第二次世界大戦末期

のイギリス、ロンドンを襲ったドイツ軍の空爆が有名です。ミサイルの着弾頻度には地域ごとのかたよりがみられたため、「ドイツ軍のミサイルは特定の地域を狙い撃ちできる」と一部のロンドン市民は恐れしました（下の図）。しかし後の分析で、ミサイルの着弾点はランダムとみなせることがわかっています。人は、ランダムな出来事の積み重ねから、意味がありそうなパターンを見だしてしまいがちなのです。



特定の地域が狙い撃ちされている？

上はロンドン中心部の地図で、赤い印はドイツ軍のV-1ミサイルによる空爆の着弾地点です。楕円で囲んだ地域に多く着弾し、ほかの地域は比較的安全なようにもみえます。しかし、当時のV-1ミサイルにはそこまでの精度はなく、より広い領域を対象にした後の分析によって、着弾地点はランダムとみなせることがわかっています。イラストはJohnson, D. (1982) "V-1 V-2: Hitler's vengeance on London" の図を参考にして作成しました。

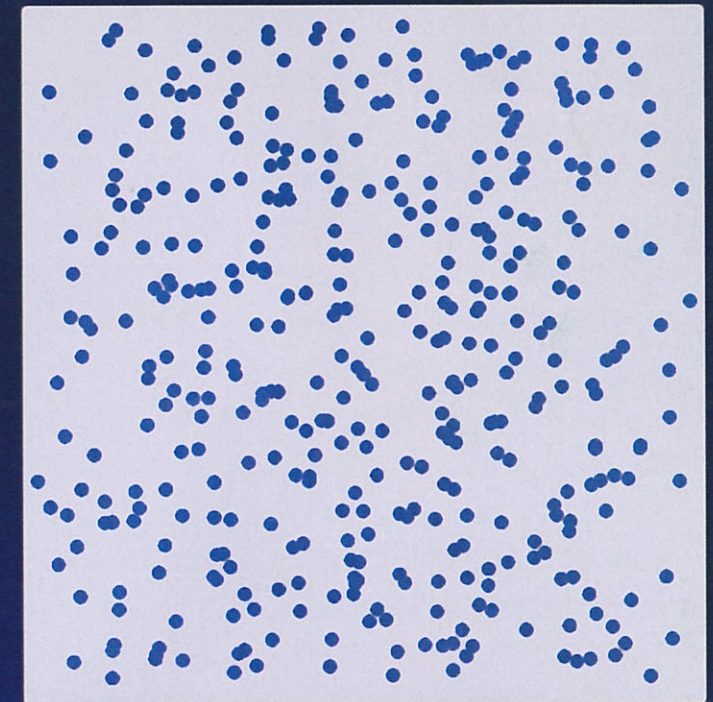
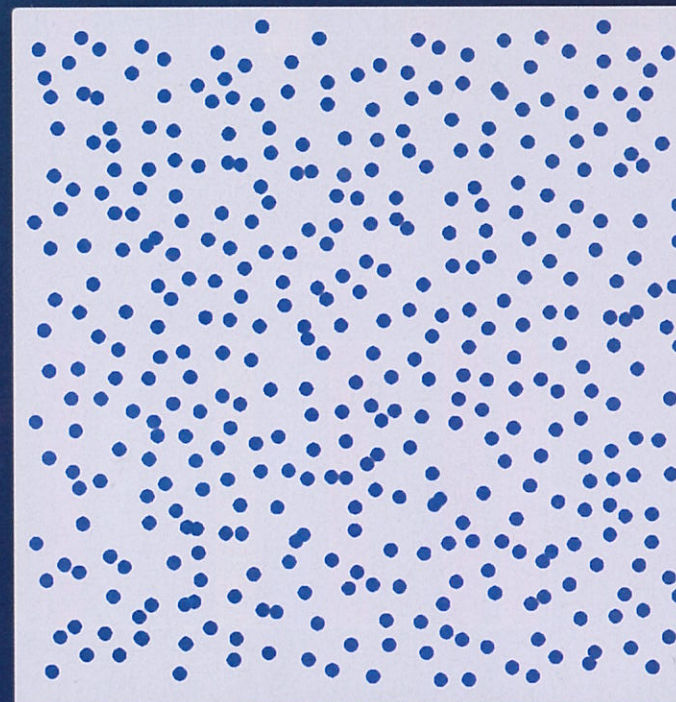
ランダムな点の分布はどちら？

クラスター錯覚のもう一つの例を紹介しましょう。古生物学者・科学史家のスティーブン・J・グールドは、エッセーの中で次のような話題を紹介しました。

「下の二つの図のうち、ランダムな点の分布はどちら？」と聞くと、多くの人は「左がランダム」と答えます。しかし実際は、左は点どうしが重ならないように意図的に配置したもので、右がランダムな分布です。左の図からは、意味がありそうなパターンが読み取れないために、「よりランダムだ」と判断して

しまいがちなのです。

人間の直感がおちいりやすい錯覚を逆に利用すれば、“より自然なランダムさ”を演出できます。デジタル音楽プレイヤーや音楽ストリーミングサービスの「シャッフル再生（ランダム選曲）」には、特定のアーティストの曲がつづけて選ばれにくくするなど、意図的にランダムさを減らすことで、“より自然なランダムさ”が感じられるようにくふうされているものがあります。同様のくふうは、コンピューターゲームにも応用されています。



どちらがランダム？

左は、点どうしが重ならないように意図して点を配置したものです。右は、乱数で決めたランダムな点の配置です。スティーブン・J・グールド（1941～2002）は著書『がんばれカミナリ竜』でこのような図をあげ、人はランダムな点の配置にパターンを感じとってしまいがちであることを論じています。こうした錯覚は心理学でよく研究され、人間の心理傾向から経済現象をときあかす「行動経済学」とよばれる分野に応用されています。

円周率 $\pi = 3.141592\cdots$ に並ぶ数字は、乱数か？

あなたはゲームプログラマーだとします。ゲームに出てくる敵キャラクターの動きをプレイヤーが予測できないようにするために、乱数を使って敵の動きを決めることにしました。0から9までの乱数をコンピューターが生み出す手取り早い方法として、あなたは円周率を使うことを思いつきました。

円周率 ($\pi = 3.141592\cdots$) を計算で求めると、小数点以下の数字が無限につづきます。これらの数字の並びには、いっさいの規則性がみつかっていません。円周率にあらわれる数字の並びは、果たして乱数といえるのでしょうか？

出やすい数字・出にくい数字はある？

「1が出やすいサイコロ」は、乱数づくりには向きません。これと同様に、乱数にはそれぞれの数字が等しい頻度で(同じ確率で)出ることが求められます。では、円周率の数字はどうでしょうか？

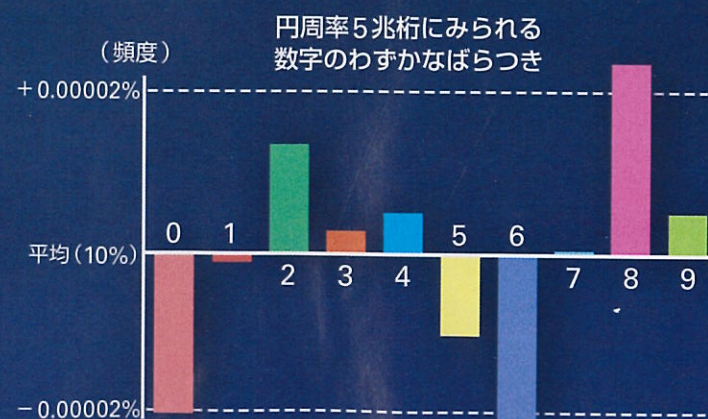
右のグラフは、円周率の小数点以下5兆桁について、0から9までの数字の出現頻度を調べたものです。最多は「8」で、最少は「6」でした。しかし、その差はごくわずかであり、0から9までの数字が出る頻度はほぼ等しいとみなせそうです。

では、円周率に並ぶ数字は、どこまでいっても0から9までの数字があらわれる頻度にかたよりのないといえるのでしょうか？ 小数表示したときの数字の並びに、0から9までの数字のかたよりのまったくない数のことを数学では「正規数」といいます。そして、円周率や2の平方根 ($\sqrt{2} = 1.414\cdots$) が正規数かどうかは、数学的には未解決の問題です。そのため、円周率や2の平方根に並ぶ数が真の乱数かどうかはまた、未解決の謎といえるのです。

■円周率5万桁の数字を色分けすると？

右は、円周率5万桁の0から9までの数字を色分けした図です。横1行が200桁分、縦に250行が並んでいます。小数点以下762桁目からは6連続で9が並んでいます。ノーベル物理学賞受賞者のリチャード・ファインマン(1918～1988)がこの6連続の9までを暗唱できたことから「ファインマン・ポイント」とよばれています。このような数字のかたよりのあちらこちらにみられるように感じられるかもしれませんが、それはランダムが生む自然なかたよりとみなせます。

【数字の色分け】



$\pi = 3.141592\cdots$

ファインマン・ポイント ...999999...



コンピューターはどうやって乱数を生み出すのか？

ゲームやモンテカルロ法のプログラムを動かすために、コンピューターはたくさんの乱数を必要とします。コンピューターは、自分で乱数をつくれるのでしょうか？ コンピューターが数の列をつくるためには、プログラムやアルゴリズムなどの規則が必要です。「次の数を決める規則がいっさい存在しない数」が乱数ですから、コンピューターが規則に従って生

みだす数の列は、真の意味での乱数とはよべません。真の乱数の代わりになるのが、あたかも乱数に見える「疑似乱数」です。疑似乱数をはじめて考案したのは、コンピューターの発明者として知られるハンガリー出身の数学者ジョン・フォン・ノイマン（1903～1957）です。フォン・ノイマンが考案した「平方採中法」が、疑似乱数の元祖です（下に解説）。

さまざまな疑似乱数

フォン・ノイマンが考案した最初の疑似乱数である「平方採中法」と、現在でも使われている「線形合同法」、そして現時点で最良の疑似乱数といわれる「メルセンヌ・ツイスター法」を示しました。

平方採中法（4桁の場合）

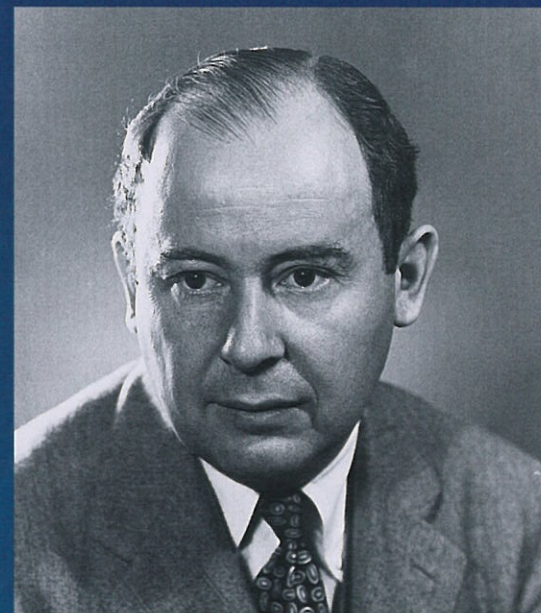
時刻の値などからシード（乱数のタネ）となる4桁の数を拾います（例では下線の1234）。シードを2乗（平方）した8桁の数（桁数が足りないときは0をおぎないます）の中間の4桁の数（5227）を採用して1個目の疑似乱数とします。この5227から同じ操作によって次の疑似乱数（3215）を得ます。これをくりかえして疑似乱数を次々につくるのが平方採中法です。

$$\begin{array}{l} \underline{1234} \times 1234 = 01 \boxed{5227} 56 \\ \boxed{5227} \times 5227 = 27 \boxed{3215} 29 \\ \boxed{3215} \times 3215 = 10 \boxed{3362} 25 \\ \boxed{3362} \times 3362 = 11 \boxed{3030} 44 \\ \vdots \end{array}$$

線形合同法（4桁の場合）

シードとなる数（例では下線の1234）に、あらかじめ決めておいた定数（例では567）を掛け算し、そこに別の定数（例では89）を足します（699767）。この数を、また別の定数（例では9973）で割ったときの余りを求めます。この余り（例では1657）を1個目の疑似乱数とします。この1657から同じ操作によって次の疑似乱数（2146）を得ます。これをくりかえして疑似乱数を次々につくるのが線形合同法です。足し算に使う定数を0にした場合は「乗算合同法」とよべます。

$$\begin{array}{l} \underline{1234} \times 567 + 89 = 699767 \rightarrow 699767 \text{ を } 9973 \text{ で割った余り} = \boxed{1657} \\ \boxed{1657} \times 567 + 89 = 939608 \rightarrow 939608 \text{ を } 9973 \text{ で割った余り} = \boxed{2146} \\ \boxed{2146} \times 567 + 89 = 1216871 \rightarrow 1216871 \text{ を } 9973 \text{ で割った余り} = \boxed{165} \\ \vdots \end{array}$$



ジョン・フォン・ノイマン。現在のコンピューターの原型をつくった科学者であり、モンテカルロ法の考案者の一人でもあります。ゲーム理論から原子爆弾の開発まで、幅広い分野に影響をあたえました。

最強の疑似乱数「メルセンヌ・ツイスター法」

フォン・ノイマン以降、「線形合同法」をはじめ、さまざまな疑似乱数が開発されました。疑似乱数の良し悪しは、「真の乱数にどれだけ近いか」や「疑似乱数の発生速度」などで決まります。現時点で最良の疑似乱数と評価され、世界中のプログラマーに広く採用されているのが、広島大学の松本眞教授と山形大学の西村拓士准教授らが1998年に考案した「メ

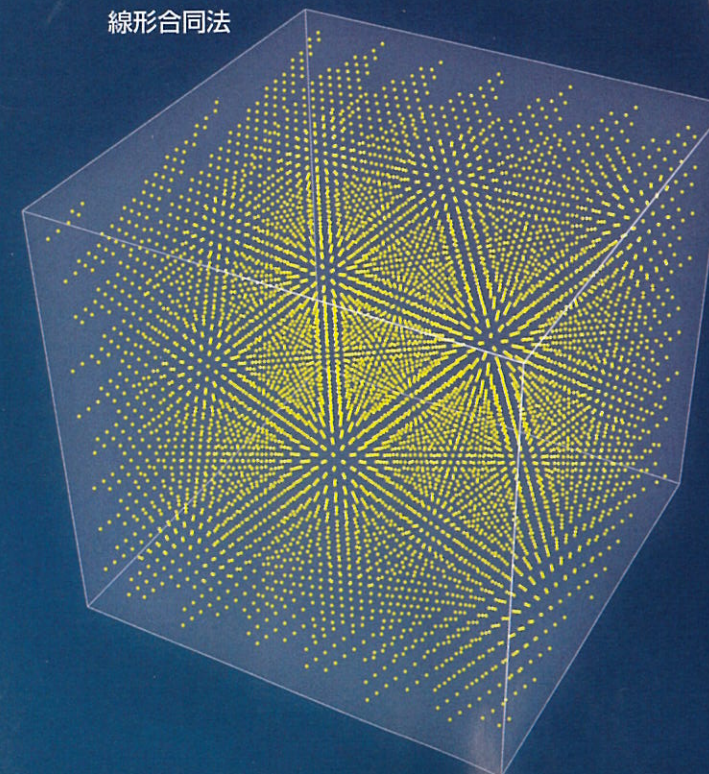
ルセンヌ・ツイスター法」です。その理解にはベクトルや行列といった数学の知識が必要なため、くわしい説明ははぶきますが、「メルセンヌ素数」とよばれる特殊な素数の性質がたくみに利用されています。

この最良の方法が生みだす乱数も、あくまでも疑似乱数であって、「真の乱数」ではありません。疑似乱数の父であるフォン・ノイマンも、「数式で乱数をつくることはある種の罪だ」とのべたといいます。

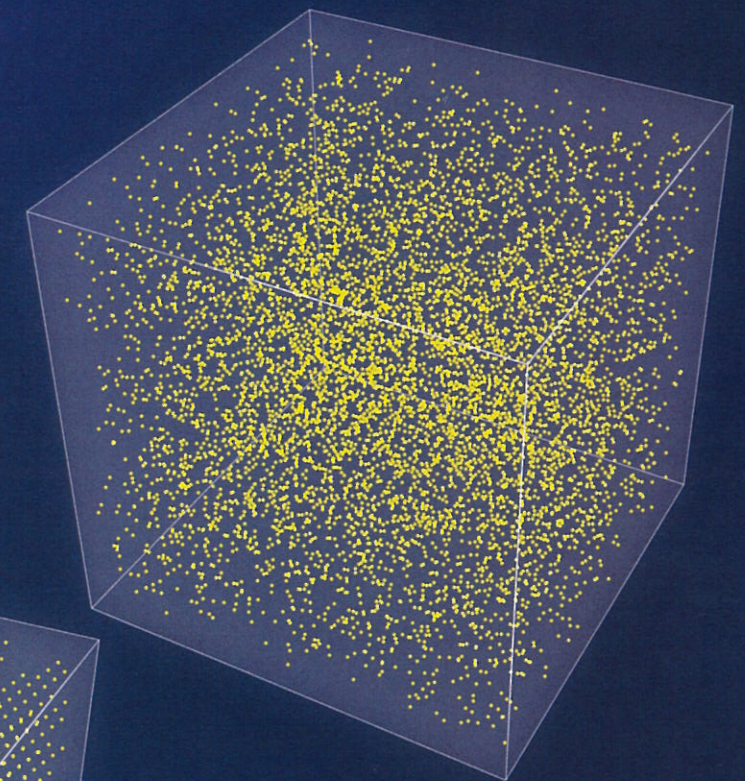
メルセンヌ・ツイスター法

左下は、線形合同法によってつくられた3個の疑似乱数を座標としてもつ点を、3次元空間に $(2^{31}-1)$ 個置いた結果です（原点付近の70倍拡大図）。点の並びには、規則的な模様やすきまがみられます。右上はメルセンヌ・ツイスター法によって同じ個数の点を置いた結果で、規則的な模様やすきまはみられません。（画像提供：広島大学 松本眞教授）

線形合同法



メルセンヌ・ツイスター法



不規則な物理現象を使えば「真の乱数」がつかれる

コンピューターが生み出す疑似乱数は、実用上は乱数として十分に役立つものの、厳密な意味では真の乱数とはいえません。では、真の乱数、すなわち「いっさいの規則性がみられない乱数」をつくるには、どうすればよいのでしょうか。

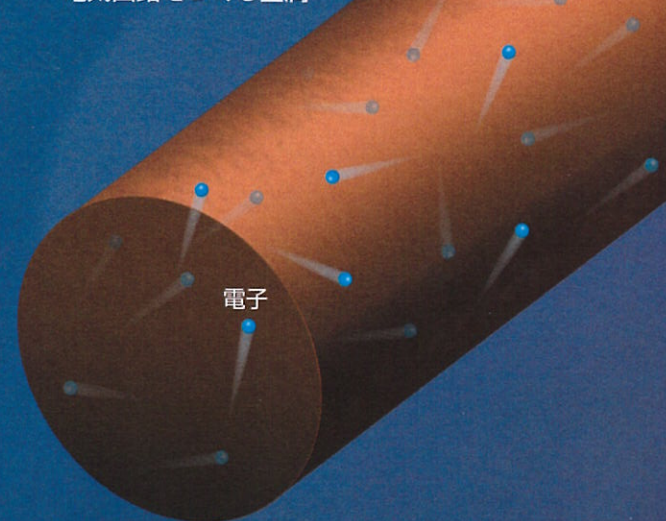
その方法は、現実世界の物理現象にあらわれる不規則性を利用して乱数をつくるというものです。これは「物理乱数発生法」とよばれます。最も実用化が進んでいる物理乱数発生法は、電気回路にあらわれるノイズを利用する方式です。電気回路にかける電圧の値を一定に保とうとしても、どうしても小刻みなノイズ（熱雑音）が生じます。このノイズにはいっさいの規則性がみられないため、これを数値として取りだせば乱数として活用できるのです。

光や電子をコインがわりに使う方法も

物理乱数発生法には、光の素粒子である「光子」を使う方式もあります。ある特殊なハーフミラーにやってきた1個の光子は、「偏光（振動方向のかたより）」とよばれる性質にしたがって、50%の確率で透過し、50%の確率で反射します。コインの裏表のように、光子が透過したら「0」、反射したら「1」とすれば、2進数の乱数になるのです。

さらには、電子がもつ「スピン」という性質を利用する方式も研究されています。地球に自転軸があるように、電子などの素粒子にも自転軸に相当するものがあります。それが「スピンの向き」です。ミクロな世界の物理学である「量子論（量子力学）」によれば、1個の電子は「スピンが上向きの状態」と「スピンが下向きの状態」が重ね合わされており、どちらの状態かは観測するまで決まりません。電子を観測して、スピンが上向きなら「0」、下向きなら「1」とすれば、やはり2進数の乱数として使えるというわけです。

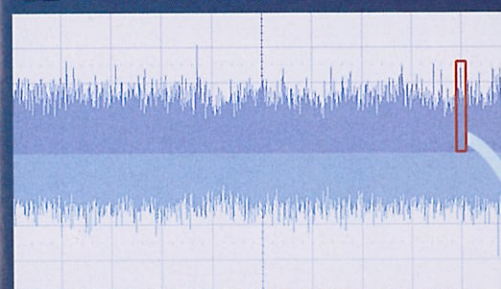
電気回路をつくる金属



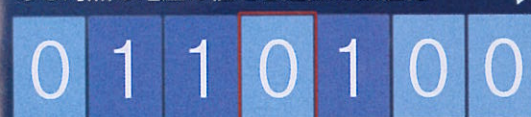
ノイズから乱数をつくる

電気回路の電圧を測定すると、小さなノイズ（熱雑音）があらわれます。一定時間ごとに電圧を測定して数値化し、それを2進数の数としてあらわします。その一部の桁を取り出すことで、0と1からなる2進数の乱数となります。一部の高性能のコンピューターには、熱雑音方式の物理乱数発生チップが搭載されています。

電圧のノイズ（熱雑音）



ある時点の電圧の値を2進数で数値化



特定の桁の数字を乱数として出力

物理現象から乱数をつくる

現在までに実用化あるいは研究されている、三つの方式の物理乱数発生法を示しました。これら以外にも、さまざまな方式の物理乱数発生法が研究されています。



光源

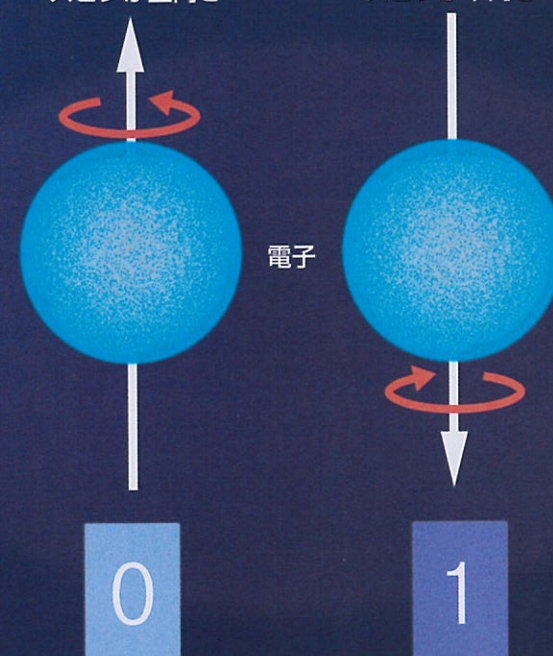
ハーフミラー

透過

反射

スピンが上向き

スピンが下向き



電子のスピンを測定して乱数をつくる

産業技術総合研究所では、電子がもつスピンの向きを乱数発生に利用する「量子ダイス」（ダイスはサイコロの意味）の研究を進めています。電子のスピンの向きを0と1に対応させる方法は、従来のコンピューターをはるかにこえる計算速度の「量子コンピューター」にも活用されています。

光の粒から乱数をつくる

光源からハーフミラーに向かって光子を1個ずつ発射し、ハーフミラーで透過したら0、反射したら1とすることで乱数をつくることができます。光方式による乱数発生装置も、すでに製品化されています。

高度に進化した“サイコロ”が、私たちの暮らしを支える

スマホ・ネット社会を支える乱数

疑似乱数や物理乱数は、コンピューターゲームのプログラムや、モンテカルロ法を使った科学研究のシミュレーション、AIの開発などに欠かすことができません。そして、スマートフォン（スマホ）やパソコン、インターネットが広く普及した今日の社会においても、乱数は非常に重要な役割を果たしています。

電子メールやLINEなどのアプリ、ネットショッピングでの情報通信には、「暗号化」という技術が使われています。通信の内容を暗号化したり、暗号化された文章を解読するための「鍵」をつくる際に、その都度発生させた乱数が使われているのです。

また、銀行のオンラインバンキング、TwitterやFacebookなどのSNS（ソーシャル・ネットワーキング・サービス）の2段階認証などには、一度しか使うことができない「ワンタイム・パスワード」が使われることがあります。このワンタイム・パスワードを発行するしくみにも乱数が使われています。スマホ



昭和20年代に使われた乱数発生用のコイン
昭和20年代に、乱数表をつくるために統計数理研究所で使われていたコインです。つぼの中に多数のコインを入れ、かきまわして取りだすことをひたすらくりかえすことで乱数表がつくられました。現在の乱数表は、メルセンヌ・ツイスター法による疑似乱数や、物理乱数発生器でつくった乱数などを使うことが日本工業規格（JIS）によって定められています。

やインターネットによって便利になった私たちの暮らしは、まさに乱数が支えているのです。

サイコロは、ほんとうにランダムか？

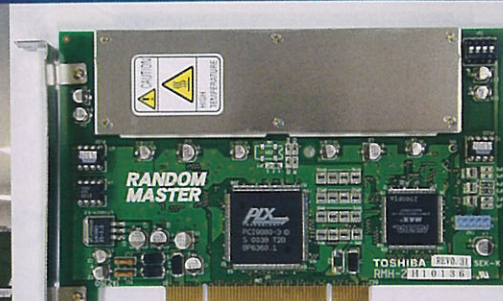
古代から、世界各地の人々はサイコロを使った^{うらな}占いや遊びを行っていました。コインが登場してからは、コインの裏表で勝敗や順序を決める「コイントス」が古くから行われ、現代でもサッカーやテニスの試合で採用されています。これは私たちが、サイコロの出る目やコインの裏表はランダムに決まり、あらかじめ予測できないと信じているからにほかなりません。

しかし、ほんとうに予測できないのでしょうか？
空中に投げられたサイコロを、究極に高い計算能力をもつAIが、^{ちやう}超高性能のカメラで観察したとします。放物運動するサイコロの軌道や回転のようす、^{きどう}空気の流れなどを^{しゅんじ}瞬時に把握し、^{はあく}地面に落ちるタイミングや^{かんべき}転がりかたを完璧に計算できるAIは、どの目が出るかをあらかじめ言いあてられるでしょう。こう考えると、すでに投げられたサイコロの目はあらかじめ決まっている（これを「決定論的」といいます）のであり、究極的には予測可能だといえます。

こうした問題をはじめて考えたのは、フランスの科学者ピエール・ラプラス（1749～1827）でした。上の例のAIのような全知全能の存在は「ラプラスの魔物」とよばれます。サイコロの目やコインの裏表がランダムに見えるのは、ラプラスの魔物もつ高い計算能力を私たち人間がもたないからだ、といえます。このように、決定論的にもかかわらず、複雑な過程によって将来の予測がむずかしくなる現象は「カオス」とよばれ、ランダムとは区別される場合もあります。

「究極の乱数」を求めて

統計数理研究所の^{たむらよしやす}田村義保博士は、乱数発生の研究を長年進めてきた一人です。田村博士らがめざすのは、何かの規則にもとづく疑似乱数ではなく、いっさいの



最新の物理乱数発生システム

（左）統計数理研究所の物理乱数発生システム。スーパーコンピューターのようにも見えるこの装置は、乱数発生器という意味で、現在の最高性能のサイコロの一つといえます。

（上）システムが搭載する熱雑音方式の物理乱数発生ボード。1秒間に640メガバイトの乱数を出力することができます。

規則をもたない「真の乱数」です。そのために田村博士は、民間の技術者と共に、電気回路のノイズ（熱雑音）を使った物理乱数発生器を研究してきました。

原子力発電にかかわる測定器など、特別な用途に^{ようど}使われる電気回路には、不要なノイズを極限まで減らす技術が求められます。ノイズを減らすための研究からは、乱数に使える“よいノイズ”をつくるための知見も得られます。こうした知見を活用することで、高速に乱数を発生させる電気回路の開発に成功しました（上の写真）。

「アナログ情報であるノイズはランダムなのですが、それを0か1というデジタル情報に変換するときに、^{へんかん}少しだけ規則性が出てしまいます。それをどう取りのぞくのが最もむずかしいところでした」（田村博士）。

“神のサイコロ”さえも利用する

物理学者アルバート・アインシュタイン（1879～1955）がのべた有名な言葉に、「神はサイコロを振らない」というものがあります。量子論では、1個の電子は複数の状態（たとえばスピンが上向きか下向きか）が重なり合っているとみなします。そして、そのどち

らの状態が観測されるかは偶然によって支配され、確率的にしか予測できない、と考えます。これはまるで、すべての物事を決める神が、サイコロを振って出た目に応じて電子の状態を決めているかのようです。アインシュタインは、こうした奇妙な考え方に反発して、「神はサイコロを振らない」とのべたのです。これに対して、量子論の創始者の一人である物理学者ニールス・ボーア（1885～1962）は「神がいかに世界を支配されるべきかを指図することは、われわれの課題ではありません」と反論したといえます。

科学者や数学者たちは、サイコロやコインにはじまった乱数発生器を、今もなお改良しつつづけています。メルセンヌ・ツイスター法に代表される高性能な疑似乱数は、少なくとも実用上は、真の乱数とほとんど見分けがつきません。そして、よりよい物理乱数発生器を求める研究者たちは、アインシュタインがいうところの“神のサイコロ”、すなわち量子論が支配する確率的な現象さえも利用しはじめています。こうして日々生みだされる無数の乱数が、今日の科学研究や私たちの暮らしを支えているのです。

（執筆：編集部 板倉 龍）

※：W. ハイゼンベルク著『部分と全体 私の生涯の偉大な出会いと対話』（山崎和夫 訳）より引用。

ギャンブルで確実にもうけることは可能？ 不可能？

“ツキ”は存在するのか？

ギャンブルで連続して勝った場合、私たちは「ツキがあった」と喜ぶ。逆に負けがこむと「ツキがなかった」とくやしがる。「ツキ」などというものが実際に存在するのだろうか？

結論からいえば、私たちがツキとよんでいるものは、結果の「かたより」のことである。102～103ページで紹介した、コインを1000回投げた実験を思いだしてほしい。あの実験では、表が9回連続で出た場合もあった。ランダムなできごととは、私たちの想像以上にかたよることがあるのだ。もしあのコイン投げのギャンブルで、9回連続で負けがつづいたら、きっと私たちはツキがないと思うだろう。しかし実

際には、9回連続で表がつづくのは、 $(1/2)^9 = 1/512$ の確率である。1000回投げたのだから、確率1/512のことがおきることもあるだろう。

結局、ツキとは、あとになって検証した際に、かたよった結果に注目してツキだと感じているにすぎない。たとえばギャンブルで絶好調がつづいていたとしても、それはあくまでも偶然の結果であり、同条件であるかぎり確率そのものは変化していない。次の勝負で勝つ確率は本来の確率のままなのである。

ツキをコントロールすることはできないのだから、ギャンブルの期待値が100%を下まわっているときには、運よく勝っている間に“勝ち逃げ”するの

が正解ということになる。

逆に、「ツイているから、もう一回！」と賭をつづけるのは、胴元の思う壺だ。胴元は、客に賭をつづけさせれば、大数の法則により理論的な確率に収束して、損を取り返せるとわかっている。だからこそ、胴元は勝っている客に寛大に振るまうのである。

ギャンブルで理論的にもうけた例

このように、ルーレットなどのギャンブルで確実にもうけることは、理論的には不可能だ。しかし実は、胴元の予想を裏切り、もうけた例がある。1873年、ジョセフ・ジャガーズというイギリス人が、モンテカルロのルーレットで大もうけしたことが知られている。

実はジャガーズは、前日にカジノに部下を送り、ルーレットの出目を記録させていた。そして、あるルーレットの出目が理論的な確率からずれていることを見抜き、翌日からそこで出やすい目につづけたのだ。

実在するサイコロやルーレットは、いかに精密につくられているといっても、その出目の確率は理論的な確率からわずかにずれている。そのずれが無視できるほど小さく、期待値が100%以下なら、ギャンブラーはもうけられない。しかし、もしある程度以上ずれが大きければ、胴元をだしぬくことができるのだ。

しかし、ジャガーズの異常な“ツキ”に気づいた胴元は、すぐにルーレットの出目が理論値からずれていることに気づいた。そして、ルーレットを毎日

点検するようになったという。

宝くじを買い占めた投資グループ

134～135ページでは、日本の宝くじは非常に期待値が低いことを解説した。ところが、オーストラリアの投資グループは、1992年に莫大な資金を投じてくじを買い、みごとに利益を上げた。合理的に投資を行い、利益を上げるプロが、なぜくじに資金を投じたのだろうか。

アメリカ、ヴァージニア州のロトくじは、1～44の数字の内、6つを選ぶというルールだった。あり得る組み合わせは、 $44!/6!38! = 705万9052$ 通りである。くじは1枚1ドルなので、約700万ドルですべてのくじが買える。そして、一等の当たりくじの当選金は、2700万ドル。1枚当たりの当選金の期待値は約3.8ドルなので、費用1ドルを大きくこえていたのだ。

もしすべてのくじを買い占めることができれば、確実に当選し、もうけることができるはずである。投資グループはこれを実行した。時間が足りず、500万枚しかくじを買えなかったが、当たりくじはその500万枚に含まれていた。また、一般の購入者がたまたま同じ番号の組み合わせを買っているということもなかった。こうして投資グループは、賞金を手にしたのである。

ポーカーのプロは何故強い？

上の二つの例のように、よほど特殊な条件がそろわない限り、ギャンブルで確実にもうけることは不可能だ。



ポーカープロはなぜ強い？

しかしポーカーは、ほかのギャンブルとはちがい、プロのプレイヤーが存在し、トーナメントが行われている。ポーカーの手が出る確率が理論的に決まっているとしたら、ポーカーのプロはなぜ強いのだろうか。

ポーカープロといえども、次にどんなカードがまわってくるかはわからない。もしわかったらそれはいかさまだ。しかしポーカープロは、机上にどんなカードが出ているかを見ることで、次にどんなカードが出てくる確率が高いかを推測できる。素人よりも確率に精通し、さらに相手プレイヤーのゲームの癖を読むことで、高い勝率を保っているのだ。

じゃんけんの癖は弱さのもと

じゃんけんの勝率は、どの手も1/3である。にもかかわらず、もし「自分はじゃんけんが弱い」と思うなら、そ

れは相手にあなたの癖を読まれているからかもしれない。

統計数理研究所のHPに公開されている「じゃんけんゲーム」で、自分のじゃんけんの癖を知ることができる。このゲームのプログラムは、はじめはランダムに手を決めているが、やがてあなたが出す手の癖を学習して、次の手を予測し、あなたに勝つようになる。ランダムに手を出せるようになれば、勝率は理論的な確率である50%になる。コンピューターと引き分けられるよう、練習してみよう。

じゃんけんゲームのHP

http://appli.ism.ac.jp/janken/janken_support.html



ツキは実在するのか？