確率論で見る自然現象*

原　隆
名古屋大学多元数理科学研究科
e-mail: hara@math.nagoya-u.ac.jp
http://www.math.nagoya-u.ac.jp/~hara/

2003 年 8 月 21 日

概 要

確率論を通して，自然現象の一端を捉えてみる．特に，コイン投げの問題を中心に据えて，ランダムな中に見られる規則性を考えていく．この基本的な例をとおして「大数の法則」「中心極限定理」「ランダムウォーク（とブラウン運動）」の初歩的な部分に触れることをめざす．これらは自然現象・社会現象に確率論が登場する際に重要な役割を果たす．

（講義時からの改良点）これはアゴラの時に配ったノートに改良を加えたものである．主な改良点は，実際に講義を行った際に出た質問などをより詳しく説明したこと — 特に，中心極限定理に関する 2.6 節〜2.7 節を充実したこと — である．ただし，改良に際してはこの講義ノートを単独で読んだ場合に読みやすくなることを心がけた．従って，実際の講義と順序がすこし変わっているところや，講義では時間が無くて触れられなかった材料なども存在する．なお，アゴラの講義時にこの完成版を用意すべきであったのではあるが，正直どの辺りに質問が集中するかの予想がはずれたために結果的に不可能であったことをお詫びする．

目 次

*数学アゴラ，2003 年 8 月 4〜6 日，於名古屋大学

1 はじめに：考える問題

日常,「確率」と言う言葉を耳にすることは多い — ほとんど毎日聞かされるのは「今日の降水確率は...」だろう．また，宝くじに当たる確率は○○，トランプのポーカーでこの役ができる確率は○○，なども耳にする．

このように「確率」は不規則な（ランダムな）現象，確実には結果を予測しがたい（でも何らかの予測ができる）場合を扱う際に使われている．そしてまた，確率論の初歩ではいろいろな確率を計算することに重きが置かれる．いろいろな確率を計算できることはそれ自身重要であるし，常識に反した結果を出すものも多々あるので非常に面白い．

しかし，この講義では少し異なった観点から確率を眺め，そこに潜む規則性を探っていく．特に,「大数の法則」「中心極限定理」などが中心となるだろう．

さて，確率論は単なる数学上のお遊びではなく，確率の絡んだ現象はいろいろなところに顔を出している．いくつかの例を挙げてみよう：

a. 物理や化学の実験では「測定には誤差が付き物だから何回か測定して測定値の平均をとるように」と教わっていると思う．この考えは日常的にも頷ける（何回も実験をくり返すと「真の」値に近づく）ものである．

b. ある高校の一学年の男子をとりだし（３００人くらい），身長を測定してその結果をヒストグラムにした（横軸に身長，縦軸にその身長の人が何人くらいいるかを書く）．その結果はなだらかなベルのようなカーブになるだろう．これは身長に限らない — 体重についても似たようなグラフが出るだろう．また（生臭くて申し訳ない）この学年の生徒の数学の期末テストの成績についても，似たような結果になるかもしれない[1]．

c. 拡散現象．容器に臭素の結晶と空気を入れ，密閉して放置すると，段々と臭素が容器中に拡がっていくのがわかる.（中学校などで実験をした人がいるかもしれないね.）これを拡散現象と言うが，臭素の色ついた部分は，時間とともにどのように拡がっていくだろうか[2]？

c′. ブラウン運動．たばこの煙などを顕微鏡で見ると，煙の粒子がフラフラと動いているのが見えるだろう．これは煙の粒子に空気の分子がいろいろな方向からぶつかって，不規則な運動をしているのであるが，この粒子は時間とともに，どのように動いていくだろうか？

c″. 気体の密度．空気は酸素と窒素の分子からできていることは知っているだろう．これらの分子は熱運動で激しく動いているはずだが，気体の密度はいつも一定に見える．これはなぜか？

d. 株価の変動．株価は日によって（又，同じ日のうちでも時間によって）不規則に動いている．非常に不規則に見えるのだが，ある程度ならして見ると，何らかの規則性が見えるようにも思う．

[1]ただし，成績の分布については身長や体重ほど話は単純ではない．その理由も後で少しだけ理解できるかもしれない

[2]類似の現象は水に食塩の結晶を溶かす場合などでも見られるが，日常生活で塩や砂糖を溶かす場合は掻き回してしまうからここで問題にしている現象は見えにくい（そもそも，食塩や砂糖では色がついていないから見えないが，インクなどを使ってもちょっとした液体の運動にかき消されてしまうので難しい）．また，液体中の拡散は気体中の拡散に比べて非常に遅いので，液体の場合は密閉した容器でも観測は簡単ではない

e. 溶媒の中の高分子． DNA のように鎖状になった高分子を溶媒に入れると，高分子は周りの溶媒の分子との熱運動でいろいろと形を変え，ある程度クシャクシャにまるまった形になる．このとき，高分子の長さと高分子の拡がり（丸まった高分子の端から端までの長さ）には，どのような関係があるか？

これらの現象は，一見，無関係なように見えるが，奥の方ではつながっている．a と b は「中心極限定理」という確率論の重要な定理，c と d は確率論の重要な研究課題であるランダムウォーク（ブラウン運動）というものの現れである[3]． e は統計力学の未解決問題の一つであるが，ランダムウォークとも密接な関連がある．さらに，中心極限定理とランダムウォーク自身にも関連がある．この講義では上のような現象を理想化・簡単化した状況を考えることで，このような現象がなぜ見られるのか，その一般的原理を理解することを目的とする．同時にこのような考察を通して，現代数学の持つ美しさの一端を紹介できれば幸いである．

1.1 記号の約束と「オーダー」の概念

不等号：

$a \le b$ は $a \leqq b$ と，$a \ge b$ は $a \geqq b$ と同じ意味.

和の記号：

$x_1+x_2+x_3$ の事を $\displaystyle\sum_{i=1}^{3} x_i$ と書く．同様に $a_1+a_2+\cdots+a_n$ を $\displaystyle\sum_{i=1}^{n} a_i$ と書く．このように $\displaystyle\sum_{i=1}^{N}$ はこの記号の後にあるものの i を 1 から N まで変えたものの和を表す．この際，i の代わりに j や k を使っても構わない．例を挙げると：

$$X_1+X_2+X_3+\cdots+X_N=\sum_{i=1}^{N} X_i=\sum_{j=1}^{N} X_j \tag{1.1.1}$$

などと書ける．この和の記号は慣れると便利で曖昧さがないので，以下でも多用する．

$N \to \infty$：「N が限りなく大きくなる極限」の概念：

この講義では 「N がどんどんと大きくなっていったときに何が起こるか」という問題をよく考える.「N がどんどんと限りなく大きくなる」ことを数学では 「N が無限大（の極限）に行く」と言い，$N \to \infty$ と書く．アゴラの参加生に高校一年生が多かったことを考え，この「極限」の用語や記法はできるだけ使わないようにするが，既に極限を知っている人のためにこの注意を設けた.

「オーダー」の概念：

（これは講義録では陽には使いませんが，知っておくと読みやすくなると思うのでここで説明します.）$f(N)$ を正の整数 N の関数とする（例：$f(N)=N^2$ とか，$f(N)=\frac{2}{N}$ とか）．この講義では N が大きくなっていったときに $f(N)$ がどのくらいの速さで大きく（小さく）なるか，に注目するので,「オーダー」という概念が便利である.

例をまず挙げると，$f(N)=N^2$ も $f(N)=5N^2$ も，$f(N)=\frac{1}{100}N^2$ も，全部 N^2 のオーダーと言う．つまり，N が大きくなっていくときに $f(N)$ が大きく（または小さく）なっていく<u>一番主要なところ</u>を，定数倍は無視して N の関数として表したものが「オーダー」である.「定数倍は無視」というのがミソで，要するに N が非常に大きく（無限大に）なった場合の状況を考えている.

別の例では $f(N)=\frac{1}{N}$ も $f(N)=\frac{3}{2N}$ も，ともに $\frac{1}{N}$ のオーダーだ．一方，$f(N)=\frac{1}{N^2}$ は N^{-2} のオーダーになる.

この講義では N が大きくなったときにある量がどのくらいの速さで大きく（または小さく）なるか，の問題が頻出するが，これは要するにその量のオーダーを訊いていることになる[4].

---

[3]ただし，上で挙げたような実際の自然現象，社会現象は様々な要因が絡み合って起こるから，a〜d はこれらの定理やモデルそのものではない．特に d には他の要素も大きい．ここはあくまで，ある程度の大ざっぱな話と思っていただきたい

[4]「オーダー」の定義には少し混乱があって，数学でよく使う定義は以下である：「$f(N)$ が N^α のオーダーである」とは，定数 C があっ

2　コイン投げの数理：大数の法則と中心極限定理

上に述べた問題 a〜e のとっかかりとして，コイン投げを考える：１０円玉を投げて，表が出るか裏が出るかを考えるのだ．ただし一回投げただけでは面白くないので，何回も投げ（一万回とか），そのうちのどのくらいが表になるか，を考えてみる．

直感的に「そりゃあ，投げた回数の半分くらいは表でしょ」と言いたくなるし，これは間違いではないのだが，もう少し定量的にも深く考えてみたい．

2.1　実際にやってみる

実際に４回ほど，コインを投げてもらい，その結果（n 回表になった人は何人）を集計した．その結果は大体，以下のようになった．

表の出た回数	0	1	2	3	4
その人数	10	18	20	18	3
人数／全人数	0.143	0.257	0.300	0.257	0.043

４回とも表であった人も，４回とも裏だった人もいるね．

2.2　少し解析する．N 回のうちに m 回表になる確率は？

では，上の結果がどのように解釈できるか，考えていこう．この講義では条件 A が実現される確率を $\mathbb{P}[A]$ と書く．例えば　$\mathbb{P}[$コインを一回投げた結果が表$]$　は文字通り「コインを一回投げた結果が表」である確率を表す．

（余分な注）本論に入る前に確率の背景についての注を２つ述べておく．

- 確率とはいったい何か，特に「現実の問題で確率をどのように決めるか」と言うのはそれほど簡単な問題ではない．１７世紀頃から延々と議論がくり返されてきたにもかかわらず，明快な解答は得られていない．むしろ，数学としての確率論はこの問いをうまく回避することで成立した経緯がある．

  この講義でもこの問いに直接取り組むことはせず，$\mathbb{P}[A]$ を「何回も同じ実験をやった場合に A が実現される割合」というくらいの認識で出発する．ただし，幾分トートロジーめくが，この決め方が矛盾のないものであることは後の大数の法則で見るだろう．

- コインを一回投げたとき，表が出るか，裏が出るか，は古典力学の問題である．つまり，コインの材質，質量分布，表面の様子・弾力，コインを受ける面の様子（摩擦，弾力など），そして何よりコインを投げる様子（コインに与える初速度），などをすべて与え，空気の抵抗や重力の作用を考慮して計算すれば，どのようにコインが着地するかを予言することは理論的に可能なはずである．

  このように考えると，確率論は必要ないようであるが，そうではない．コイン投げの場合，条件（コインをはじく強さ，はじく位置，コインの温度による弾性，etc）の微妙な差によって表裏の結果が異なる．かつ，これらの微妙な条件を生身の人間がコントロールすることはほとんど不可能であるので，微妙に異なった条件の結果として，表裏がランダムにでているように見えてくる．この意味で確率論は有効である — 毎回同じように投げる「コイン投げマシーン」を使った場合は結果は同じはずで確率論の出番はないだろう．

て，$|f(N)| \le CN^\alpha$ がすべての $N \ge 1$ で成り立つこと．つまり，この定義によれば $f(N)$ が N^α よりずっと小さくても良い（極端な場合，$f(N) = 1$ でも $f(N)$ は N のオーダーである，といえる）．しかしこの定義はこの講義程度ではかえってわかりにくいので採用しない．なぜ，厳密な定義がこうなっているかというと，ここで採用した定義のなりたたない関数も扱えるようにするためである．(例：$f(N)$ の定義が「N が偶数なら N^2，奇数なら N」となっている場合，この講義ノートでの定義ではオーダーが定義できずに困ってしまう．一方厳密な定義ではこの関数のオーダーは N^2 だ）

このように古典力学の世界では，確率論は我々の側のある種の「情報の欠如」（コイン投げならコインの初速度などがコントロールできない）に伴って登場することが多い．なお，量子力学では「情報の欠如」とは本質的に異なった意味で確率論が登場する．

（余分な注終わり）

コインを N 回投げたときの i 回目の結果によって決まる確率変数（ランダムな数）X_i を定義しよう（$i = 1, 2, 3, \dots, N$）．ここで

$$X_i = \begin{cases} 1 & i \text{ 回目が表の時} \\ 0 & i \text{ 回目が裏の時} \end{cases} \tag{2.2.1}$$

と決めておく.（$0, 1$ を使うのは,「表」「裏」と書くのがじゃまくさいからであるが，後で見るように別の効用もある.）そして１回目からの結果を並べて (X_1, X_2, X_3, X_4) などのように書こう.

この記法に従うと，１回目から４回目まで表だけが出るのは $(1, 1, 1, 1)$ と書かれる．同様に，$(1, 1, 0, 1)$ は３回目だけが裏で残りは表，の場合を表している．時にはスペースを省略するため，$(1, 1, 0, 1)$ の代わりに 1101 などと書くこともある.

これから，上のような出方のそれぞれが，どのくらいの割合で起こるか，その確率を計算していこう．それにはコイン投げについて２つの重要な仮定を行う必要がある.

１つ目の仮定：

一つ目の仮定は，コインを１回投げた場合の表と裏の出やすさの割合である．通常のコインは表裏がほとんど同じに作ってあるし，材質も均一だろうから，表と裏はほとんど同じくらい出やすいだろうと思われる．そこで我々はコインの表と裏は同じくらい出やすいと仮定し，$\mathbb{P}[\,\text{コインを一回投げた結果が表}\,] = \frac{1}{2}$ ととる事にする．ただし，実際にはコインのひずみによって $\mathbb{P}[\,\text{コインを１回投げた結果が表}\,] = \frac{51}{100}$，$\mathbb{P}[\,\text{コインを１回投げた結果が裏}\,] = \frac{49}{100}$，などととるのが良いのかもしれない．この取り方が良かったかどうかは後で実験をしてみないとわからない．後で表と裏の出やすさが違う場合を考えるが，その際には $\mathbb{P}[\,\text{コインを１回投げた結果が表}\,] = p$（$p$ は $0 < p < 1$ なる決まった数）とおいて計算していく．以下ではより一般の場合でもできるように，$\mathbb{P}[\,\text{コインを１回投げた結果が表}\,] = p$ として進むが，特に断らない限りは $p = \frac{1}{2}$ と思って良い.

なお，投げ方によってはコインが端で立つような事もあり得るが，簡単のためにそのような場合は起こりえないとして進む.

２つ目の仮定：

上の仮定はコインを一回投げた場合の確率を言っているだけで，２回以上投げた場合にどうなるかには新しい仮定が必要である．それがコイン投げの独立性に関する以下の仮定である：

> （普通の人がフェアに投げた場合）コイン投げの結果をコントロールする（表か裏を選択的に出す）ことはほとんど不可能である．すなわち，表を出してやろうとか，裏を出してやろうとか思っても，自分の意志でそのようにすることはできない．特に，i 回目までの結果を見て，$i+1$ 回目以降の結果を左右しようとしても，それは不可能である．その結果，$i \neq j$ の場合，i 回目の結果と j 回目の結果の間には何の影響力も働いていない.

これは X_i の言葉に直すと，どうなるだろうか?手始めに $\mathbb{P}[\,X_1 = 1 \text{ かつ } X_2 = 1\,]$（１回目，２回目ともに表になる確率）について考えてみよう．１回目に表が出るのは全体の p の割合である．２回目も表になるのは１回目が表だったうちの p の割合のはず（ここで独立性を使った[5]）．結局，$\mathbb{P}[\,X_1 = 1 \text{ かつ } X_2 = 1\,] = \mathbb{P}[\,X_1 = 1\,]\,\mathbb{P}[\,X_2 = 1\,] = p^2$ となる．同様に，$\mathbb{P}[\,X_1 = 1 \text{ かつ } X_2 = 0\,] = \mathbb{P}[\,X_1 = 1\,]\,\mathbb{P}[\,X_2 = 0\,] = p(1-p)$ となる．一般に ϵ_1, ϵ_2 を 0 か 1 のどちらか（どっちでも良い）として

$$\mathbb{P}[\,X_1 = \epsilon_1 \text{かつ } X_2 = \epsilon_2\,] = \mathbb{P}[\,X_1 = \epsilon_1\,]\,\mathbb{P}[\,X_2 = \epsilon_2\,] \tag{2.2.2}$$

[5]もし独立でなく，例えば１回目と同じ結果が出やすい場合は，$\mathbb{P}[\,X_1 = 1 \text{ かつ } X_2 = 1\,] > \mathbb{P}[\,X_1 = 1\,]\,\mathbb{P}[\,X_2 = 1\,]$ となるだろう

が成立するはずである[6].

このような事情は3回以上の結果についても同様に成立するから，結果として

$$\mathbb{P}[\,X_1=\epsilon_1, X_2=\epsilon_2,\cdots,X_N=\epsilon_N\,]=\mathbb{P}[\,X_1=\epsilon_1\,]\,\mathbb{P}[\,X_2=\epsilon_2\,]\,\cdots\,\mathbb{P}[\,X_N=\epsilon_N\,] \tag{2.2.3}$$

となる[7]．ここで ϵ_i は 0 でも 1 でも，勝手な値でよい．これを $\mathbb{P}[\,X_i=1\,]=p, \mathbb{P}[\,X_i=0\,]=1-p$ を代入して書き直すと

$$\mathbb{P}[\,X_1=\epsilon_1, X_2=\epsilon_2,\cdots,X_N=\epsilon_N\,]=p^{(\text{表の数})}\,(1-p)^{(\text{裏の数})} \tag{2.2.4}$$

となることもわかる．要するに，表が出る確率は p，裏が出る確率は $1-p$ だから，それを表と裏の個数分だけかければよいのだ[8].

これを元に,「N 回投げたときに m 回表が出る」確率を求めよう．後のために

$$S=S_N=X_1+X_2+\cdots+X_N=\sum_{i=1}^{N}X_i=(\text{表の出た回数}) \tag{2.2.5}$$

を定義しておく.

簡単なところからいこう．$N=1$ の時は仮定そのもので

$$\mathbb{P}[\,S_1=1\,]=p, \qquad \mathbb{P}[\,S_1=0\,]=1-p \tag{2.2.6}$$

で面白くない．$N=2$ の時,

$$\mathbb{P}[\,S_2=2\,]=p^2, \qquad \mathbb{P}[\,S_2=0\,]=(1-p)^2 \tag{2.2.7}$$

は両方とも表，両方とも裏，だから良いだろう．$S_2=1$ の場合はどうか？この場合，$(1,0)$（初めに表，次に裏）と $(0,1)$（初めが裏，次に表）の2通りの出方があり，どちらも確率は $p(1-p)$ である．よってこの2通りを足して,

$$\mathbb{P}[\,S_2=2\,]=p^2, \qquad \mathbb{P}[\,S_2=1\,]=2p(1-p), \qquad \mathbb{P}[\,S_2=0\,]=(1-p)^2 \tag{2.2.8}$$

となる（他の場合も比較のために書いた）.

$N=3$ も同様に計算できる．全部表，全部裏は良いとして，$S_3=2$ の場合を考えると，$110, 101, 011$ の3通りの出方があり，それぞれの確率は $p^2(1-p)$ である．従って（全部表や全部裏，の場合も書くと）,

$$\mathbb{P}[\,S_3=3\,]=p^3, \quad \mathbb{P}[\,S_3=2\,]=3p^2(1-p), \quad \mathbb{P}[\,S_3=1\,]=3p(1-p)^2, \quad \mathbb{P}[\,S_3=0\,]=(1-p)^3 \tag{2.2.9}$$

となる.

このへんで一般に「N 回投げて m 回表」の確率を考えよう．何通りの出方があるか，と言うのが問題だが，これは「N 個の結果の中で丁度 m 個だけ $X_i=1$ となる」なり方の個数である．これを

N 個から m 個をとる組み合わせの数

---

[6] ϵ はイプシロンと読むギリシャ文字である

[7] この式，およびその元になった独立性の仮定を当たり前だと思ってはいけない．それは以下の問いを考えるとわかる：「普通の（表裏が同じ確率で出るだろう）コインを 100 回投げたら 100 回とも表だった．101 回目も表の確率は何か？」独立性を仮定するなら答えは $\frac{1}{2}$ であるが，なんとなく「100 回も表が続いたんだから次は裏が出やすいだろう」と考えたくならないだろうか？

今は既に 100 回も表が出てしまった場合を考えているので,「100 回とも表だった」という条件のもとでの「次は裏」を考える必要がある．独立性の仮定はここで,「条件が付いていてもいなくても確率は同じで $\frac{1}{2}$」と主張するものであり,（コイン投げをコントロールすることは実質的に不可能であることなどを考えると）今までに説明したようにこの独立性の主張が正しい（現実に近い）と思われる．この辺りの考えは数学 B の「条件付き確率」で明快になるだろう.

しかし，これは「100 回表だったので次は裏」とは反する考えであることには十分に注意した方が良い.「100 回表だったので次は裏」と考えがちなのは,「既に 100 回も表が出てしまった」という条件が付いていることをきちんと考えていないためだろうが，我々はどうしてもこのような方向に引きずられやすい.（なお，途中から見たので良くわからなかったが，ここのところを思いっきり勘違いしているテレビ番組が最近あったようである．番組の意図が良くわからなかったが，あれがギャグやネタのつもりでないのなら，かなり恥ずかしいと思う.）

なお，コインを 100 回投げて 100 回とも表だったら，この問いの前提を疑って「このコインはイカサマだ，または投げ方がイカサマだ」とする方が良いかもしれないが，それは別の話である

[8] しつこいが，このようになるのは「独立性」のおかげで (2.2.3) がなりたつからである

といって，${}_NC_m$ で表す．上の考察から ${}_3C_3={}_3C_0=1$，${}_3C_2={}_3C_1=3$ などがわかったが，一般には

$$ {}_NC_m=\frac{N!}{m!\,(N-m)!},\qquad N!=N\times(N-1)\times(N-2)\times\cdots\times3\times2\times1 \tag{2.2.10}$$

であることが（少し考えると）わかる[9]．これを認めると

$$\mathbb{P}[\,S_N=m\,]={}_NC_m\,p^m\,(1-p)^{N-m} \tag{2.2.11}$$

が得られる．さてさて，皆さんに投げてもらった結果と比較すると

表の出た回数 m	0	1	2	3	4
その人数	10	18	20	18	3
人数／全人数	0.143	0.257	0.300	0.257	0.043
確率 $\mathbb{P}[\,S_4=m\,]$	0.0625	0.250	0.375	0.250	0.0625

となる．３行目と４行目を比較すべきだが，当たらずといえども遠からず，というところかな[10]．

これからの予告を兼ねて $p=\frac{1}{2}$ の場合， いろいろな N の値に対して $\mathbb{P}[\,S_N=m\,]$ を計算したグラフを図１に載せる．なぜこんなことが起こるのか，以下で見ていこう．

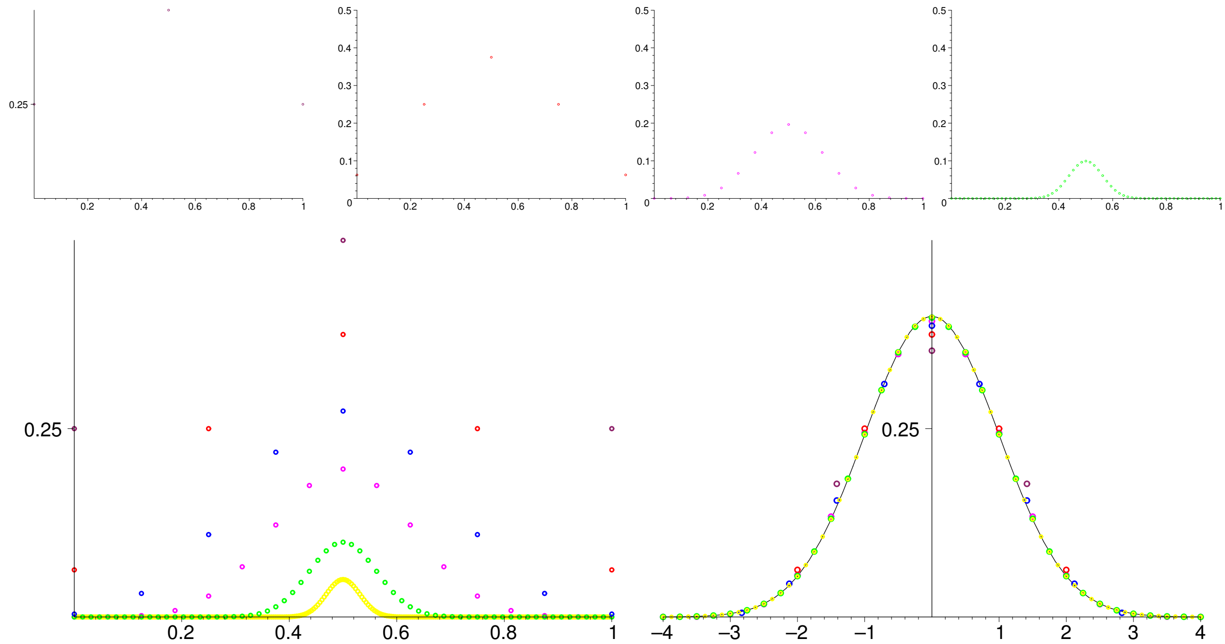


図 1: N 回投げて m 回が表の確率 $\mathbb{P}[\,S_N=m\,]$ のグラフ．いろいろ書いてみた．一行目の４つのグラフは $N=2,4,16,64$ のそれぞれを描いたもので，横軸が $\frac{m}{N}$，縦軸は $\mathbb{P}[\,S_N=m\,]$ である．２行目の左はこの４つ，および $N=8$ と $N=256$ を重ねて描いたもの（軸の取り方は同じ）．２行目の右は左のグラフを $\frac{m}{N}=\frac{1}{2}$ を中心にして縦軸，横軸をうまく伸び縮みさせたものである — どのように伸び縮みさせたのか，また，実線で描いてある曲線は何なのか，は後のお楽しみ．

[9] ここのところは「順列と組み合わせ」として高校一年でやるはず

[10] 確率というのはたくさん（無限に多く）の人に実験をやってもらった結果，というつもりだから，７０人くらいの実験ではバラツキが出て，人数比が一番下の行の理論値に一致しないのは仕方ない．何人くらいの人に実験してもらったら理論値とのズレがどのくらい小さくなるか，というのは今やっていることの延長上の問題である

2.3　N が大きくなったら？ I. 大数の法則

本題に戻ろう．前節では「コインを N 回投げて，そのうちの m 回が表」の確率を（表，裏が同じ確率で出るとして）計算した．結果は

$$\mathbb{P}[\,S_N = m\,] = {}_N C_m \, 2^{-N} \tag{2.3.1}$$

というもので，その結果をグラフで見せた．それを再録すると図 2 の左になっている（ただし，余りたくさん点があるとわかりにくいので $N = 4, 16, 64, 256$ の 4 通りに制限した）．$p = \frac{1}{2}$ だけでは説得力がないので，$p = \frac{3}{4}$ もやってみたのが図の右である．

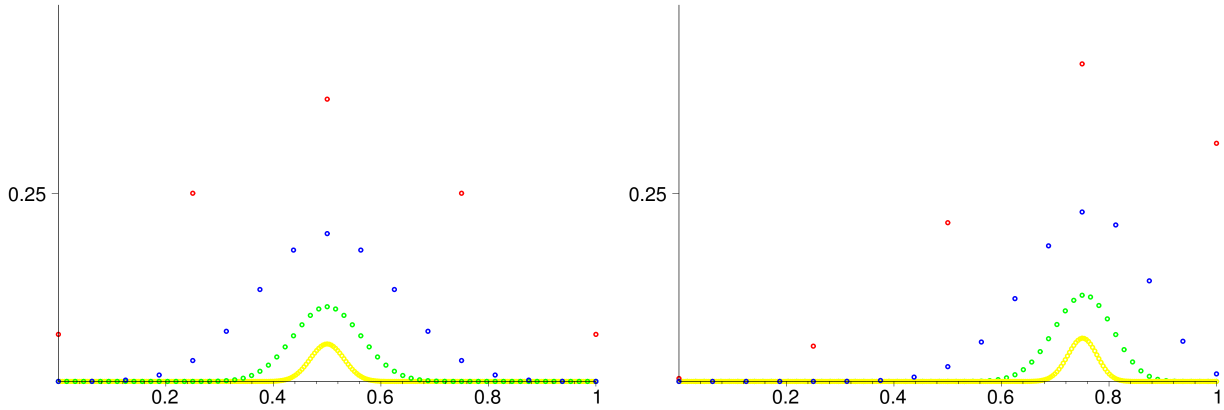


図 2: 左：表と裏が同程度に出やすいコインを N 回投げたときの確率．横軸は m/N，縦軸はその $\mathbb{P}[\,S_N = m\,]$ を表している． 4 種類の点は上から $N = 4$（赤），16（青），64（緑），256（黄）．N が大きくなるにつれて確率が $\frac{S_N}{N} = \frac{1}{2}$ に集中していく．
右：同様の計算を表が $\frac{3}{4}$ で出るコインで行った結果．今度は $\frac{S_N}{N} = \frac{3}{4}$ に集中が見られる．

図 2 では N を大きくすると，$\frac{S_N}{N}$ の分布が p のところに集中していくことが非常に綺麗に現れている．この背後にある定理を述べると以下のようになる（証明は 2.5 節）．

（大数の弱法則）表の出る確率が p のコインを投げた場合，N 回投げたときに表の出る回数を S_N と書く（$\frac{S_N}{N}$ が表の出る割合）．このとき，「$\frac{S_N}{N}$ が p からずれる確率」は N が無限大になるとゼロに近づく．もっと詳しく言うと，勝手な正の数 a に対して，

$$\mathbb{P}\Big[\,\Big|\frac{S_N}{N} - p\Big| > a\,\Big] \le \frac{p(1-p)}{a^2\,N} \tag{2.3.2}$$

が成り立つ．

（細かい注）通常,「大数の弱法則」というのは上の箱の中の前半部分だけを言い，後半の (2.3.2) は含まない．ここでは定理の主張がわかりやすくなるように，後半まで含めて書いた．

とても大ざっぱに言うと，N が大きくなるにつれ，$\frac{S_N}{N}$ が p に近づいていく，と言うことだ．ただし，この言い方は不正確なので注意すべきである．すなわち，N が有限である限り，どんなに大きな N でも,「$\frac{S_N}{N}$ が p からかなり離れている」ことは起こりえる（例えば N 回ともすべて表，つまり $\frac{S_N}{N} = 1$，になる確率は p^N であって，これはゼロではない）．上の定理の主張は「このような変態な可能性は否定できないが，N が大きければ大きいほど，その変態なことが起こる確率はゼロに近づく」と言うものである[11]．

2.5 節の証明を見ればわかるように，この定理はもっともっと広いモデルに対してなりたつ（例：サイコロを N 回，転がして 1 の目が何回出たか，を訊く）．

[11]「大数の強法則」と言うものもあって，それならもう少しだけ強いことが言えるのだが，それは大学でのお楽しみ

2.4 N が大きくなったら？ II. 中心極限定理

さて，大数の法則だけでは N が大きいときに $\frac{S_N}{N} - p$ がどのようにふるまっているのかが良くわからない（N が大きくなると確率的にゼロになる，ことはわかったが，もう少し詳しいことを知りたい）．この答えは「中心極限定理」で与えられるのだが，その説明には少し準備が必要である．まずは 2.2 節で見せたグラフ（図 1）を少し手直しして見せよう（図 3）．

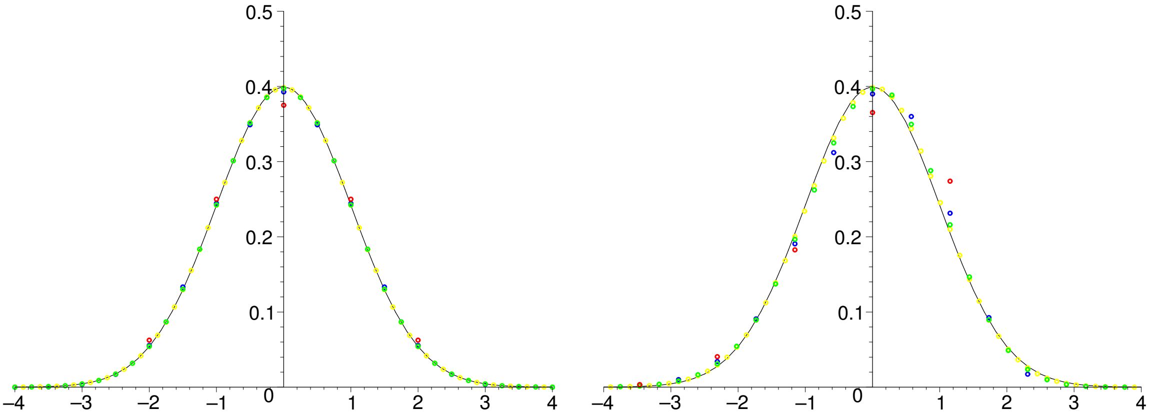


図 3: 図 2 の座標軸を取り替えたもの（横軸方向にずらした後，縦横ともに拡大；図 1 の右下の図に相当）．実線は $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフで，４種類の点は $N = 4$（赤），16（青），64（緑），256（黄） の場合の確率を表す．座標軸の取り方は，横軸は $\sqrt{\frac{N}{p(1-p)}} \left(\frac{m}{N} - \frac{1}{2}\right)$，縦軸は $\mathbb{P}[\,S_N = m\,] \times \sqrt{p(1-p)N}$.
左の図は $p = \frac{1}{2}$ のコインの場合で，右は $p = \frac{3}{4}$ の場合である．
左右ともに，N が大きくなるとこれらの点が急速に実線のグラフの上に乗って行くことがわかる．(p の値が違う右と左が，両方とも同じ関数 $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフに近づいていくことに注意.）

上のグラフを数学的な定理の形で述べるのが，以下の定理である．なお，上では図 2 のグラフを伸び縮みさせたが，本来は「縦軸に確率，横軸に $S_N - pN$」をとったグラフをまず書いて，それを縦軸は $\sqrt{p(1-p)N}$ 倍，横軸は $1/\sqrt{p(1-p)N}$ 倍にする，と考えるのが自然である（この点は 2.6 節でより詳しく説明する）．

表の出る確率が p であるコイン投げを考えよう．このとき，新しい確率変数（ランダムな数）

$$Z_N = \frac{S_N - pN}{\sqrt{p(1-p)N}} = \frac{\sqrt{N}}{\sqrt{p(1-p)}} \left(\frac{S_N}{N} - p\right) \tag{2.4.1}$$

を定義する．この Z_N は図 3 の横軸そのものである．このとき：

（中心極限定理）上の Z_N 自身はランダムであるが，N が大きくなると,「標準正規分布」とよばれるランダムな変数に収束する．つまり，N が大きくなった時，確率 $\mathbb{P}[\,a \le Z_N \le b\,]$ は,

グラフ $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ と 3 直線 $x = a, x = b, y = 0$ で囲まれた部分の面積

に収束する．

ここでいくつかの注を付ける必要があろう．

- $e^{-x^2/2}$ と言うのは，以下のような関数である（この注は高校一年生以下の人向き）．まず $e = 2.71828\ldots$ は「自然対数の底」とよばれる特別の実数である（どのように特別かは微分積分と指数関数・対数関数を学習すればわかる）．次に，e^y というのは，この数 e の y 乗（e を y 回かけたもの）を表す — y が有理数なら良いが，無理数の時の定義には少し工夫が必要だが，ここでは立ち入らない．最後に，この y を $\frac{x^2}{2}$ で置き換えたものが $e^{-x^2/2}$ である．図 4 の左に $y = e^x$，右に $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフを掲げる．

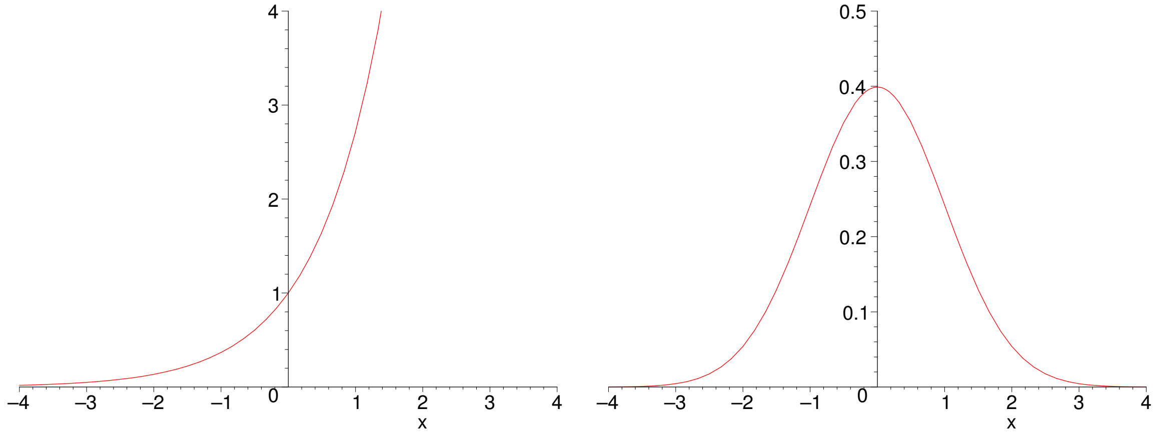


図 4: 左： $y = e^x$ のグラフ.　　右： $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフ

- 標準正規分布 z とは，実数の値をとるランダムな変数で，その分布が
$$\mathbb{P}\left[\, a \le z \le b \,\right] = \left(\text{グラフ } y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ と 3 直線 } x = a, x = b, y = 0 \text{ で囲まれた部分の面積} \right) \tag{2.4.2}$$
で与えられるものである[12].
- (2.4.2) の右辺の面積は「積分」を用いると
$$\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \tag{2.4.3}$$
と書けるのだが，積分をまだ習っていない人も多いと思われるので，ここではこれ以上立ち入らない．積分を知っている人は「ああ，そうだね」と納得してくれればよいし，マダの人は「積分は面積なんだな」とここでは思ってくれればよい[13].

以上がコイン投げの問題に対する，一応の数学的な解答 — 特に我々が直感的に考える「大体半分は表が出るでしょ」の定量的な意味 — である．

ここまでは話をコイン投げに限定してきたが,「大数の法則」や「中心極限定理」はより広い範囲の問題に対しても成り立つ（一般にある程度の性質を満たす「独立」なランダム変数の和について成立する；この事情や上で出てきた $p(1-p)$ などの意味は次節で大数の法則の“ 証明 ”をやると少し見えてくるだろう）．これらの定理はある種の「独立な」現象に関して普遍的に成り立つ非常に一般的なものなので，数学的に非常に美しく，また重要である[14]．同時に，イントロの a, b, c′′ の問題の背景を説明してくれる．

2.5　大数の法則の“ 証明 ”

大数の法則は「チェビシェフの不等式」を用いるとあっけなく証明できる．この威力を堪能するため，少し一般に話を進めてみよう．一般論にするのには，もっと切実な理由もある．生半可なやり方では，以下のような問題に立ち向かえないのだ．

（問題）コインではなく，サイコロを N 回，転がして，出た目の数の合計を S_N とする． $\frac{S_N}{N}$ はどのような値になるだろうか？(または，どのような分布になるだろうか？)

[12] この辺りは意図的にぼかして書いてあるから，わかる範囲で大体の感じをつかんでもらえれば良い
[13] ううむ，こんな事書いてたら石が飛んできそう...　まあ，後で積分はしっかりやってください
[14]（余談）我々が物事を「わかった」「理解した」と感じるのは，一見バラバラな物事にある種の規則性が見えた場合や，様々な局面で統一的に（普遍的に）成り立つ法則を実感した場合が多い．これが僕が「普遍性」に拘る理由である

この問題はコインの問題よりも手強い．一回ごとの結果が 1 から 6 の６通りもあるため，確率としては「N 回の内で，１の目が m_1 回，２の目が m_2 回，３の目が m_3 回，４の目が m_4 回，５の目が m_5 回，６の目が m_6 回でる」ものを考えないといけない（$m_1 + m_2 + \cdots + m_6 = N$）が，この計算はかなり大変（「多項分布」と呼ばれるものになる）．世の中には正１２面体や正２０面体のサイコロもある．そればかりか，実際にはサイコロよりもっともっと複雑な現象も考えたいわけで，何か良い方法がないと苦しくなる.「チェビシェフの不等式」は正にその方法を与えてくれる[15].

2.5.1　確率変数，期待値と分散

まず,「確率変数」と言う概念を正式に導入する．これは一言で言うと,「その値が確率的に決まるような変数」のことであって，コイン投げでの X_i や S_N，Z_N などが例である．

確率変数を定義するには,（1）その確率変数のとりうる値 $x_1, x_2, \ldots$,（2）それぞれの値をとる確率，つまり $p_i = \mathbb{P}[\,X = x_i\,]$（$i = 1, 2, \ldots$）を決めればよい（この２つが同じなら，同じ確率変数とみなす）．つまり，以下のような表を与えることが確率変数を決めることになる．また，このような x_i と p_i の対応を X の分布という．

確率変数のとりうる値	x_1	x_2	x_3	$\ldots$	x_n
それぞれをとる確率	p_1	p_2	p_3	$\ldots$	p_n

なお，上では n 個の値しかとらない確率変数を考えたが，実際には連続無限個の値をとるような確率変数もたくさんある（中心極限定理で出てきた標準正規分布はその例）．連続的な値をとる確率変数の扱いは数学的に少し厄介だが，この講義では有限個の値の場合からの類推で誤魔化すことにする．

ある確率変数があるとき，これをどのように特徴づければよいか，考えてみよう．勿論，確率変数 X を完全に決めるには上のような表を与えればよいのだが，これは実際にはなかなか大変である（正２０面体のサイコロや，X が 10^8 とおりもの値をとる場合を想像してみよ）．たとえそれができたとしても，10^8 個もの場合のそれぞれの確率 $p_1, p_2, \ldots$ を教えてもらっても，何かわかった気になるだろうか[16]？

この困難を排して「直感的」に確率変数の分布を知るため，いろいろな方法が考えられてきた．その代表的なものが期待値と分散である[17].

確率変数 X が $x_1, x_2, \ldots, x_n$ の値を，確率 $p_1, p_2, \ldots, p_n$ でとるとき，X の期待値（平均値）$\langle X \rangle$ を

$$\langle X \rangle = \sum_{i=1}^{n} p_i\, x_i = p_1 x_1 + p_2 x_2 + p_3 x_3 + \cdots + p_n x_n \tag{2.5.1}$$

で定義する．また，

$$\mathrm{Var}[X] = \Big\langle \big(X - \langle X \rangle\big)^2 \Big\rangle = \sum_{i=1}^{n} p_i \left(x_i - \langle X \rangle\right)^2 \tag{2.5.2}$$

を X の分散と言い，$\sqrt{\mathrm{Var}[X]}$ を X の標準偏差と言う.（標準偏差は σ で表すことが多い.）

このうち,「平均値」の方はよく知っているはずだ．X をあるクラスの生徒の数学のテストの点数としてみると，上で定義した「期待値」はこのクラスの点数の「平均値」に他ならない．つまり，X の期待値というのは X の分布の“ 中心 ”をだいたい表している．

これに対して,「分散」は X の分布の“ 広がり ”を示す．より正確には標準偏差 σ が，X の分布の大体の拡がりを示す．テストの点数の例で言うと，以下のようになる：いま，同じテストをしたところ，クラス A もクラス

[15]（余談）結果が簡単，または普遍的なものであるのにその証明が複雑である場合は，何か本質的なものを見逃している可能性もある．この意味で，より簡単な（明瞭な）証明を探すことは数学の発展上も大切である

[16]（余談）物事を「わかる」ためには多すぎる情報をうまく縮約することも大切だ，という例

[17]期待値や分散には確率変数の分布を特徴づける以上の意味もある．と言うのは，期待値や分散を計算する方が確率そのものの計算よりも簡単な場合が多いのだ（期待値の計算が簡単な理由の一つは以下の (2.5.5)–(2.5.6) などの性質）．このため，最前線の研究の場では，期待値や分散（その仲間としての「特性関数」）などの計算を如何にうまく行って，それから確率の解析に持っていけるか，が問題となることも多い

B も平均点は同じだった．しかし，クラス A ではみんながほとんど同じ点数だったので，分散が小さい（ゼロに近い）．一方，クラス B ではできる人とできない人の差が非常に大きかったので，分散も大きい．

（進んだ話題）勿論，期待値と分散だけを見ても，元の分布は決まらない．つまり，期待値と分散が同じでも異なった分布を持つような確率変数の例を考えることができる（実際に例を作ってみよ）．テストの点数に例をとれば，期待値と分散が同じでも異なる点数分布になるようなクラスがあり得る．元の分布をもっと限定していくためには ℓ 次のモーメントと呼ばれる量

$$\left\langle \big(X - \langle X \rangle\big)^{\ell} \right\rangle \tag{2.5.3}$$

をすべて見ていく必要がある（$\ell = 2, 3, 4, \ldots$）[18]．ごく大ざっぱに言うと，期待値とすべての次数のモーメントを知れば，元の確率変数の分布を決定することができる．つまり，確率変数 X, Y に対して

$$\langle X \rangle = \langle Y \rangle, \qquad \left\langle \big(X - \langle X \rangle\big)^{\ell} \right\rangle = \left\langle \big(Y - \langle Y \rangle\big)^{\ell} \right\rangle \quad (\ell = 2, 3, 4, \ldots) \tag{2.5.4}$$

ならば，X と Y は同じ分布に従っている．

2.5.2 期待値と分散の基本的な性質

さて，場合によっては期待値や分散が（確率そのものよりも）計算しやすいのは，以下のような関係式が成り立つからである．まず，いつでも成り立つ性質として以下がある：

一般の確率変数 X, Y と勝手な実数 a に対して，

$$\big\langle X + Y \big\rangle = \langle X \rangle + \langle Y \rangle \tag{2.5.5}$$

$$\big\langle aX \big\rangle = a \langle X \rangle \tag{2.5.6}$$

$$\mathrm{Var}[a\,X] = a^2\,\mathrm{Var}[X] \tag{2.5.7}$$

が成立する．

また，X, Y が独立の場合は，以下も成立する：

また，X, Y が独立の場合には

$$\langle X\,Y \rangle = \langle X \rangle \langle Y \rangle \tag{2.5.8}$$

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] \tag{2.5.9}$$

が成り立つ．

（注意）(2.5.8) や (2.5.9) は，X, Y が独立でない場合はなり立たないことが多い．

これらの性質は複雑な量の期待値や分散を，簡単な量の期待値や分散に分解して計算する手段を与えてくれる（その具体例を 2.5.4 節で見るだろう）．以下，これらの性質の証明を簡単に述べる．

(2.5.6) の証明

まず準備として，確率変数 aX の確率分布がどうなるかを書いてみると

X のとりうる値	x_1	x_2	x_3	$\ldots$	x_n
aX のとりうる値	ax_1	ax_2	ax_3	$\ldots$	ax_n
上のそれぞれの値をとる確率	p_1	p_2	p_3	$\ldots$	p_n

[18] 2 次のモーメントは上で定義した分散と同じである

となっている（ほとんどアタリマエのことを丁寧にやってしまったが）．だから期待値の定義から

$$\langle aX\rangle = \sum_{i=1}^{n} p_i\, ax_i = a\sum_{i=1}^{n} p_i\, x_i = a\,\langle X\rangle \tag{2.5.10}$$

となって，(2.5.6) が証明できた． □

(2.5.7) **の証明**

まず，(2.5.6) を用いると

$$\bigl(aX - \langle aX\rangle\bigr)^2 = \bigl(aX - a\,\langle X\rangle\bigr)^2 = a^2\bigl(X - \langle X\rangle\bigr)^2 \tag{2.5.11}$$

が得られるので，分散の定義から

$$\mathrm{Var}[aX] = \Bigl\langle \bigl(aX - \langle aX\rangle\bigr)^2 \Bigr\rangle = \Bigl\langle a^2\bigl(X - \langle X\rangle\bigr)^2 \Bigr\rangle = a^2\,\Bigl\langle \bigl(X - \langle X\rangle\bigr)^2 \Bigr\rangle = a^2\,\mathrm{Var}[X]. \tag{2.5.12}$$

が得られる．真ん中の等号ではまたもや (2.5.6) を用いた． □

(2.5.5) **の証明**

X, Y は独立とは仮定していないので，X は $x_1, x_2, \ldots, x_n$，Y は $y_1, y_2, \ldots, y_m$ の値をとるとして，$\mathbb{P}[\,X = x_i, Y = y_j\,] = p_{ij}$ と書くことにしよう．このとき，

$$\sum_j p_{ij} = \mathbb{P}[\,X = x_i\,], \qquad \sum_i p_{ij} = \mathbb{P}[\,Y = y_j\,] \tag{2.5.13}$$

である（why?）．これを用いると[19]，

$$\begin{aligned}\langle X+Y\rangle &= \sum_{i,j} p_{ij}(x_i + y_j) = \sum_{i,j} p_{ij}x_i + \sum_{i,j} p_{ij}y_j = \sum_i \Bigl(\sum_j p_{ij}\Bigr)x_i + \sum_j \Bigl(\sum_i p_{ij}\Bigr)y_j \\ &= \sum_i \mathbb{P}[\,X = x_i\,]\,x_i + \sum_j \mathbb{P}[\,Y = y_j\,]\,y_j = \langle X\rangle + \langle Y\rangle \end{aligned} \tag{2.5.14}$$

となって (2.5.5) が証明された． □

(2.5.8) **の証明**

いま X, Y が独立だと仮定しているので，

$$p_{ij} = \mathbb{P}[\,X = x_i \text{ かつ } Y = y_j\,] = \mathbb{P}[\,X = x_i\,]\,\mathbb{P}[\,Y = y_j\,] \tag{2.5.15}$$

が成立している．だから，

$$\langle XY\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}x_iy_j = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{P}[\,X = x_i\,]\,\mathbb{P}[\,Y = y_j\,]\,x_iy_j \tag{2.5.16}$$

が成立するが，右辺の和は i についての和と j についての和の積になっているので

$$= \sum_{i=1}^{n} \mathbb{P}[\,X = x_i\,]\,x_i \sum_{j=1}^{m} \mathbb{P}[\,Y = y_j\,]\,y_j = \langle X\rangle\,\langle Y\rangle \tag{2.5.17}$$

となって証明された．しつこいが，上では確率が (2.5.15) のように積に分解した事が本質的だったことを重ねて注意しておく． □

[19]最初のところ，1ステップ抜いてある．定義通りやるには，まず，$X+Y$ の分布の表を作って，そこから $\langle X+Y\rangle$ を書く必要がある

(2.5.9) **の証明**

最後に (2.5.9) であるが，これは今までの式を動員すればできる．定義から

$$\mathrm{Var}[X+Y]=\Big\langle\big(X+Y-\langle X+Y\rangle\big)^2\Big\rangle=\Big\langle\big(X-\langle X\rangle+Y-\langle Y\rangle\big)^2\Big\rangle \tag{2.5.18}$$

であるが，２乗のところを展開して

$$=\Big\langle\big(X-\langle X\rangle\big)^2+\big(Y-\langle Y\rangle\big)^2+2\big(X-\langle X\rangle\big)\big(Y-\langle Y\rangle\big)\Big\rangle \tag{2.5.19}$$

更に (2.5.5) と (2.5.6) を用いて和の期待値を期待値の和になおしたりすると

$$=\Big\langle\big(X-\langle X\rangle\big)^2\Big\rangle+\Big\langle\big(Y-\langle Y\rangle\big)^2\Big\rangle+2\big\langle\big(X-\langle X\rangle\big)\big(Y-\langle Y\rangle\big)\big\rangle \tag{2.5.20}$$

となる．ここで最初の２項はそれぞれ $\mathrm{Var}[X]$ と $\mathrm{Var}[Y]$ である．また，最後の項は $X-\langle X\rangle$ と $Y-\langle Y\rangle$ が独立であることから[20]

$$\big\langle\big(X-\langle X\rangle\big)\big(Y-\langle Y\rangle\big)\big\rangle=\langle X-\langle X\rangle\rangle\,\langle Y-\langle Y\rangle\rangle=0\times 0=0 \tag{2.5.21}$$

となる． □

2.5.3 S_N などの期待値や分散の計算

これからの証明に必要になるし，ある種の見通しも与えてくれるので，S_N などの期待値や分散を計算してみよう．上で述べたように，期待値は確率変数の分布のある種の中心，分散（正確には標準偏差）は分布の拡がりの目安を与えるものだから，この種の計算は直感的な理解のためにも重要である．

まず準備として X_i の期待値と分散を計算すると，定義から

$$\big\langle X_i\big\rangle=p,\qquad \mathrm{Var}[X_1]=p(1-p) \tag{2.5.22}$$

となる．

S_N の期待値からやってみよう．期待値の線形性 (2.5.5) をくり返し使うと，

$$\begin{aligned}\langle S_N\rangle&=\langle X_1+X_2+X_3\cdots+X_N\rangle=\langle X_1\rangle+\langle X_2+X_3+\cdots+X_N\rangle=\cdots=\langle X_1\rangle+\langle X_2\rangle+\cdots+\langle X_N\rangle\\&=\sum_{i=1}^{N}\langle X_i\rangle\end{aligned} \tag{2.5.23}$$

が成り立つ．ここで (2.5.22) から $\langle X_i\rangle=p$ であるので，(2.5.23) から

$$\langle S_N\rangle=N\,\langle X_1\rangle=Np \tag{2.5.24}$$

が得られる．次に S_N の分散であるが，X_i が独立であるために (2.5.9) が使える．(2.5.23) と同じノリで進むと

$$\mathrm{Var}[S_N]=\mathrm{Var}[X_1+X_2+\cdots+X_N]=\mathrm{Var}[X_1]+\mathrm{Var}[X_2]+\cdots+\mathrm{Var}[X_N] \tag{2.5.25}$$

となるが，(2.5.22) から $\mathrm{Var}[X_i]=p(1-p)$ であるので，(2.5.25) から

$$\mathrm{Var}[S_N]=N\,\mathrm{Var}[X_1]=Np(1-p) \tag{2.5.26}$$

が得られる．

これだけの情報でも役に立つ．つまり，(2.5.24) は S_N の分布は大体 Np を中心にしていること，また (2.5.26) はその分布の拡がりは大体 $\sqrt{Np(1-p)}$ くらいであること，を示唆している[21]．

[20] X,Y が独立なら，勝手な（ただしランダムでない）数 a,b に対して $X-a,Y-b$ も独立であることを使った

[21] 段々と見ていくように，一番大事なのは $\sqrt{N}$ の部分である．つまり，1.1 節の言葉を借りると，分布の拡がりが $\sqrt{N}$ のオーダーであることが重要なのだ

この点については後で戻ってくることにして，計算を続ける．

次に，$\frac{S_N}{N}$ について考えてみる．これは上の結果を用いると簡単で，(2.5.6) から直ちに，

$$\left\langle \frac{S_N}{N} \right\rangle = \frac{1}{N} \langle S_N \rangle = \frac{1}{N} \times Np = p \tag{2.5.27}$$

が出る．同様に，(2.5.7) から

$$\mathrm{Var}\Big[\frac{S_N}{N}\Big] = \Big(\frac{1}{N}\Big)^2 \mathrm{Var}[S_N] = \Big(\frac{1}{N}\Big)^2 \times Np(1-p) = \frac{p(1-p)}{N} \tag{2.5.28}$$

が得られる．この 2 式は $\frac{S_N}{N}$ の分布の中心が p 付近で，その拡がりが $\sqrt{\dfrac{p(1-p)}{N}}$ くらいであることを示唆している[22]．ここで実際に図 2 のグラフが大体上の拡がりを持っていることを確かめよう．

2.5.4 大数の弱法則の証明

証明には以下の不等式を用いる．

（チェビシェフの不等式）期待値と分散が有限である確率変数 Y に対して以下が成立する：

$$\mathbb{P}\big[\, |Y - \langle Y \rangle| > a \,\big] \ \le\ \frac{\mathrm{Var}[Y]}{a^2} \qquad (a \text{ は勝手な正の数}). \tag{2.5.29}$$

この不等式の証明は後回しにして，上の不等式を $Y = \frac{S_N}{N}$ に対して適用してみよう．単に代入すると

$$\mathbb{P}\Big[\Big|\frac{S_N}{N} - \Big\langle \frac{S_N}{N} \Big\rangle\Big| > a\Big] \le \frac{\mathrm{Var}[\frac{S_N}{N}]}{a^2} \tag{2.5.30}$$

となるので，$\langle \frac{S_N}{N} \rangle$ と $\mathrm{Var}[\frac{S_N}{N}]$ を計算すればなにか良いことがあるかもしれない．

実はこれらの量は (2.5.27) と (2.5.28) で計算してある．これらをチェビシェフの不等式 (2.5.30) に代入すると，任意の正の数 a に対して

$$\mathbb{P}\Big[\Big|\frac{S_N}{N} - p\Big| > a\Big] \le \frac{\frac{p(1-p)}{N}}{a^2} = \frac{p(1-p)}{a^2\, N} \tag{2.5.31}$$

が成立することがわかる．

この不等式を解釈しよう．左辺は，「$\frac{S_N}{N}$ がその期待値 $\langle \frac{S_N}{N} \rangle = p$ から a 以上ずれる確率」である．右辺は，この確率が，N が大きくなると $1/N$ のように減少していくことを主張している．これは我々の日常感覚と一致している．すなわち，N を大きくしても（たくさんの回数，コインを投げても），表の出る割合がきっちり p になるとは言い切れないが，表の割合は非常に高い確率で p の近くに来る，とは言えるのである．

さて，上の解析を振り返ると，コイン投げであることはどこにも使っていない．正確には，期待値と分散の計算で使ったが，ここは別の値であっても上の解析 — 特に不等式 (2.5.31) の類似の式 — は成り立つ．重要なのは $\frac{S_N}{N}$ の分散が $1/N$ のように（大きい N に対して）小さくなっていくことであった．そこで，少し一般に拡張すると，以下のような結果になる．(興味のある人は証明の細部を書き出してみて欲しい.)

[22] ここも 1.1 節の言葉では，$\frac{S_N}{N}$ の分布の拡がりが $1\sqrt{N}$ のオーダーである，と言える

（大数の弱法則）期待値が μ，分散が σ^2 である独立（かつ同分布）な確率変数 $X_1, X_2, \ldots$ に対して，$S_N = \sum_{i=1}^{N} X_i$ を定義する．このとき，$\frac{S_N}{N}$ が μ からずれる確率は N が無限大になるとゼロに近づく．もっと詳しくは，任意の正の数 a に対して，

$$\mathbb{P}\Big[\,\Big|\frac{S_N}{N} - \mu\Big| > a\,\Big] \le \frac{\sigma^2}{a^2\,N} \tag{2.5.32}$$

が成り立つ．

このように，チェビシェフの不等式と期待値や分散の基本的性質を用いると，我々が日常考えていることを基礎づけることができた．また，これは「確率とは何か」にもある程度の答えを与えてくれる．すなわち，コイン投げの場合に表が出る確率 p とは,「コインを何回も投げたときに表の出る割合」と定義したくなるのであるが，上の大数の法則はこの決め方に矛盾がない，ことを示している．

註： 上では X_i が同分布（同じ試行の繰り返し）だとしたが，同分布でない場合にもある程度拡張して成り立つことがわかる（大学に入ってのお楽しみ）．

2.5.5 チェビシェフの不等式の証明

では，チェビシェフの不等式の証明をする．びっくりするくらい簡単だよ．

証明を見ればわかるように，チェビシェフの不等式の類似物は一杯作れる（マルコフの不等式，などと名前がついているのもある）．少し一般に考えた方が原理がわかりやすいだろうから，a, b を勝手な数（$a > 0$）として

$$\mathbb{P}\,[\,|X - b| \ge a\,] \;\le\; \frac{\big\langle (X-b)^2 \big\rangle}{a^2} \tag{2.5.33}$$

を証明し，最後に $b = \langle X \rangle$ としよう．

確率変数 X は x_i の値を，確率 p_i でとるものとする（$i = 1, 2, 3, \ldots, n$）．$\big\langle (X-b)^2 \big\rangle$ の表式から出発し，i の和を $|x_i - b| \ge a$ を満たすものに限定すると，和の値は小さくなる（和の中身が非負なので，足す項が減れば和は減る）：

$$\big\langle (X-b)^2 \big\rangle = \sum_{i=1}^{n} p_i\,(x_i - b)^2 \ge \sum_{i:|x_i - b| \ge a} p_i\,(x_i - b)^2 \tag{2.5.34}$$

最後の和の記号は，$|x_i - b| \ge a$ を満たしているような i についてのみ和をとる，ということである．ところが，右辺の和の中身はいつでも $|x_i - b| \ge a$ を満たしている．つまり，和の中ではいつでも $(x_i - b)^2 \ge a^2$ であるので，この和は a^2 以上のものを足していることになる．従って和の値は $a^2\times$（和の個数）よりも大きい：

$$\ge \sum_{i:|x_i - b| \ge a} p_i\,a^2 = a^2 \sum_{i:|x_i - b| \ge a} p_i \tag{2.5.35}$$

ところが，この右辺の和は，$|x_i - b| \ge a$ なる x_i の実現確率を足しているのだから，$\mathbb{P}\,[\,|X - b| \ge a\,]$ そのものである：

$$= a^2\,\mathbb{P}\,[\,|X - b| \ge a\,]. \tag{2.5.36}$$

両辺を a^2 で割ると (2.5.33) を得る．また，b は何でも良かったので特に $b = \langle X \rangle$ ととると，チェビシェフの不等式になる． □

2.6 中心極限定理の“ 説明 ”

さて，ここで中心極限定理の証明をすべきだが，時間の関係で講義ではあまり触れられなかった．触れられなかった事も含めて，いくつかの要点を述べておく．その前に，少し一般の形で定理を述べる．

独立かつ同分布な確率変数 $X_1, X_2, \ldots$ を考える．X_1 の期待値 $\mu = \langle X_1 \rangle$，標準偏差 $\sigma = \sqrt{\mathrm{Var}[X_1]}$ を用いて確率変数

$$Z_N = \frac{S_N - N\mu}{\sigma\sqrt{N}} \tag{2.6.1}$$

を定義すると，N が無限大になるとき，Z_N は標準正規分布に“ 収束 ”する．特に，N が無限大になる極限では，確率 $\mathbb{P}[\, a \le Z_N \le b\,]$ は,

$$\left(\text{グラフ } y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ と 3 直線 } x = a, x = b, y = 0 \text{ で囲まれた部分の面積} \right) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \tag{2.6.2}$$

に収束する.（極限や積分を習った人向けに書くと）

$$\lim_{N\to\infty} \mathbb{P}[\, a \le Z_N \le b\,] = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \tag{2.6.3}$$

ということになる.

上の定理に現れているように，大数の法則と同じく，中心極限定理も非常に広い範囲で成り立つ定理である．

以下ではアゴラの期間中に寄せられた質問に答える形で，この定理についての説明を与えたい．答えようとしている質問は以下の通り.

- グラフの横軸はなぜ，あのように伸び縮みさせるのか？（関連質問）なぜ上のように Z_N を定義するのか？
- グラフの縦軸はなぜ，あのように伸び縮みさせるのか？（関連質問）なぜ確率が面積で表されるのか？
- このように伸び縮みさせるとなぜ，あの曲線に行くのか？（関連質問）極限で出てくる実数のグラフ $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ はどのような原理で決まるのか？
- （これは質問は出なかったが）実際の証明はどうするのか？

これらの質問に完全に答えるには大学程度の知識が必要であり，それを延々と展開しても余り意味はないだろう.（それなら，僕のノートよりも大学生向けの確率論の教科書を読んだ方がよい.）以下では「高校の知識だけでもある程度は理解できる」範囲に限って，部分的な解答を試みる.

2.6.1 グラフの横軸はどう決めたのか？つまり Z_N はなぜ，このように決めるのか？

まず気になるのが Z_N の取り方やグラフの伸び縮みのさせ方をどのように決めたのか，と言うことであろう．これについてはある程度の解答を与えることができる（縦軸の伸び縮みは次の小節で扱う）．この問題の解答には，横軸を（$\frac{S_N}{N}$ でなく）S_N にとったグラフから出発する方がわかりやすいだろうから，グラフをいろいろと伸び縮みさせる課程を図 5 に示す（いろいろな p の例があった方が良かろうから，今度は $p = \frac{1}{5}$ としてみた）.

まず，図の解説をしよう.

(1) 左上のグラフ：横軸は m，縦軸は $\mathbb{P}[\, S_N = m\,]$ にとった．黄色（$N = 256$）は図の範囲外（右側）に分布の中心があるので（$Np = 51.2$），よく見えていない.

(2) 右上のグラフ：横軸は $m - pN$，縦軸は $\mathbb{P}[\, S_N = m\,]$ にとった．上の (1) では N とともに右の方に中心が移動していたので規則性が見えにくいから，S_N の期待値（Np）の位置を中心に持ってくるように各曲線を平行移動した．これで分布の中心は大体そろったが，N が大きくなるにつれて高さは低く，幅は広くなっている.

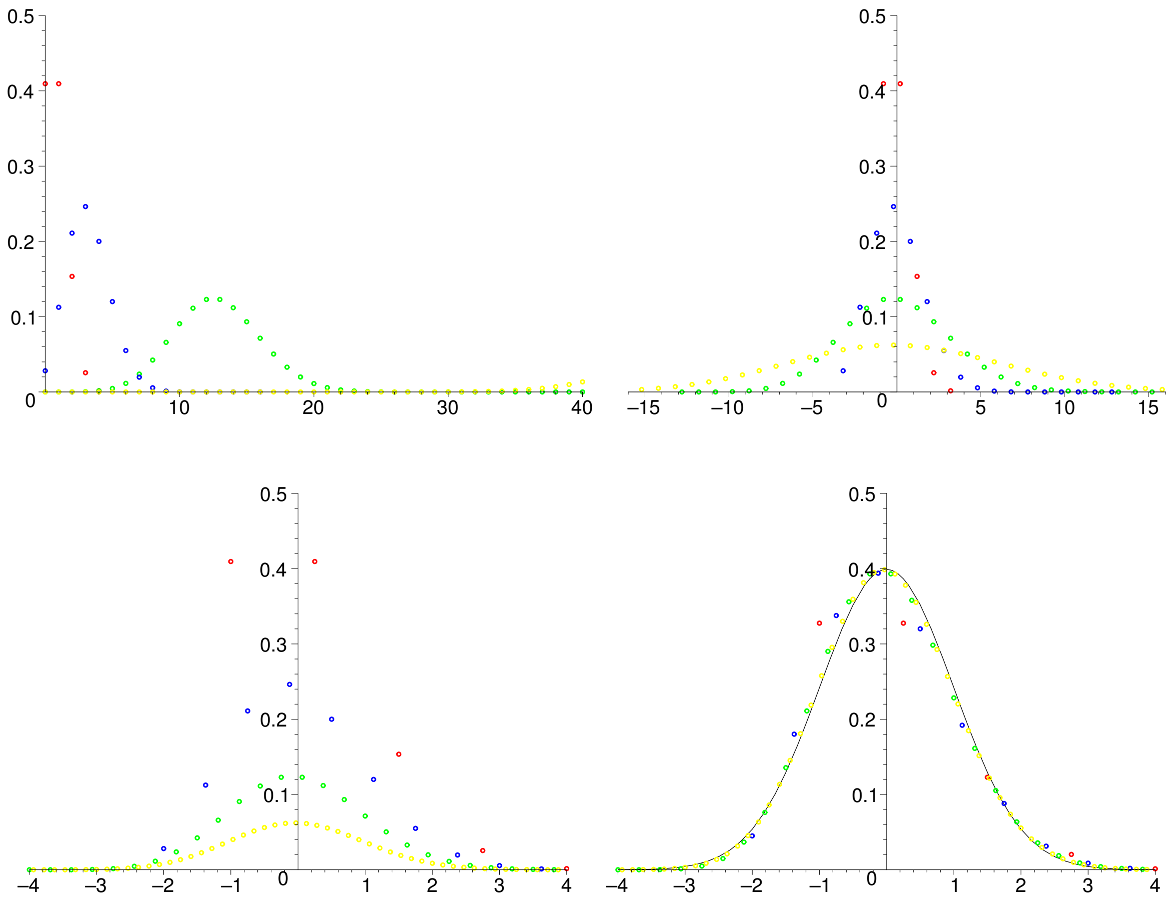


図 5: これらはどれも $p=\frac{1}{5}$ の時に，$N=4$（赤），16（青），64（緑），256（黄）に対する $\mathbb{P}[\,S_N=m\,]$ の様子を図示したものであるが，グラフによって，縦軸と横軸の取り方が違う（詳細は本文）.

(3) 左下のグラフ：横軸は $(m-pN)/\sqrt{p(1-p)N}$，縦軸は $\mathbb{P}[\,S_N=m\,]$ にとって，(2) のグラフを横軸方向に $1/\sqrt{p(1-p)N}$ に縮めた．幅はどの N でも同じようになってきたが，高さがそろってない．

(4) 右下のグラフ：横軸は $(m-pN)/\sqrt{p(1-p)N}$，縦軸は $\mathbb{P}[\,S_N=m\,]\times\sqrt{p(1-p)N}$ にとって，(3) のグラフを縦方向に引き延ばした．これで漸く，高さも幅もそろうようになった．実線は $y=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ のグラフである．

最終段階の (4) のグラフを表現したのが中心極限定理であった．この小節では上の (1) から (3) の取り方を説明し，次の小節で (2) から (4) を説明する．

まず，(1) から (2) への変換は大数の法則から示唆される．つまり，$\frac{S_N}{N}$ が p の周りに集中していくわけだから，S_N そのものを見てたら分布の中心は右の方へ動いてしまう．よって，分布の中心（Np）を常にグラフの中心にするように平行移動しただけのことだ．

問題は (2) から (3) である．(2) のグラフでは分布の中心は y-軸であるものの，N とともに幅が大きくなったのでこれを縮めたいのである．どのくらい縮めればよいだろうか？

以前，2.5.1 節で，確率変数 X の拡がりの目安は $\sigma=\sqrt{\mathrm{Var}[X]}$ であると言った．また，大数の弱法則の証明のところで（(2.5.26) 参照），$\mathrm{Var}[S_N]=p(1-p)N$ である事も見た．これは $\sigma=\sqrt{p(1-p)N}$ であって，S_N の拡がりが N とともに $\sigma=\sqrt{p(1-p)N}$ くらいで増えていくことを示唆する（ここのところ，図 5 の右上のグラフで確かめよう！）．これを打ち消すように $\sigma=\sqrt{p(1-p)N}$ で横軸を割ってやれば幅が大体一定の分布ができるだろう．これが (2) から (3) への変換の理由である．

なお，この２つの変換を続けてやることは

$$Z_N = \frac{S_N - Np}{\sqrt{p(1-p)\,N}} = \frac{1}{\sqrt{p(1-p)}} \times \frac{S_N - Np}{\sqrt{N}} \tag{2.6.4}$$

を定義するのと同じである．いままで天下りに定義していた Z_N の定義が少し身近に感じられただろうか？ここで分子の $S_N - Np$ は (1) から (2) への平行移動を，分母の $\sqrt{p(1-p)N}$ は (2) から (3) への横軸方向の縮めを表している．

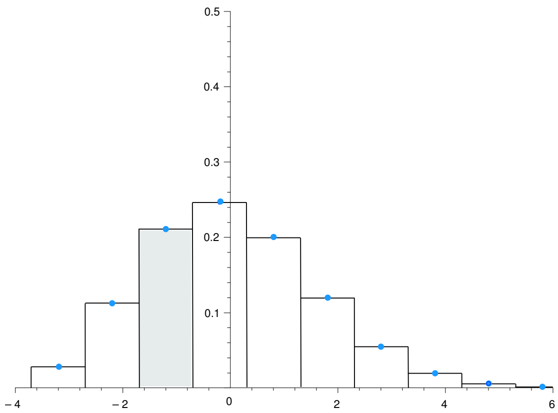
註： 上の (2.6.4) の伸び縮みのファクター $\sqrt{p(1-p)N}$ の内，一番大事なのは $\sqrt{N}$ の部分である[23]．このお陰でいろいろな N でも大体同じ幅を持つようになる．もう一方の $\sqrt{p(1-p)}$ の方は，異なる p でも同じ幅になるように働いている．別の言い方をすると，$\sqrt{p(1-p)}$ は，できた Z_N の分散をそろえるため，つまり

$$\mathrm{Var}[Z_N] = 1 \tag{2.6.5}$$

となるように，導入したものである.（ただし，このようにとったら異なる p のグラフがすべて同じ曲線に行くことは決して自明ではない！）

2.6.2　グラフの縦軸はどう決めたのか？またはなぜ，確率が面積で与えられるのか？

上では横軸の伸び縮みの理由が大体わかった．今度は縦軸の方を考えよう．図 5 では (3) から (4) の過程にあたるが，以下の議論のためには (2) から (4) を考える必要が出てくる．結論から言うと，グラフの縦軸は確率が曲線の下の面積で与えられるように決めている．そこで，まず，確率とグラフの下の面積の関係を有限の $N = 16$ を例にとって考えよう．



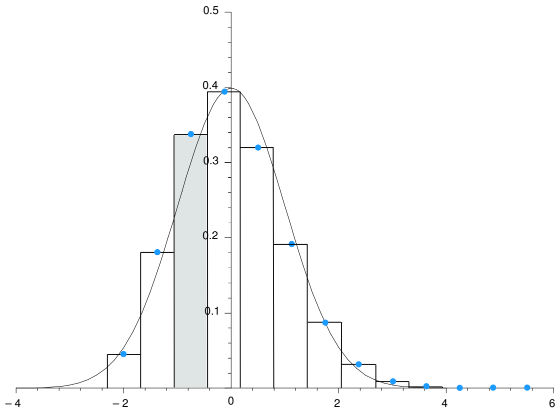


図 6: $p = \frac{1}{5}, N = 16$ のグラフ（本文中での説明のために，従来の青点に加えて，ヒストグラムの形にした）．
左：縦軸は $\mathbb{P}[\,S_N = m\,]$，横軸は $m - Np$（図 5 の (2)）．スペースの関係で $m - Np \le 6$ のみを図示している．
右：縦軸は $\mathbb{P}[\,S_N = m\,] \times \sqrt{p(1-p)N}$，横軸は $(m - Np)/\sqrt{p(1-p)N}$（図 5 の (4)）．実線はいつも通り，$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフである．

図 6 の左には $\mathbb{P}[\,S_N = m\,]$ のグラフを，横軸を $m - Np$ にとって描いた．これは図 5 の (2) そのものであるが，各 m での $\mathbb{P}[\,S_N = m\,]$ の値のみならず，それぞれの m の周りに幅 1 の短冊をとって，ヒストグラムのようにしてみた．こうすると，$\mathbb{P}[\,S_N = m\,]$ の値はそれぞれの m のところにある長方形の面積そのものになる — 例えば，影を付けた $m = -1$ のところの長方形の面積は $\mathbb{P}[\,S_N = -1\,]$ に等しい.（なぜなら，この長方形は幅が 1，高さが $\mathbb{P}[\,S_N = m\,]$ なので，面積は $\mathbb{P}[\,S_N = m\,]$ になるから.）

[23] 1.1 節の言葉では，$\sqrt{N}$ のオーダーであることが大事

この考えに基づくと,（$A < B$ は勝手な整数）確率

$$\mathbb{P}[\,A \le S_N - Np \le B\,] = \sum_{m=A}^{B} \mathbb{P}[\,S_N = m\,] \tag{2.6.6}$$

は図 6 の左の折れ線と x-軸の間の，$x = A - \frac{1}{2}$ から $x = B + \frac{1}{2}$ の部分の面積と考えることもできる．これが図の右ではどうなるかを次に考えよう．

図 6 の右は左の図を,「横軸は $1/\sqrt{p(1-p)N}$ だけ縮め，縦軸は $\sqrt{p(1-p)N}$ 倍に引き延ばした」ものである（今は $p = \frac{1}{5}, N = 16$ なので $\sqrt{p(1-p)N} = \frac{8}{5}$ である）．このとき，横軸を縮めたのと同じ割合だけ縦軸を引き延ばしているのがミソで，こうすると，左側に出ている長方形（短冊）のそれぞれは（横が縮んだ分だけ縦が伸びたから），面積が同じ対応する短冊に移る．例えば左で影を付けた長方形の面積は $\mathbb{P}[\,S_N = -1\,]$ であったが，対応する（右でも影をつけた）長方形の面積も $\bigl(\mathbb{P}[\,S_N = m\,] \times \sqrt{p(1-p)N}\bigr) \times \bigl(1/\sqrt{p(1-p)N}\bigr) = \mathbb{P}[\,S_N = m\,]$ となって，面積は変わらない．

従って，(2.6.6) を表す面積は，確率

$$\mathbb{P}\Bigl[\,\frac{A}{\sqrt{p(1-p)N}} \le Z_N \le \frac{B}{\sqrt{p(1-p)N}}\,\Bigr] = \sum_{m=A}^{B} \mathbb{P}[\,S_N = m\,] \tag{2.6.7}$$

とも書けるが，これは図 6 の右図の折れ線と x-軸の間の，$x = \frac{A-1/2}{\sqrt{p(1-p)N}}$ から $x = \frac{B+1/2}{\sqrt{p(1-p)N}}$ の部分の面積と考えることもできる．

これで漸く，この小節の問いに答えることができる．今まで見てきたように，N が小さくても，確率を面積で表すことは可能であった．ただし，そのときに出てくる曲線はヒストグラムに出てくるような折れ線になっている．ところが,（少なくとも図 5 の (4) のような例を見る限り）図 6 の右の短冊の頂点（青点）は N が大きくなるにつれて実線のグラフの上に乗っていく．また，短冊の幅もどんどん小さくなっていく．ということは N が大きくなった極限では (2.6.7) の面積は実線のグラフ $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフと x-軸の間の部分の面積に近づきそうである．これが中心極限定理で $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフの下の面積が出てくる理由である．また縦軸を伸ばす割合は，この面積の解釈ができるように「横軸を縮めた分を打ち消すように引き延ばす」ことで決められていたのである．

註： ここでは横軸を縮める割合と縦軸を伸ばす割合が丁度同じであった．これは実は S_N のとりうる値がたまたま間隔 1 で分布していたための，幸運な事情である．元々の確率変数のとりうる値が間隔 a で分布していたら，上で説明した「確率と面積」の解釈を少し変更する必要が出る．またその結果，縦軸の引き延ばしは $\sqrt{p(1-p)N}/a$ にすべきであることがわかる.（興味のある人は細部を詰めてみると，上の議論がより良く納得できるだろう.）

2.6.3　なぜ，あの曲線に“ 収束 ”するのか？

次に気になるのは，このような変換をしたものがなぜ，$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ の曲線に近づくのか，と言うことである．これは非常に重要な問いなのだが，余り簡単な答えを示すことはできない．大学３年程度の数学（例えば次の 2.6.4 節で紹介している「特性関数の方法」）を駆使すれば完全な答えを与えられるが，高校程度でわかるようにはなかなかできない．

そこで目標を少し下げて，

> N が大きくなったものが何かの曲線に近づくならば，それは $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ の曲線だ

と言うことを説明（完全な証明ではない）しようと思う．当初の問いに答えるためには，この後に「N が大きくなったときに，実際に上の極限に近づくこと」も示す必要があるが，これはなかなか大変なので諦めた．ただし，この説明自身も高校数学の枠内ではなかなか難しく，大学１年程度の数学も用いるので，この講義ノートの他の部分より程度が高く，読みにくくなってしまった．そこで，この説明全体を 2.7 節に移したので，興味のある人は読んでみて欲しい．

2.6.4　中心極限定理の証明は実際にはどうするのか？（お話しだけ）

中心極限定理の実際の証明方法について，お話しだけ述べる．何通りかの方法がある．

（1）コイン投げの問題なら $\mathbb{P}[\,S_N = m\,]$ の具体形がわかっているから（2.2 節），これが（上で決めたスケール変換を行った後で） N を大きくしたときにどのような値に近づくか，計算してみると良い． $\mathbb{P}[\,S_N = m\,]$ の中に出てくる ${}_N C_m$ の計算が曲者で,「スターリングの公式」と言うものを使ってガリガリ計算するとできるが, かなり大変である[24]．ただし，この方法はコイン投げのように，一回の試行の結果が 2 通りしかない場合にのみ使える．サイコロなど一回の試行の結果が 6 つもあるものになると，この種の方法でやるのはほとんど絶望的だ.

（2）通常，大学の数学でやるのは「特性関数」と言うものを使って証明する方法である．これは非常にエレガントで，かつ，適用範囲も大変に広い．良いことづくしなのだが，問題はこの方法を理解するには「積分」「フーリエ変換」などが必要になることである.（いろいろなことを天下りに認めてもらって概要を説明するつもりだったがまあ，講義の中では全くできなかった.）積分を知っている人のための大筋は以下の通り[25]：

1. まず，関数 $f(x)$ に対してその「フーリエ変換」 $\hat{f}(k)$ を，
$$\hat{f}(k) = \int_{-\infty}^{\infty} f(x)\, e^{ikx} dx \tag{2.6.8}$$
で定義する．ここで k は勝手な実数，$i = \sqrt{-1}$ は虚数単位.
2. 上の式は f から $\hat{f}$ を決める式だったが，驚くことに逆の式も成り立つ.
$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(k)\, e^{-ikx} dk \tag{2.6.9}$$
これを「フーリエ逆変換」と言うが，大事なことは $f(x)$ と $\hat{f}(k)$ が一対一に対応していることである．つまり，$f(x)$ の代わりに $\hat{f}(k)$ を計算できれば，それから $f(x)$ を逆算できる.
3. 一般の確率変数 X に対し，
$$\hat{f}(k) = \left\langle e^{ikX} \right\rangle \tag{2.6.10}$$
を X の分布の「特性関数」と呼ぶ．特に X の分布がある関数 $f(x)$ を使って $\mathbb{P}[\,a \le X \le b\,] = \displaystyle\int_a^b f(x)dx$ と書ける場合には特性関数は (2.6.8) で定義したフーリエ変換 $\hat{f}(k)$ に一致する．我々は Z_N の分布を知りたいのだから Z_N の特性関数，すなわち $\hat{f}_N(k) = \left\langle e^{ikZ_N} \right\rangle$ を計算したい．これができれば上の 2 から Z_N の分布を決められるはずだ[26].
4. さて,「期待値や分散の計算は確率そのものの計算よりも楽な場合が多い」と以前に注意したが， $\left\langle e^{ikZ_N} \right\rangle$ についても正にその通りである．指数関数の性質 $e^{a+b} = e^a\, e^b$ をくり返し使うと（指数関数 e^a は $\exp(a)$ とも書く；μ と σ は X_i の期待値と標準偏差）
$$e^{ikZ_N} = \exp\biggl(\sum_{i=1}^{N} \frac{ik(X_i - \mu)}{\sigma\sqrt{N}} \biggr) = \prod_{i=1}^{n} \exp\biggl(\frac{ik(X_i - \mu)}{\sigma\sqrt{N}} \biggr) \tag{2.6.11}$$
となる．ここで X_i が互いに独立であったことを思い出すと，$\exp\bigl(\frac{ik(X_i-\mu)}{\sigma\sqrt{N}}\bigr)$ も互いに独立である．独立な確率変数の積の期待値は期待値の積に分解したから（(2.5.8)），結局
$$\hat{f}_N(k) = \biggl\langle \prod_{i=1}^{n} \exp\biggl(\frac{ik(X_i - \mu)}{\sigma\sqrt{N}} \biggr) \biggr\rangle = \prod_{i=1}^{n} \biggl\langle \exp\biggl(\frac{ik(X_i - \mu)}{\sigma\sqrt{N}} \biggr) \biggr\rangle = \biggl\langle \exp\biggl(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}} \biggr) \biggr\rangle^N \tag{2.6.12}$$
となる．N が大きくなる極限ではこの量は（指数関数をテイラー展開することで）計算できて，$e^{-k^2/2}$ に近づくことがわかる．これをフーリエ逆変換して，$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ が得られる.

[24] 興味のある人は挑戦すると良いが,「極限」に関するある程度の勘が必要になるだろう

[25] こんな事を書いて意味があるのかどうかわからないが，何人かの人には高校の数学のその先が少しは見えるかもしれない．なお，この大筋ではいろいろ細かい条件はすべて無視した．例えば (2.6.9) は $f(x)$ がある種の条件を満たさないと成り立たないが，ここには書いていない

[26] この辺りはいよいよええ加減に書いているので，詳しくは大学でやってくださいね

（３）確率変数の分布を特徴づけるには，高次のモーメントを全部計算すれば（大体）十分である．従って，Z_N の高次のモーメントを計算し，これが標準正規分布の高次のモーメントに収束していく（N が大きくなると）ことを示す手もある[27]．自分で納得するには，3 次，4 次くらいのモーメントを計算してみるのが良いかもしれない.「独立」性が非常に効いて，分散の時の類似の式が成り立ち，高次のモーメントが簡単になるのがわかるだろう（類似の計算は (2.6.12) の右辺の計算でも使われる）.

（４）統計力学や情報理論などで使われる「エントロピー」を使う方法もある（最近の発展）．これは非常に面白い方法だと思うものの，僕がまだ勉強中なのでここではちゃんと書けません．申し訳なし.

2.7 完全なおまけ：中心極限定理に出てくる曲線を求めよう

ここでは 2.6.3 節の予告通り，中心極限定理にでてくる極限の曲線がなぜ $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ であるのか，に部分的に答えることにする．ただし，いろいろ工夫はしたが，この節の内容は高校３年〜大学一年程度の知識を必要とする[28]ものになってしまったことをお断りする．このようなことを書いても読者がいないのではないかという気はするが,「特性関数の方法」とは少し違った視点としての存在意義はあると考え，書いておく．なお，以下の方法は「安定分布」「無限分解可能分布」などでてくる考えを，我々の目的に合うように書き直したものである（例えば文献 [5] 参照）.

まず，用語を導入しておく．一般の確率変数 X について，ある関数 $f(x)$ が存在して

$$\mathbb{P}[\, a \le X \le b \,] = \int_a^b f(x)dx \qquad \text{（すべての } a < b \text{ に対して）} \tag{2.7.1}$$

と書けるとき，$f(x)$ を X の（確率）分布密度関数という．このような $f(x)$ が存在しないような確率変数もいっぱいある — 例えばコイン投げの X_i や S_N のように，有限個の値しかとらないものはダメ — が，中心極限定理ででてくる極限分布のような，連続的にいろいろな値をとる確率変数なら存在するかもしれない.

この小節では，以下の主張：

> 中心極限定理の Z_N の極限が $\displaystyle\lim_{N\to\infty} \mathbb{P}[\, a < Z_N < b \,] = \int_a^b f(z)dz$ の意味で分布密度関数 $f(z)$ を持つと仮定するならば，$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ しかあり得ない.

を説明（完全な証明ではない）する．中心極限定理を満足に説明するには，うえでは仮定した「求めた $f(z)$ に対して $\displaystyle\lim_{N\to\infty} \mathbb{P}[\, a < Z_N < b \,] = \int_a^b f(z)dz$ が実際に成立すること」も示す必要があるが，これはもっと大変なので諦めた．そこまで理解するには 2.6.4 節で紹介している「特性関数の方法」などを詳しく勉強してほしい.

中心極限定理で出てくる極限は,

$$Z_N = \frac{X_1 + X_2 + \cdots + X_N}{\sigma\sqrt{N}} \tag{2.7.2}$$

の（$N \to \infty$ の）分布として出現するが，以下では σ で割るのをやめてしまって，もう少し一般に,

$$Y_N = \frac{X_1 + X_2 + \cdots + X_N}{\sqrt{N}} \tag{2.7.3}$$

の極限分布として現れるものを求めてみよう[29]．$N \to \infty$ の極限では (2.7.3) の Y_N の分布が

$$\lim_{N\to\infty} \mathbb{P}[\, a \le Y_N \le b \,] = \int_a^b f(y)dy \tag{2.7.4}$$

と表されると仮定して，$f(y)$ を求めていくのである．説明は何段階かに分けて行う.

[27]ただし，確率変数の分布を確実に決めるには，付加的な条件も必要であり，厳密な証明にしようとしてそれらを確かめていると（２）の「特性関数」の方法とほとんど同じになってしまう

[28]必要な知識は，極限，微分，積分（高校３年程度）; 微分方程式の初歩（高校３年〜大学１年），少しだけ重積分（大学１年程度）

[29]2.6.1 節で詳しく説明したように，σ で割る割らないは単に横軸の伸び縮みだから割らないのをやっておけば十分である

2.7.1 行き先の確率変数の満たすべき性質は何か？

まず，(2.7.3) をもう少し変形しておく．M, N を（大きな）勝手な正の整数とすると，(2.7.3) の定義から

$$Y_{M+N} = \frac{1}{\sqrt{M+N}} \sum_{i=1}^{M+N} X_i = \frac{\sqrt{M}}{\sqrt{M+N}} \frac{1}{\sqrt{M}} \sum_{i=1}^{M} X_i + \frac{\sqrt{N}}{\sqrt{M+N}} \frac{1}{\sqrt{N}} \sum_{i=M+1}^{M+N} X_i \tag{2.7.5}$$

と書ける．ここで，X_i （$i = 1, 2, \ldots, M+N$）が独立・同分布であるから $\frac{1}{\sqrt{M}} \sum_{i=1}^{M} X_i$ と $\frac{1}{\sqrt{N}} \sum_{i=M+1}^{M+N} X_i$ はそれぞれ Y_M，Y_N と同じ分布に従うはずである[30]．つまり，標語的には

$$Y_{M+N} \overset{\mathcal{D}}{=} \sqrt{\frac{M}{M+N}}\, Y_M \; + \; \sqrt{\frac{N}{M+N}}\, Y_N \tag{2.7.6}$$

と書ける．ここで $\overset{\mathcal{D}}{=}$ は，両辺に出てくる確率変数の分布が等しい，事を意味する記号であるが，ここでは余り気にしないことにしよう．

さて，ここで，$\frac{M}{M+N} = \epsilon$ （$0 < \epsilon < 1$）を一定にしたまま N と M を無限大にしよう．Y_{M+N}, Y_N, Y_M の 3 つとも， その分布密度関数が $f(x)$ で与えられる極限の確率変数に行くはずである（Y_N の極限が $f(x)$ に従う，と言うのがそもそも我々の仮定）．また，このとき

$$\sqrt{\frac{M}{M+N}} = \sqrt{\epsilon}, \qquad \sqrt{\frac{N}{M+N}} = \sqrt{1-\epsilon} \tag{2.7.7}$$

である．そこで Y_N の極限として出てくる確率変数を Y と書くと，Y は

$$Y \overset{\mathcal{D}}{=} \sqrt{\epsilon}\, Y \; + \; \sqrt{1-\epsilon}\, Y \qquad \text{（任意の } 0 < \epsilon < 1 \text{ に対して）} \tag{2.7.8}$$

を満たすはずである．この式はあくまで，左右両辺の確率変数の分布が等しい，と読むべきである．(2.7.8) が核心の関係式であり，以下ではこれを元に解析を行っていく．

2.7.2 行き先の分布密度 $f(x)$ の満たすべき性質は何か？ — 積分方程式

では，上の関係式 (2.7.8) を Y の分布密度関数 $f(y)$ の言葉に焼き直そう．左辺は Y だからその分布密度は $f(y)$ である．右辺のは何だろうか？これには以下の 2 つの関係式を用いる．

確率変数 X, Y の分布密度をそれぞれ $f(x), g(x)$ とすると，勝手な正の実数 α に対して以下が成立：
(i) $Z = X + Y$ の分布密度関数を $h_1(z)$ とすると，

$$h_1(z) = \int_{-\infty}^{\infty} f(x)\, g(z-x)\, dx. \tag{2.7.9}$$

(ii) $W = \alpha X$ の分布密度関数を $h_2(w)$ とすると，

$$h_2(w) = \frac{1}{\alpha}\, f\big(\frac{w}{\alpha}\big). \tag{2.7.10}$$

(i) の性質の証明には大学一年でやる「重積分」を使うのが自然である（高校数学でも理解できないわけではないが，少し苦しい）．(ii) の方はたんなる積分変数の変換なので高校数学で十分に理解できる．これらの証明は略．

[30] $\frac{1}{\sqrt{M}} \sum_{i=1}^{M} X_i$ は Y_M そのものだが，$\frac{1}{\sqrt{N}} \sum_{i=M+1}^{M+N} X_i$ は X_{M+1} から X_{M+N} の和だから Y_N そのものではない．しかし，X_i が独立・同分布なので，分布は等しいはずだ，という主張

これを (2.7.8) の右辺に用いよう．まず上の (ii) から，$\sqrt{\epsilon}Y$ の分布密度は $\frac{1}{\sqrt{\epsilon}}f\bigl(\frac{w}{\sqrt{\epsilon}}\bigr)$ であり，$\sqrt{1-\epsilon}\,Y$ の分布密度は $\frac{1}{\sqrt{1-\epsilon}}f\bigl(\frac{w}{\sqrt{1-\epsilon}}\bigr)$ であることがわかる．次に，(i) をもちいると，$\sqrt{\epsilon}Y+\sqrt{1-\epsilon}\,Y$ の分布密度が

$$h(z)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{\epsilon}}f\bigl(\frac{x}{\sqrt{\epsilon}}\bigr)\,\frac{1}{\sqrt{1-\epsilon}}f\bigl(\frac{z-x}{\sqrt{1-\epsilon}}\bigr)\,dx=\int_{-\infty}^{\infty}f(y)\,\frac{1}{\sqrt{1-\epsilon}}f\bigl(\frac{z-\sqrt{\epsilon}\,y}{\sqrt{1-\epsilon}}\bigr)\,dy \tag{2.7.11}$$

であることがわかる（第 2 の等号は $x=\sqrt{\epsilon}y$ と置換積分して得られた）．ところが (2.7.8) によれば，この分布が $f(z)$ に等しいはずである．つまり（少し式を整理して）$f(x)$ は

$$f(z)=\frac{1}{\sqrt{1-\epsilon}}\int_{-\infty}^{\infty}f(y)\,f\Bigl(\frac{1}{\sqrt{1-\epsilon}}z-\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}y\Bigr)\,dy\qquad\text{（すべての } z \text{ と } 0<\epsilon<1 \text{ に対して）} \tag{2.7.12}$$

を満たすべきである，という結論になった．我々はこのような $f(x)$ を求めたいのだ．

2.7.3 行き先の分布密度 $f(x)$ の満たすべき性質は何か？ — 微分方程式

後は (2.7.12) を解けば良いのであるが，これは高校数学では簡単ではない．右辺の積分は畳み込み（convolution）と呼ばれるもので,「フーリエ変換」を用いて解くのが一般的である．しかし，これをやると 2.6.4 節の「特性関数の方法」と同じになってしまう．ここでは敢えて泥臭い方法でやってみる（と言っても線形微分方程式の性質をすこし使うので高校の範囲を超えてしまっているが）．

(2.7.12) の右辺は f が 2 つ出てきて積分だから始末が悪い．幸い，(2.7.12) は $0<\epsilon<1$ を満たす勝手な ϵ で成り立つので，両辺を ϵ の級数に展開してみよう（ここで「テイラー展開」の考えを使っている）．級数に展開した場合の各次数の係数が互いに等しいはずだ，ということから f の満たすべき方程式を導くのである[31]．

ϵ の級数としてみたとき，左辺はそのまま $f(z)$ である．右辺はと言うと，

$$\sqrt{1-\epsilon}=1-\frac{1}{2}\epsilon+O(\epsilon^2),\qquad \frac{1}{\sqrt{1-\epsilon}}=1+\frac{1}{2}\epsilon+O(\epsilon^2) \tag{2.7.13}$$

を用いて強引に

$$\frac{1}{\sqrt{1-\epsilon}}z-\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}y=z+\frac{1}{2}\epsilon z-\sqrt{\epsilon}y+O(\epsilon^{3/2}) \tag{2.7.14}$$

と思って，$f\bigl(\frac{z-\sqrt{\epsilon}\,y}{\sqrt{1-\epsilon}}\bigr)$ をテイラー展開すると（f' は f の 1 階導関数，f'' は f の 2 階導関数），

$$f\Bigl(\frac{1}{\sqrt{1-\epsilon}}z-\frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}y\Bigr)=f(z)+f'(z)\Bigl(\frac{1}{2}\epsilon z-\sqrt{\epsilon}y\Bigr)+\frac{f''(z)}{2}\epsilon y^2+O(\epsilon^{3/2}) \tag{2.7.15}$$

を得る[32]．これを (2.7.12) に代入して整理すると

$$\begin{aligned}f(z)&=\Bigl[1+\frac{1}{2}\epsilon+O(\epsilon^2)\Bigr]\int_{-\infty}^{\infty}f(y)\Bigl[f(z)+f'(z)\Bigl(\frac{1}{2}\epsilon z-\sqrt{\epsilon}y\Bigr)+\frac{f''(z)}{2}\epsilon y^2+O(\epsilon^{3/2})\Bigr]\,dy\\&=\Bigl(1+\frac{\epsilon}{2}\Bigr)f(z)\int_{-\infty}^{\infty}f(y)\,dy+\frac{f'(z)}{2}\epsilon z\int_{-\infty}^{\infty}f(y)\,dy-f'(z)\sqrt{\epsilon}\int_{-\infty}^{\infty}y\,f(y)\,dy\\&\quad+\frac{f''(z)}{2}\epsilon\int_{-\infty}^{\infty}y^2\,f(y)\,dy+O(\epsilon^{3/2})\end{aligned} \tag{2.7.16}$$

ここで，

$$\int_{-\infty}^{\infty}f(y)\,dy=1 \tag{2.7.17}$$

[31] ここのところは多項式間の等式において,「左右両辺で x^n の係数を等置して」とやるのとノリは全く同じ

[32]（注意深い人への注）ここで「おいおい，ちょっと待ってくれ．y はいろいろな値（特に非常に大きな値）をとるから ϵ をかけても小さくならないかもしれないよ」と疑問を感じたあなた，正解です．実際に y は積分変数だから正負の無限大になり，$\epsilon^2 y$ などもいくらでも大きくなり得るので，この項を単純に $O(\epsilon^2)$ と書くのは少し問題だ．しかし，もう少し考えるとこうやってもよいことがわかる．理由は以下の通り：このような項は常に $f(y)$ がかかって積分されるが，$f(y)$ は最終的には e^{-y^2} のような関数だとわかる．つまり，$|y|$ の大きいところは $f(y)$ の値が非常に小さいので，実際に $|y|$ が大きいところからの積分への寄与はほとんどない．つまり，実質的に $|y|$ は余り大きくない数と思っても良く，従って $\epsilon^2 y$ を $O(\epsilon^2)$ とみなしても良い．

一応，このように理屈をつけたけれども，これは結果を先取りして使っているからちょっと気持ちは悪い

であることに注意する（これは Y が何かの勝手な値をとる確率だから，全確率に等しく 1）．また，

$$\int_{-\infty}^{\infty} y\, f(y)\, dy = 0 \tag{2.7.18}$$

であろう（この量は Y の期待値であり，もともとの Y_N の定義から Y_N の期待値はゼロになっているので，極限でもゼロだろう）．最後に

$$\int_{-\infty}^{\infty} y^2\, f(y)\, dy = \alpha \tag{2.7.19}$$

とおくと（この量 α は何かの正の値のはず），

$$f(z) = \Bigl(1+\frac{\epsilon}{2}\Bigr) f(z) + \frac{f'(z)}{2}\epsilon z + \frac{f''(z)}{2}\epsilon\, \alpha + O(\epsilon^{3/2}) \tag{2.7.20}$$

となるので，ϵ の係数を比較して，

$$f(z) + zf'(z) + \alpha f''(z) = 0 \tag{2.7.21}$$

が得られた．つまり，極限の分布密度 $f(z)$ は（適当な正の定数 α に対して）上の微分方程式を満たす必要があることがわかった．

2.7.4 $f(x)$ の微分方程式を解く

最後に (2.7.21) の微分方程式を解けば，話は完結する．この方程式を綺麗に解くことは高校の範囲を超えているが，大体の大筋を説明する．

かなり一般的に使える微分方程式の解法の一つに「級数展開で解く」やり方がある：

$$f(z) = \sum_{n=0}^{\infty} a_n\, z^n \tag{2.7.22}$$

の形に級数で書けることを仮定して微分方程式に代入し，係数 a_n の間の関係式（漸化式）を求め，それから a_n を求めていく方法である.（そして最後に級数展開が実際にできることを確かめる）.

ともかくやってみよう．(2.7.22) を (2.7.21) に代入すると

$$0 = f(z) + zf'(z) + \alpha f''(z) = \sum_{n=0}^{\infty} a_n\, z^n + z \times \frac{d}{dz}\Biggl[\sum_{n=0}^{\infty} a_n\, z^n\Biggr] + \alpha \frac{d^2}{dz^2}\Biggl[\sum_{n=0}^{\infty} a_n\, z^n\Biggr] \tag{2.7.23}$$

となるが，和全体の微分を和の各項を微分することで置き換えると（ここのところ — 微分と級数の順序交換 — はまったく自明な問題ではない．大学一年生でこの辺りを詳しくやるだろう．今はこのような問題には目をつぶる）

$$\begin{aligned} &= \sum_{n=0}^{\infty} a_n\, z^n + z \times \sum_{n=0}^{\infty} a_n\, nz^{n-1} + \alpha \sum_{n=0}^{\infty} a_n\, n\,(n-1)\, z^{n-2} = \sum_{n=0}^{\infty} a_n\,(n+1) z^n + \alpha \sum_{n=0}^{\infty} a_n\, n\,(n-1)\, z^{n-2} \\ &= \sum_{n=0}^{\infty} \Bigl[a_n\,(n+1) + \alpha\, a_{n+2}(n+2)(n+1)\Bigr]\, z^n \end{aligned} \tag{2.7.24}$$

となった（最後の行では級数の添え字を付け替えた）．この級数がすべての z の値に対してゼロなのだから，z^n の係数がすべて 0 である：すべての $n \ge 0$ に対して

$$a_n\,(n+1) + \alpha\, a_{n+2}(n+2)(n+1) = 0 \qquad \Longrightarrow \qquad a_{n+2} = -\frac{1}{\alpha(n+2)} a_n. \tag{2.7.25}$$

これは a_n と a_{n+2} の間の関係式（漸化式）だから高校数学で十分に解ける．n の小さい方から順番にやれば良く，結果は $n \ge 0$ に対して

$$a_{2n} = \Bigl(\frac{-1}{2\alpha}\Bigr)^n \frac{1}{n!}\, a_0, \qquad a_{2n+1} = \Bigl(\frac{-1}{\alpha}\Bigr)^n \frac{1}{(2n+1)!!}\, a_1 \tag{2.7.26}$$

となる．ここで $(2n+1)!! = (2n+1)\cdot(2n-1)\cdot(2n-3)\cdots 5\cdot 3\cdot 1$ であり，$0! = 1$ と解釈する．

これを見ると，a_n の添え字が偶数と奇数で 2 つの系統に分かれているのがわかる．そこで，偶数・奇数毎にこの結果をまとめて

$$f_{\rm even}(z) = \sum_{n=0}^{\infty}\Big(\frac{-1}{2\alpha}\Big)^n \frac{1}{n!}\, z^{2n} = \sum_{n=0}^{\infty}\Big(\frac{-z^2}{2\alpha}\Big)^n \frac{1}{n!} = \exp\Big(\frac{-z^2}{2\alpha}\Big) \tag{2.7.27}$$

$$f_{\rm odd}(z) = \sum_{n=0}^{\infty}\Big(\frac{-1}{\alpha}\Big)^n \frac{1}{(2n+1)!!}\, z^{2n+1} \tag{2.7.28}$$

を定義すると（(2.7.27) の最後の等式は指数関数のテイラー展開を逆に使った），微分方程式の解 $f(z)$ は勝手な定数 a_0 と a_1 を用いて，

$$f(z) = a_0\, f_{\rm even}(z) + a_1\, f_{\rm odd}(z) \tag{2.7.29}$$

と書けることがわかった[33].

上で定義した $f_{\rm even}(z)$ は正に求める形をしているのでメデタイ[34]．問題は $f_{\rm odd}(z)$ の方であって，問題の条件を満たさないなどの理由で $f_{\rm odd}(z)$ を排除できたら話は完結する．

そのためには $f_{\rm odd}(z)$ がどのような関数かを調べる必要がある．答えを言ってしまうとこれは

$$f_{\rm odd}(z) = \exp\Big(-\frac{z^2}{2\alpha}\Big)\int_0^z \exp\Big(\frac{t^2}{2\alpha}\Big)dt \tag{2.7.30}$$

となっているのである.（この等式の証明には，基本的に両辺の級数展開が一致することを示せばよい．ただし，どのようにしてこの等式に導かれるかは秘密.）この積分の形にするといろいろなことがわかる．特に,（1）$f_{\rm odd}(z)$ は z の奇関数で，z が正なら $f_{\rm odd}(z)$ の値も正,（2）$z \to \infty$ で $f_{\rm odd}(z)$ は $O(\frac{1}{z})$ くらい，などがわかる．このどちらの性質も

$$\int_{-\infty}^{\infty} y\, f(y)\, dy = 0 \tag{2.7.31}$$

には不向きである —（1）からはこの積分が正であることがわかる.（2）からはこの積分がそもそも収束しないことがわかる．

と言うわけで，$f_{\rm odd}(z)$ は極限分布の $f(z)$ の成分としては非常に不適切であることがわかった．従って，$f(z)$ として許されるのは $f_{\rm even}(z)$ のみ，つまり

$$f(z) = a_0\, \exp\Big(\frac{-z^2}{2\alpha}\Big) \tag{2.7.32}$$

という結論になった．メデタシメデタシ.（なお，a_0 の値そのものは $\int_{-\infty}^{\infty} f(y)\, dy = 1$ であるべき事から $a_0 = \frac{1}{\sqrt{2\pi\alpha}}$ と決まるが，本質ではないので省略する.）

[33]「わかった」と書いたが，これ以外に解がないこと，またそもそも $f_{\rm even}$ や $f_{\rm odd}$ の級数がきちんと収束して定義できていること，は証明すべきである．これ以外に解がないことは「微分方程式の解の存在と一意性」をやれば納得できる．また，級数の存在などは大学一年でイヤというほどやるはずだ

[34] α が決まってないじゃないかと言う人がいるかもしれないが，今は分散の値を決めていない (2.7.3) をやっているので，この自由度は残って当然である

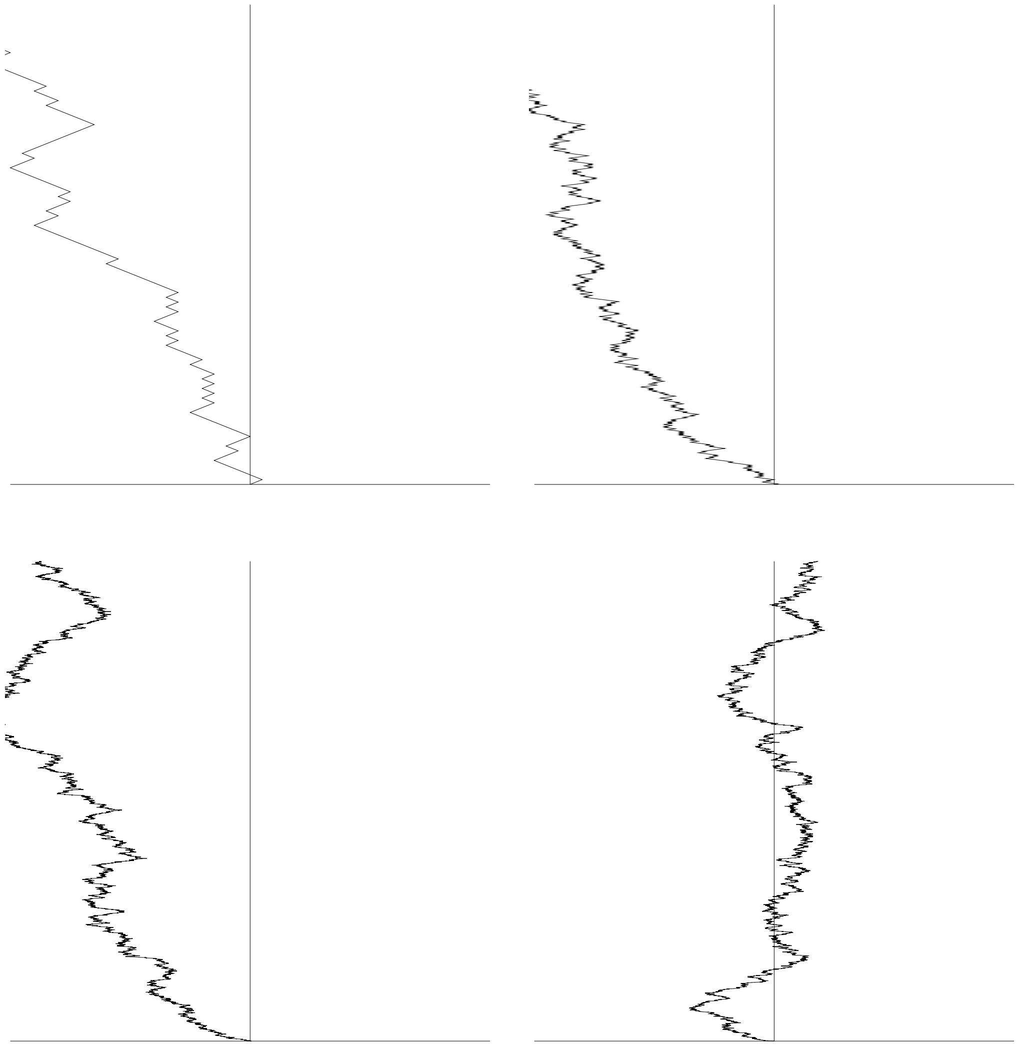


図 7: １次元ランダムウォークの例：第１行は左から全歩数が $N = 10^2, 10^3$ steps の場合，第２行は左から全歩数が $N = 10^4, 10^5$ steps の場合．横軸は x, 縦軸は歩数で，図示している範囲は縦軸は 0 から N，横軸は $\pm 2\sqrt{N}$.

3　ランダムウォーク

3.1　１次元ランダムウォーク

以下のような問題を考えよう：人が数直線上の原点にいる．コインを投げ，結果が表なら数直線の正の方向（右）に一歩進む；結果が裏なら，負の方向（左）に一歩進む．一歩進んだ後は今いる地点を出発点にしてコインを投げて進むことをくり返す．これを N 回くり返したとき，この人はどの付近にいるだろうか？このような運動は１次元単純ランダムウォーク（または酔歩）と呼ばれる．図７に横軸を数直線，縦軸を歩数にとって，１次元単純ランダムウォークの例を示す.

この問題に対する部分的な答えは, 今までやってきた大数の法則や中心極限定理から直ちに得られる．というのは，この問題を以下のように定式化できるからである．

まず，i 歩めでも数直線上の動きを X_i で表すことにする：

$$X_i = \begin{cases} 1 & （右に動いたとき） \\ -1 & （左に動いたとき） \end{cases} \tag{3.1.1}$$

この X_i は確率変数で，まともなコインを投げる場合は $+1$ と -1 を同じ確率でとるはずだ．つまり，$\mathbb{P}[\,X_i = 1\,] = \mathbb{P}[\,X_i = -1\,] = \frac{1}{2}$．また，コイン投げの時と同様，$X_i$ と X_j は独立なはずである．そして，N 歩めの位置は $S_N = \sum_{i=1}^{N} X_i$ となっている．

このように考えると，この単純ランダムウォークの問題は，コイン投げで表の割合を考えたのとほとんど同じで，違いは（1）$X_i = 0, 1$ の代わりに $X_i = -1, +1$ となっていること，（2）今回は N 歩の後の位置，つまり，S_N そのものに興味があること，の２つだけである．（1）の違いにより，今回は

$$\langle X_i \rangle = 0, \qquad \mathrm{Var}[X_i] = 1 \tag{3.1.2}$$

となる．

従って，大数の弱法則の証明（(2.5.26) 付近）を思い出すと，

$$\langle S_N \rangle = 0, \qquad \mathrm{Var}[S_N] = N \tag{3.1.3}$$

と結論できる．

これは何を言っているのだろうか？まず，ランダムウォークの場合の $\mathrm{Var}[S_N]$ の意味を考えよう．$\langle S_N \rangle = 0$ だから，分散の定義から

$$N = \mathrm{Var}[S_N] = \left\langle |S_N|^2 \right\rangle \tag{3.1.4}$$

である．$|S_N|$ は N 歩後の位置の，原点からの距離そのものであるから，上の分散には「原点からの距離の２乗の平均」という意味がつく（このため，この量を平均２乗変位 (mean square displacement) とよぶ）．つまり，(3.1.3) や (3.1.4) は「原点からの距離の２乗の平均は N である」と言っているわけ．距離の２乗の平均が N だから，距離そのものの平均は大体 $\sqrt{N}$ と思っても良いだろう[35]．これがランダムウォークの著しい特徴である[36]．

コイン投げの時にも強調したように，このような確率現象では，たまたまコインが「すべて表」になる可能性もあり，その場合には正の方向に N 離れたところまで行ってしまっている．しかしこのようなのは例外的な場合であり，その確率は非常に小さい．そして「典型的」な場合には，原点から $\sqrt{\mathrm{Var}[S_N]} = \sqrt{N}$ くらい（これ以内）の距離のところにいる，と期待されるわけである．これは要するに，表と裏が入り交じって出たために，原点付近で行きつ戻りつした結果である．この事情は「典型的」ないくつかの例を見ると，より良く納得できる（図 7 を振り返ろう）．

3.2 高次元ランダムウォーク

同じような問いを，この人の運動の次元を上げて考えることもできる．平面上の原点に人がいて，確率 $\frac{1}{4}$ ずつで東西南北の一つの向きに一歩進む，ことを考える．（一歩目の後は，今いる位置を出発点に同じ事をくり返す．）この時，N 歩の後にはこの人はどの付近にいるだろうか？（このモデルは２次元単純ランダムウォークと呼ばれる．）図 8 に２次元単純ランダムウォークの例を示す．

この問題に対しても１次元の時と全く同じように解析を進めることができる．特に，N 歩目の位置について，１次元と全く同様にアインシュタインの関係式

$$\Big\langle\ (N\ \text{歩後の原点からの距離})^2\ \Big\rangle = N \tag{3.2.1}$$

が成り立つ．

[35] $\left\langle |X|^2 \right\rangle \geq \langle |X| \rangle^2$ ではあるが，逆向きの不等式は一般にはなりたたない．特に，$\left\langle |X|^2 \right\rangle$ は非常に大きいのに $\langle |X| \rangle^2$ は小さいような X はいくらでも作ることができる．従って，$\left\langle |X|^2 \right\rangle = N$ だからと言っても $\langle |X| \rangle^2$ も N くらい大きいとは一般には言い切れない．これが「大体」と書いた理由である．ただし，ランダムウォークに関しては $\langle |X| \rangle$ が大体 $\sqrt{N}$ くらいであることはわかっている

[36] なお，(3.1.3) や (3.1.4) に相当する式は 1905 年のアインシュタインの論文で初めて明確に指摘されたので，$\left\langle |S_N|^2 \right\rangle = N$ は「アインシュタインの関係式」と呼ばれている．（アインシュタイン自身の仕事は３次元のブラウン運動に関するものだが，本質は同じ．）余談だが 1905 年というのはアインシュタインの当たり年で，「特殊相対性理論」「光電効果」「ブラウン運動」の３つの重要な論文が発表された．アインシュタインはその後，「光電効果」の仕事でノーベル賞を受けた

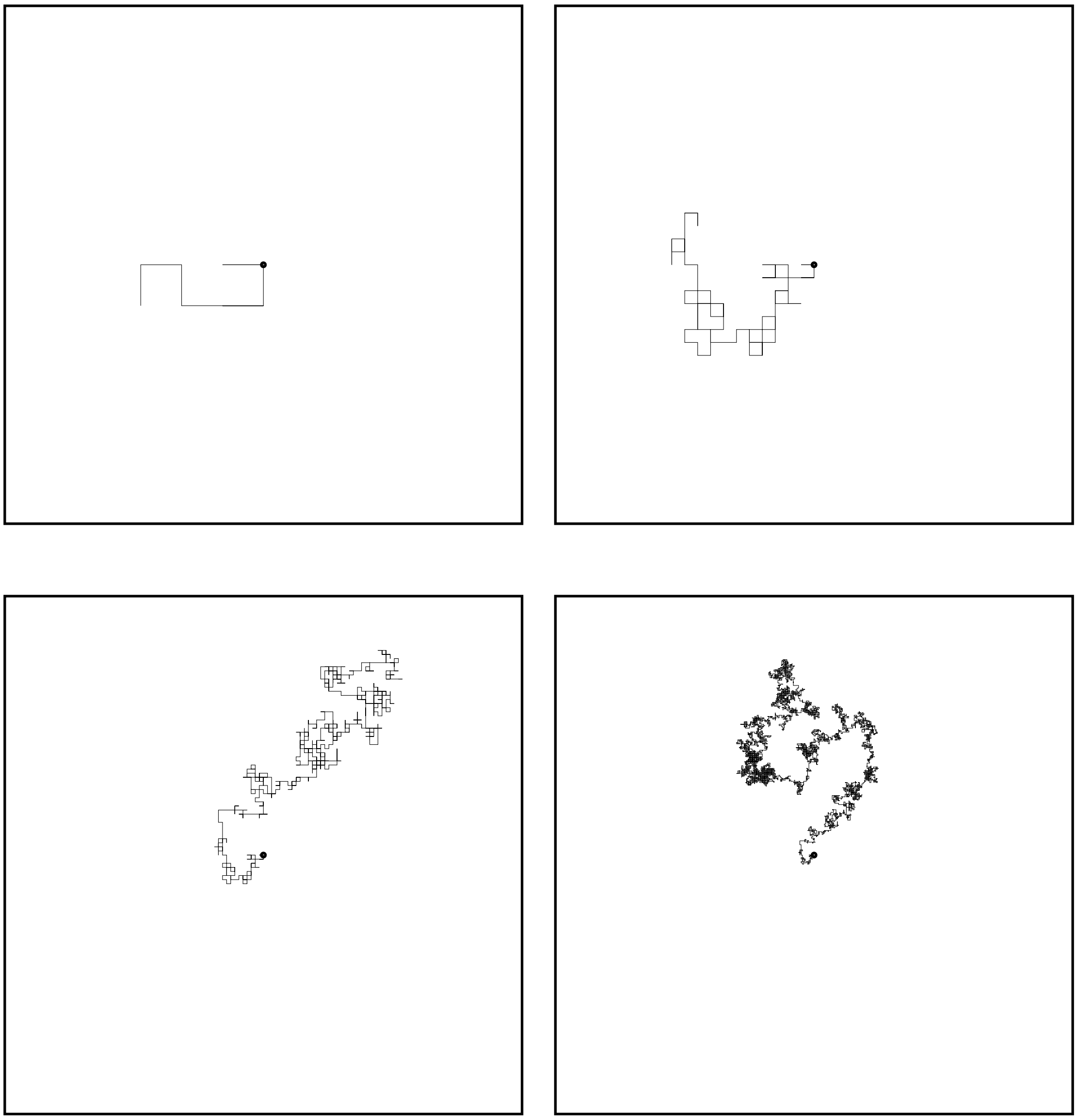


図 8: 2 次元ランダムウォークの例：第 1 行は左から $N = 10, 10^2$ steps，第 2 行は左から $N = 10^3, 10^4$ steps までの軌跡を表す．図示している範囲は $\pm 2\sqrt{N}$ で，原点（出発点）に小さな黒丸をうってある．

実はアインシュタインの関係式，つまり (3.2.1) は３次元以上の単純ランダムウォークでも成り立つことが容易にわかる．また，ランダムウォークのモデル（この人の進み方）を少々変えても — 例えば，時々２歩分進むことにするとか — 概要は変わらない：この場合は適当な定数 c を用いて

$$\Big\langle \, (N \text{ 歩後の原点からの距離})^2 \, \Big\rangle \approx cN \tag{3.2.2}$$

と書ける．つまり，1.1 節の言葉で言うと，「平均２乗変位が N のオーダーである」ことはやはり成り立っている．

この意味で，ランダムウォークにおけるアインシュタインの関係は，普遍的に見られる現象の一つであると言える．また，これがイントロの問題 c や c′ のヒントを与えてくれる．

4　まとめと未解決問題

今まで考えてきたのは，数学的には解決された問題である．その中心にあったのはたくさんの独立な確率変数の和，であった．そしてこの場合（N 個の確率変数の和を S_N と書く），独立性のために $\mathrm{Var}[S_N]$ が大体 N くらいになる[37]という大きな特徴が見られた．中心極限定理もランダムウォークも，すべてはこの「独立性」と $\mathrm{Var}[S_N] \approx N$ から来ている，と言っても過言ではない．

また，「独立性」が鍵だったのだから，独立性の仮定が満たされれば（今までに述べてきた定理の条件が完全には満たされなくても — 例： X_i が同分布でなくても）これらの現象（または類似の現象）が見られることが予想される．実際，今までにも少し注意したように X_i が完全に同分布でない場合にも大数の法則や中心極限定理が成り立つことは証明されている．また，独立性の仮定も（ホンの少しなら）破っても大丈夫であることも証明されている．

ところが，「独立性」の仮定がある程度破られると問題は一気に難しくなる．このような場合に何が普遍的に言えるのか，を考えるのは現代確率論の大きな未解決問題の一つである．以下では代表的なものを２つ紹介して結びとしたい．

（１）中心極限定理の拡張に関して：

中心極限定理は「独立な」確率変数に対して成り立つ，ということは何回か強調した．又，その際，$S_N - \langle S_N \rangle$ を $\sqrt{N}$ で割ることで，N が無限大になった場合に良い極限に収束するような確率変数（Z_N）を作り出せた．

では，独立でない場合はどうなるのだろうか？この場合，独立性が少しだけ破られている場合は実質的に独立な場合と同じであり，$\sqrt{N}$ で割ると行き先は正規分布になる．しかし，独立性がもっと破られている場合は未解決である．この場合中心極限定理に類似の結果[38]が成り立つと期待されているが，詳細はよくわかっていない．この問題は統計物理学の「臨界現象」とも密接な関連がある．

（２）ランダムウォークの拡張に関して：

ランダムウォークにおいて，人が動くときに「今までにいた場所には行けない（自分の足跡は踏んではいけない)」という条件を付けてみよう（このような条件をつけたモデルは self-avoiding walk （自己回避酔歩）と言われる)．「自分の足跡を踏まない」ためには，各ステップが独立ではダメだ（もし独立にステップを踏んでいると，容易に自分の足跡を踏んでしまう）．つまり，「自分の足跡を踏まない」条件は各ステップの独立性を壊してしまうので，問題が非常に難しくなるのである．

特に，「N が大きいときの平均二乗距離が N とともにどう増えるか（どのようなオーダーか)」すら $2, 3, 4$ 次元では未解決のママである[39]．なお，イントロの問題 e はこの self-avoiding walk の $d = 3$ のものと実質的に同じである．

[37] 1.1 節の言葉では $\mathrm{Var}[S_N]$ が N のオーダーである

[38] ただし，$\sqrt{N}$ ではなくて N^α などで割る必要があるだろう（ここで，α は未知の，$\frac{1}{2}$ より大きな定数である．また割って作った確率変数も正規分布にはならないだろう

[39] 単純ランダムウォークでは平均二乗距離は N のオーダーだったが，self-avoiding walk では $N^{2\nu}$ （$\nu > \frac{1}{2}$）のオーダーだろうと期待されているが，よくわからない．ただし，5 次元以上なら $\nu = \frac{1}{2}$ は証明されている

これらは確率論，統計物理学の大きな未解決問題で，研究が続けられている．僕個人は，これらの問題がこれからの確率論の中心的なテーマの一つになると思っているが，僕のような考えを持っている人は今のところ,（日本国内では）少数派のようであるから鵜呑みにしないように.

A　文献案内

高校生向けの確率論の本はいろいろと出ているが，この講義で扱った内容について詳しく書いてある本は（ちょっと探した範囲では）見つけることができなかった．そこで，余り程度には拘らず，よいと思うものを挙げておくことにする.

まず，以下の [1] は非常に良いのでお奨めである（読みこなすには高校３年くらいの知識が期待される）．ただ，残念ながら絶版になっている.

次に，[2] は「数学セミナー」という月刊誌の特集記事である．何人かの人が書いていて，基本的には大学一年生辺りがターゲットと思われるが，熱意のある高校生なら読める部分も多い．又，一つ一つの記事は長くないので，大体の感じを掴むのにも適していると思う.（この後でもっと本格的な本を読めばよい.）

また [3] と [4] は大学生向けの教科書ではあるが，高校生でも読める部分もある．[2] の後に読むのに適していると思う.

最後に，[5] は大学生向けの本格的な教科書である．高校生で読むのはかなり困難であろうが，将来の参考として挙げておく.

なお，僕の web page の「講義」のところ（http://math.nagoya-u.ac.jp/ hara/lectures/lectures-j.html）には過去の講義ノートの一部が置いてある．2001 年度の「数学展望 II」のものはこのアゴラの延長上にあるもので，参考になるかもしれない．また 2002 年度の「確率論」はその更に延長上にある.

参考文献

[1] 楠岡成雄：「確率・統計」森北出版，新数学入門シリーズ７ (1995)

[2] 特集「確率」：数学セミナー 1995 年 6 月号 (1995)

[3] 福島正俊：「確率論」裳華房，数学シリーズ (1998)

[4] 小針あき（日へんに見）宏：「確率・統計入門」岩波書店 (1973)

[5] 西尾真喜子：「確率論」実教出版 (1978)