

確率論の極限定理（と臨界現象）*

原 隆

九州大学数理学研究院

e-mail: hara@math.kyushu-u.ac.jp

http://www.math.kyushu-u.ac.jp/~hara/

概要

確率論は、身の回りにある「ランダム」な現象を理解するために発展した学問です。ランダムな現象には一見、なんの法則性も無いように見えるのですが、もう少し調べると「大数の法則」「中心極限定理」などに代表される、非常に美しい法則性が発見されます。これらは数学として大変に面白いのみならず、統計学、物理学、数理生物学など、様々な学問分野でも有効です。

本講義では、「大数の法則」「中心極限定理」などの確率論の極限定理を中心にしたお話しをします。また、関係する話題として、「ランダムウォーク」や統計力学における「臨界現象」などにも少し触れたいと考えています。

目次

1	はじめに：考える問題	2
1.1	記号の約束と「オーダー」の概念	3
2	コイン投げの数理：大数の法則と中心極限定理	4
2.1	実際にやってみました	4
2.2	少し解析する。N 回のうちに m 回表になる確率は？	4
2.3	N が大きくなったら？ I. コイン投げにおける大数の法則	8
2.4	N が大きくなったら？ II. コイン投げにおける中心極限定理	9
3	一般の確率変数と極限定理	10
3.1	一般の確率変数	10
3.2	大数の法則	16
3.3	中心極限定理	19
3.4	中心極限定理の証明	23
3.5	完全なおまけ：中心極限定理に出てくる曲線を求めよう	29
4	ランダムウォーク	34
4.1	1次元ランダムウォーク	34
4.2	高次元ランダムウォーク	35
5	臨界現象へ	36
5.1	臨界現象とは	37
5.2	スピン系の臨界現象：スピンの和のふるまいから	41
5.3	くりこみ群の考え	42
A	文献案内	43

*九州大学公開講座，2019 年 8 月 10 日

1 はじめに：考える問題

日常、「確率」という言葉を耳にすることはよくあります — ほとんど毎日、「今日の降水確率は…」を聞かされているように思います。また、宝くじに当たる確率は〇〇，トランプのポーカーでこの役ができる確率は〇〇，なども耳にします。

このように「確率」は不規則な（ランダムな）現象，確実に結果を確実に予測しがたい（でも何らかの予測ができる）場合を扱う際に使われています。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ます。いろいろな確率を計算できることはそれ自身重要ですし，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わけです。

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こうと思います。特に，たくさんの試行を繰り返した場合の結果を予言する「大数の法則」「中心極限定理」などにまず，焦点を当てましょう。その後，これらの法則から良く理解できるランダムウォークに進み，物理の世界とのつながりをつけます。更に，物理の世界で普遍的に見られる「臨界現象」に話を進めたいのですが，時間の関係でちょっと難しいかもしれません。

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ます。いくつかの例を挙げますと：

- 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」と教わる。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。が，この根拠は何なのだろう？
- b. ある学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。何が原因でこうなるのか？
- 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，だんだんと臭素が容器中に拡がっていくのがわかる。これを拡散現象と言うが，臭素の色がついた部分は，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。
- e. 溶媒中の高分子。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（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？
- f. 統計力学の臨界現象。磁石を熱していくとある温度（臨界温度 T_c ）以上では磁石ではなくなる。 T_c 付近ではどんな現象が見られるのか？実は，このような現象は磁石に限らず，いろいろな物質で見られるものである。

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ありません。その理由も後で少しだけ理解できると思います

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ますが，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のです（そもそも，食塩や砂糖では色がついていないから見えませんが，インクなどを使ってもちよつとした液体の対流運動にかき消されてしまうので難しいです）。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ありません

これらの現象は、一見、無関係なように見えますが、奥の方ではつながっています。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す³。e, f は統計力学の未解決問題の一つですが、ランダムウォークや中心極限定理と密接な関連があります。中心極限定理とランダムウォーク自身にも関連がありますし、**中心極限定理は「統計学」の基礎**にもなっています。

本講義では上のような現象を理想化・単純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します。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す。

1.1 記号の約束と「オーダー」の概念

不等号： $a \leq b$ は $a \leq b$ と、 $a \geq b$ は $a \geq b$ と同じ意味です。

和の記号： $x_1 + x_2 + x_3$ の事を $\sum_{i=1}^3 x_i$ と書きます。同様に $a_1 + a_2 + \dots + a_n$ を $\sum_{i=1}^n a_i$ と書きます。このように $\sum_{i=1}^n$ 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します。この際、 i の代わりに j や k を使っても構いません。例を挙げると：

$$X_1 + X_2 + X_3 + \dots + X_n = \sum_{i=1}^n X_i = \sum_{j=1}^n X_j \quad (1.1.1)$$

などと書けます。 \sum は英語の S に相当するギリシャ文字の大文字で、(英語では) sum の頭文字 S を表しています。

積の記号： 和と同じノリで、積を表す記号も導入します。

$$X_1 \times X_2 \times \dots \times X_n = \prod_{j=1}^n X_j \quad (1.1.2)$$

と書きます。 \prod はギリシャ文字 π の大文字で、(英語では) product の頭文字 P を表しています。

$N \rightarrow \infty$: 「 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ます。「 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、 $N \rightarrow \infty$ と書きます。

「オーダー」の概念：

$f(N)$ を正の整数 N の関数とします (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す。

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」と言います。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) になっていく**一番主要なところ**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」です。「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「増え方の主要部分」を考えているのです。

別の例では $f(N) = \frac{1}{N}$ も $f(N) = \frac{3}{2N}$ も、ともに $\frac{1}{N}$ のオーダーです。一方、 $f(N) = \frac{1}{N^2}$ は N^{-2} のオーダーです。

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、の問題が頻出しますが、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りませう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ただけ起こるから、a~d はこれらの定理やモデルそのものではありません。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だけ思ってくださいと幸いです。

⁴「オーダー」の定義には少し混乱があって、数学での正しい定義は以下のようなものです:「 $f(N)$ が N^2 のオーダーである」とは、定

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~f のとっかかりとして、コイン投げを考えましょう：10 円玉を投げて、表が出るか裏が出るかを考察します。ただし一回投げただけでは面白くないので、「何回も（一万回とか）投げたうちのどのくらいが表になるか」を考えてみます。

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ですが、もう少し定量的に考えてみたいと思います。

2.1 実際にやってみました

名古屋大学で 2003 年の夏に、高校生向け（約 70 人）に公開講座をしました。その際、実際に 4 回ほど、コインを投げてもらいました。その結果 (n 回表になった人は何人か) は大体、以下のようになりました。

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ます。これだけでは規則性はあまり見えていません。

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、上の結果を解析していきましょう。この講義では条件 A が実現される確率を $\mathbb{P}[A]$ と書きます。例えば $\mathbb{P}[\text{コインを一回投げた結果が表}]$ は文字通り「コインを一回投げた結果が表」である確率を表します。

(余分な注) 本論に入る前に確率の背景についての注を 2 つ述べておきます。

- 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ありません。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おりません。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ります。

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（ちょっと曖昧な）認識で出発します。ただし、幾分トートロジーめきますが、この決め方に矛盾がないことは後の大数の法則でわかるでしょう。

- 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（物理）の問題です。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、初角速度）、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。

このように考えると、確率論は必要ないようですが、もちろん、そうではありません。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ります。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なので、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ます。この意味で確率論は有効です — もしも、いつでも完全に同じように投げる「コイン投げマシン」ができた場合には、投げた結果はいつも同じはずだから、確率論の出番はないでしょう。

数 C があって、 $|f(N)| \leq CN^2$ がすべての $N \geq 1$ で成り立つことです。つまり、この定義によれば $f(N)$ が N^2 よりずっと小さくても良いのです（極端な場合、 $f(N) = N$ でも、 $f(N) = 1$ でも、ともに「 $f(N)$ は N^2 のオーダーである」といえます）。このように、数学の正しい定義では、 $F(N)$ の「大きさ」を忠実に再現しない場合が含まれるので、この講義では採用しません。なお、なぜ厳密な定義がこうなっているかという、ここで採用した定義のなりたない関数も扱えるようにするためです。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います。一方厳密な定義ではこの関数のオーダーは N^2 （またはそれ以上のオーダー）になります

このように古典力学の世界での確率論は、我々の側のある種の「情報の欠如」（コインの初速度などがコントロールできない、など）に伴って登場することが多いです。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しますが、この講義ではそれは避けて通ることにします。

（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 を定義しましょう ($i = 1, 2, 3, \dots, N$)。ここで

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

と決めます。(0,1を使うのは、「表」「裏」と書くのがじゃまくさいからですが、後で見るように別の効用もあります。)そして1回目からの結果を並べて (X_1, X_2, X_3, X_4) などのように書きます。例えば、「1回目から4回目まで表だけが出る」のは $(1, 1, 1, 1)$ と書きます。同様に、 $(1, 1, 0, 1)$ は「3回目だけが裏で残りは表」の場合を表します。時にはスペースを省略するため、 $(1, 1, 0, 1)$ の代わりに1101などと書くかもしれません。

記号を整理したので、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、その確率を計算しましょう。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ります。

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についてです。通常のコインは表裏がほとんど同じに作ってあるし、材質も均一でしょうから、表と裏はほとんど同じくらい出やすいだろうと思われまふ。そこで普通は**コインの表と裏は同じくらい出やすい**と仮定し、 $\mathbb{P}[\text{コインを1回投げた結果が表}] = \frac{1}{2}$ と仮定します。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ませんが、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りません。このように表と裏の出やすさが違う場合も後で考えますが、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きます。

以下ではより一般の場合でもできるように、 $\mathbb{P}[\text{コインを1回投げた結果が表}] = p$ として進みますが、特に断らない限りは $p = \frac{1}{2}$ とと思ってくださって結構です。また、投げ方によってはコインが端で立つような事もあり得ますが、簡単のためにそのような場合は起こらないとして進みましょう。

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す。それがコイン投げの**独立性**に関する以下の仮定です：

（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、どうなるでしょうか？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）について考えてみます。1回目に表が出るのは全体の p の割合です。2回目も表になるのは1回目表だったうちの p の割合のはずです（ここで独立性を使いました⁵）。結局、独立性の仮定から $\mathbb{P}[X_1 = 1 \text{ かつ } X_2 = 1] = \mathbb{P}[X_1 = 1] \mathbb{P}[X_2 = 1] = p^2$ となります。同様に、 $\mathbb{P}[X_1 = 1 \text{ かつ } X_2 = 0] = \mathbb{P}[X_1 = 1] \mathbb{P}[X_2 = 0] = p(1-p)$ となります。このように考えていくと、 ϵ_1, ϵ_2 を0か1のどちらか（どっちでも良い）として

$$\mathbb{P}[X_1 = \epsilon_1 \text{ かつ } X_2 = \epsilon_2] = \mathbb{P}[X_1 = \epsilon_1] \mathbb{P}[X_2 = \epsilon_2] \quad (2.2.2)$$

が成立するはず⁶です。

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、 $\mathbb{P}[X_1 = 1 \text{ かつ } X_2 = 1] > \mathbb{P}[X_1 = 1] \mathbb{P}[X_2 = 1]$ となるでしょう

⁶ ϵ はイプシロンと読むギリシャ文字

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しますから、結果として

$$\mathbb{P}[X_1 = \epsilon_1, X_2 = \epsilon_2, \dots, X_N = \epsilon_N] = \mathbb{P}[X_1 = \epsilon_1] \mathbb{P}[X_2 = \epsilon_2] \cdots \mathbb{P}[X_N = \epsilon_N] \quad (2.2.3)$$

が得られます⁷。ここで ϵ_i は 0 でも 1 でも、勝手な値でよいです。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

$$\mathbb{P}[X_1 = \epsilon_1, X_2 = \epsilon_2, \dots, X_N = \epsilon_N] = p^{(\text{表の出た数})} (1 - p)^{(\text{裏の出た数})} \quad (2.2.4)$$

となります。要するに、表が出る確率は p , 裏が出る確率は $(1 - p)$ ですから、それを表と裏の個数分だけかければよいわけです⁸。

これを元にして、「 N 回投げたときに m 回表が出る」確率を求めましょう。後のために

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

を定義しておきます。

簡単なところから出発します。 $N = 1$ の時は仮定そのもので

$$\mathbb{P}[S_1 = 1] = p, \quad \mathbb{P}[S_1 = 0] = 1 - p \quad (2.2.6)$$

で面白くありません。 $N = 2$ の時、

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

は両方とも表、両方とも裏、だから納得です。ね。 $S_2 = 1$ の場合はどうでしょう？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 $p(1 - p)$ です。よってこの 2 通りを足して、

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

となります (他の場合も比較のために書きました)。

$N = 3$ も同様に計算できます。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 $p^2(1 - p)$ です。従って (全部表や全部裏、の場合も書くと)、

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

とわかります。

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

N 個から m 個をとる組み合わせの数

⁷この式、およびその元になった独立性の仮定は当たり前ではありません。それは以下の問いを考えるとわかります：「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 $\frac{1}{2}$ ですが、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」と考えたくありませんか？

今は既に 100 回も表が出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」を考える必要があります。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、(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) と思われま。

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十分に注意してください。「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でしょうが、我々はどうしてもこのような方向に引きずられやすいように思います。(条件付き確率に関連した話題には、Monty Hall problem など、我々の直感に反する感じのする面白い問題がたくさんあります。)

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ません (つまり、表が出る確率 p が $1/2$ より大きいと考える)。実際、「100 回も表だったら 101 回めも表」と思ってしまうのは、経験的に独立性を疑うというよりも「コインがイカサマ」と判断しているためかもしれません

⁸しつこいですが、このようになるのは「独立性」のおかげで (2.2.3) がなりたつからです

といて、 ${}_N C_m$ または $\binom{N}{m}$ で表します。上の考察から ${}_3 C_3 = {}_3 C_0 = 1$, ${}_3 C_2 = {}_3 C_1 = 3$ などですが、一般には

$$\binom{N}{m} = {}_N C_m = \frac{N!}{m!(N-m)!}, \quad N! = N \times (N-1) \times (N-2) \times \cdots \times 3 \times 2 \times 1 \quad (2.2.10)$$

であることがわかります⁹。これを認めると

$$\mathbb{P}[S_N = m] = {}_N C_m p^m (1-p)^{N-m} \quad (2.2.11)$$

が得られます。さてさて、名古屋大の公開講座での結果と比較すると

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

となります。3行目と4行目を比較すべきですが、当たらずといえども遠からず、というところでしょうか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 に載せました。なぜこんなことが起こるのか、以下で考えます。

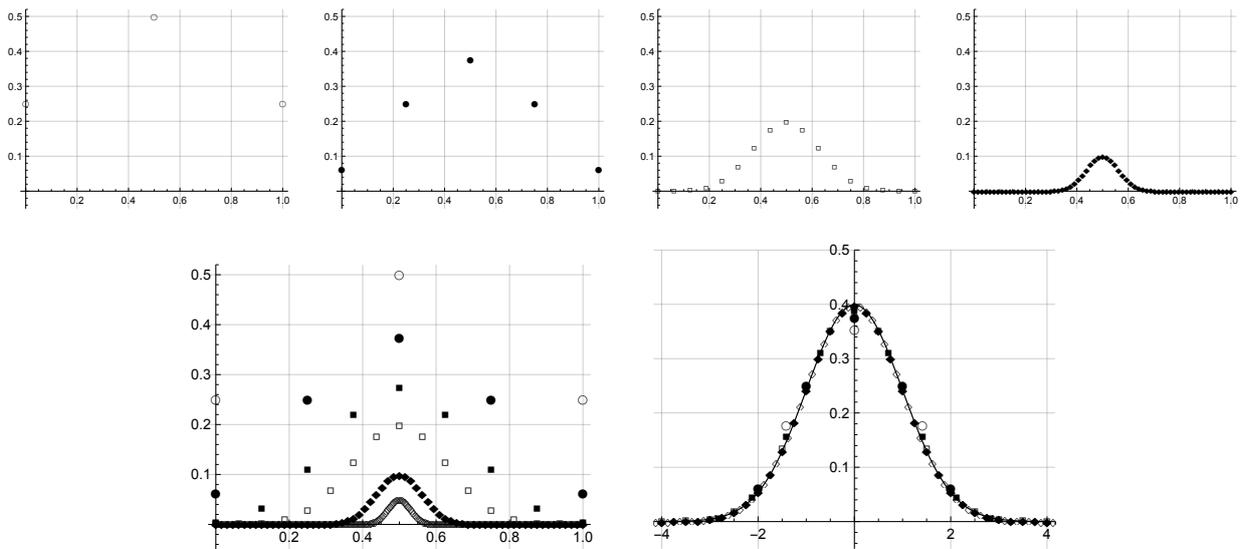


図 1: N 回投げて m 回が表の確率 $\mathbb{P}[S_N = m]$ のグラフ。いろいろ描いてみました。一行目の 4 つのグラフは $N = 2, 4, 16, 64$ のそれぞれを描いたもので、横軸が $\frac{m}{N}$ 、縦軸は $\mathbb{P}[S_N = m]$ です。2行目の左はこの 4 つ、および $N = 8$ と $N = 256$ を重ねて描いたもの（軸の取り方は同じ）。2行目の右は左のグラフを $\frac{m}{N} = \frac{1}{2}$ を中心にして縦軸、横軸をうまく伸び縮みさせたものです — どのように伸び縮みさせたのか、また、実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

⁹いわゆる「順列と組み合わせ」の計算ですね

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、というつもりですから、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ありません。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す

2.3 N が大きくなったら？ I. コイン投げにおける大数の法則

本題に戻ります。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）計算しました。結果は

$$\mathbb{P}[S_N = m] = {}_N C_m 2^{-N} \quad (2.3.1)$$

というもので、その結果をグラフでお見せしました。それを再録すると図 2 の左のようになっています（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 通りに制限しました）。 $p = \frac{1}{2}$ 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す。

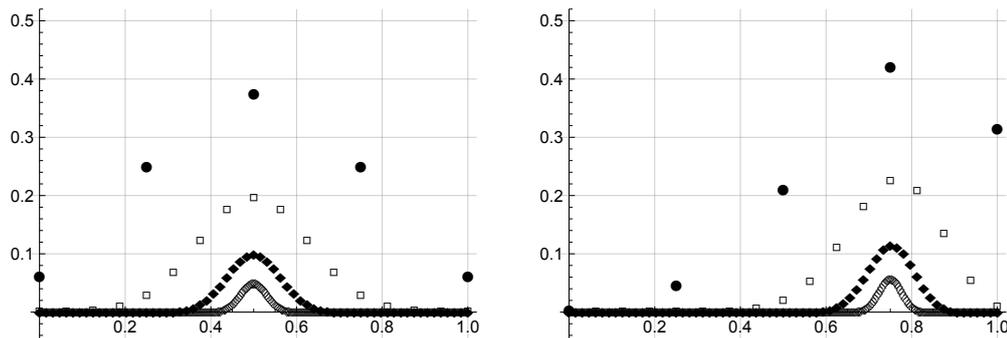


図 2: 左：表と裏が同程度に出やすいコインを N 回投げたときの確率。横軸は m/N 、縦軸はその $\mathbb{P}[S_N = m]$ を表しています。4 種類の点は上から $N = 4$ (●), 16 (□), 64 (◆), 256 (◇)。 N が大きくなるにつれて確率が $\frac{S_N}{N} = \frac{1}{2}$ に集中していきます。

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 $\frac{S_N}{N} = \frac{3}{4}$ に集中が見られます。

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ます。この背後にある定理を述べると以下ようになります（一般化と証明は 3.2 節）。

定理 2.3.1 (コイン投げにおける大数の弱法則) 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、以下が成り立つ：

$$\mathbb{P}\left[\left|\frac{S_N}{N} - p\right| > a\right] \leq \frac{p(1-p)}{a^2 N} \quad (2.3.2)$$

(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) は含みません。ここでは定理の主張がより具体的になるように、後半まで含めて書きました。

この定理の意味するところは、 N が大きくなるにつれ、 $\frac{S_N}{N}$ が p に近づいていく、ということです。ただし、この言い方は**不正確**です。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」ことは起こりえます（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、これはゼロではありません）。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」というものです¹¹。

(予告) 3.2 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ちます（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」というものもあって、それならもう少しだけ強いことが言えるのですが、それはこの公開講座の範囲を超えています

2.4 N が大きくなったら？ II. コイン投げにおける中心極限定理

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りません (N が大きくなると確率的にゼロになる、ことはわかりましたが、もう少し詳しいことを知りたいのです)。この答えは「中心極限定理」なのですが、その説明には少し準備が必要です。まずは2.2節で見たグラフ (図1) を少し手直ししてお見せします (図3)。

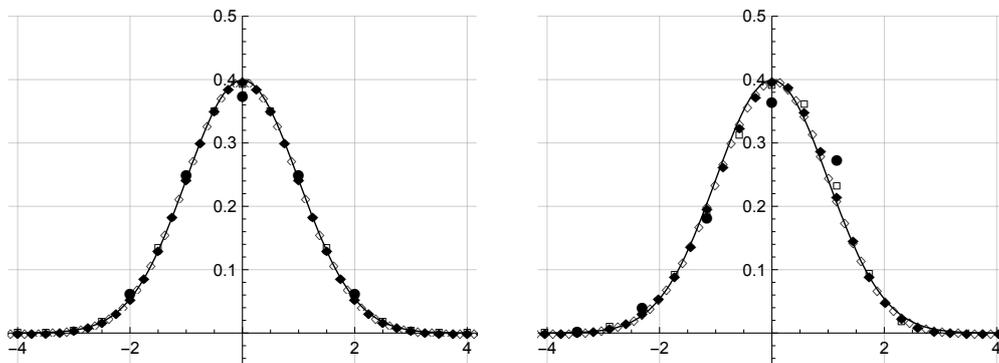


図3: 図2の座標軸を取り替えたもの (横軸方向にずらした後、縦横ともに拡大; 図1の右下の図に相当)。実線は $y = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ のグラフで、4種類の点は $N = 4$ (●), 16 (□), 64 (◆), 256 (◇) の場合の確率を表します。座標軸の取り方は、横軸は $\sqrt{\frac{N}{p(1-p)}}(\frac{m}{N} - \frac{1}{2})$, 縦軸は $\mathbb{P}[S_N = m] \times \sqrt{p(1-p)N}$ です。左の図は $p = \frac{1}{2}$ のコインの場合で、右は $p = \frac{3}{4}$ の場合です。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ります。(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目してください。)

上のグラフを数学的な定理の形で述べるのが、以下の定理です。なお、上では図2のグラフを伸び縮みさせましたが、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す (この点はこの後で、より詳しく説明します)。

定理を述べるために、まず状況を設定します。表の出る確率が p であるコイン投げを考え、新しい確率変数 (ランダムな数)

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

を定義します。この Z_N は図3の横軸そのものです。このとき、以下が成り立ちます (一般化と証明は

定理 2.4.1 (コイン投げにおける中心極限定理) 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 $\mathbb{P}[a \leq Z_N \leq b]$ は、

$$\left(\text{グラフ } y = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \text{ と 3直線 } x = a, x = b, y = 0 \text{ で囲まれた部分の面積} \right) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \tag{2.4.2}$$

に収束する。

少し注を付けておきます。

- $e^{-x^2/2}$ と言うのは、以下のような関数です。まず $e = 2.71828\dots$ は「自然対数の底」とよばれる特別の実数です。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) を表します — y が無理数の時の定義には少し工夫が必要ですが、ここでは立ち入りません。最後に、この y を $-\frac{x^2}{2}$ で置き換えたものが $e^{-x^2/2}$ です。図4の左に $y = e^x$, 右に $y = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ のグラフを掲げました。

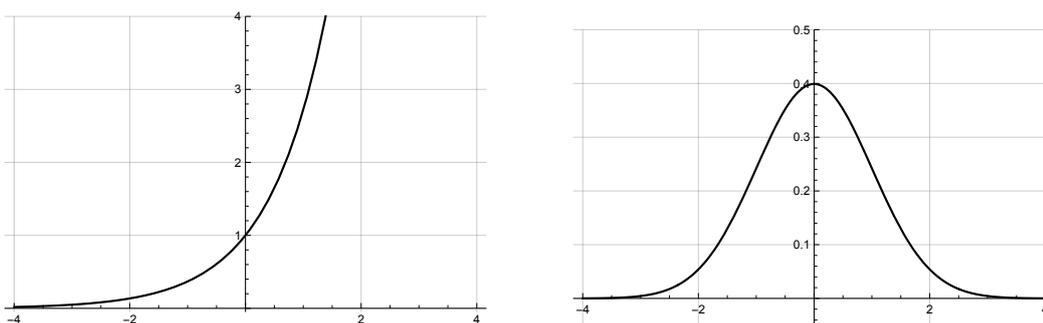


図 4: 左: $y = e^x$ のグラフ. 右: $y = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ のグラフ

- 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

$$\mathbb{P}[a \leq z \leq b] = \left(\text{グラフ } y = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \text{ と 3 直線 } x = a, x = b, y = 0 \text{ で囲まれた部分の面積} \right) \quad (2.4.3)$$

で与えられるものです.

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 — です.

次の節では, これらの極限定理をもっと一般化するとともに証明します.

3 一般の確率変数と極限定理

ここまでは話をコイン投げに限定してきましたが, 「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ちます (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します; 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と思います). 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, また重要です¹². 同時に, この定理はイントロの a, b, c'' の問題の背景を説明してくれます.

3.1 一般の確率変数

一般の場合を考えるため, この節では少し準備をします. 一般論にするのには, もっと切実な理由もあります. 生半可なやり方 (とまかく計算) では, 以下のような問題に立ち向かえないのです.

(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?)

この問題はコインの問題よりも手強いです. 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 回出る」ものを考えないといけません ($m_1 + m_2 + \dots + m_6 = N$) が, この計算はかなり大変です (「多項分布」と呼ばれるものになります). 正 6 面体 (立方体) のサイコロでさえこんなに大変なのに, 世の中には正 12 面体や正 20 面体のサイコロもあるし, 全ての面が同じ確率で出ないようなイカサマサイコロもあります. さらに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ります. 「大数

¹² (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と僕は思います. これが僕が「普遍性」に拘る理由です

の法則」や「中心極限定理」、さらに大数の法則の証明にもちいる「チェビシェフの不等式」は正にその方法を与えてくれるものです¹³。

3.1.1 確率変数

まず、「確率変数」(random variable, 略して rv) という概念を正式に導入します。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 などが例です。

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, \dots , (2) それぞれの値をとる確率、つまり $p_i = \mathbb{P}[X = x_i]$ ($i = 1, 2, \dots$) を決めます (この 2 つが同じなら、同じ確率変数とみなす)。つまり、以下ののような表を与えることが確率変数を決めることになります。また、このような x_i と p_i の対応を X の**分布**といいます。

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

上では n 個の値しかとらない確率変数 (離散型確率変数) を考えましたが、実際には連続無限個の値をとるような確率変数 (連続型確率変数) もたくさんあります (中心極限定理で出てきた標準正規分布はその例)。そんな場合、確率変数 X について、ある函数 $f(x)$ が存在して

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x) dx \quad (\text{すべての } a < b \text{ に対して}) \quad (3.1.1)$$

と書けるとき、 $f(x)$ を X の**確率密度函数** (probability density function) といいます。この場合の X の期待値は

$$\langle X \rangle = \int_{-\infty}^{\infty} x f(x) dx \quad (3.1.2)$$

と定義します¹⁴。

なお、以下では「積分」に慣れていない人のことを考慮して、かなりの議論を離散型確率変数に限定して行います。同様の議論が連続型確率変数に対しても成り立つことは、「積分」を理解すれば納得できるので、省略します。

3.1.2 期待値と分散

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、考えてみましょう。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ですが、これは実際にはなかなか大変です (正 20 面体のサイコロや、 X が 10^8 通りの値をとる場合を想像してください)。たとえそれができたとして、 10^8 個もの場合のそれぞれの確率 p_1, p_2, \dots を教えてもらっても、何かわかった気になるでしょうか¹⁵?

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ました。その代表的なものが期待値と分散です¹⁶。

¹³ (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ります。この意味で、より普遍的な定理、その簡単な (明瞭な) 証明を探すことは数学の発展上も大切だと思います

¹⁴ さらに、離散型と連続型が入り混じったような確率変数もあります。例えば、 $\mathbb{P}[X = 0] = \frac{1}{2}$, かつ $0 < a \leq 1$ に対して $\mathbb{P}[0 < X < a] = \frac{a}{2}$ となっているようなものです。この類を数学的に扱うには、**累積分布函数** (cumulative distribution function) または単に **分布函数** (distribution function) と呼ばれる $F(x) := \mathbb{P}[X \leq x]$ を用いると便利です (数学では単に分布函数といった場合、確率密度函数ではなく累積分布函数をさすことに注意)。連続型確率変数の場合、累積分布函数と確率密度函数の間には $F(x) = \int_{-\infty}^x f(y) dy$ の関係があります。

¹⁵ (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、という例です

¹⁶ 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ります。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です (期待値の計算が簡単な理由の一つは以下の (3.1.11)–(3.1.12) 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「特性函数」) 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きけるか、が問題となることもしばしばです

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

$$E[X] = \langle X \rangle = \sum_{i=1}^n p_i x_i = p_1 x_1 + p_2 x_2 + p_3 x_3 + \dots + p_n x_n \quad (3.1.3)$$

で定義します¹⁷. また,

$$\text{Var}[X] = \langle (X - \langle X \rangle)^2 \rangle = \sum_{i=1}^n p_i (x_i - \langle X \rangle)^2 \quad (3.1.4)$$

を X の**分散** (variance) と言い, $\sqrt{\text{Var}[X]}$ を X の**標準偏差** (standard deviation) と言います. (標準偏差は σ で表すことが多いです.)

このうち, 「期待値」の方はおなじみです. 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」に他なりません. 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ます.

これに対して, 「分散」は X の**分布の広がり**を表します. より正確には標準偏差 σ が, X の分布の大体の広がりを示します. テストの点数の例で言うと, 以下ようになります: いま, 同じテストをしたところ, クラス A もクラス 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, 元の分布は決まりません. 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ます. 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ます. 元の分布をもっと限定していくためには ℓ 次 (期待値の周りの) **モーメント** (moment) と呼ばれる量

$$\langle (X - \langle X \rangle)^\ell \rangle \quad (3.1.5)$$

をすべて見ていく必要があります ($\ell = 2, 3, 4, \dots$)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ます. つまり, 確率変数 X, Y に対して

$$\langle X \rangle = \langle Y \rangle, \quad \langle (X - \langle X \rangle)^\ell \rangle = \langle (Y - \langle Y \rangle)^\ell \rangle \quad (\ell = 2, 3, 4, \dots) \quad (3.1.6)$$

ならば, X と Y は同じ分布に従っています¹⁹.

3.1.3 「独立な」確率変数

大数の法則や中心極限定理のキーになるのは, 「独立な」確率変数という概念です. コイン投げの例を思い出しましょう.

コイン投げの場合, 1 回目の結果と 2 回目の結果と 3 回目の結果, ... は「独立」であると考えるのが自然だろうと説明しました. なぜなら, 「1 回目表だったから, 2 回目は裏にしてやろう」などと結果をコントロールできる人は, まずいないだろうからです. この「独立性」を数式にすれば

$$\mathbb{P}[1 \text{ 回目表}, 2 \text{ 回目裏}, 3 \text{ 回目表}] = \mathbb{P}[1 \text{ 回目表}] \mathbb{P}[2 \text{ 回目裏}] \mathbb{P}[3 \text{ 回目表}] \quad (3.1.7)$$

さらに一般に (2.2.3) で書いたように,

$$\mathbb{P}[X_1 = \epsilon_1, X_2 = \epsilon_2, \dots, X_N = \epsilon_N] = \mathbb{P}[X_1 = \epsilon_1] \mathbb{P}[X_2 = \epsilon_2] \dots \mathbb{P}[X_N = \epsilon_N] \quad (3.1.8)$$

¹⁷ 数学の本では期待値を $E[X]$ と書くことが普通ですが, 物理などでは $\langle X \rangle$ と書くことも多いです. この講義では主に後者を使います

¹⁸ 2 次のモーメントは上で定義した分散と同じです

¹⁹ ここはちょっと不正確で, 数学的に厳密なことをいうためには, 確率変数 X, Y に対して, もう少し条件が必要です

がすべての $\epsilon_j = 0, 1$ についてなりたつこと、つまり、『積事象の確率』が『各事象の確率の積』に分解することでした。

これを一般の確率変数に拡張して、以下の定義を行います。

定義 3.1.1 (独立な確率変数) 確率変数 X と Y が、それぞれ、 $X = x_1, x_2, \dots, x_m$ 、および $Y = y_1, y_2, \dots, y_n$ の値をとるとする。このとき、すべての $i = 1, 2, \dots, m$ かつ $j = 1, 2, \dots, n$ に対して

$$\mathbb{P}[X = x_i \text{ かつ } Y = y_j] = \mathbb{P}[X = x_i] \mathbb{P}[Y = y_j] \quad (3.1.9)$$

が成り立つ時、 X と Y は**独立な確率変数** (independent random variables) であるという。より一般に、 L 個の確率変数 $X^{(1)}, X^{(2)}, \dots, X^{(L)}$ があり、それぞれが $x_j^{(l)}$ の値をとるとき、

$$\begin{aligned} \mathbb{P}[X^{(1)} = x_i^{(1)} \text{ かつ } X^{(2)} = x_j^{(2)} \text{ かつ } \dots \text{ かつ } X^{(L)} = x_k^{(L)}] \\ = \mathbb{P}[X^{(1)} = x_i^{(1)}] \mathbb{P}[X^{(2)} = x_j^{(2)}] \dots \mathbb{P}[X^{(L)} = x_k^{(L)}] \end{aligned} \quad (3.1.10)$$

が成り立つなら、 $X^{(1)}, X^{(2)}, \dots, X^{(L)}$ は**独立な確率変数**であるという。

コイン投げの例で説明したように、「独立」な確率変数が表しているのは、それらの確率変数の間に関係（相関）がない、ということです。大数の法則や中心極限定理が成り立つのは、以下に述べるように、独立（または独立に近い）確率変数に対してです。

3.1.4 期待値と分散の基本的な性質

さて、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ことにより、考えている確率変数を、もっと簡単なものに分解して考えられるからです。まず、いつでも成り立つ性質として：

命題 3.1.2 (確率変数の期待値と分散) 一般の確率変数 X, Y と勝手な実数 a に対して、以下が成立する：

$$\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle \quad (3.1.11)$$

$$\langle aX \rangle = a \langle X \rangle \quad (3.1.12)$$

$$\text{Var}[aX] = a^2 \text{Var}[X] \quad (3.1.13)$$

(3.1.11) と (3.1.12) を**期待値の線型性**といいます。(3.1.13) も含めて、自然な性質だと思います。

また、 X, Y が**独立**の場合は、以下も成立します：

命題 3.1.3 (独立な確率変数の積の期待値と分散) 独立な確率変数 X, Y に対しては、以下も成り立つ：

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle \quad (3.1.14)$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \quad (3.1.15)$$

こちらの方は、**独立性**が非常に効いています（以下の証明参照）。

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ます（その具体例を 3.2.1 節で見ると予定です）。以下、これらの性質の証明を簡単に述べます。

(3.1.12) 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	\dots	x_n
aX のとりうる値	ax_1	ax_2	ax_3	\dots	ax_n
上のそれぞれの値をとる確率	p_1	p_2	p_3	\dots	p_n

となっています (ほとんどアタリマエのことですが、定義通りに期待値を計算するため、丁寧に書きました)。だから期待値の定義から

$$\langle aX \rangle = \sum_{i=1}^n p_i ax_i = a \sum_{i=1}^n p_i x_i = a \langle X \rangle \quad (3.1.16)$$

となって、(3.1.12) が証明できます。□

(3.1.13) の証明

まず、(3.1.12) を用いると

$$(aX - \langle aX \rangle)^2 = (aX - a \langle X \rangle)^2 = a^2 (X - \langle X \rangle)^2 \quad (3.1.17)$$

が得られるので、分散の定義から

$$\text{Var}[aX] = \langle (aX - \langle aX \rangle)^2 \rangle = \langle a^2 (X - \langle X \rangle)^2 \rangle = a^2 \langle (X - \langle X \rangle)^2 \rangle = a^2 \text{Var}[X]. \quad (3.1.18)$$

が得られます。真ん中の等号ではまたもや (3.1.12) を用いました。□

(3.1.11) 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, \dots, x_n , Y は y_1, y_2, \dots, y_m の値をとるとして、 $\mathbb{P}[X = x_i, Y = y_j] = p_{ij}$ と書くことにします。このとき、

$$\sum_j p_{ij} = \mathbb{P}[X = x_i], \quad \sum_i p_{ij} = \mathbb{P}[Y = y_j] \quad (3.1.19)$$

です (why?)。これを用いると²⁰,

$$\begin{aligned} \langle X + Y \rangle &= \sum_{i,j} p_{ij} (x_i + y_j) = \sum_{i,j} p_{ij} x_i + \sum_{i,j} p_{ij} y_j = \sum_i \left(\sum_j p_{ij} \right) x_i + \sum_j \left(\sum_i p_{ij} \right) y_j \\ &= \sum_i \mathbb{P}[X = x_i] x_i + \sum_j \mathbb{P}[Y = y_j] y_j = \langle X \rangle + \langle Y \rangle \end{aligned} \quad (3.1.20)$$

となって (3.1.11) が証明されました。 X, Y が独立でなくても (3.1.11) が成り立つことは、それほど当たり前でないかもしれませんが (特に、以下の分散の場合と対比すると)。□

(3.1.14) の証明

いま X, Y が**独立**だと仮定しているので、

$$p_{ij} = \mathbb{P}[X = x_i \text{ かつ } Y = y_j] = \mathbb{P}[X = x_i] \mathbb{P}[Y = y_j] \quad (3.1.21)$$

が成立しています。だから、

$$\langle XY \rangle = \sum_{i=1}^n \sum_{j=1}^m p_{ij} x_i y_j = \sum_{i=1}^n \sum_{j=1}^m \mathbb{P}[X = x_i] \mathbb{P}[Y = y_j] x_i y_j \quad (3.1.22)$$

が成立しますが、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

$$= \sum_{i=1}^n \mathbb{P}[X = x_i] x_i \sum_{j=1}^m \mathbb{P}[Y = y_j] y_j = \langle X \rangle \langle Y \rangle \quad (3.1.23)$$

²⁰最初のところ、1ステップ抜いてあります。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 $\langle X + Y \rangle$ を書く必要があります

となって証明されました。しつこいですが、上では確率が (3.1.21) のように積に分解した事 (**独立性の仮定**) が本質的だったことを重ねて注意しておきます。 □

(3.1.15) の証明

最後に (3.1.15) ですが、これは今までの式を動員すればできます。定義から

$$\text{Var}[X + Y] = \langle (X + Y - \langle X + Y \rangle)^2 \rangle = \langle (X - \langle X \rangle + Y - \langle Y \rangle)^2 \rangle \quad (3.1.24)$$

ですが、2乗のところを展開して

$$= \langle (X - \langle X \rangle)^2 + (Y - \langle Y \rangle)^2 + 2(X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \quad (3.1.25)$$

更に (3.1.11) と (3.1.12) を用いて和の期待値を期待値の和になおしたりすると

$$= \langle (X - \langle X \rangle)^2 \rangle + \langle (Y - \langle Y \rangle)^2 \rangle + 2\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \quad (3.1.26)$$

となります。ここで最初の2項はそれぞれ $\text{Var}[X]$ と $\text{Var}[Y]$ なので、独立でない場合も含めた**一般**の公式は

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \quad (3.1.27)$$

となることがわかりました。

ところが最後の項は、 $(X - \langle X \rangle)$ と $(Y - \langle Y \rangle)$ が**独立である**ことから²¹、(3.1.14) を用いて

$$\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle X - \langle X \rangle \rangle \langle Y - \langle Y \rangle \rangle = 0 \times 0 = 0 \quad (3.1.28)$$

となります。この最後のステップで**独立性の仮定**が本質的だったことを重ねて注意しておきます。 □

3.1.5 S_N などの期待値や分散の計算

コイン投げの時に出てきたように、期待値が μ 、標準偏差が σ であるような、独立同分布な確率変数 X_1, X_2, \dots, X_N に対して

$$S_N = \sum_{j=1}^N X_j \quad (3.1.29)$$

と定義します。今さっき証明した性質を用いて、 S_N などの期待値や分散を計算してみましょう。すでに述べたように、期待値は確率変数の**分布の中心**、分散 (正確には標準偏差) は**分布の拡がり**の目安を与えるから、この種の計算は直感的な理解のためにも重要です。

まず S_N の期待値を計算しましょう。期待値の線型性 (3.1.11) をくり返し使うと、

$$\begin{aligned} \langle S_N \rangle &= \langle X_1 + X_2 + X_3 \cdots + X_N \rangle = \langle X_1 \rangle + \langle X_2 + X_3 + \cdots + X_N \rangle = \cdots = \langle X_1 \rangle + \langle X_2 \rangle + \cdots + \langle X_N \rangle \\ &= \sum_{i=1}^N \langle X_i \rangle \end{aligned} \quad (3.1.30)$$

が成り立ちます。ここで仮定から $\langle X_i \rangle = \mu$ であるので、(3.1.30) から

$$\langle S_N \rangle = N \langle X_1 \rangle = N\mu \quad (3.1.31)$$

が得られます。次に S_N の分散ですが、 X_i が独立であるために (3.1.15) が使えます。(3.1.30) と同じノリで進むと

$$\text{Var}[S_N] = \text{Var}[X_1 + X_2 + \cdots + X_N] = \text{Var}[X_1] + \text{Var}[X_2] + \cdots + \text{Var}[X_N] \quad (3.1.32)$$

²¹ 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いました

となりますが, 仮定から $\text{Var}[X_i] = \sigma^2$ であるので, (3.1.32) から

$$\text{Var}[S_N] = N \text{Var}[X_1] = N \sigma^2 \quad (3.1.33)$$

が得られます.

(3.1.31) は「 S_N の分布は大体 $N\mu$ を中心にしていること」, また (3.1.33) は「その分布の拡がりは大体 $\sqrt{N}\sigma$ くらいであること」, を示唆しています²².

大事な性質なので, まとめておきます.

補題 3.1.4 (S_N の期待値と分散) 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, \dots

に対して, $S_N = \sum_{i=1}^N X_i$ を定義すると,

$$\langle S_N \rangle = N \langle X_1 \rangle = N\mu, \quad \text{Var}[S_N] = N \text{Var}[X_1] = N\sigma^2 \quad (3.1.34)$$

が成り立つ.

3.2 大数の法則

以上の準備の下に, 大数の弱法則を定式化, かつ証明できます.

定理 3.2.1 (大数の弱法則, Weak Law of Large Numbers, wLLN) 期待値が μ である**独立**かつ同分布な確率変数 X_1, X_2, \dots に対して, $S_N = \sum_{i=1}^N X_i$ を定義する. このとき, $\frac{S_N}{N}$ は μ に**確率収束**する (converges in probability). すなわち

$$\text{任意の正の数 } \epsilon \text{ に対して, } \lim_{N \rightarrow \infty} \mathbb{P} \left[\left| \frac{S_N}{N} - \mu \right| > \epsilon \right] = 0 \quad (3.2.1)$$

が成立する (上の式が確率収束の定義).

特に, X_i が有限な分散 σ^2 を持つ場合には, 任意の正の数 ϵ に対して,

$$\mathbb{P} \left[\left| \frac{S_N}{N} - \mu \right| > \epsilon \right] \leq \frac{\sigma^2}{\epsilon^2 N} \quad (3.2.2)$$

が成り立つ.

註:

- 上では X_i が同分布 (同じ試行の繰り返し) だとしてきましたが, 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証明されています. 文献の [5][6] などをご覧ください.
- 通常, 大数の弱法則という場合には, 定理の前半部分のみをいいますが, わかりやすい評価を与えるという意味で, 後半も書きました.

期待値の存在だけを仮定して定理を証明するのはそこそこ厄介なので, この講義では後半部分 (分散の存在を仮定) を証明することにします. 前半部分の証明には, 例えば文献 [6] の Theorem 5.2.2 や Theorem 6.4.3 の証明を参照してください (ごく簡単な概略は 3.4.4 節に載せました).

²²段々と見ていくように, 一番大事なのは \sqrt{N} の部分です. つまり, 1.1 節の言葉を借りると, 「 S_N の分布の拡がり \sqrt{N} のオーダーである」ことが重要なのです

なお、大数の強法則というのもあり、以下のように表現されます：

定理 3.2.2 (大数の強法則, Strong Law of Large Numbers, sLLN) 期待値が μ である独立かつ同分布な確率変数 X_1, X_2, \dots に対して, $S_N = \sum_{i=1}^N X_i$ を定義する. このとき, $\frac{S_N}{N}$ は μ に, 概収束する (converges with probability one, converges almost surely). すなわち,

$$\mathbb{P} \left[\lim_{N \rightarrow \infty} \frac{S_N}{N} = \mu \right] = 1 \quad (3.2.3)$$

である (上の式が概収束の定義).

上の定義で出てきた確率変数 $\lim_{N \rightarrow \infty} \frac{S_N}{N}$ は, 無限個の X_j を決めないと決まらない量です. このような「無限個の X_i を決めて初めて決まる量」に対する確率分布を考えることは自然ですが, 数学的にはそれほど簡単ではありません. なので, この講義では強法則にはこれ以上踏み込みません. ただし, 物理学への応用などを考えると, 強法則まで理解したいところではあります.

3.2.1 大数の弱法則の証明と解釈

大数の弱法則の証明には, 以下の不等式を用います.

定理 3.2.3 (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：

$$\mathbb{P}[|Y - \langle Y \rangle| > a] \leq \frac{\text{Var}[Y]}{a^2} \quad (a \text{ は勝手な正の数}). \quad (3.2.4)$$

この証明は後回しにして, まずは大数の弱法則を証明しましょう.

チェビシェフの不等式を $Y = \frac{S_N}{N}$, $a = \epsilon$ に対して適用してみます. 単に代入すると

$$\mathbb{P} \left[\left| \frac{S_N}{N} - \left\langle \frac{S_N}{N} \right\rangle \right| > \epsilon \right] \leq \frac{\text{Var}[\frac{S_N}{N}]}{\epsilon^2} \quad (3.2.5)$$

となるので, $\langle \frac{S_N}{N} \rangle$ と $\text{Var}[\frac{S_N}{N}]$ を計算します.

まず, $\frac{S_N}{N}$ については, 期待値の線型性 (3.1.12) と S_N の期待値に関する補題 3.1.4 から

$$\left\langle \frac{S_N}{N} \right\rangle = \frac{1}{N} \langle S_N \rangle = \frac{1}{N} \times N\mu = \mu \quad (3.2.6)$$

が出ます. 同様に, (3.1.13) と補題 3.1.4 から

$$\text{Var} \left[\frac{S_N}{N} \right] = \left(\frac{1}{N} \right)^2 \text{Var}[S_N] = \left(\frac{1}{N} \right)^2 \times N\sigma^2 = \frac{\sigma^2}{N} \quad (3.2.7)$$

が得られます. この 2 式は $\frac{S_N}{N}$ の分布の中心が μ 付近で, その拡がり $\frac{\sigma}{\sqrt{N}}$ くらいであることを示唆しています²³.

これらをチェビシェフの不等式 (3.2.5) に代入すると, 任意の正の数 ϵ に対して

$$\mathbb{P} \left[\left| \frac{S_N}{N} - \mu \right| > \epsilon \right] \leq \frac{\frac{\sigma^2}{N}}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 N} \quad (3.2.8)$$

が成立することがわかります. これで大数の弱法則が証明されました. \square

²³ここも 1.1 節の言葉では, 「 $\frac{S_N}{N}$ の分布の拡がり $\frac{1}{\sqrt{N}}$ のオーダーである」と言えます

確率の評価 (3.2.2) の解釈

大数の弱法則に現れている不等式 (3.2.2), (3.2.8) を解釈しましょう. 不等式の左辺は, 「 $\frac{S_N}{N}$ がその期待値 $\langle \frac{S_N}{N} \rangle = \mu$ から ϵ 以上ずれる確率」です. 不等式の右辺は, 「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」ことを主張しています.

これは我々の日常感覚と一致しています. コイン投げを例にとれば, N を大きくすると (たくさんの回数, コインを投げると), 表の出る割合がきっちり p になるとは言い切れないが, 表の出た割合は非常に高い確率で $\mu = p$ の近くに来る, とは言えるのです.

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て, 我々が日常考えていることを基礎づけることができました. 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ます. 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すが, 上の大数の法則はこの決め方に矛盾がない, ことを示しています.

3.2.2 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しますが, びっくりするくらい簡単です.

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ます (「マルコフの不等式」などと名前がついているのもあります). 少し一般に考えた方が原理がわかりやすいので, a, b を勝手な数 ($a > 0$) として

$$\mathbb{P}[|X - b| \geq a] \leq \frac{\langle (X - b)^2 \rangle}{a^2} \quad (3.2.9)$$

を証明し, 最後に $b = \langle X \rangle$ としましょう.

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

$$\langle (X - b)^2 \rangle = \sum_{i=1}^n p_i (x_i - b)^2 \geq \sum_{i:|x_i - b| \geq a} p_i (x_i - b)^2 \quad (3.2.10)$$

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

$$\geq \sum_{i:|x_i - b| \geq a} p_i a^2 = a^2 \sum_{i:|x_i - b| \geq a} p_i \quad (3.2.11)$$

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

$$= a^2 \mathbb{P}[|X - b| \geq a]. \quad (3.2.12)$$

両辺を a^2 で割ると (3.2.9) を得ます. 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ります. \square

3.2.3 確率変数の収束概念について (おまけ)

すでに見たように, 確率論では確率変数の収束を考えるのですが, その強さにはいくつかのレベルがあります. 参考までにまとめておきます.

確率変数の列 Y_n ($n = 1, 2, 3, \dots$) と, その収束先 (の候補) の確率変数 Z があるとします.

- Y_n が Z に**法則収束**または**分布収束**する (converges in law, converges in distribution) とは, Y_n と Z の累積分布関数をそれぞれ F_{Y_n} と F_Z と書いたときに

$$F_Z \text{ が連続な全ての } x \text{ において } \lim_{n \rightarrow \infty} F_{Y_n}(x) = F_Z(x) \quad (3.2.13)$$

となること (大まかに言って, Y_n の分布が Z のそれに収束すること) です.

- Y_n が Z に**確率収束**する (converges in probability) とは,

$$\text{全ての } \epsilon > 0 \text{ に対して, } \lim_{n \rightarrow \infty} \mathbb{P}[|Y_n - Z| > \epsilon] = 0 \quad (3.2.14)$$

が成り立つことです.

- Y_n が Z に**概収束**する (converges with probability one, converges almost surely) とは,

$$\mathbb{P}\left[\lim_{n \rightarrow \infty} Y_n = Z\right] = 1 \quad \text{または同値ですが} \quad \mathbb{P}\left[\lim_{n \rightarrow \infty} Y_n \neq Z\right] = 0 \quad (3.2.15)$$

が成り立つことです.

確率収束と概収束は大数の法則で既に登場しました. 中心極限定理では法則収束が出てきます.

これらの収束概念の強さには, 以下の関係があります:

$$\text{概収束} \implies \text{確率収束} \implies \text{法則収束} \quad (3.2.16)$$

すなわち, 概収束する場合は確率収束も法則収束もします. 確率収束する場合には概収束しない例がありますが, 法則収束はします. 法則収束は, まともな収束概念の中では一番「緩い」もので, これだけでは確率収束も概収束も保証できません. さらに詳しいことは, 文献の [5][6][7] などを参照してください.

3.3 中心極限定理

上では「大数の法則」を見たので, 今度はもう一つの重要な極限定理, 「中心極限定理」について述べます.

定理 3.3.1 (中心極限定理, Central Limit Theorem, CLT) 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, \dots を考える. 確率変数

$$Z_N = \frac{S_N - N\mu}{\sigma\sqrt{N}} \quad (3.3.1)$$

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**法則収束**または**分布収束**する (converges in law, converges in distribution). すなわち,

$$\lim_{N \rightarrow \infty} \mathbb{P}[a \leq Z_N \leq b] = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (3.3.2)$$

が成立する (上の式が法則収束の定義). グラフの面積で表現すれば, N が無限大になる極限では, 確率 $\mathbb{P}[a \leq Z_N \leq b]$ は,

$$\left(\text{グラフ } y = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \text{ と 3 直線 } x = a, x = b, y = 0 \text{ で囲まれた部分の面積} \right) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (3.3.3)$$

に収束するということだ.

註：

- 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**定理です。
- 大数の法則と同様に， X_i が**同分布でない**場合でも， X_i が適当な条件を満たせば中心極限定理が成り立つことが，色々研究されています（特に有名なのは Lindelöf の条件）．文献の [5][6] などをご覧ください。
- X_i の**独立性の仮定**を弱めるのはもっと微妙な，難しい問題です。「下手に弱めると中心極限定理が成り立たない」ことの一部を，5 節で見たいと思います。

この定理の証明は次の節に回し，この節ではまずコイン投げの場合に戻って，2.4 節で紹介したグラフの振る舞いを説明します．特に以下の質問に答えようと思います。

- 図 3 のグラフの横軸はなぜ，あのよう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？
- 図 3 のグラフの縦軸はなぜ，あのよう伸び縮みさせるのか？
（関連質問）なぜ確率が面積（積分）で表されるのか？
- このよう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？

その後，3.4 節で，より一般の場合も含めて，中心極限定理の証明（説明）を行います。

なお，数式を少し簡単にするため，また一般の中心極限定理との関係をつけるため，コイン投げの個々の X_j の期待値と分散を

$$\mu = \langle X_j \rangle = p, \quad \sigma^2 = \text{Var}[X_j] = p(1-p) \quad (3.3.4)$$

と書いて，以下では使います。

3.3.1 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，ということでしょう。これについてはある程度の解答を与えられます（縦軸の伸び縮みは次の小節で扱います）。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過程を図 5 に示します（いろいろな p の例があった方が良くと思って今度は $p = \frac{1}{5}$ としました）。

まず，図の解説をします。

- (1) 左上のグラフ：横軸は $m = \sum_{j=1}^N X_j$ ，縦軸は $\mathbb{P}[S_N = m]$ にとりました。 $N = 256$ (\diamond) は図の範囲外（右側）に分布の中心があるので ($Np = 51.2$)，よく見えていません。
- (2) 右上のグラフ：横軸は $m - N\mu$ ，縦軸は $\mathbb{P}[S_N = m]$ にとりました。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ました。これで分布の中心は大体そろいましたが， N が大きくなるにつれて高さは低く，幅は広がっています。
- (3) 左下のグラフ：横軸は $\frac{m - N\mu}{\sigma\sqrt{N}}$ ，縦軸は $\mathbb{P}[S_N = m]$ にとって，(2) のグラフを横軸方向に $\frac{1}{\sigma\sqrt{N}}$ に縮めました。幅はどの N でも同じようになってきましたが，高さがそろっていません。
- (4) 右下のグラフ：横軸は $\frac{m - N\mu}{\sigma\sqrt{N}}$ ，縦軸は $\mathbb{P}[S_N = m] \times \sigma\sqrt{N}$ にとって，(3) のグラフを縦方向に引き延ばしました。これで漸く，高さも幅もそろそろようになりました。実線は $y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ のグラフです。

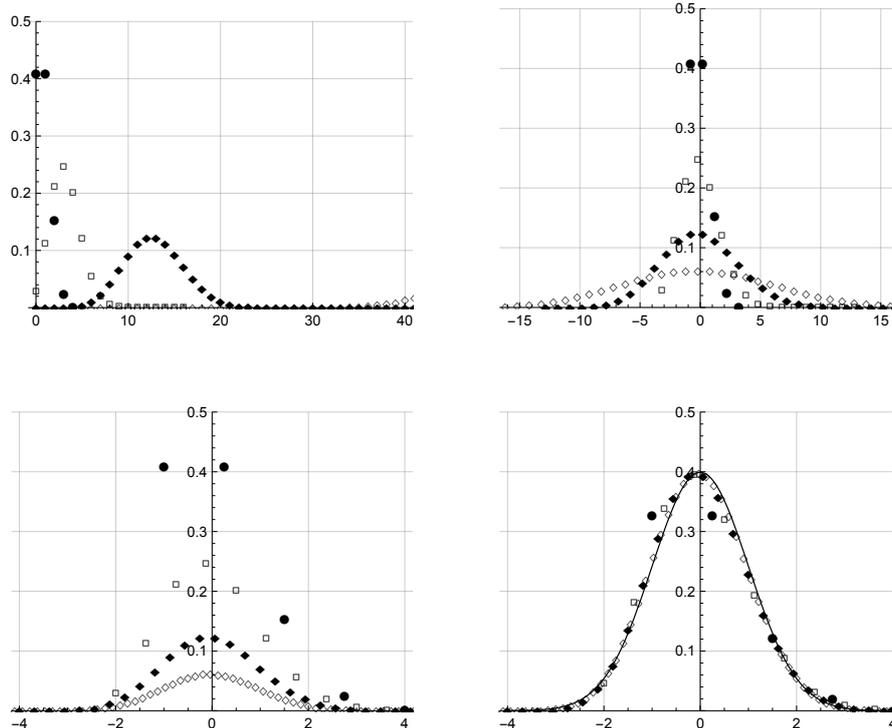


図 5: これらはどれも $p = \frac{1}{5}$ の時に, $N = 4$ (\bullet), 16 (\square), 64 (\blacklozenge), 256 (\circ) に対する $\mathbb{P}[S_N = m]$ の様子を图示したのですが, グラフによって, 縦軸と横軸の取り方が違います (詳細は本文).

最終段階の (4) のグラフを表現したのが中心極限定理でした. 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) を説明します.

まず, (1) から (2) への変換は大数の法則から示唆されます. つまり, $\frac{S_N}{N}$ が p の周りに集中していくわけですから, 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います. 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** だけのことです.

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

すでに, 確率変数 X の拡がりの目安は $\sqrt{\text{Var}[X]}$ であることを見ました. また, 大数の弱法則の証明のところで ((3.1.33) 参照), $\text{Var}[S_N] = p(1-p)N = N\sigma^2$ である事も見ました. S_N の標準偏差で見れば $\sqrt{p(1-p)N} = \sigma\sqrt{N}$ となっていて, S_N の拡がりが N とともに $\sigma\sqrt{N}$ くらいで増えていくことを示唆します (このところ, 図 5 の右上のグラフで確かめてください. 今は $p = \frac{1}{5}$ なので, $\sigma = \frac{2}{5}$ です). これを打ち消すように $\sigma\sqrt{N}$ で **横軸を割ってやれば, 幅が大体一定の分布** ができるでしょう. これが (2) から (3) への変換の理由です.

なお, この 2 つの変換を続けてやることは

$$Z_N = \frac{S_N - Np}{\sqrt{p(1-p)N}} = \frac{1}{\sigma} \times \frac{S_N - N\mu}{\sqrt{N}} \tag{3.3.5}$$

を定義するのと同じで, これは正に (2.4.1) です. いままで天下に定義していた Z_N の定義が少し身近に感じられたでしょうか? ここで分子の $S_N - N\mu$ は (1) から (2) への平行移動を, 分母の $\sigma\sqrt{N}$ は (2) から (3) への横軸方向の縮めを表しています.

註: 上の (3.3.5) の伸び縮みのファクター $\sigma\sqrt{N}$ の内, 一番大事なのは \sqrt{N} の部分です²⁴. 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ります. もう一方の $\sigma = \sqrt{p(1-p)}$ の方は, **異なる p でも同じ幅になるよ**

²⁴1.1 節の言葉では, \sqrt{N} のオーダーであることが大事

うに働いています. 別の言い方をすると, $\sigma = \sqrt{p(1-p)}$ は, **できた Z_N の分散をそろえるため**, つまり

$$\text{Var}[Z_N] = 1 \quad (3.3.6)$$

となるように, 導入したものです. (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ありません! その「自明でないことが起こる」のが中心極限定理の主張です.)

3.3.2 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りました. 今度は縦軸の方を考えましょう. 図 5 では (3) から (4) の過程ですが, 以下の議論のためには (2) から (4) を考える必要が出てきます. 結論から言うと, **グラフの縦軸は確率が曲線の下での面積で与えられるように決めています**. そこで, まず, 確率とグラフの下の面積の関係を有限の $N = 16$ を例にとって考えましょう.

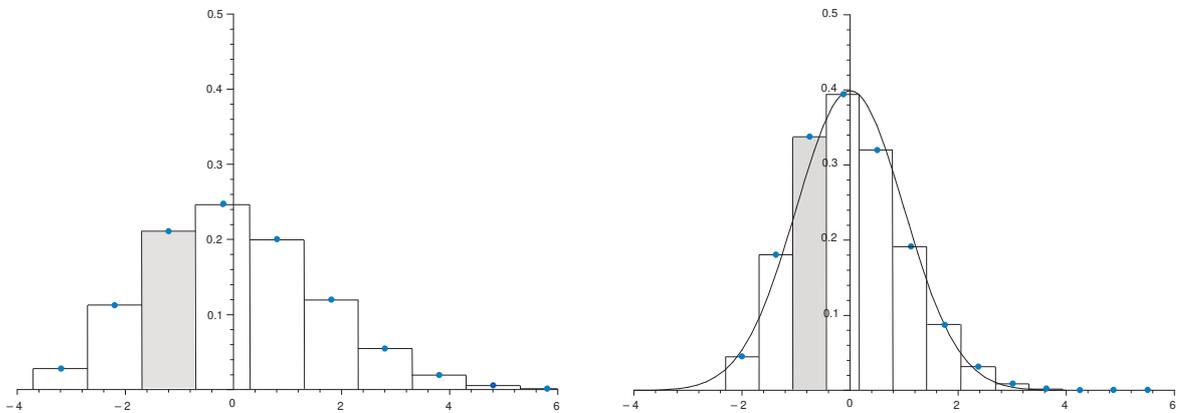


図 6: $\mu = p = \frac{1}{5}$, $N = 16$ のグラフ (本文中での説明のために, 従来の青点に加えて, ヒストグラムの形にしました).

左: 縦軸は $\mathbb{P}[S_N = m]$, 横軸は $m - N\mu$ (図 5 の (2)). スペースの関係で $m - N\mu \leq 6$ のみを図示しています.
右: 縦軸は $\mathbb{P}[S_N = m] \times \sigma\sqrt{N}$, 横軸は $(m - Np)/(\sigma\sqrt{N})$ (図 5 の (4)). 実線はいつも通り, $y = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ のグラフです.

図 6 の左には $\mathbb{P}[S_N = m]$ のグラフを, 横軸を $m - Np$ にとって描きました. これは図 5 の (2) そのものですが, 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ものです. 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**になります — 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.)

この考えに基づくと, ($A < B$ は勝手な整数) 確率

$$\mathbb{P}[A \leq S_N - Np \leq B] = \sum_{m=A}^B \mathbb{P}[S_N = m] \quad (3.3.7)$$

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ます. これが図の右ではどうなるかを次に考えましょう.

図 6 の右は左の図を, 「横軸は $1/(\sigma\sqrt{N})$ だけ縮め, 縦軸は $\sigma\sqrt{N}$ 倍に引き延ばした」ものです (今は $p = \frac{1}{5}$, $N = 16$ なので $\sigma\sqrt{N} = \sqrt{p(1-p)N} = \frac{8}{5}$ です). 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**対応する短冊に移ります. 例えば左で影を付けた長方形の面積は $\mathbb{P}[S_N = -1]$ でしたが, 対応する (右

でも影をつけた) 長方形の面積も $(\mathbb{P}[S_N = m] \times \sigma\sqrt{N}) \times (\sigma\sqrt{N})^{-1} = \mathbb{P}[S_N = m]$ となって, 面積は変わりません.

従って, (3.3.7) を表す面積は, 確率

$$\mathbb{P}\left[\frac{A}{\sigma\sqrt{N}} \leq Z_N \leq \frac{B}{\sigma\sqrt{N}}\right] = \sum_{m=A}^B \mathbb{P}[S_N = m] \quad (3.3.8)$$

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

これで漸く, この小節の問いに答えることができます. 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した. 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ます. 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きます. また, 短冊の幅もどんどん小さくなっていきます. ということは N が大きくなった極限では (3.3.8) 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す. 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す. 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す.

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した. 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, 幸運な事情です. 元々の確率変数のとりうる値が間隔 a で分布していたなら, 上で説明した「確率と面積」の解釈を少し変更する必要があります. またその結果, 縦軸の引き延ばしは $\sigma\sqrt{N}/a$ にすべきであることがわかります.

3.3.3 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, ということですが, これは非常に重要な問いですが, 余り簡単な答えを示すことはできません.

大学 3 年程度の数学 (例えば「特性関数の方法」) を駆使すれば完全な答えを与えられるので, これは 3.4.2 節で紹介しました.

また目標を少し下げて,

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, 3.5 節で説明 (完全な証明ではない) します.

3.4 中心極限定理の証明

この節では中心極限定理の証明をします.

3.4.1 中心極限定理の証明のいろいろ

中心極限定理の証明には, 何通りかの方法があります.

(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わけです. $\mathbb{P}[S_N = m]$ の中に出てくる ${}_N C_m$ の計算 (その定義の中の「階乗」の評価) が曲者です. これは「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ますが、かなり大変です。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)使えます。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(不可能ではないにせよ)ほとんど絶望的なのは、既に述べた通りです。

(2) 通常、大学の数学でやるのは「特性函数」というものを使って証明する方法です。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ですが、問題はこの方法を理解するには「フーリエ変換」などが必要になることです。大筋は、この後の3.4.2節にまとめました。

(3) 確率変数の分布を特徴づけるには、高次のモーメント(またはそれに関連した**キュムラント**と呼ばれる量)を全部計算すれば(大体)十分です。従って、 Z_N の高次のモーメントを計算し、(N が大きくなると)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ことを示す手もあります²⁵。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でしょう²⁶。「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納得できると思います。

(4) 統計力学や情報理論などで使われる「エントロピー」を使う方法もあります(最近の発展)。これも非常に面白い方法です。

3.4.2 特性函数による中心極限定理の証明(概要)

参考までに、中心極限定理の、特性函数を用いた証明の概略を説明します²⁷。既に述べた通り、定理は「期待値と分散が有限である独立同分布な確率変数」について成り立ちますが、ここでは証明を簡単にするため、

$$C_3 := \langle |X_i - \mu|^3 \rangle \quad \text{が有限である} \quad (3.4.1)$$

という仮定を付け加えて証明します。「フーリエ変換」「特性函数の収束と確率変数の収束の関係」および「ガウス積分」を認めれば、高校数学の範囲で理解できるように工夫しました。一般の場合の証明については、例えば最後の文献[5]のp.147や文献[6]のTheorem 6.4.4を参照してください(概略は3.4.4節に載せました)。

Step 1. 引数が純虚数の場合の指数函数

引数が純虚数の場合の指数函数を($i = \sqrt{-1}$ は虚数単位)、

$$\text{勝手な実数 } y \text{ に対して} \quad e^{iy} := \cos y + i \sin y \quad \text{と定義} \quad (3.4.2)$$

することにしましょう(オイラーの公式)。このように定義すれば(実部と虚部を別々に計算して)、この指数函数の導函数は

$$\frac{d}{dy} e^{iy} = \frac{d}{dy} (\cos y + i \sin y) = -\sin y + i \cos y = i e^{iy} \quad (3.4.3)$$

が成り立つこともわかります。(つまり、引数が実数の場合の指数函数の導函数の公式

$$\frac{d}{dy} e^{ay} = a e^{ay} \quad (3.4.4)$$

において、形式的に $a = i$ とおいたものが成立しています²⁸。)

なお、 e^{ix} を用いずに、 $\sin x, \cos x$ だけを用いて以下の議論を展開することも可能ですが、却って見通しが良くないと考えて、敢えて e^{ix} を導入しました。

²⁵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認していると(2)の「特性函数」の方法とほとんど同じになってしまいますが...

²⁶ただし、正規分布のモーメントの計算には、「ガウス積分」が必要になります

²⁷ところどころ、細かい条件は無視しています。例えば(3.4.6)は $f(x)$ がある種の条件を満たさないと成り立ちませんが、ここには書いていません

²⁸オイラーの公式に意味をつける(「なぜこの定義が良いのか」を考える)一つの方法は、このように「引数が一般の複素数でも、引数が実数の場合と同じような微分の表式が成り立つようにする」ことです

Step 2. フーリエ変換と逆変換

さて、上で定義した指数函数を用いて、一般の函数 $f(x)$ に対してその**フーリエ変換** (Fourier transform) $\hat{f}(k)$ を、

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(x) e^{ikx} dx \quad (3.4.5)$$

で定義します。ここで k は勝手な実数です。

上の式は f から \hat{f} を決める式ですが、驚くことに逆の式も成り立ちます。

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(k) e^{-ikx} dk \quad (3.4.6)$$

これを**フーリエ逆変換** (inverse Fourier transform) と呼びますが、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す。つまり、 $f(x)$ の代わりに $\hat{f}(k)$ を計算できれば、それから $f(x)$ を逆算できるのです。以上のフーリエ変換と逆変換の関係は、ここでは天下りで認めてください。**(天下り要素その 1)**

Step 3. 特性函数

一般の確率変数 X に対し、

$$\phi_X(k) = \langle e^{ikX} \rangle \quad (3.4.7)$$

を X の分布の**特性函数** (characteristic function) と呼びます。特に X の分布確率が、(3.1.1) のように、確率密度函数 $f(x)$ を持つ場合には、期待値の定義から

$$\phi_X(k) = \langle e^{ikX} \rangle = \int_{-\infty}^{\infty} e^{ikx} f(x) dx = \hat{f}(k) \quad (3.4.8)$$

となり、特性函数は (3.4.5) で定義したフーリエ変換 $\hat{f}(k)$ に一致します²⁹。さらに、かなり緩い条件の下で、

A. 二つの確率変数 X, Y の特性函数が一致すれば、 X, Y の分布は同じである

B. 確率変数の列 (Y_1, Y_2, Y_3, \dots) と確率変数 Z があり、 $\lim_{n \rightarrow \infty} \phi_{Y_n} = \phi_Z$ が成り立つなら、 (Y_n) は Z に法則収束する

ことも成り立ちます。これらが特性函数を用いる利点です。

中心極限定理の証明では Z_N の極限の分布を知りたいので、我々は Z_N の特性函数、すなわち $\phi_N(k) = \hat{f}_N(k) = \langle e^{ikZ_N} \rangle$ (の $N \rightarrow \infty$ での極限) を計算したいのです。これができれば上の 2 のフーリエ逆変換や 3 の B の性質を用いて、 Z_N の極限の分布を決められるはずですが。

Step 4. Z_N の特性函数の計算その 1

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」と以前に述べましたが、 $\langle e^{ikZ_N} \rangle$ についても正にその通りです。指数函数の性質 $e^{a+b} = e^a e^b$ をくり返し使う³⁰と (指数函数 e^a は $\exp(a)$ と書く； μ と σ は X_i の期待値と標準偏差)

$$e^{ikZ_N} = \exp\left(\sum_{i=1}^N \frac{ik(X_i - \mu)}{\sigma\sqrt{N}}\right) = \prod_{i=1}^N \exp\left(\frac{ik(X_i - \mu)}{\sigma\sqrt{N}}\right) \quad (3.4.9)$$

と積の形になります。ここで X_i が**互いに独立**であったことを思い出すと、 $\exp\left(\frac{ik(X_i - \mu)}{\sigma\sqrt{N}}\right)$ も互いに独立です。独立な確率変数の積の期待値は期待値の積に分解するので ((3.1.14))、結局

$$\phi_{Z_N}(k) = \hat{f}_N(k) = \left\langle \prod_{i=1}^N \exp\left(\frac{ik(X_i - \mu)}{\sigma\sqrt{N}}\right) \right\rangle = \prod_{i=1}^N \left\langle \exp\left(\frac{ik(X_i - \mu)}{\sigma\sqrt{N}}\right) \right\rangle = \left\langle \exp\left(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}}\right) \right\rangle^N \quad (3.4.10)$$

²⁹ただし、コイン投げのように、 Z_N が確率密度函数を持たない場合でも、特性函数 $\phi(k)$ は定義できます。これが特性函数の方法の凄いところでもあります

³⁰今は指数函数の肩が純虚数なので、この指数法則が成り立つかどうか、すぐには明らかではありません。しかし、指数法則が今の場合にもなりたつことは、オイラーの公式と三角函数の加法定理から簡単に証明できます

となります (最後のところでは, X_i の分布が同分布であることを用いました).

この段階でも, N が大きい時, 何か変なことが起こりそうな感じはします. 指数函数の中身では, 分母に \sqrt{N} があって, これが無限大に行けば, 指数函数の中身はゼロになります. つまり $N \rightarrow \infty$ の極限では指数函数が 1 に行くので, 期待値も 1 に行くことが期待されます. 一方で, 期待値全体が N 乗されているので, 期待値が 1 に行っても, その速さが遅ければ, 期待値の N 乗は無限大やゼロになるかもしれません.

この辺りの事情をしっかりと解析するため, 「テイラー展開」の概念を使います.

Step 5. テイラーの公式

一般に, 函数 f が $(n+1)$ -階微分可能の時, 以下の (残項が積分形の) **テイラーの公式** (Taylor's formula) が成り立ちます:

$$\begin{aligned} f(x) &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \int_a^x \frac{f^{(n+1)}(y)}{n!}(x-y)^n dy \\ &= \sum_{k=0}^n \frac{f^{(k)}(a)}{k!}(x-a)^k + \int_a^x \frac{f^{(n+1)}(y)}{n!}(x-y)^n dy \end{aligned} \quad (3.4.11)$$

この公式はこれ自身でも大変に面白く, かつ高校の知識でも導けるので, 3.4.3 節に証明を載せます.

ともかくこの公式を $f(x) = e^{ix}$ に対して, $n=2$, $a=0$ として用いると ((3.4.3) の導函数の表式を使う)

$$e^{ix} = 1 + ix - \frac{x^2}{2} - \frac{i}{2} \int_0^x e^{iy} (x-y)^2 dy \quad (3.4.12)$$

が得られます. ここで右辺最後の項については, 積分の性質³¹

$$a < b \quad \text{ならば} \quad \left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx \quad (3.4.13)$$

を用いて積分の中身を絶対値で置き換えて (まずは $x > 0$ とします. また $|e^{iy}| = 1$ を用いました)

$$\left| \int_0^x e^{iy} (x-y)^2 dy \right| \leq \int_0^x (x-y)^2 dy = \frac{x^3}{3} \quad (3.4.14)$$

が得られます. $x < 0$ の場合には, 上の右辺を $|x|^3/3$ に置き換えれば良いこともわかるので, 両方合わせて (3.4.12) から

$$e^{ix} = 1 + ix - \frac{x^2}{2} + g(x), \quad |g(x)| \leq \frac{|x|^3}{6} \quad (3.4.15)$$

が得られます. これが今の場合に特化させた, 指数函数のテイラーの公式です.

なお, (3.4.15) に類似した式は, 「テイラー展開」などと言わなくても, 高校のレベルで導出できる不等式 ($x > 0$)

$$x - \frac{x^3}{6} \leq \sin x \leq x, \quad 1 - \frac{x^2}{2} \leq \cos x \leq 1 - \frac{x^2}{2} + \frac{x^4}{24} \quad (3.4.16)$$

から導くことができます. ただ, 「テイラー展開」は一般的に成り立つ重要なものなので, 上では導入しました.

Step 6. Z_N の特性函数の計算その 2

以上を用いて (3.4.10) に出ている期待値を計算しましょう. (3.4.15) を用いて定義通りに書けば

$$\left\langle \exp\left(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}}\right) \right\rangle = \left\langle 1 + \left(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}}\right) - \frac{1}{2} \left(\frac{k(X_1 - \mu)}{\sigma\sqrt{N}}\right)^2 + g\left(\frac{k(X_1 - \mu)}{\sigma\sqrt{N}}\right) \right\rangle \quad (3.4.17)$$

となり, 期待値の線型性より, これは

$$= 1 + \frac{ik}{\sigma\sqrt{N}} \langle X_1 - \mu \rangle - \frac{k^2}{2\sigma^2 N} \langle (X_1 - \mu)^2 \rangle + \left\langle g\left(\frac{k(X_1 - \mu)}{\sigma\sqrt{N}}\right) \right\rangle \quad (3.4.18)$$

に等しい. ここで

$$\langle X_1 - \mu \rangle = \mu - \mu = 0, \quad \langle (X_1 - \mu)^2 \rangle = \sigma^2 \quad (3.4.19)$$

³¹ 被積分函数が実数の場合は高校でも習うはずですが, 被積分函数が複素数になっても, 同様に証明できます

に注目すると、第 2, 第 3 項はかんたんになります。また、最後の項に関しては、(3.4.15) での g の性質から

$$\left\langle g\left(\frac{k(X_1 - \mu)}{\sigma\sqrt{N}}\right)\right\rangle \leq \left\langle \frac{1}{6} \left| \frac{k(X_1 - \mu)}{\sigma\sqrt{N}} \right|^3 \right\rangle = \frac{|k|^3}{6\sigma^3 N^{3/2}} \langle |X_1 - \mu|^3 \rangle = \frac{C_3 |k|^3}{6\sigma^3 N^{3/2}} \quad (3.4.20)$$

となります。ここで $C_3 = \langle |X_1 - \mu|^3 \rangle$ は、仮定 (3.4.1) のおかげで、 X_1 の分布のみで決まる単なる定数です。

以上から、問題の期待値が

$$\left\langle \exp\left(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}}\right)\right\rangle = 1 - \frac{k^2}{2N} + h, \quad |h| \leq \frac{C_3 |k|^3}{6\sigma^3 N^{3/2}} \quad (3.4.21)$$

となっていることがわかり、(3.4.10) より特性関数が

$$\phi_{Z_N}(k) = \left\langle \exp\left(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}}\right)\right\rangle^N = \left(1 - \frac{k^2}{2N} + h\right)^N, \quad |h| \leq \frac{C_3 |k|^3}{6\sigma^3 N^{3/2}} \quad (3.4.22)$$

を満たすことがわかります。

この右辺の N -乗については

$$\left(1 - \frac{k^2}{2N} + h\right)^N = \left(1 - \frac{k^2}{2N}\right)^N \times \left(1 + \frac{h}{1 - \frac{k^2}{2N}}\right)^N = \left(1 + \frac{-k^2/2}{N}\right)^N \times \left(1 + \frac{1}{N} \frac{hN}{1 - \frac{k^2}{2N}}\right)^N \quad (3.4.23)$$

と変形してみましょう。 $N \rightarrow \infty$ の極限で、右辺の第一因子は $e^{-k^2/2}$ に行きます (数学 III で習うはずの公式

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a \quad (3.4.24)$$

を用いました)。また第 2 因子については、

$$\lim_{n \rightarrow \infty} \frac{hN}{1 - \frac{k^2}{2N}} = 0 \quad (3.4.25)$$

であるので、同様の理由によって、 $e^0 = 1$ に行きます。

以上から、最終的に、

$$\phi(k) := \lim_{N \rightarrow \infty} \phi_{Z_N}(k) = \lim_{N \rightarrow \infty} \left(1 - \frac{k^2}{2N} + h\right)^N = e^{-k^2/2} \quad (3.4.26)$$

が得られました。

Step 7.

Step 3 の B の性質、および「標準正規分布の特性関数が $e^{-k^2/2}$ であること³²⁾」を用いれば、上の Step 6 が完了した時点で「 Z_N の極限分布が標準正規分布である」こと、つまり中心極限定理が証明されました。

これで十分なのですがもう少し続けて、確率密度関数 $f(x)$ を求めようとする、以下ようになります。

$\phi(k)$ を、Step 2 の知識でもってフーリエ逆変換すれば、我々の知りたかった $f(x)$ がわかります。具体的には

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(k) e^{-ikx} dk = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx - k^2/2} dk \quad (3.4.27)$$

という積分をやればよろしい。これは**ガウス積分**というもので、大学初年度に習います (**天下り要素その 2**)³³⁾。結果だけを書くと

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (3.4.28)$$

となり、まさに、我々の欲しかった正規分布の確率密度関数が得られました。

³²⁾ただし、この事実の証明には以下で述べる「ガウス積分」が必要になるので、以下で述べるのと本質的には同じことです

³³⁾概要は以下の通り：まず、指数関数の肩について、 $-ikx - \frac{k^2}{2} = -\frac{1}{2}(k+ix)^2 - \frac{x^2}{2}$ と変形できることに注目すると、被積分関数の指数関数は $e^{-(k+ix)^2/2} \times e^{-x^2/2}$ となります。次に、(ix が純虚数であるにも関わらず) 普通の積分の変数変換のつもりで、 $u = k+ix$ とおきます。すると、問題の積分は $e^{-x^2/2} \int_{-\infty+ix}^{\infty+ix} e^{-u^2/2} du$ となり、さらに議論して最後の積分は $\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}$ であることがわかります。ただし、「なぜ $u = k+ix$ と変数変換して良いのか」「なぜそのような積分の値になるのか」などには大学レベルの理解が必要なので、ここに書いたことはあくまで大筋です

3.4.3 おまけ：テイラーの公式 (3.4.11) の証明

最後に、テイラーの公式 (3.4.11) の証明を述べます。大変に美しく汎用性の高い定理ですが、数学的帰納法を用いれば、高校数学の範囲内で十分に証明できます。

第 1 段. $n = 0$ および $n = 1$. (本来, $n = 0$ だけやれば十分ですが, 一般の n の場合への導入として, $n = 1$ もやります.)

$n = 0$ の場合は簡単で、微分と積分の関係式

$$f(x) - f(a) = \int_a^x f'(y) dy \quad (3.4.29)$$

の左辺から $f(a)$ を右辺に移行すればできます。

$n = 1$ の場合は、恣意的に $1 = \frac{d}{dy}(y-x)$ と思って、上の右辺を部分積分します：

$$\begin{aligned} \int_a^x f'(y) dy &= \int_a^x f'(y) \times \frac{d}{dy}(y-x) dy = \left[f'(y) \times (y-x) \right]_a^x - \int_a^x f''(y) \times (y-x) dy \\ &= f'(a) \times (x-a) + \int_a^x f''(y) \times (x-y) dy \end{aligned} \quad (3.4.30)$$

第 n 段. $(n-1)$ までできたとして、 n のとき。 $(n-1)$ までできたとすると、帰納法の仮定から

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (x-a)^k + \int_a^x f^{(n)}(y) \frac{(x-y)^{n-1}}{(n-1)!} dy \quad (3.4.31)$$

が成り立っています。またもや恣意的に

$$\frac{(x-y)^{n-1}}{(n-1)!} = -\frac{d}{dy} \frac{(x-y)^n}{n!} \quad (3.4.32)$$

だと思って、(3.4.31) の第 2 項を部分積分すれば終わりです。 \square

3.4.4 余分なことを仮定しない、大数の弱法則と中心極限定理の証明

今まで、大数の弱法則や中心極限定理を、定理の仮定よりも強い仮定の下で証明しました (大数の弱法則なら分散の存在, 中心極限定理なら 3 次のモーメントの存在, をそれぞれ仮定)。ここでは、それらの定理を余分な仮定なしで証明する方法の概略を述べます³⁴。(完全に原の趣味の反映なので、記述はかなり不親切です。)

まず、テイラーの公式の一つのバージョンとして、以下が成り立つことに注意する： $f(x)$ が $x = a$ で n -階微分可能であるとき、

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + o(|x-a|^n) \quad (3.4.33)$$

が成り立つ。ここで $o(|x-a|^n)$ とは

$$\lim_{x \rightarrow a} \frac{o(|x-a|^n)}{|x-a|^n} = 0 \quad (3.4.34)$$

となる函数である³⁵。

中心極限定理の証明には、(3.4.10) まではこれまでと同様に進む：

$$\phi_{X_N}(k) = \{g(k)\}^N, \quad g(k) := \left\langle \exp\left(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}}\right) \right\rangle \quad (3.4.35)$$

右辺にでてくる $g(k)$ については、

$$g(0) = 1, \quad g'(0) = \left\langle \frac{ik(X_1 - \mu)}{\sigma\sqrt{N}} \right\rangle = 0, \quad g''(0) = \left\langle \left(\frac{ik(X_1 - \mu)}{\sigma\sqrt{N}}\right)^2 \right\rangle = -\frac{k^2}{N\sigma^2} \quad (3.4.36)$$

³⁴そもそも、これらの定理の主張には期待値や分散しか出ていないですから、余分な仮定なしで証明できてもおかしくないわけです

³⁵証明には、例えば、杉浦光夫「解析入門 I」(岩波書店) の定理 2.12 を参照

がわかる³⁶ので, (3.4.33) より

$$g(k) = 1 - \frac{k^2}{2N\sigma^2} + o\left(\frac{k^2}{N\sigma^2}\right) \quad (3.4.37)$$

が結論できる. 従って

$$\phi_{X_N}(k) = \left[1 - \frac{k^2}{2N\sigma^2} + o\left(\frac{k^2}{N\sigma^2}\right)\right]^N \xrightarrow{N \rightarrow \infty} e^{-k^2/2} \quad (3.4.38)$$

を得る.

大数の弱法則の証明も同様で, S_N/N の特性関数を ψ として,

$$\psi(k) = \left\langle \exp\left(\frac{ikX_1}{N}\right) \right\rangle^N = \left[1 + \frac{ik\mu}{N} + o\left(\frac{ik}{N}\right)\right]^N \xrightarrow{N \rightarrow \infty} e^{ik\mu} \quad (3.4.39)$$

を得る. この右辺は, $Y \equiv \mu$ (確率変数は, $Y = \mu$ の値しかとらない恒等函数) である確率変数の特性函数だから, S_N/N は μ という定数に法則収束することが結論できる. しかし, 行き先が定数の場合は法則収束も確率収束も同じなので, これで大数の弱法則が証明された.

3.5 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 3.3.3 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, に部分的に答えてみます. ただし, 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³⁷ものになってしまいました. でも, 前小節で紹介した「特性函数の方法」とは少し違った視点としての存在意義はあると考え, 書くことにしました. なお, 以下の方法は「安定分布」「無限分解可能分布」などに関連した考えを, 我々の目的に合うように書き直したものです (例えば文献 [5] 参照).

この小節では, 以下の主張:

中心極限定理の Z_N の極限が $\lim_{N \rightarrow \infty} \mathbb{P}[a < Z_N < b] = \int_a^b f(z)dz$ の意味で確率密度函数 $f(z)$ 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) します. 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 $\lim_{N \rightarrow \infty} \mathbb{P}[a < Z_N < b] = \int_a^b f(z)dz$ が実際に成立すること」も示す必要がありますが, これはもつと道具立てが大仰になるので諦めました. そのまで理解するには 3.4.2 節で紹介している「特性函数の方法」などを用いるのが良いでしょう.

中心極限定理でてくる極限は,

$$Z_N = \frac{X_1 + X_2 + \cdots + X_N}{\sigma\sqrt{N}} \quad (3.5.1)$$

の ($N \rightarrow \infty$) の分布として出現しますが, 以下では σ で割るのをやめてしまって, もう少し一般に,

$$Y_N = \frac{X_1 + X_2 + \cdots + X_N}{\sqrt{N}} \quad (3.5.2)$$

の極限分布として現れるものを求めてみます³⁸. $N \rightarrow \infty$ の極限では (3.5.2) の Y_N の分布が

$$\lim_{N \rightarrow \infty} \mathbb{P}[a \leq Y_N \leq b] = \int_a^b f(y)dy \quad (3.5.3)$$

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

³⁶形式的に微分すれば出るが, そのように微分して良いかには少し吟味が必要

³⁷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). これでも「特性函数」を用いるよりは, 必要な道具立ては少ない

³⁸3.3.1 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ですから, 割らないのをやっておけば十分です

3.5.1 極限の確率変数の満たすべき性質は何か？

まず, (3.5.2) をもう少し変形します. M, N を (大きな) 勝手な正の整数とすると, (3.5.2) の定義から

$$Y_{M+N} = \frac{1}{\sqrt{M+N}} \sum_{i=1}^{M+N} X_i = \frac{\sqrt{M}}{\sqrt{M+N}} \frac{1}{\sqrt{M}} \sum_{i=1}^M X_i + \frac{\sqrt{N}}{\sqrt{M+N}} \frac{1}{\sqrt{N}} \sum_{i=M+1}^{M+N} X_i \quad (3.5.4)$$

と書けます. ここで, X_i ($i = 1, 2, \dots, M+N$) が独立・同分布ですから $\frac{1}{\sqrt{M}} \sum_{i=1}^M X_i$ 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³⁹. つまり, 標語的には

$$Y_{M+N} \stackrel{D}{=} \sqrt{\frac{M}{M+N}} Y_M + \sqrt{\frac{N}{M+N}} Y_N \quad (3.5.5)$$

と書けるはず⁴⁰. ここで $\stackrel{D}{=}$ 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す.

さて, ここで, $\frac{M}{M+N} = \epsilon$ ($0 < \epsilon < 1$) を一定にしたまま N と M を無限大にしましょう. Y_N の極限として出てくる確率変数を Y と書くと⁴⁰, Y_{M+N}, Y_M の 2 つも, 同じ確率変数 Y に行くでしょう. またこのとき

$$\sqrt{\frac{M}{M+N}} = \sqrt{\epsilon}, \quad \sqrt{\frac{N}{M+N}} = \sqrt{1-\epsilon} \quad (3.5.6)$$

です. 従って, (3.5.5) の両辺で $N \rightarrow \infty$ としたと思えば, 極限の Y は

$$Y \stackrel{D}{=} \sqrt{\epsilon} Y + \sqrt{1-\epsilon} Y \quad (\text{任意の } 0 < \epsilon < 1 \text{ に対して}) \quad (3.5.7)$$

を満たすはず⁴⁰. (3.5.7) 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きます.

3.5.2 極限の確率密度関数 $f(x)$ の満たすべき性質は何か? — 積分方程式

「 Y_N の極限が $f(x)$ に従う」と言うのがそもそもの我々の仮定だったので, 極限 Y の確率密度関数は $f(x)$ そのものになるはず⁴⁰. これを用いて, 上の関係式 (3.5.7) を Y の確率密度関数 $f(y)$ の言葉に焼き直しましょう. 左辺は Y だから, 今述べたことにより, その確率密度関数は $f(y)$ です. 右辺のは何でしょう? これには以下の 2 つの関係式を用います.

命題 3.5.1 (連続型確率変数の確率密度関数) (i) 確率変数 X, Y の確率密度関数をそれぞれ $f(x), g(x)$ とする. $Z = X + Y$ の確率密度関数を $h_1(z)$ とし, さらに X と Y は**独立**であると仮定すると,

$$h_1(z) = \int_{-\infty}^{\infty} f(x) g(z-x) dx \quad (3.5.8)$$

(ii) 確率変数 X の確率密度関数を $f(x)$ とし, 勝手な正の実数 α に対して, 新しい確率変数 $W = \alpha X$ を定義する. W の確率密度関数を $h_2(w)$ とすると, 勝手な正の実数 α に対して

$$h_2(w) = \frac{1}{\alpha} f\left(\frac{w}{\alpha}\right) \quad (3.5.9)$$

(i) の性質の証明には大学一年でやる「重積分」を使うのが自然です (高校数学でも理解できないわけではないですが, 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ます. これらの証明は略.

³⁹ $\frac{1}{\sqrt{M}} \sum_{i=1}^M X_i$ は Y_M そのものですが, $\frac{1}{\sqrt{N}} \sum_{i=M+1}^{M+N} X_i$ は X_{M+1} から X_{M+N} の和だから Y_N そのものではありません. しかし, X_i が独立・同分布なので, 分布は等しいはずだ, という主張です

⁴⁰ 中心極限定理的なものを仮定しているので, Y_{M+N}, Y_N, Y_M の極限のような確率変数が存在すると期待するのは自然です (が, 自明ではありません)

これを (3.5.7) の右辺に用います。まず上の (ii) から、 $\sqrt{\epsilon}Y$ の確率密度関数は $\frac{1}{\sqrt{\epsilon}}f\left(\frac{w}{\sqrt{\epsilon}}\right)$ であり、 $\sqrt{1-\epsilon}Y$ の確率密度関数は $\frac{1}{\sqrt{1-\epsilon}}f\left(\frac{w}{\sqrt{1-\epsilon}}\right)$ であることがわかります。次に、(i) を用いると、 $\sqrt{\epsilon}Y + \sqrt{1-\epsilon}Y$ の確率密度関数が

$$h(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\epsilon}}f\left(\frac{x}{\sqrt{\epsilon}}\right) \frac{1}{\sqrt{1-\epsilon}}f\left(\frac{z-x}{\sqrt{1-\epsilon}}\right) dx = \int_{-\infty}^{\infty} f(y) \frac{1}{\sqrt{1-\epsilon}}f\left(\frac{z-\sqrt{\epsilon}y}{\sqrt{1-\epsilon}}\right) dy \quad (3.5.10)$$

であることがわかります (第 2 の等号は $x = \sqrt{\epsilon}y$ と置換積分して得られました)。ところが (3.5.7) によれば、この分布が $f(z)$ に等しいはずですが、つまり (少し式を整理して) $f(x)$ は

$$f(z) = \frac{1}{\sqrt{1-\epsilon}} \int_{-\infty}^{\infty} f(y) f\left(\frac{1}{\sqrt{1-\epsilon}}z - \frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}y\right) dy \quad (\text{すべての } z \text{ と } 0 < \epsilon < 1 \text{ に対して}) \quad (3.5.11)$$

を満たすべきである、という結論になりました。我々はこのような $f(x)$ を求めたいのです。

3.5.3 極限の確率密度関数 $f(x)$ の満たすべき性質は何か? — 微分方程式

後は (3.5.11) を解けば良いのですが、これは高校数学では簡単に解けません。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すが、これをやると 3.4.2 節の「特性関数の方法」と同じになってしまいます。そこで、ここでは「特性関数の方法」を使わず、敢えて泥臭い方法でやってみます。

(3.5.11) の右辺は f が 2 つ出てきて積分だから始末が悪いわけです。幸い、(3.5.11) 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ましょう (ここで「テイラー展開」の考えを使いました)。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す⁴¹。

ϵ の級数としてみたとき、左辺はそのまま $f(z)$ です。右辺はと言うと、

$$\sqrt{1-\epsilon} = 1 - \frac{1}{2}\epsilon + O(\epsilon^2), \quad \frac{1}{\sqrt{1-\epsilon}} = 1 + \frac{1}{2}\epsilon + O(\epsilon^2) \quad (3.5.12)$$

を用いて強引に

$$\frac{1}{\sqrt{1-\epsilon}}z - \frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}y = z + \frac{1}{2}\epsilon z - \sqrt{\epsilon}y + O(\epsilon^{3/2}) \quad (3.5.13)$$

とあって、 $f\left(\frac{z-\sqrt{\epsilon}y}{\sqrt{1-\epsilon}}\right)$ 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)、

$$f\left(\frac{1}{\sqrt{1-\epsilon}}z - \frac{\sqrt{\epsilon}}{\sqrt{1-\epsilon}}y\right) = f(z) + f'(z)\left(\frac{1}{2}\epsilon z - \sqrt{\epsilon}y\right) + \frac{f''(z)}{2}\epsilon y^2 + O(\epsilon^{3/2}) \quad (3.5.14)$$

を得ます⁴²。これを (3.5.11) に代入して整理すると

$$\begin{aligned} f(z) &= \left[1 + \frac{1}{2}\epsilon + O(\epsilon^2)\right] \int_{-\infty}^{\infty} f(y) \left[f(z) + f'(z)\left(\frac{1}{2}\epsilon z - \sqrt{\epsilon}y\right) + \frac{f''(z)}{2}\epsilon y^2 + O(\epsilon^{3/2}) \right] dy \\ &= \left(1 + \frac{\epsilon}{2}\right) f(z) \int_{-\infty}^{\infty} f(y) dy + \frac{f'(z)}{2}\epsilon z \int_{-\infty}^{\infty} f(y) dy - f'(z)\sqrt{\epsilon} \int_{-\infty}^{\infty} y f(y) dy \\ &\quad + \frac{f''(z)}{2}\epsilon \int_{-\infty}^{\infty} y^2 f(y) dy + O(\epsilon^{3/2}) \end{aligned} \quad (3.5.15)$$

となります。ここで、

$$\int_{-\infty}^{\infty} f(y) dy = 1 \quad (3.5.16)$$

⁴¹ 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です

⁴² (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」と疑問を感じた方、その通りです! 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です。しかし、もう少し考えようやってもよいことがわかります。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、 $f(y)$ は最終的には e^{-y^2} のような関数だとわかります。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ありません。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 $O(\epsilon^2)$ とみなしても良いのです。

一応、このように理屈をつけたいけれども、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のは事実ですね。もう少し注意深く議論すれば、何も問題ないことは証明できます

であることに注意します (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 であるべきだ). また,

$$\int_{-\infty}^{\infty} y f(y) dy = 0 \quad (3.5.17)$$

でしょう (この量は Y の期待値で, もととの 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

$$\int_{-\infty}^{\infty} y^2 f(y) dy = \alpha \quad (3.5.18)$$

とおくと (この量 α は何かの正の値のはず),

$$f(z) = \left(1 + \frac{\epsilon}{2}\right) f(z) + \frac{f'(z)}{2} \epsilon z + \frac{f''(z)}{2} \epsilon \alpha + O(\epsilon^{3/2}) \quad (3.5.19)$$

となるので, ϵ の係数を比較して,

$$f(z) + z f'(z) + \alpha f''(z) = 0 \quad (3.5.20)$$

が得られます. つまり, 極限の確率密度関数 $f(z)$ 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りました.

3.5.4 $f(x)$ の微分方程式を解く

最後に (3.5.20) の微分方程式を解けば, 話は完結します. この方程式を綺麗に解くことは高校の範囲を超えています, 大体の大筋を説明します⁴³.

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」やり方があります:

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \quad (3.5.21)$$

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す. (そして最後に級数展開が実際にできることを確かめます).

これをやってみましょう. (3.5.21) を (3.5.20) に代入すると

$$0 = f(z) + z f'(z) + \alpha f''(z) = \sum_{n=0}^{\infty} a_n z^n + z \times \frac{d}{dz} \left[\sum_{n=0}^{\infty} a_n z^n \right] + \alpha \frac{d^2}{dz^2} \left[\sum_{n=0}^{\infty} a_n z^n \right] \quad (3.5.22)$$

となりますが, 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ありません. 大学一年生で詳しくやるところです. 今はこのような問題には目をつぶります)

$$\begin{aligned} &= \sum_{n=0}^{\infty} a_n z^n + z \times \sum_{n=0}^{\infty} a_n n z^{n-1} + \alpha \sum_{n=0}^{\infty} a_n n(n-1) z^{n-2} = \sum_{n=0}^{\infty} a_n (n+1) z^n + \alpha \sum_{n=0}^{\infty} a_n n(n-1) z^{n-2} \\ &= \sum_{n=0}^{\infty} \left[a_n (n+1) + \alpha a_{n+2} (n+2)(n+1) \right] z^n \end{aligned} \quad (3.5.23)$$

となります (最後の行では級数の添え字を付け替えました). 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 であるべきです. つまり, すべての $n \geq 0$ に対して

$$a_n (n+1) + \alpha a_{n+2} (n+2)(n+1) = 0 \quad \implies \quad a_{n+2} = -\frac{1}{\alpha(n+2)} a_n. \quad (3.5.24)$$

⁴³以下では解の形がわからない時にそれを見つける方法を示します. 解の形 (3.5.31) を最初から知っているなら, 以下の議論の大半は無視して, (1) これが問題の微分方程式を満たすこと, (2) 微分方程式の一般論 (特に解の存在と一意性の定理の成立条件) を満たすためにこれ以外の解がないこと, を示せば十分です

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ます. n の小さい方から順番にやれば良く, 結果は $n \geq 0$ に対して

$$a_{2n} = \left(\frac{-1}{2\alpha}\right)^n \frac{1}{n!} a_0, \quad a_{2n+1} = \left(\frac{-1}{\alpha}\right)^n \frac{1}{(2n+1)!!} a_1 \quad (3.5.25)$$

となります. ここで $(2n+1)!! = (2n+1) \cdot (2n-1) \cdot (2n-3) \cdots 5 \cdot 3 \cdot 1$ であり, $0! = 1$ と解釈します.

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ります. そこで, 偶数・奇数毎にこの結果をまとめて

$$f_{\text{even}}(z) = \sum_{n=0}^{\infty} \left(\frac{-1}{2\alpha}\right)^n \frac{1}{n!} z^{2n} = \sum_{n=0}^{\infty} \left(\frac{-z^2}{2\alpha}\right)^n \frac{1}{n!} = \exp\left(\frac{-z^2}{2\alpha}\right) \quad (3.5.26)$$

$$f_{\text{odd}}(z) = \sum_{n=0}^{\infty} \left(\frac{-1}{\alpha}\right)^n \frac{1}{(2n+1)!!} z^{2n+1} \quad (3.5.27)$$

を定義すると ((3.5.26) 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,

$$f(z) = a_0 f_{\text{even}}(z) + a_1 f_{\text{odd}}(z) \quad (3.5.28)$$

と書けることがわかります⁴⁴.

上で定義した $f_{\text{even}}(z)$ は正に求める形をしているのでメデタイわけです⁴⁵. 問題は $f_{\text{odd}}(z)$ の方であって, 問題の条件を満たさないなどの理由で $f_{\text{odd}}(z)$ を排除できたら話は完結します.

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

$$f_{\text{odd}}(z) = \exp\left(-\frac{z^2}{2\alpha}\right) \int_0^z \exp\left(\frac{t^2}{2\alpha}\right) dt \quad (3.5.29)$$

となっています. (この等式を証明するには, 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ります. 特に, (1) $f_{\text{odd}}(z)$ は z の奇函数で, z が正なら $f_{\text{odd}}(z)$ の値も正, (2) $z \rightarrow \infty$ で $f_{\text{odd}}(z)$ は $O(\frac{1}{z})$ くらい, などがわかります. このどちらの性質も

$$\int_{-\infty}^{\infty} y f(y) dy = 0 \quad (3.5.30)$$

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

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

$$f(z) = a_0 \exp\left(\frac{-z^2}{2\alpha}\right) \quad (3.5.31)$$

という結論です. メデタシメデタシ. (なお, a_0 の値そのものは $\int_{-\infty}^{\infty} f(y) dy = 1$ であるべきことから $a_0 = \frac{1}{\sqrt{2\pi\alpha}}$ と決まりますが, 本質ではないので省略します.)

⁴⁴ 「わかります」と書きましたが, これ以外に解がないこと, またそもそも f_{even} や f_{odd} の級数がきちんと収束して定義できていること, は証明すべきです. 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」をやれば納得できます. また, 級数の存在なども大学一年程度の数学を用いれば納得できます.

⁴⁵ α が決まっていなくてもいいかと思いますが, 今は分散の値を決めていない (3.5.2) をやっているのだから, この自由度は残って当然なのです. 分散の値を指定すると, α も決まります.

4 ランダムウォーク

この節では統計力学への橋渡しとして、ランダムウォークを考えます。

4.1 1次元ランダムウォーク

以下のような問題を考えます：

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, random walk）と呼ばれます。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しました。

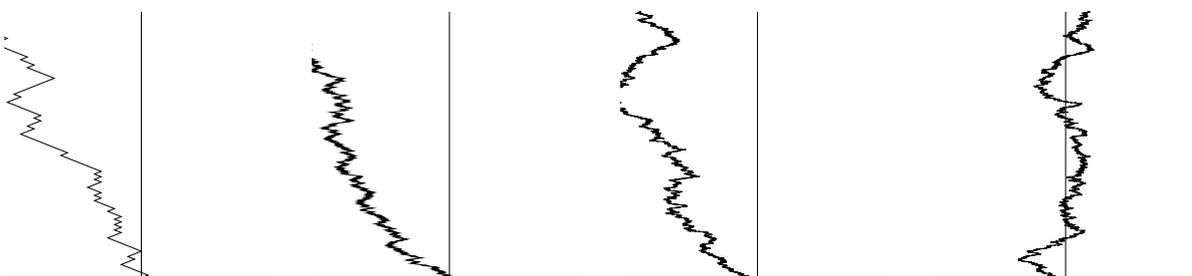


図 7: 1次元ランダムウォークの例：左から全歩数が $N = 10^2, 10^3, 10^4, 10^5$ steps の場合。横軸は x ，縦軸は歩数で、図示している範囲は縦軸は 0 から N ，横軸は $\pm 2\sqrt{N}$ 。

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ます。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す。

まず、 i 歩めの数直線上の動きを X_i で表します：

$$X_i = \begin{cases} 1 & (\text{右に動いたとき}) \\ -1 & (\text{左に動いたとき}) \end{cases} \quad (4.1.1)$$

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で、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、そして、 N 歩めの位置は $S_N = \sum_{i=1}^N X_i$ となっています。

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同様で、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 つだけです。(1) の違いにより、今回は

$$\langle X_i \rangle = 0, \quad \text{Var}[X_i] = 1 \quad (4.1.2)$$

となります。

従って、大数の弱法則の証明 ((3.1.33) 付近) を思い出すと、

$$\langle S_N \rangle = 0, \quad \text{Var}[S_N] = N \quad (4.1.3)$$

と結論できます。

これは何を言っているのでしょうか？まず、ランダムウォークの場合の $\text{Var}[S_N]$ の意味を考えます。 $\langle S_N \rangle = 0$ だから、分散の定義から

$$N = \text{Var}[S_N] = \langle |S_N|^2 \rangle \quad (4.1.4)$$

です。 $|S_N|$ は N 歩後の位置の、原点からの距離そのものですから、上の分散には「原点からの距離の 2 乗の平均」という意味がつきます（このため、この量を**平均 2 乗変位** (mean square displacement) とよびます）。つまり、(4.1.3) や (4.1.4) は「原点からの距離の 2 乗の平均は N である」と言っているわけです。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}** と思っても良いでしょう⁴⁶。これがランダムウォークの著しい特徴です⁴⁷。

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ます。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わけです。そして「典型的」な場合には、原点から $\sqrt{\text{Var}[S_N]} = \sqrt{N}$ 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す。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**した結果です。この事情は「典型的」ないくつかの例を見ると、より良く納得できます。(図 7 参照)。

4.2 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ます。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、ことを考えましょう。(一歩目の後は、今いる位置を出発点と同じ事をくり返します。) この時、 N 歩の後にはこの人はどの付近にいますでしょうか？(このモデルは 2 次元単純ランダムウォークと呼ばれます。) 図 8 に 2 次元単純ランダムウォークの例を示しました。

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ます。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

$$\langle (N \text{ 歩後の原点からの距離})^2 \rangle = N \quad (4.2.1)$$

が成り立ちます。

実はアインシュタインの関係式、つまり (4.2.1) 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ります。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 — 概要は変わりません：この場合は適当な定数 c を用いて

$$\langle (N \text{ 歩後の原点からの距離})^2 \rangle \approx cN \quad (4.2.2)$$

と書けます。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ます。

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ます。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ます。

⁴⁶ $\langle |X|^2 \rangle \geq \langle |X| \rangle^2$ ですが、逆向きの不等式は一般にはなりたちません。特に、 $\langle |X|^2 \rangle$ は非常に大きいのに $\langle |X| \rangle^2$ は小さいような X はいくらでも作れます。従って、 $\langle |X|^2 \rangle = N$ 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ません。これが「大体」と書いた理由です。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ます。

⁴⁷なお、(4.1.3) や (4.1.4) 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、 $\langle |S_N|^2 \rangle = N$ は「アインシュタインの関係式」と呼ばれています。(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ですが、本質は同じ。) 更に余談ですが 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ました。そのため、2005 年は、これら 3 大論文を記念して「国際物理年」になっていました。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ました。

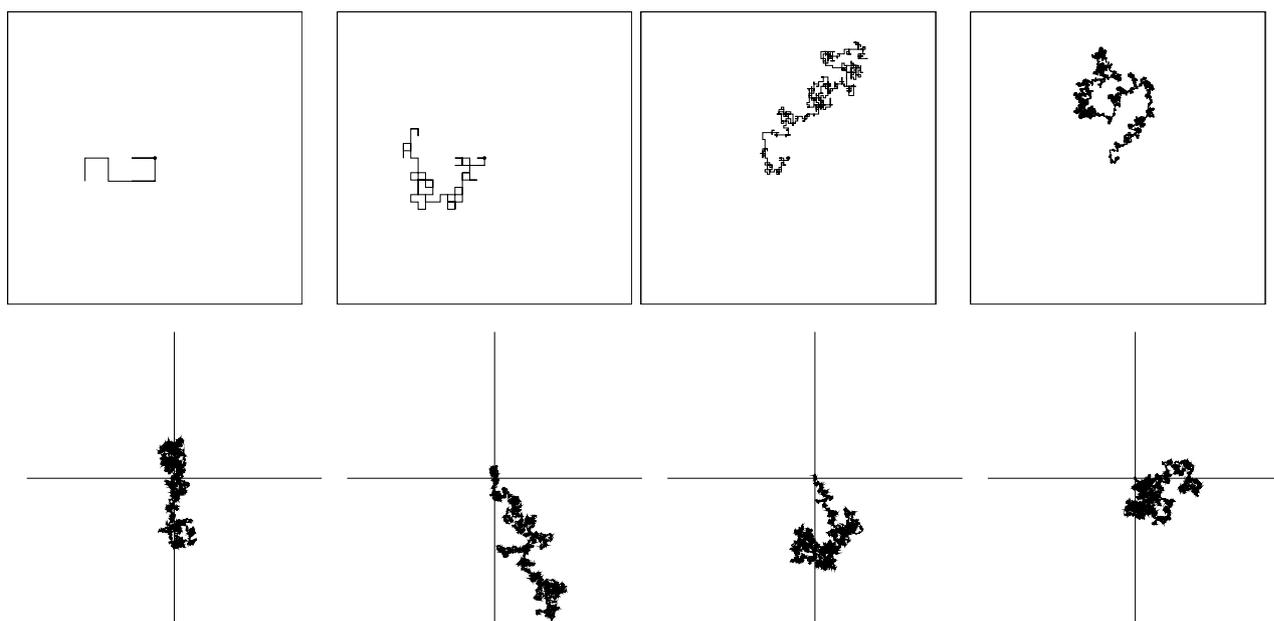


図 8: 2次元ランダムウォークの例:1行目は左から $N = 10, 10^2, 10^3, 10^4$ steps, 2行目は左から $10^5, 10^6, 10^7, 5 \times 10^7$ steps までの軌跡を表す. 図示している範囲は $\pm 2\sqrt{N}$ で, 原点 (出発点) に小さな黒丸をうってある.

5 臨界現象へ

今まで考えてきたのは, 数学的には解決された問題です. その中心にあったのはたくさんの**独立な**確率変数の和でした. そしてこの場合 (N 個の確率変数の和を S_N と書く), **独立性のおかげで $\text{Var}[S_N]$ が大体 N くらいになる**⁴⁸という大きな特徴が見られました. Var の平方根 (つまり標準偏差) がその量のバラツキの目安になることを思い出すと, これは「 N この独立な確率変数の和は \sqrt{N} くらいのバラツキをもっている」とまとめることができます. そして, 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「 N この和のバラツキは \sqrt{N} 」から来ている, と言っても過言ではありません.

また, 「独立性」が鍵でしたから, 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ます. 実際, 今までも少し述べたように X_i が完全には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ます. 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ます. この意味で, 中心極限定理は非常に**普遍的**に成り立つものと言えます.

ところが, 「独立性」の仮定がある程度以上破られると, 問題は一気に難しくなります. 一見なんでもありの世界になってしまいそうですが, そのような場合でも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す. そしてまた, このテーマに対する, 非常に面白くかつ具体的な問題が, **統計力学における臨界現象**という形で現れてきます.

この公開講座の最後では統計力学における臨界現象に焦点をあてて, この未解決問題を解説したいと思います. より具体的には, 以下の (一見ことなる) 2つの問題を考えます.

(1) 中心極限定理の拡張に関して (磁石のモデルの臨界現象):

中心極限定理は「独立な」確率変数に対して成り立つ, ということは何度も強調した通りです. 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) を作り出せました.

では, 独立でない場合はどうなるのでしょうか? 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

⁴⁸1.1 節の言葉では $\text{Var}[S_N]$ が N の「オーダー」であるということ

じであり、 \sqrt{N} で割ると行き先は正規分布になります。しかし、独立性がもっと破られている場合は未解決です。この場合、**中心極限定理に類似の結果**⁴⁹が成り立つと期待されていますが、詳細はよくわかっていません。

さて一方、このような典型的な振る舞いを示す例が、物理学に（日常生活に）存在します。それは磁石です。ご存知のように、永久磁石を熱していくと、ある温度（キュリー点）以上では鉄を引きつけなくなります（磁石ではなくなる）。しかし（少し工夫しつつ）温度を下げてくるとキュリー点より下ではまた磁石に戻ります。この現象を磁石になっている「相」と磁石でない「相」の間の**相転移**と呼びます。更に、このキュリー点付近では「臨界現象」と呼ばれる、普遍的な現象が見られることもわかっています。

(2) ランダムウォークの拡張に関して（高分子の臨界現象）：

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ましょう（このような条件をつけたモデルは self-avoiding walk（自己回避酔歩）と言われます）。「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です（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す。

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す。

言葉の定義

d -次元の正方格子とは、 d -次元の空間のなかで、座標の各成分が整数値をとっているような点の全体をいいます。 $d = 1$ なら数直線上の整数の点の全体、 $d = 2$ なら平面上の点で、 x, y 座標がともに整数の点の全体、 etc. となります。

このとき、 d -次元正方格子の点の一つ一つを**サイト**、隣り合ったサイトのペア（つまり、隣り合ったサイトをつなぐ線分）を**ボンド**といいます。

5.1 臨界現象とは

では、臨界現象とは何なのか、上に簡単に述べたランダムウォークや磁石のモデルを説明しましょう。以下では d -次元正方格子の上でモデルを考えていきます。

5.1.1 自己回避ランダムウォーク (SAW)

今まで考えたランダムウォークは、過去の履歴に関係なく、座標軸の方向に一步ならどこでも動けるものでした。これは空気中の分子の運動などを表すには自然な設定です。しかし、ここで問題を少し変えて、溶液中に長い線のような高分子が入っている場合を考えてみます。高分子を形作っている分子や原子は溶液の分子にぶつかられてあちこち動きます。結果として、高分子は非常にギザギザに折曲がりながら、溶液中を漂うことになるでしょう。そこで問題は、高分子がどのくらいギザギザしているのか、また、まっすぐのぼした時の高分子の長さ、溶液中の長さにはどのような関係があるのか、ということです。

このような問題も、ランダムウォークをもとにしてモデル化することができます。すなわち、今まで考えてきたランダムウォークの軌跡（つまり、時刻ゼロから粒子の動いた跡）を見るのです。これは一般に非常にギザギザな線になっているでしょうから、高分子の問題にぴったりでしょう。ただし、これだけでは大きな欠陥があります。溶液中の高分子は、決して**自分自身とは交わりませんが**、いままで考えてきたランダムウォークでは、軌跡は自分自身と交わることも許しています。

そこで、高分子を考えるために、今までの問題を以下のように変形しましょう。

⁴⁹ただし、 \sqrt{N} ではなくて N^α などで割る必要があるでしょう（ここで、 α は未知の、 $\frac{1}{2}$ より大きな定数。また割って作った確率変数も正規分布にはならないでしょう）

今まで考えてきたランダムウォークの問題において、「その軌跡が自分自身と交わってはいけない」の条件を付加してみる。(このようなランダムウォークを、**Self-Avoiding Walk (SAW)** という。そして、 N -ステップの SAW の端から端までがどのくらいの距離になっているか (つまり、平均 2 乗変位がどのくらいのオーダーか) を考えたい。

新しく付け加えた条件は、一見、大したものではないように見えます。しかし、「軌跡が自分自身と交わってはいけない」ことを保証するには、各ステップが独立ではいけません (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いますから)。更に、過去の自分と交わらないためには、過去の履歴をすべて覚えておく必要があります。ウォークが長くなればなるほど、自分の過去も長くなりますから、この条件は大変になってきます。こんなわけで、問題が非常に難しくなるのです。

この問題は 1950 年代には知られていましたが、見かけによらず難問でした。現在でも、完全に解決されている訳ではありません。ただし、その後の研究によって、以下のようなことが段々とわかってきました。考えている空間の次元を d 、原点から出発する n -ステップの SAW の数を c_n 、 n -ステップの SAW の平均二乗変位を $(\ell_n)^2$ と書きます (ℓ_n が n -ステップの平均の拡がりを表します)。すると以下が証明されます。

- 次元に依存する定数 μ があって、 $n \rightarrow \infty$ では $c_n \approx \mu^n$ と書ける。(より正確には、 $\lim_{n \rightarrow \infty} (c_n)^{1/n} = \mu$)。
- 更に、 $\mu^n \leq c_n \leq \mu^n e^{n^\alpha}$ がなりたつ。ここで α は 1 より小さい定数である

以上は割合早く、1960 年代半ばまでには証明されていました。ところが、これ以上の厳密な結果となると、20 年近くも進歩が止まってしまいました。その間、物理学者、化学者の努力により、数学的には厳密でないながらも、以下のような予想がたてられました。

- a. 実際には、ある定数 γ, ν があって、 $n \rightarrow \infty$ で $c_n \approx \mu^n n^{\gamma-1}$ および $\ell_n \approx n^\nu$ が成り立っているだろう。(この γ, ν を **臨界指数** とよびます。)
- b. 定数 μ はモデルを少し変える (例: 1 ステップで遠くまで跳べるようにする、など) と値が変わる。しかし、 γ, ν は考えている次元のみで決まる非常に安定な量だろう。(これを **臨界指数の普遍性** と言います)。
- c. 一般に臨界指数の値は次元による。しかし、臨界指数の間には $(2 - \eta)\nu = \gamma$ のような関係式が成り立つ。(これを **スケーリングの関係式** と言います。 η はここでは定義していませんが、別の臨界指数です。)
- d. 4 次元より上では、 $\gamma = 1, \nu = \frac{1}{2}, \eta = 0$ だろう。(この簡単な値を **平均場的な臨界指数の値** と呼びます。また、「 $\circ\circ$ 次元より上では平均場的」となる次元 $\circ\circ$ のことを **臨界次元** と呼びます。) つまり、SAW の臨界次元は 4 だろう。
- e. $d = 2$ では $\gamma = \frac{43}{32}, \nu = \frac{3}{4}, \eta = \frac{5}{24}$ だろう。

現在、予想 a,b,c,e は 2, 3, 4 次元では未だに予想であり、証明されている訳ではありません。ただし、2 次元については 21 世紀になって非常に大きな進歩があり、予想 a,b,c,e が証明されるのも時間の問題と思われま⁵⁰。また、5 次元以上での予想 a~d は 1985 年から 1990 年にかけて証明されました。

5.1.2 パーコレーション

ランダムウォークと磁石のモデルの橋渡しのために、パーコレーションというモデルについて簡単に触れます。パーコレーションとは例えば、大きな岩に小さな亀裂がたくさん入った場合に、その亀裂がつながって岩の中に水が浸透 (percolate) していくような現象をモデル化したものです。

数学のモデルとして定義する際には、以下のようにします。格子上の隣り合った点の間を「ボンド」と呼びます。そして、

⁵⁰実は 2005 年に公開講座をした時、まさに「時間の問題」と書いたのですが、今だに未解決なようです...

- それぞれのボンドはランダムに亀裂になったり, ならなかったりする.
- あるボンドが亀裂になる (水が通る) 確率は p (亀裂にならない確率は $1-p$)
- 各ボンドの状態は他のボンドとは独立に決める

と取り決めて, このようにボンドの状態 (亀裂) が決まった時に端から端まで亀裂だけをつないだ通り道ができるかどうかなどを問題にします. 特に, 原点から無限遠まで亀裂の道でつながるかどうか重要です. (つながれば岩の表面から深部まで水がしみ込める. つながらないなら, しみ込めない.)

すぐに予想されるように, p が小さい (亀裂ができにくい) 時には, 原点から亀裂を通していける場所は非常に限られてしまいます. 当然, この場合には無限遠にもつながりません. 一方, p が 1 に近い場合にはそこら中が亀裂な訳ですから, 亀裂を伝わってどこまでも行くことができます. つまり, この場合は原点が正の確率で無限遠とつながるでしょう.

実際, これらは数学的に厳密に証明されています. 原点が無限遠につながる確率を $\theta(p)$ と書きます. また, 原点が平均して何個くらいの点とつながっているか, その期待値を $\chi(p)$ と書きます. 更に, 原点がこのくらいの距離までつながっているという距離の目安を $\xi(p)$ と書きます. すると, 証明された事実として

- p が小さいところでは, 原点が無限遠までつながる確率 $\theta(p)$ はゼロである.
- 2次元以上では, p が 1 に近い場合, 正の確率で原点が無限遠につながる, つまり $\theta(p) > 0$.
- 2次元以上では $0 < p_c < 1$ なる「臨界確率」があつて,
 - $p < p_c$ では $\theta(p) = 0$, $\chi(p) < \infty$, $\xi(p) < \infty$.
 - $p > p_c$ では $\theta(p) > 0$, $\chi(p) = \infty$, $\xi(p) = \infty$.

更に, 厳密には証明されていない部分もあるものの, 以下の予想があります.

- ある定数 (臨界指数) γ, ν, β があつて,
 - $p \nearrow p_c$ では $\chi(p) \approx (p_c - p)^{-\gamma}$, $\xi(p) \approx (p_c - p)^{-\nu}$.
 - $p \searrow p_c$ では $\theta(p) \approx (p - p_c)^\beta$
 となっているだろう.
- 臨界確率 p_c はモデルを少し変える (例: 斜めにも水がしみ通れるようにする, など) と値が変わる. しかし, γ, ν, β は考えている次元のみで決まる, 非常に安定な量だろう (臨界指数の普遍性).
- 一般に臨界指数の値は次元による. しかし, 臨界指数の間には $(2 - \eta)\nu = \gamma$ のような関係式が成り立つ.
- 6次元より上では, $\gamma = 1$, $\nu = \frac{1}{2}$, $\beta = 1$, $\eta = 0$ だろう (平均場的な臨界指数の値; 臨界次元は 6).
- $d = 2$ では $\gamma = \frac{43}{18}$, $\nu = \frac{4}{3}$, $\beta = \frac{5}{36}$ だろう.

定性的には SAW と同じような予想になっていることに注意してください. SAW の場合と同じく, これらは未だに証明されていません. ただし, 系の次元が十分に高いとき (15次元以上, 適当な条件をつけると 7次元以上) には証明されています. また, 2次元での最近の進歩には目覚ましいものがあります.

5.1.3 磁石のモデル（スピン系）

身の回りにある磁石（永久磁石）は熱していくと、ある温度以上で磁力を失います。この現象を数学的にモデル化したのが「イジングスピン」系とよばれるものです。

このモデルでは実際の磁性体を理想化して、結晶の各原子が小さな磁石（スピン）からなっていると考えます。それぞれのスピンは熱運動によってランダムに向きを変えますが、同時に、互いに並び合いたいという傾向も持っています。スピン同士が並び合いたい傾向と、熱運動によりランダムになる傾向のどちらが強いかで、磁石になったりならなかったりすると考えるのです。

もう少し詳しく言うと、以下ようになります。まず、結晶をモデル化して d -次元の格子を考えます。格子の各点 x にスピン (ϕ_x) が乗っていると、スピン変数 ϕ_x は $+1$ と -1 の値をとると考えます。そしてスピン同士の相互作用によるスピン系のエネルギー（ハミルトニアン）は

$$H = -J \sum_{\langle x,y \rangle} \phi_x \phi_y \quad (5.1.1)$$

となると仮定します ($\langle x,y \rangle$ は隣り合った格子点のペアについての和)。さて、格子上のスピンはそれぞれの ϕ_x が ± 1 の値をとれる訳ですから、無数の実現可能性があります。しかし、ある特定のスピンの配位が実現される確率は、統計力学の一般原理によりボルツマンファクター

$$e^{-\beta H} \quad (5.1.2)$$

に比例すると考えるのです。ここで β は温度の逆数に相当する正の数です。

このモデルは第 2 節で考えてきた独立な確率変数の問題を、**独立でないものに拡張**したものになっています。つまり、 $\beta = 0$ （物理の言葉では超高温の極限）では H の値がなんであってもスピンの配位の実現確率は同じです。つまり、この場合には各スピンが独立にランダムな値をとっていることになり、まさに第 2 節の問題そのものになります。

一方、 $\beta = +\infty$ （物理の言葉では絶対零度の極限）では、 H の値が最小値をとるようなスピンの配位の実現確率だけが、その他の配位の実現確率よりも無限に大きくなってしまいます。つまり、この場合には H が最小値をとるようなスピンの配位だけが実現されるのです。今考えているモデルでは H が最小値をとるのは、すべての $\phi = +1$ となったものか、すべての $\phi = -1$ となったもののどちらかです。いずれの場合もスピンは完全にそろっており、スピンの和が巨視的な磁力となって現れる訳です。

問題はこの中間でどうなっているかですが、パーコレーションの時と同じく、以下のようなことが数学的に証明されています。スピン変数の値の期待値を $M(\beta)$ と書いて、これが磁石の強さを表していると考えます。また、下に出てくる χ, ξ はパーコレーションでの対応物に類似した量です。

- β が小さいところでは、結晶の持つ磁力 $M(\beta)$ はゼロである。
- 2次元以上では、 β が十分に大きい場合、 $M(\beta) > 0$ である。
- 2次元以上では $0 < \beta_c < 1$ なる「臨界逆温度」があつて、
 - $\beta < \beta_c$ では $M(\beta) = 0$, $\chi(\beta) < \infty$, $\xi(\beta) < \infty$,
 - $\beta > \beta_c$ では $M(\beta) > 0$, $\chi(\beta) = \infty$, $\xi(\beta) = \infty$.

更に、厳密には証明されていない部分もあるものの、以下の予想があります。

- a. ある定数（臨界指数） γ, ν, β があつて、
 - $\beta \nearrow \beta_c$ では $\chi(\beta) \approx (\beta_c - \beta)^{-\gamma}$, $\xi(\beta) \approx (\beta_c - \beta)^{-\nu}$.
 - $\beta \searrow \beta_c$ では $M(\beta) \approx (\beta - \beta_c)^\beta$

となっているだろう。

- b. 臨界逆温度 β_c はモデルを少し変える (例: 隣り合っていないスピン同士にも相互作用が働く) と値が変わる. しかし, γ, ν, β は考えている次元のみで決まる, 非常に安定な量だろう (臨界指数の普遍性).
- c. 一般に臨界指数の値は次元による. しかし, 臨界指数の間には $(2 - \eta)\nu = \gamma$ のような関係式が成り立つ.
- d. 4次元より上では, $\gamma = 1, \nu = \frac{1}{2}, \beta = \frac{1}{2}, \eta = 0$ だろう (平均場的な臨界指数の値; 臨界次元は 6).
- e. $d = 2$ では $\gamma = \frac{7}{4}, \nu = 1, \beta = \frac{1}{8}, \eta = \frac{1}{4}$ である.

繰り返すだけ紙の無駄とも思えるくらい, パーコレーションの時と似た結果になっています. 異なるのは臨界次元と臨界指数の値のみです. 今までのモデルと同じく, $d > 4$ の結果は証明されています. また, $d = 2$ のモデルは正確に解けるので, やはり証明されています.

5.1.4 まとめ

以上, 3つの異なるモデルでの臨界現象について, 数学的に証明されたことと予想されていることを列挙しました. 3つのモデルとも, 良く似た振る舞いを示しています. 特に, パーコレーションとイジングモデルはほとんど同じです. 特徴を整理すると

- 「臨界値」(p_c, β_c) があって, 臨界値を境にして系の定性的振る舞いが変わる.
- 臨界値に近づく際の系の振る舞いは少数の「臨界指数」で特徴づけられる
- 臨界値そのものはモデルの定義に大きく依存するが, 臨界指数は系の次元などで決まる, 非常に安定な (普遍的な) 量である.
- 臨界指数の間にはある種の恒等式 (スケーリングの関係式) がなりたつ.
- 系によって決まるある次元 (臨界次元) があって, この次元より上では臨界指数は非常に簡単な値をとる

となります.

以上の特徴は, 特に臨界指数の普遍性は, 実は第 2 節の中心極限定理にも見られたことです. あの場合, N の確率変数の和は $\sqrt{N} = N^{1/2}$ のオーダーの拡がりを持っていました. この性質は (確率変数が独立なら) かなり一般に成り立つので「臨界指数が $1/2$ 」が普遍的に成立していることとなります.

ただし, ここのお話と第 2 節のお話には**根本的な違い**もあります. 第 2 節の話は非常に広い確率変数に対して (独立性を少々破っても) 成り立つものでした. 一方, ここでの臨界現象は, 系のパラメーターを臨界値に近づけないと観察できません. つまり, パラメーターを非常にうまく調節してやる必要がある訳で, 闇雲にパラメーターを選んだら観測できない可能性が高いのです. この意味で, ここでの臨界現象は中心極限定理の成り立つようなものよりも起こりにくいと言えます.

以下では, このような臨界現象 (特にその普遍性) をどのようにしたら解釈できるのか, 時間の許す限り述べたいと思います.

5.2 スピン系の臨界現象: スピンの和のふるまいから

これまでに見てきた臨界現象を中心極限定理の成り立つ場合と比較するために, スピン系を例にとって, スピン全体の和がどのように振る舞っているのか, わかっているところを述べておきます. β_c という臨界点を境にして系の振る舞いが大きく変わることは既に述べました. このとき, N このスピンの和 (を適当に規格化したもの) がどのように振る舞うかには, 以下のような予想 (一部は証明済み) があります. 実のところ, このような予想とその証明は後から紹介する「くりこみ群」の考えに基づいているのですが, こちらを先に述べた方がわかりやすいだろうと考えて, この順序にしました.

以下, B_L とは, 一辺の長さが L (この中に L^d このサイトが入っている) の立方体を表します. このとき, 大体以下のようなことがわかって / 予想されています.

1. $\beta < \beta_c$ では, $\Phi_L = \frac{1}{\sqrt{\chi L^d}} \sum_{x \in B_L} \phi_x$ を考えると, Φ_L は $L \rightarrow \infty$ で正規確率変数に収束する. つまり, $\beta < \beta_c$ では中心極限定理が成り立つ.
2. $\beta > \beta_c$ では磁石になっているので単純に上のようなことはなりたない. ただし, Φ_L の定義を修正して, $\Phi_L = \frac{1}{\sqrt{\chi' L^d}} \sum_{x \in B_L} (\phi_x - M)$ を考えると, これは正規確率変数に収束する (だろう). この意味で, $\beta > \beta_c$ でも (スピン変数の期待値を修正すれば) 中心極限定理がなりたつ. なお, χ' とは $\beta < \beta_c$ での χ に相当する, 有限な量である.
3. 問題は $\beta = \beta_c$ であるが, これについては良くわかっていない. ただし, β が β_c に近づくと χ が無限大になっていくから, 上のような単純な結果は成り立たないと思われる.
 - $d > 4$ では $\Phi_L = \frac{1}{\sqrt{L^{d+2}}} \sum_{x \in B_L} \phi_x$ がある正規確率変数にいくと思われ, 証明も割合に良いところまで来ている.
 - $d < 4$ でどうなるかは全くわからない. 「くりこみ群」の考えがある種の示唆を与えてくれるが, マダマダである.

ここで $\beta \neq \beta_c$ の場合の Φ_L の定義に注目してください. 和の記号の前の係数は $1/\sqrt{L^d}$ ですが, L^d というのはこのブロックの中に入っているスピンの総数です. この意味で, Φ_L というのは第 2 節で考えてきた S_N/\sqrt{N} と本質的に同じものです. そして上の 1 は, スピン同士の相互作用があったとしても, $\beta < \beta_c$ では相互作用は大して重要ではない, ことを主張しています.

ところが, $\beta = \beta_c$ では $\chi = \infty$ なので, 従来の意味での中心極限定理は期待できません. $\chi = \infty$ になる分, \sqrt{N} よりももっと大きな数でスピンの和を割ってやる必要があります. これであまく行く保証はありませんが, $d > 4$ での結果は, これがある程度うまく行くことを示しています.

5.3 くりこみ群の考え

さて, このような現象を理解する道具となりうるのが「くりこみ群」(renormalization group) の考え方です. この考え方では, 上に考えたような Φ_L の分布に注目するのですが, 大きな L に対する Φ_L の分布を一気に計算するのではなく, $L = 2$ のような変換を積み重ねることで最終的に大きな L の Φ_L の分布に到達することを目指します.

もう少し具体的に言うと, 以下ようになります. (簡単のために 1 次元のスピン系を考えて, スピン変数に端から $1, 2, 3, \dots$ と番号をふっておきます.) 我々の目的は大きな (無限大の) L に対する $\Phi_L = C_L \sum_{i=1}^L \phi_i$ の分布を調べることです. しかし, これを一気にやる代わりにブロックスピンと呼ばれる新しい確率変数

$$\phi_j^{(1)} = C_1(\phi_{2j-1} + \phi_{2j}) \quad (5.3.1)$$

を導入し, $\{\phi_j^{(1)}\}$ がどのような分布に従っているかを考えるのです (C_1 は解析がうまく行くようにとる定数). $\{\phi_j^{(1)}\}$ の分布がわかったら, また新しい変数

$$\phi_j^{(2)} = C_2(\phi_{2j-1}^{(1)} + \phi_{2j}^{(1)}) \quad (5.3.2)$$

を導入し, こいつの分布を調べます. 以下同様に

$$\phi_j^{(n+1)} = C_n(\phi_{2j-1}^{(n)} + \phi_{2j}^{(n)}) \quad (5.3.3)$$

を順次 ($n = 1, 2, \dots$) 導入して, これらの分布を調べます. 作り方から, $\phi^{(n)}$ は 2^n この ϕ の和 (に定数をかけたもの) になっていますから, $n \rightarrow \infty$ における $\phi^{(n)}$ の分布をみてやれば, 欲しいものが大体得られます. そして, このように $\phi^{(n)}$ の分布から $\phi^{(n+1)}$ の分布を作る変換を「くりこみ変換」と言います.

くりこみ群の特徴とメリットは何でしょうか? わざわざややこしいことをやっているように見えるだろうと思うのですが, 以下の 2 点は挙げてよいでしょう.

- くりこみ変換は $\phi^{(n)}$ の分布 $f^{(n)}$ から $\phi^{(n+1)}$ の分布 $f^{(n+1)}$ への変換を与えるから、これはスピンの分布関数の空間の中での**力学系**を定義する。元々の統計力学の問題が、分布の空間での力学系の問題に書き換えられた。
- この場合、「時間」に相当するものは n （どのスケールでものを見るか）である。くりこみ変換はものを見るスケールを変えた時に、系がどのように変わって見えるかを教えてくれる。また、それをもとにして元々の系の性質を知ることができる。実際にくりこみ変換をやってみると、元々異なった系がスケールを大きくすると同じようなものに見えてくることもある。これは普遍性の根拠を与える。これらは従来になかった画期的なものの方である。
- 臨界現象を示すような系は大抵複雑で、とてもではないけども解ききることはできない。しかし、くりこみ群の方法が目指すような、「ちょっとずつ変換する」方法ならやっていけることが多々ある。実用的にもくりこみ群は重要である。

時間があれば、このあたりのことを実例を通して説明したいと考えています。

A 文献案内

確率論の本はいろいろと出ていますが、この講義で扱った内容について適度に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ませんでした。そこで、程度には余り拘らず、良いと思うものを挙げます。

まず、以下の [1] は非常に良いのでお奨めです（高校 3 年くらいの知識で読みこなせます）。ただ、残念ながら現在は絶版のようです。

次に、[2] は「数学セミナー」という月刊誌の特集記事です。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います。

また [3][4] は大学生向けの教科書ですが、それほど難しくありません。[2] の後に読むのに適していると思います。

[5][6] は大学生向けの本格的な教科書で、特に [6] は非常によく書かれていて、お薦めです。[7] は大数の完全法則に焦点を当てた、文字通りマニアックな本です。

なお、僕の web page の「講義」のところ (<http://math.kyushu-u.ac.jp/hara/lectures/lectures-j.html>) からのリンクには過去の講義ノートの一部が置いてあります。2001 年度の「数学展望 II」のものはこの公開講座の延長上にあるもので、参考になるかもしれません。また 2002 年度の「確率論」はその更に延長上にあります。

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)
6. Kai Lai Chung：A Course in Probability Theory Academic Press (1974)
7. 服部哲弥：「確率変数の収束と大数の完全法則 - 少しマニアックな確率論入門 -」共立出版 (2019)

臨界現象やくりこみ群に関しては、

8. 特集：「くりこみ群」数学セミナー 1997 年 4 月号 (1997)
9. 服部哲弥：「ランダムウォークとくりこみ群」共立出版 (2004)
10. 田崎清明+原隆：「相転移と臨界現象の数理」共立出版 (2015)

を挙げておきます。