MCS.T312 講義資料

確率論基礎

中野 張

2026 年 5 月 19 日

目次

第 1 章　はじめに　4
1.1　組合せ確率の復習　4
1.2　現代確率論の応用　6
1.3　確率論の歴史　9
1.4　確率の解釈　10
1.5　参考図書　11

第 2 章　有限試行　12
2.1　試行と確率空間　12
2.2　確率変数　15
2.3　期待値　16
2.4　独立性　19
2.5　条件付き確率　22
2.6　大数の法則　24
章末問題　26

第 3 章　確率論の基本的枠組み　28
3.1　無限集合おさらい　28
3.2　確率が意味をもつ集合　30
3.3　確率測度・確率分布　32
3.4　確率変数　36
3.5　期待値　48
3.6　期待値の収束　60
3.7　無限回コイン投げ空間　65
章末問題　66

第 4 章　大数の法則　70
4.1　ボレル・カンテリの補題　70
4.2　大数の弱法則　71
4.3　大数の強法則　74
章末問題　75

第 5 章　中心極限定理 77
5.1　特性関数 77
5.2　法則収束 86
5.3　中心極限定理 88
章末問題 91

参考文献 93

記法

- $\mathbb{N} = \{1, 2, \ldots\}$.
- $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$.
- $\mathbb{R}^d$: d 次元ユークリッド空間.
- $\mathbb{Z}^d = \{(x^1, \ldots, x^d) : x^i \in \mathbb{Z},\ \ 1 \leq i \leq d\}$.
- $\mathbb{R}^{m \times d}$: 実 $m \times d$ 行列全体.
- $\mathbb{C}$: 複素数全体.
- $|x|$: $x \in \mathbb{R}^d$ のとき，x のユークリッドノルム.
- $|S|$: S が集合のとき，S の要素数.
- $x^+ = \max\{x, 0\}$, $x \in \mathbb{R}$.
- $x^- = \max\{-x, 0\}$, $x \in \mathbb{R}$.
- $\mathbb{E}[X]$: 確率変数 X の期待値.
- $\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$: 確率変数 X の分散.
- I_m: $\mathbb{R}^{m \times m}$ の単位行列.

第 1 章

はじめに

確率の理論とは基本的に計算に還元された常識にすぎない.

ラプラス「確率の哲学的試論」

本授業では，現代確率論の基礎を学ぶ．ここで言う「現代確率論」とは，コルモゴロフによって定式化された公理論的確率論を指す．この「公理論的」とは，例えば確率の σ-加法性のような性質を公理として定め，その上に理論体系を構築する立場を意味する．確率に関する基本的な性質を公理として受け入れることで，特定の解釈（頻度説や主観説など）に依存しない数学的な枠組みが確立されている．このように「確率」を公理的に取り扱うという姿勢は，これまで組合せ論的確率論に親しんできた読者にとっては，やや抽象的で戸惑いを感じるかもしれない．高校までの確率論では，しばしば「確率＝数えるもの」という理解が中心だったが，ここから先は「確率＝測るもの」という視点が必要となる．本授業では、そのために必要となる数学的道具を順を追って整備していく．一つの目的は、それらの道具を一通り使いこなせるようになることであり，もう一つの目的は，大数の法則や中心極限定理といった極限定理を理解し，その背後にある理論構造を把握することである.「ランダムであるがゆえに成り立つ法則」が存在する．このことは確率論の魅力の一つである.

1.1 組合せ確率の復習

例題から始める.

例題 1.1.1 ある箱に，ひらがな一字が書かれた紙が 12 枚入っている．この箱から 1 枚ずつ紙を取り出す．箱に入っている紙は

「う」「す」「ち」「の」「も」「も」「も」「も」「も」「も」「も」「も」

である．このとき，取り出した順に左から並べ,「すもももももももものうち」になる確率は？

解答. 1 枚ずつ取り出すことを**独立試行**といい，独立試行の確率は掛け算で求めたことを思い出そう.

例えば,

$$\Pr(1\text{ 回目に「す」}, 2\text{ 回目に「ち」}) = \Pr(1\text{ 回目に「す」}) \times \Pr(2\text{ 回目に「ち」}).$$

ここで $\Pr(A)$ は A が起こる確率を表す．この考え方に従って，求める確率は

$$\frac{1}{12} \times \underbrace{\frac{8}{11} \times \frac{7}{10} \times \frac{6}{9} \times \frac{5}{8} \times \frac{4}{7} \times \frac{3}{6} \times \frac{2}{5} \times \frac{1}{4}}_{=1/{}_{11}C_8} \times \frac{1}{3} \times \frac{1}{2} = 0.00008417508.$$

> **例題 1.1.2 (誕生日問題)** ある年（うるう年ではない）に生まれた k 人のうち，最低 2 人以上が同じ誕生日である確率を求めよ．ただし以下を仮定する．
>
> - k 人の誕生日は無関係である（特に，双子はいない）．
> - 各日が同等に誕生日になり得る（日による出生率の違いは無視する）．
> - $k \leq 365$（そうでないと，必ず 2 人以上の誕生日が同じになる）．

解答. k 人それぞれ 365 通りあるので，起こり得る結果は 365^k 通りである．k 人すべての誕生日が異なるのは,

$$365 \times 364 \times 363 \times \cdots \times (365 - k + 1)$$

通り．これより，k 人すべての誕生日が異なる確率は

$$\frac{365 \times 364 \times 363 \times \cdots (365 - k + 1)}{365^k}.$$

よって求める確率 p は

$$p = 1 - \frac{365!}{(365 - k)!365^k}.$$

表 1.1.1 はいくつかの値の k に対する最低 2 人以上が同じ誕生日である確率 p を並べたものである．これによれば，$p \geq 0.5$ となるためには 23 人いれば十分ということになる．また，100 人のときは $p = 0.9999997$ であり，ほぼ 100% で 2 人以上いると言ってよい．

表 1.1.1　最低 2 人以上が同じ誕生日である確率

k	p	k	p
5	0.027	25	0.569
10	0.117	30	0.706
15	0.253	40	0.891
20	0.411	50	0.970
22	0.476	60	0.994
23	0.507		

ここで**組合せ確率**の定義を思い出しておこう．組合せ論的確率論では，A が起こる確率 $\Pr(A)$ は

$$\Pr(A) = \frac{A\text{が起こる場合の数}}{\text{起こりうるすべての場合の数}}$$

により定められている．この定義が機能するための前提として，「根元事象は同程度に起こりやすい」という仮定を置いていたのだった．このとき，次のような疑問が思いつくだろう．

- 起こりうるすべての場合の数 $= \infty$ のときは？
- 根元事象の起こりやすさに違いがあるときは？

このような状況に対応できるよう整備されているのが現代確率論である．すなわち，組合せ確率を一般化，抽象化したのが現代確率論であると言える．

1.2　現代確率論の応用

ここでは現代確率論の応用例をいくつか紹介する．

初等確率論の応用

例 1.2.1 (People v. Collins **事件**)　裁判において確率論が適切に用いられた例として，1968 年のアメリカ合衆国カリフォルニア州控訴裁判所の有名な判例を取り上げよう．事件の概要は以下の通り．

- ロサンゼルスで財布が盗まれた事件が発生した．
- 「ブロンドヘアをポニーテールに束ねた若い女性とあごひげを生やした黒人運転手が黄色い車に乗って逃走した」という証言があった．
- 特徴に合致するカップルが事件の数日後に逮捕された．
- 証言以外に物証はなく，無作為抽出したカップルがこの特徴に合致する可能性は $8.3 \times 10^{-8} \approx$ 1200 万分の 1.
- この確率の小ささと他に物証がないことから，陪審員は有罪の評決を下した．

しかし，裁判所は，この確率計算には根拠がないとし有罪判決を棄却した．以下がより有用な確率計算である．特徴に合致するカップルが 1 組存在するという条件の下で，2 組目が存在する確率を考える．カップル自体は n 組存在し，先の特徴に合致するカップルが抽出される確率を p とすると，

$$\begin{aligned}&\Pr(\text{最低 2 組のカップルが特徴に合致} \mid \text{最低 1 組のカップルが特徴に合致})\\&= \frac{1-(1-p)^n - np(1-p)^{n-1}}{1-(1-p)^n} \approx 0.2966.\end{aligned} \tag{1.2.1}$$

ここで $\Pr(A \mid B)$ は事象 B で条件付けた A の条件付き確率である（これについては次章で学ぶ）．0.2966 という確率は被告人を有罪とすることに疑念を持つには十分な大きさである．

統計的推測

統計的推測とは，一般的には，観測データをもとに関心のある分布に関して確率論の概念を用いた叙述を行うことを指す．これはもちろん確率論の重要な応用である．

例 1.2.2 (**臨床試験**)　医療における次のような状況を想定しよう．40 人の患者がおり，患者それぞれ

について有効か無効かの 2 つの結果があり得るとする．i 番目の患者が有効であれば，$X_i = 1$，無効なら $X_i = 0$ とする．有効率 P は患者次第で確率的だが，$P = p$ のもとでの $X_i = 1$ となる確率は p であると仮定する．このとき，有効となる患者の数を X とすると，$E[X \mid P] = 40P$ である．ベイズの定理（後述）を用いて患者の有効／無効のデータから期待有効率を推定すると，

$$E[P \mid X] = \frac{X+1}{42}$$

となる．ここで左辺は X が与えられたときの P の条件付き期待値を表している．

例 1.2.3 (**信頼区間**) ある工場で製造される製品の長さを X とし，$X \sim N(m, v)$ と仮定する（$v > 0$ は既知，$m \in \mathbb{R}$ は未知）．長さを n 回測定して独立な観測値 $X_1, \ldots, X_n$ を得たとき，m の推定値として標本平均 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ を考えるのは自然である．

このとき正規分布の再生性（あるいは中心極限定理）より $\bar{X}_n \sim N(m, v/n)$ であるから，

$$\mathbb{P}\left(m \in \left[\bar{X}_n - 1.96\sqrt{v/n},\ \bar{X}_n + 1.96\sqrt{v/n}\right]\right) \approx 0.95$$

が成り立つ．この区間を m の 95% **信頼区間**と呼び，観測データから真の値の含まれる範囲を確率的に評価する典型的な統計的推測である．

生成モデル

例 1.2.4 (**大規模言語モデル**) 近年の人工知能の進展，特に**大規模言語モデル**（Large Language Models, LLMs）の成功は，確率分布の推定とサンプリングという現代確率論の基本的枠組みが，極めて高次元かつ複雑な空間においても有効に機能しうることを示している．

具体的には，語彙集合 $\mathcal{V}$ に基づく列空間 $\mathcal{V}^\infty$ 上で，訓練データに基づいて確率測度 $\mathbb{P}_\theta$ を学習し，これを用いて任意の系列 $(x_1, \ldots, x_n)$ に対する生成確率

$$\mathbb{P}_\theta(x_1, \ldots, x_n) = \prod_{t=1}^{n} \mathbb{P}_\theta(x_t \mid x_1, \ldots, x_{t-1})$$

をモデル化・推定する．ここで，θ はモデルパラメータであり，実際には最尤推定等により近似的に決定される．

このようにして得られた確率分布 $\mathbb{P}_\theta$ からのサンプリングにより，人間の言語に類似したテキストが自動的に生成される．すなわち，LLM は「経験的分布から確率測度を構成し，そこから確率変数を生成する」という現代確率論の中心的アイデアの応用例である．

例 1.2.5 (**拡散モデル，**Flow Matching) **拡散モデル**（Diffusion Models）と **Flow Matching** は，連続的なデータ空間（たとえば画像空間）上の複雑な分布 μ_{data} からサンプリングを行うための代表的な深層生成モデルである．いずれも，$N(0, I_d)$ のような単純な初期分布 μ_0 とデータ分布 $\mu_1 = \mu_{\mathrm{data}}$ をつなぐ確率測度の経路 $\{\mu_t\}$ を考え，その経路を実現する確率過程あるいは決定的フローを学習する．

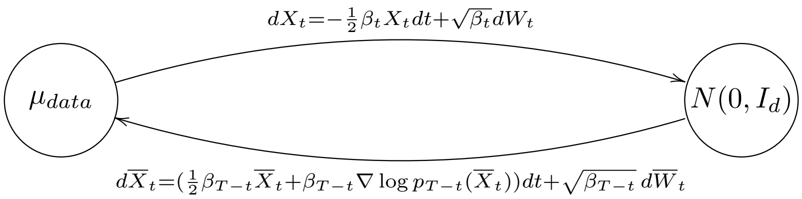
拡散モデルは，時間 $t \in [0, T]$ に対しデータをノイズへ変換する順時間 SDE

$$dX_t = -\frac{1}{2}\beta_t X_t dt + \sqrt{\beta_t}\, dW_t$$

と，その時間反転である逆時間 SDE

$$d\overline{X}_t = \left(\frac{1}{2}\beta_{T-t}\overline{X}_t + \beta_{T-t}\nabla \log p_{T-t}(\overline{X}_t)\right) dt + \sqrt{\beta_{T-t}}\, d\overline{W}_t$$

を用い，後者を $\overline{X}_0 \sim N(0, I_d)$ から出発させて $\overline{X}_T \sim \mu_{\mathrm{data}}$ を得る．



Flow Matching は，例えば $X_0 \sim \mu_0$, $X_1 \sim \mu_{\mathrm{data}}$ から $X_t := (1-t)X_0 + tX_1$ として確率測度の経路 $\mu_t = \mathrm{Law}(X_t)$ を構成し，これを実現する常微分方程式

$$\frac{dx_t}{dt} = v_t(x_t)$$

のベクトル場 v_t をニューラルネットワークで近似する．拡散モデルが確率的な SDE, Flow Matching が ODE に基づく点が対照的で，後者はサンプリングが速い等の利点があり，Stable Diffusion 3 など最新の画像生成モデルで採用されている．

これらは (1) 未知の高次元確率分布を推定し，(2) それに従う乱数を生成するという現代確率論の中心的な営みを極めて高次元な空間で実現するという点で，現代的確率論の応用例として捉えられる．

確率論的手法

例 1.2.6 (**モンテカルロ積分**) 物理学や統計学，ファイナンスにおいて広く用いられる**モンテカルロ積分**も，現代確率論の典型的な応用例である．モンテカルロ法は確率変数 X と関数 f に対し，$f(X)$ の期待値 $E[f(X)]$ を近似するために，独立同分布な乱数列 $X_1, \cdots, X_n$ を用いて

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k)$$

という経験平均を計算する手法である．とくに高次元積分において，従来の数値積分が破綻する場合でも，モンテカルロ法はその確率的性質により安定した近似を提供する．現在では，ベイズ推論や生成モデルの学習アルゴリズムにおいて不可欠な手法となっている．

理論的には，例えば，X が 100 次元の一様分布に従う場合，これと同分布の確率変数 $X^{(k)}$, $k = 1, \ldots, n$, を独立に発生させるならば十分大きい n に対し

$$E\left|\frac{1}{n}\sum_{k=1}^{n} f(X^{(k)}) - \int_0^1 \cdots \int_0^1 f(x_1, \ldots, x_{100})\, dx_1 \cdots dx_{100}\right| = O(n^{-1/2})$$

となることが分かる．これについては第 5 章でみる．

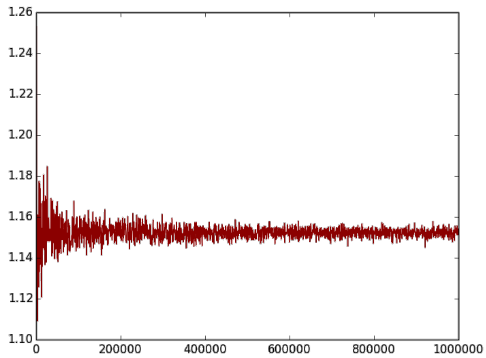


図 1.1　縦軸は積分の近似値（$\sum_{k=1}^{n} f(X^{(k)})/n$）．横軸は乱数の数（$n$）

例 1.2.7 (マルコフ連鎖モンテカルロ法) モンテカルロ法では関心のある分布から独立な乱数を生成する必要があるが，高次元の複雑な分布では直接サンプリングが困難なことが多い．**マルコフ連鎖モンテカルロ法**（Markov Chain Monte Carlo, MCMC）は，目標分布 π を不変分布にもつようなマルコフ連鎖 $\{X_n\}_{n=0}^{\infty}$ を構成し，これを十分長く走らせることで X_n の分布を π に近づけ，さらに大数の法則

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to \int f\, d\pi$$

により π に関する期待値を近似する手法である．代表的なアルゴリズムとして Metropolis-Hastings 法や Gibbs サンプリングがあり，ベイズ統計，統計物理，機械学習などで広く用いられている．

1.3　確率論の歴史

確率論の生みの親は賭け事である．古くは，ダンテの『神曲』についてのある注釈書の中に，3 個のサイコロを振って出る目の生起確率に関する記述がある．次に，カルダーノ (1444-1524) の死後だいぶ経ったのちに出版された『さいころあそびについて』（1663 年発行）がある．これは確率を論じているというより「賭博指南書」とでも呼ぶべきもので，経験則から導いた結論や，論理的に誤った議論も多い．16〜17 世紀のヨーロッパでは，サイコロやカードゲームの勝敗予測が知識人の間で知的関心の対象となり，ケプラーやガリレオも著書の中で確率の問題について言及している．特に，パスカルとフェルマーの書簡（1654 年）は，現代確率論の嚆矢と見なされている．彼らは賭けの配分問題（Problem of Points）を契機として，偶然の事象に数学的構造を見出そうと試みた．その後，18 世紀にはヤコブ・ベルヌーイが『アルス・コンジェクランディ（推測術)』（1713 年）を著し，大数の法則の初歩的な形を提示した．また，19 世紀にはラプラスが確率論を天文学や統計的推論に応用し，確率を理性に基づく不確実性の尺度として捉えた．この時期の確率論は主に解析的手法と組み合わさり，直観に基づく古典的確率論（等確率な事象の比）として発展した．しかし，20 世紀に入ると，古典的な枠組みだけでは複雑な現象や無限次元空間上の確率的対象を扱うには不十分であることが明らかになった．この課題に対し，1933 年にコルモゴロフが『確率論の基礎』を発表し、測度論を土台

とする公理的確率論を確立した．彼の理論により，確率空間は測度空間の特別な場合として定式化され，確率論は解析学，特にルベーグ積分を基礎とした厳密な数学的理論となった．

1.4　確率の解釈

「確率」とは何か．この問いに対する答えは，歴史的にも哲学的にも一様ではない．確率の解釈にはさまざまな立場があり，しばしば次のように分類される．

（物的に）客観的解釈

この立場では，確率は観測される現象そのものに内在する客観的な性質であるとされる．典型的な例が**頻度説**であり，試行を十分に繰り返したときの相対頻度の極限として確率が定義される．例えば，特定の放射性原子が一定時間内に崩壊する確率や，偏りのないコインを何度も投げたときに表が出る比率などは，観測可能な頻度として確率を解釈できると考えられている．これらは物理的・自然的な過程に起因する，客観的な確率である．

認識論的解釈

これに対して，確率を主観的な信念や不確実性に対する認識の度合いとして捉えるのが認識論的解釈である．ここでは，確率は外部世界の性質というよりも，観測者の知識や情報の不完全性に由来するものとされる．すなわち，確率は「知らないからこそそう思う」ものであり，これは**主観説**や**ベイズ的確率**の立場と関係が深い．

例 1.4.1 会社 A は株式市場に関する予測情報を提供していると称している．ある人物 B さんは、会社 A からのダイレクトメッセージ（DM）を毎週受け取り，7 週連続で株価の予測が的中していることに驚く．B さんから見れば、この出来事が偶然である確率は $2^{-7} = 0.008$ と極めて小さい．ところが，実際には会社 A は最初から $2^7 = 128$ 人の見込み客に対して以下のような方法で DM を送っていた．

- 第 1 週: 64 人に「上がる」，64 人に「下がる」と送信．
- 第 2 週: 前週の的中者 64 人のうち 32 人に「上がる」，32 人に「下がる」と送信．
- 以下同様に繰り返す．

この手法により，7 週の終わりには「7 週連続で予測が当たった唯一の人物」が必ず 1 人存在する．つまりこの現象は確率的に偶然生じたものではなく，B さんの知識が限られていることに起因する主観的誤認識によって「ありそうもない出来事」に見えているにすぎない．

このように，同じ「確率」という語で表現される現象にも，物理的必然性と情報的限定性という異なる側面がある．

現代確率論の立場

これらの解釈の多様性にもかかわらず，現代の確率論は，特定の確率解釈に依存しない形式体系として構築されており，その背後にある解釈が客観的であれ主観的であれ，数学的には同等に扱うことができる．言い換えれば，「確率をどう解釈するか」は確率論の外部の問題であり，理論の内部では中立的である．

1.5　参考図書

本授業の参考図書として以下を挙げておく．この資料を作る際にも参考にした．

- P. ブレモー著『モデルで学ぶ確率論』丸善出版，2012. [7].
- 吉田伸生『新装版 確率の基礎から統計へ』日本評論社. [11].
- 佐藤坦著『測度から確率へ』共立出版，1994. [10].
- 伊藤清著『確率論の基礎』岩波書店，1991. [8].
- W. フェラー『確率論とその応用』紀伊國屋書店. [4, 5, 3, 6].
- M. H. デグルート著，M. J. シャービッシュ著『確率と統計』[1].

[7] と [11] は本授業と同程度の難易度である．[10] や [8] は数学的に厳密に書かれており，本授業で端折る測度論的議論を補うのによいかもしれない．フェラーの本は絶版で入手が難しいが（図書館には置いてあるだろう），具体的で面白い問題が数多く載っている．[1] は測度論的取扱いは無いが確率・統計についての網羅的なテキストである．

第 2 章

有限試行

高校数学における確率の考え方では，確率計算の対象となる事象 A を日常言語で記述し，「同様に確からしい」の原理を適用することで A の確率を算出したのであった．これに対して現代確率論では，A を日常言語で記述し，さらに数学の言語に翻訳したのち確率論の諸定理を適用することで A の確率を算出する．この過程において，事象 A は集合論の枠組みで扱い，A の確率は集合関数として扱うことになる．この手続きに慣れるため，本章では，起こり得る結果は有限個と仮定し，確率論の基本的な概念を学ぶ．

2.1　試行と確率空間

試行とは，同じ条件のもとで繰り返すことができ，その結果が偶然によって決まる実験や観察などのことをいう．**有限試行**とは，有限個の異なる結果しかもたない試行．各結果を**標本点**，標本点全体の集合を**標本空間**という．記号としては標本点には ω, 標本空間には Ω を使うことが多い．

以下，本章では，試行は有限試行であるとし，現れる標本空間は全て有限集合とする．

例 2.1.1 (サイコロの 1 回投げ試行) サイコロを 1 回投げるとき，起こり得る結果は $1, 2, 3, 4, 5, 6$ の 6 通りである．よって，この試行の標本点は各 $1, 2, 3, 4, 5, 6$ であり，標本空間は $\{1, 2, 3, 4, 5, 6\}$ である．

例題 2.1.2 チーム 1, 2, 3, 4 がサッカーの試合をし，ノックアウト形式のトーナメントで優勝チームを決める．この試行に対応する標本空間を求めよ．ただし，1 回戦第 1 試合はチーム 1 対チーム 2，第 2 試合はチーム 3 対チーム 4 と決まっており，全ての対戦で引き分けはないものとする．

解答. 1 回戦第 1 試合の勝者，1 回戦第 2 試合の勝者，決勝戦の勝者を左から並べて $(1, 3, 1)$ などと書くことにすると，

$$\begin{aligned}\Omega &= \{(i, j, k) \mid k \in \{i, j\},\ i \in \{1, 2\},\ j \in \{3, 4\}\} \\ &= \{(1,3,1), (1,3,3), (1,4,1), (1,4,4), (2,3,2), (2,3,3), (2,4,2), (2,4,4)\}.\end{aligned}$$

$\mathbf{T}$ を試行とし，Ω を $\mathbf{T}$ の標本空間とする．$A, B \subset \Omega$ に対し，

- A が**起こる** $\overset{\text{def}}{\iff}$ 試行 T の結果として A に属する標本点が実現する.
- $A \subset \Omega$ を**事象**と呼ぶ.
- $A^c := \Omega \setminus A = \{\omega \in \Omega \,|\, \omega \notin A\}$ を**余事象（補集合）**という.
- $A \cup B = \{\omega \in \Omega \,|\, \omega \in A$ または $\omega \in B\}$ を**和事象（和集合）**という.
- $A \cap B = \{\omega \in \Omega \,|\, \omega \in A$ かつ $\omega \in B\}$ を**交事象（積集合）**という.
- $A \setminus B = \{\omega \in \Omega \,|\, \omega \in A$ かつ $\omega \notin B\}$ を**差事象（差集合）**という.
- $A \cap B = \emptyset$ のとき，A と B は互いに**排反**であるという.

試行 $\mathbf{T}$ がサイコロの 1 回投げのとき，結果が同程度に確からしいとすれば，事象 A の起こる確率 $\mathbb{P}(A)$ は

$$\mathbb{P}(A) = \frac{|A|}{6}.$$

確率の直観的意味を踏まえ，数学的には次のように確率を集合関数として定義する.

定義 2.1.3 Ω を試行 $\mathbf{T}$ の標本空間とする．Ω の部分集合に対して定義される集合関数 $\mathbb{P}$ で次の 3 つの性質をもつものを Ω 上の**確率測度**あるいは単に**確率**と呼ぶ.

(1) 任意の $A \subset \Omega$ に対して $\mathbb{P}(A) \geq 0$.
(2) 加法性: $A \cap B = \emptyset$ なる任意の $A, B \subset \Omega$ に対して $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
(3) $\mathbb{P}(\Omega) = 1$.

- 2^Ω で Ω の部分集合全体を表す.
- Ω の上の確率測度は $\mathbb{P} : 2^\Omega \to [0, 1]$ で加法性を満たすものである.
- 組 $(\Omega, \mathbb{P})$ を**確率空間**という.
- 試行 $\mathbf{T}$ を考えるとき確率は，数学的には，試行 $\mathbf{T}$ に対応する Ω 上の確率測度により表現される.
- $\mathbf{T}$ を調べることは，数学的には，確率空間 $(\Omega, \mathbb{P})$ を調べることに相当する.

定理 2.1.4 ある試行 $\mathbf{T}$ に対応する確率空間 $(\Omega, \mathbb{P})$ を考える．このとき以下が成り立つ.

(1) $i \neq j$ に対し $A_i \cap A_j = \emptyset$ ならば，$\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$.
(2) $A \subset B$ のとき，$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.
(3) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
(4) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
(5) $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.

証明. (1) まず, $A_1 \cap (\bigcup_{i=2}^n A_i) = \emptyset$ である. 実際, $\omega \in A_1 \cap (\bigcup_{i=2}^n A_i) \Leftrightarrow$「$\omega \in A_1$ かつ $\omega \in \bigcup_{i=2}^n A_i$」$\Leftrightarrow$「$\omega \in A_1$ かつ $\omega \in A_i$ となる $i \in \{2, \ldots, n\}$ が存在」であり，$A_1 \cap A_i = \emptyset$ よりこれは不可能である.

定義 2.1.3 (2)（確率測度の加法性）より,

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}\left(A_1 \cup \left(\bigcup_{i=2}^n A_i\right)\right) = \mathbb{P}(A_1) + \mathbb{P}\left(\bigcup_{i=2}^n A_i\right).$$

$\bigcup_{i=2}^n A_i = A_2 \cup (\bigcup_{i=3}^n A_i)$ で，上の議論と同様にして A_2 と $\bigcup_{i=3}^n A_i$ は互いに排反であるから再び確率測度の加法性より

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}\left(\bigcup_{i=3}^n A_i\right).$$

この議論を繰り返すことにより，

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)$$

を得る．

(2) $B = (B \setminus A) \cup A$ であり，$B \setminus A$ と A は互いに排反であるから，確率測度の加法性より

$$\mathbb{P}(B) = \mathbb{P}((B \setminus A) \cup A) = \mathbb{P}(B \setminus A) + \mathbb{P}(A).$$

よって $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.

(3) $\Omega = A \cup A^c$ であるから定義 2.1.3 より

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

よって $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(4) $A \cup B$ は互いに排反な 3 事象の和事象として

$$A \cup B = (A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B)$$

と表される．よって確率の加法性と (2) の結果より

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A \setminus (A \cap B)) + \mathbb{P}(B \setminus (A \cap B)) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).\end{aligned}$$

(5). $A = \{\omega_1, \ldots, \omega_n\}$ と表すとき，$\{\omega_i\}$, $i = 1, \ldots, n$ は互いに排反で，$A = \bigcup_{i=1}^n \{\omega_i\}$ である．この表現に確率の加法性を適用すればよい． □

命題 2.1.5 ある試行 $\mathbf{T}$ に対応する確率空間 $(\Omega, \mathbb{P})$ を考える．このとき以下が成り立つ．

(1) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i)$.
(2) $\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n \mathbb{P}(A_i^c)$.

証明. (1) 定理 2.1.4(4) と確率の非負性より

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \mathbb{P}\left(A_1 \cup \left(\bigcup_{i=2}^n A_i\right)\right) = \mathbb{P}(A_1) + \mathbb{P}\left(\bigcup_{i=2}^n A_i\right) - \mathbb{P}\left(A_1 \cap \left(\bigcup_{i=2}^n A_i\right)\right) \\ &\leq \mathbb{P}(A_1) + \mathbb{P}\left(\bigcup_{i=2}^n A_i\right).\end{aligned}$$

この議論を繰り返し

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}\left(\bigcup_{i=3}^{n} A_i\right) \le \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n).$$

(2) 本命題 (1) の結果より

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i^c\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i^c).$$

他方，$\bigcap_{i=1}^n A_i = (\bigcup_{i=1}^n A_i^c)^c$ である．実際，$\omega \in (\bigcup_{i=1}^n A_i^c)^c \Leftrightarrow \omega \notin \bigcup_{i=1}^n A_i^c \Leftrightarrow$ 任意の $i = 1, \ldots, n$ に対して $\omega \notin A_i^c \Leftrightarrow$ 任意の $i = 1, \ldots, n$ に対して $\omega \in A_i \Leftrightarrow \omega \in \bigcap_{i=1}^n A_i$．このことと定理 2.1.4 (3) より

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i^c\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^{n} A_i^c\right) \ge 1 - \sum_{i=1}^{n} \mathbb{P}(A_i^c).$$

□

2.2　確率変数

以下，ある試行 $\mathbf{T}$ の確率空間を $(\Omega, \mathbb{P})$ とする．$X : \Omega \to \mathbb{R}$ を（実）**確率変数**という．

例 2.2.1 $\mathbf{T}$ をサイコロの 1 回投げ試行とする．このとき，$\Omega = \{1,2,3,4,5,6\}$, $\mathbb{P}(A) = |A|/6$ である．奇数が出れば 100 円を得て，偶数が出れば 100 円を失う賭けを考える．この収益は

$$X(\omega) = \begin{cases} 100 & (\omega = 1,3,5) \\ -100 & (\omega = 2,4,6) \end{cases}$$

と確率変数を用いて記述することができる．

例 2.2.2 サイコロの 2 回投げ試行を考えるとき，対応する標本空間と確率はそれぞれ $\Omega = \{(i,j) \,|\, i,j = 1,2,3,4,5,6\} = \{1,2,3,4,5,6\}^2$, $\mathbb{P}(A) = \frac{|A|}{36}$ により与えられる．X_1 を 1 回目の出目，X_2 を 2 回目の出目，X を二つの合計とする．すなわち，

$$X_1(i,j) = i, \quad X_2(i,j) = j, \quad X(i,j) = i + j.$$

これらは全て Ω 上の関数なので，確率変数である．

$\Omega^X := X(\Omega) = \{X(\omega) \,|\, \omega \in \Omega\}$ を X の**状態空間**と呼ぶ.
$B \subset \Omega^X$ に対して，

- $X^{-1}(B) := \{\omega \in \Omega \,|\, X(\omega) \in B\}$ を X による B の**逆像**という．
- $X^{-1}(B)$ は X の値が B に入るという事象を表す．
- $\{\omega \in \Omega \,|\, X(\omega) \in B\}$ を $\{X \in B\}$ と略記することが多い．

定理 2.2.3 X を $(\Omega, \mathbb{P})$ 上の（実）確率変数とする．このとき，$P^X(B) := \mathbb{P}(X \in B)$, $B \subset \Omega^X$, は Ω^X 上の確率である．

証明．定義 2.1.3 の 3 条件を確かめればよい．まず，任意の $B \subset \Omega^X$ に対して，$P^X(B) = \mathbb{P}(X \in B) \geq 0$．よって条件 (1) は成り立つ．$A, B \subset \Omega^X$ で $A \cap B = \emptyset$ のとき，$X(\omega) \in A$ と $X(\omega) \in B$ は同時に起こらない．よって，$\{X \in A\}$ と $\{X \in B\}$ は互いに排反であるから，

$$\begin{aligned} P^X(A \cup B) &= \mathbb{P}(X \in A \cup B) = \mathbb{P}(\{X \in A\} \cup \{X \in B\}) \\ &= \mathbb{P}(X \in A) + \mathbb{P}(X \in B) = P^X(A) + P^X(B). \end{aligned}$$

ゆえに条件 (2) も成り立つ．条件 (3) については $P^X(\Omega^X) = \mathbb{P}(X \in \Omega^X) = \mathbb{P}(X \in X(\Omega)) = 1$ より従う．□

- P^X を X の**分布**または**確率法則**という．

例題 2.2.4 サイコロ 2 回投げ試行の例における確率変数 X_1, X_2, X の分布を求めよ．

解答．$\Omega^{X_1} = \Omega^{X_2} = \{1, 2, 3, 4, 5, 6\}$,

$$P^{X_i}(\{k\}) = \mathbb{P}(X_i = k) = \mathbb{P}(\{k, j\} \mid j \in \Omega^{X_i}) = \sum_{j=1}^{6} \mathbb{P}(\{k, j\}) = \sum_{j=1}^{6} \frac{1}{36} = \frac{1}{6}, \quad k \in \Omega^{X_i}, \quad i = 1, 2$$

である．X は X_1 と X_2 の和なので，$\Omega^X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ であり，

$$P^X\{k\} = \mathbb{P}(X = k) = \mathbb{P}(X_1 + X_2 = k) = \begin{cases} \dfrac{k-1}{36} & (2 \leq k \leq 7), \\ \dfrac{13-k}{36} & (8 \leq k \leq 12). \end{cases}$$

- 上の解答のように，$\mathbb{P}(\{k\})$ を $\mathbb{P}\{k\}$ としばしば略記する．

命題 2.2.5 X を確率変数，$\varphi : \Omega^X \to \mathbb{R}$ とする．このとき，$Y(\omega) := \varphi(X(\omega))$ について，

$$\Omega^Y = \varphi(\Omega^X), \quad P^Y(B) = P^X(\varphi^{-1}(B)).$$

証明．$y \in \Omega^Y \Leftrightarrow y \in Y(\Omega) \Leftrightarrow \exists \omega \in \Omega \text{ s.t. } Y(\omega) = y \Leftrightarrow \exists \omega \in \Omega \text{ s.t. } \varphi(X(\omega)) = y \Leftrightarrow y \in \varphi(X(\Omega))$．よって $\Omega^Y = \varphi(\Omega^X)$．二つ目の主張は

$$P^Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(\varphi(X) \in B) = \mathbb{P}(X \in \varphi^{-1}(B)) = P^X(\varphi^{-1}(B))$$

より従う．□

- 確率変数 X_1, X_2 に対し，$X = (X_1, X_2)$ を (2 次元) **確率ベクトル**という．
- $\Omega^X \subset \Omega^{X_1} \times \Omega^{X_2}$.
- P^X を X_1 と X_2 の**結合分布**という．

2.3 期待値

実確率変数 X に対して，

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}\{\omega\}$$

を X の**期待値**または**平均値**という．

$A \subset \Omega$ に対し，

$$\mathbb{E}[X, A] := \sum_{\omega \in A} X(\omega)\mathbb{P}\{\omega\}.$$

確率ベクトル $X = (X_1, \ldots, X_n)$ に対し，

$$\mathbb{E}X := (\mathbb{E}X_1, \ldots, \mathbb{E}X_n) \in \mathbb{R}^n.$$

すなわち，$\mathbb{E}X$ は成分ごとに定義する．$\mathbb{E}X$ を X の期待値ベクトルという．

定理 2.3.1 実確率変数 X, Y に対して以下が成り立つ．

(1) $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y] \quad (\alpha, \beta$ は定数).
(2) $\mathbb{E}[X, A \cup B] = \mathbb{E}[X, A] + \mathbb{E}[X, B] \quad (A,\ B$ は排反事象）.
(3) $\mathbb{E}X = \sum_{x \in \Omega^X} xP^X\{x\}$.
(4) $\mathbb{E}f(X) = \sum_{x \in \Omega^X} f(x)P^X\{x\}$.
(5) $X \geq Y$ ならば $\mathbb{E}X \geq \mathbb{E}Y$.
(6) $X \geq Y$ かつ $\mathbb{E}X = \mathbb{E}Y$ ならば $\mathbb{P}(X = Y) = 1$.

証明. (1).

$$\begin{aligned}\mathbb{E}[\alpha X + \beta Y] &= \sum_{\omega \in \Omega}(\alpha X + \beta Y)(\omega)\mathbb{P}\{\omega\} = \sum_{\omega \in \Omega}(\alpha X(\omega) + \beta Y(\omega))\mathbb{P}\{\omega\} \\ &= \alpha\sum_{\omega \in \Omega} X(\omega)\mathbb{P}\{\omega\} + \beta\sum_{\omega \in \Omega} Y(\omega)\mathbb{P}\{\omega\} = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y].\end{aligned}$$

(2).

$$\begin{aligned}\mathbb{E}[X, A \cup B] &= \sum_{\omega \in A \cup B} X(\omega)\mathbb{P}\{\omega\} = \sum_{\omega \in A} X(\omega)\mathbb{P}\{\omega\} + \sum_{\omega \in B} X(\omega)\mathbb{P}\{\omega\} \\ &= \mathbb{E}[X, A] + \mathbb{E}[X, B].\end{aligned}$$

(3). Ω は $\Omega = \bigcup_{x \in \Omega^X}\{X = x\}$ と排反事象の和として表されるので，直前の (2) より，

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in \Omega^X} \mathbb{E}[X, \{X = x\}] = \sum_{x \in \Omega^X}\sum_{\omega \in \{X = x\}} X(\omega)\mathbb{P}\{\omega\} \\ &= \sum_{x \in \Omega^X} x \sum_{\omega \in \{X = x\}} \mathbb{P}\{\omega\} = \sum_{x \in \Omega^X} x\mathbb{P}(X = x).\end{aligned}$$

(4). 直前の (3) と同様に，

$$\begin{aligned}\mathbb{E}[f(X)] &= \sum_{x \in \Omega^X} \mathbb{E}[f(X), \{X = x\}] = \sum_{x \in \Omega^X}\sum_{\omega \in \{X = x\}} f(X(\omega))\mathbb{P}\{\omega\} \\ &= \sum_{x \in \Omega^X} f(x)\mathbb{P}(X = x).\end{aligned}$$

(5).

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}\{\omega\} \geq \sum_{\omega \in \Omega} Y(\omega)\mathbb{P}\{\omega\} = \mathbb{E}Y.$$

(6). 仮定より，$0=\sum_{\omega\in\Omega}(X(\omega)-Y(\omega))\mathbb{P}\{\omega\}$ であり，$X\geq Y$ であるから，$\mathbb{P}\{\omega\}>0$ なる任意の $\omega\in\Omega$ に対して $X(\omega)=Y(\omega)$ である．よって，$\mathbb{P}(X\neq Y)=0$. すなわち，$\mathbb{P}(X=Y)=1-\mathbb{P}(X\neq Y)=1$. □

- $\mathbb{P}(X=Y)=1$ のとき，X と Y は**ほとんど確実に**（almost surely）等しいといい，

  $$X=Y\ \text{ a.s.}$$

  と書く．

例題 2.3.2 $X>0$ とする．次の等式は成立するか調べよ．

$$\mathbb{E}\left[\frac{1}{X}\right]\stackrel{?}{=}\frac{1}{\mathbb{E}X}. \tag{2.3.1}$$

解答．（第 1 段階）．$X(\omega)=C>0$（定数）のときは (2.3.1) は等号で成立する．実際，

$$\mathbb{E}[X]=\sum_{\omega\in\Omega}X(\omega)\mathbb{P}\{\omega\}=C\sum_{\omega\in\Omega}\mathbb{P}\{\omega\}=C\mathbb{P}(\Omega)=C.$$

同様に，$\mathbb{E}[1/X]=1/C$.

（第 2 段階）．一般に，$X,Y\geq 0$ に対して，コーシー・シュワルツの不等式

$$\mathbb{E}[XY]\leq\sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]} \tag{2.3.2}$$

が成立する．これを確認しよう．まず，$\mathbb{E}[X]=0$ のときは定理 2.3.1 (6) の証明と同じ議論により $X=0$ a.s. 従って (2.3.2) は両辺 0 で成立する．$\mathbb{E}[Y]=0$ のときも同様である．よって，$\mathbb{E}[X]>0$ かつ $\mathbb{E}[Y]>0$ を仮定して示す．実数 $x,y\geq 0$ に対して $xy\leq x^2/2+y^2/2$ であるから，$\tilde{X}:=X/\sqrt{\mathbb{E}[X^2]}$, $\tilde{Y}:=Y/\sqrt{\mathbb{E}[Y^2]}$ について

$$\tilde{X}\tilde{Y}\leq\frac{1}{2}\tilde{X}^2+\frac{1}{2}\tilde{Y}^2.$$

定理 2.3.1 (1) と (5) より，

$$\mathbb{E}[\tilde{X}\tilde{Y}]\leq\frac{1}{2}\mathbb{E}[\tilde{X}^2]+\frac{1}{2}\mathbb{E}[\tilde{Y}^2]=1.$$

再び定理 2.3.1 (1) より $\mathbb{E}[\tilde{X}\tilde{Y}]=\mathbb{E}[XY]/\sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$ であるから，(2.3.2) が従う．

（第 3 段階）．$\sqrt{X}$ と $1/\sqrt{X}$ についてコーシー・シュワルツの不等式 (2.3.2) を適用すると，

$$1=\mathbb{E}\left[\sqrt{X}\times\frac{1}{\sqrt{X}}\right]\leq\sqrt{\mathbb{E}[X]}\sqrt{\mathbb{E}\left[\frac{1}{X}\right]}.$$

よって，

$$\mathbb{E}\left[\frac{1}{X}\right]\geq\frac{1}{\mathbb{E}[X]}. \tag{2.3.3}$$

もし逆の不等式が成立すれば (2.3.1) は等号で成立するということになるが，一般にこれは成り立たない．実際，コインの 1 回投げ試行の確率空間 $\Omega=\{0,1\}$, $\mathbb{P}\{0\}=\mathbb{P}\{1\}=1/2$ において，確率変数 $X(\omega)=2\omega+2$ を考える．このとき，

$$\mathbb{E}\left[\frac{1}{X}\right]=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{4}=\frac{3}{8},\quad\mathbb{E}[X]=\frac{1}{2}\times 2+\frac{1}{2}\times 4=3.$$

よって

$$\mathbb{E}\left[\frac{1}{X}\right]=\frac{3}{8}>\frac{1}{3}=\frac{1}{\mathbb{E}[X]}.$$

（第 4 段階）．以上より (2.3.1) は不等式 (2.3.3) で成立することが分かった．最後に，等式が成立するのはどの場合か調べよう．第 1 段階で X が定数の場合に成立するのは確かめた．仮に (2.3.3) で等号が成立しているとする．このとき，

$$\tilde{X}:=\frac{\sqrt{X}}{\sqrt{\mathbb{E}[X]}},\quad \tilde{Y}:=\frac{\sqrt{1/X}}{\sqrt{\mathbb{E}[1/X]}}$$

は

$$1=\tilde{X}\tilde{Y}\leq\frac{1}{2}\tilde{X}^2+\frac{1}{2}\tilde{Y}^2$$

かつ

$$\mathbb{E}\left[\frac{1}{2}\tilde{X}^2+\frac{1}{2}\tilde{Y}^2\right]=1$$

を満たす．従って定理 2.3.1 (6) より，ほとんど確実に

$$\tilde{X}\tilde{Y}=\frac{1}{2}\tilde{X}^2+\frac{1}{2}\tilde{Y}^2.$$

これより，$(\tilde{X}-\tilde{Y})^2=0$ a.s. よって $\tilde{X}=\tilde{Y}$．すなわち $X=\mathbb{E}[X]$ a.s. これは X が確率 1 で定数であることを意味する．ゆえに，(2.3.3) で等号が成立するのは X が確率 1 で定数のとき，およびその時に限る．

2.4　独立性

公平なコインを 2 回投げるとき，2 回目の結果は 1 回目の結果の影響を受けない．言い換えると，1 回目と 2 回目は**独立試行**であり，例えば，1 回目が表で 2 回目が裏となる確率は，1 回目が表となる確率 1/2 と 2 回目が裏となる確率 1/2 の積により与えられるのであった．

定義 2.4.1 事象 A と事象 B が**独立**であるとは，次が成り立つときにいう．

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B),\quad A,B\in 2^\Omega.$$

- 2 つの事象が互いに排反であることは独立性を意味しない．実際，$A\cap B=\emptyset$, $\mathbb{P}(A)>0$, $\mathbb{P}(B)>0$ ならば，$\mathbb{P}(A\cap B)=0<\mathbb{P}(A)\mathbb{P}(B)$ である．よって A と B は独立ではない．排反事象は「同時には起こらない」という意味で互いに依存している．
- $A_1,\ldots,A_n$ の独立性は次により定義する．

$$\mathbb{P}\left(\bigcap_{i=1}^m A_{n_i}\right)=\prod_{i=1}^m\mathbb{P}(A_{n_i}),\quad \{n_1,\ldots,n_m\}\subset\{1,\ldots,n\}.$$

例題 2.4.2 サイコロの 1 回投げ試行において，出目が偶数である事象と出目が 3 の倍数である事象は独立であることを示せ．

解答．この試行に対応する確率空間 $(\Omega, \mathbb{P})$ はもちろん $\Omega = \{1,2,3,4,5,6\}$, $\mathbb{P}\{i\} = 1/6$, $i \in \Omega$, である．出目が偶数である事象を A, 出目が 3 の倍数である事象を B で表すことにすると,

$$\mathbb{P}(A \cap B) = \mathbb{P}\{6\} = \frac{1}{6}, \quad \mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \mathbb{P}\{3,6\} = \frac{1}{3}.$$

よって

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

例 2.4.3 サイコロ 2 回投げの試行を考える．このとき，対応する確率空間 $(\Omega, \mathbb{P})$ は $\Omega = \{(i,j) \mid i,j = 1,2,3,4,5,6\}$, $\mathbb{P}(A) = |A|/36$, $A \subset \Omega$, である．事象

$$\begin{aligned} A &= \{2 \text{ 回目に } 1 \text{ または } 2 \text{ または } 5 \text{ が出る}\}, \\ B &= \{2 \text{ 回目に } 4 \text{ または } 5 \text{ または } 6 \text{ が出る}\}, \\ C &= \{2 \text{ 回の出目の和が } 9\} \end{aligned}$$

を考える．このとき,

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(C) = \frac{1}{9}$$

であり,

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{36} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

他方,

$$\begin{aligned} \mathbb{P}(A \cap B) &= \frac{1}{6} \neq \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(B \cap C) &= \frac{1}{12} \neq \frac{1}{18} = \mathbb{P}(B)\mathbb{P}(C), \\ \mathbb{P}(C \cap A) &= \frac{1}{36} \neq \frac{1}{18} = \mathbb{P}(C)\mathbb{P}(A). \end{aligned}$$

よって A, B, C は独立ではない.

定義 2.4.4 二つの確率変数 X, Y が独立であるとは，次が成り立つときにいう．

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad x \in \Omega^X, \ y \in \Omega^Y.$$

- 確率変数 $X_1, \ldots, X_n$ の独立性は次により定義する．

$$\mathbb{P}\left(\bigcap_{i=1}^{m} \{X_{n_i} = x_{n_i}\}\right) = \prod_{i=1}^{m} \mathbb{P}(X_{n_i} = x_{n_i}), \quad x_{n_i} \in \Omega^{X_{n_i}}, \ \{n_1, \ldots, n_m\} \subset \{1, \ldots, n\}.$$

定理 2.4.5 確率変数 X, Y, 事象 A, B に対して以下が成り立つ．

(1) X と Y が独立 $\iff$ $\mathbb{P}(X \in E, Y \in F) = \mathbb{P}(X \in E)\mathbb{P}(Y \in F)$, $E \subset \Omega^X$, $F \subset \Omega^Y$.

(2) A と B が独立 $\iff$ 確率変数 1_A と 1_B が独立.

(3) X と Y が独立 $\Longrightarrow$ $f(X)$ と $g(Y)$ も独立.

(4) 実確率変数 X, Y が独立 $\Longrightarrow$ $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

証明. (1) ($\Rightarrow$). $E \subset \Omega^X$, $F \subset \Omega^Y$ を任意にとる．事象 $\{X \in E, Y \in F\}$ は排反事象の和として $\{X \in E, Y \in F\} = \bigcup_{x \in E,\, y \in F} \{X = x, Y = y\}$ と表されるので,

$$\begin{aligned}\mathbb{P}(X \in E, Y \in F) &= \sum_{x \in E,\, y \in F} \mathbb{P}(X = x, Y = y) = \sum_{x \in E,\, y \in F} \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &= \sum_{x \in E} \mathbb{P}(X = x) \sum_{y \in F} \mathbb{P}(Y = y) = \mathbb{P}(X \in E)\mathbb{P}(Y \in F).\end{aligned}$$

($\Leftarrow$). 任意の $x \in \Omega^X$, $y \in \Omega^Y$ に対して，$E = \{x\}$, $F = \{y\}$ を考えればよい.

(2) ($\Rightarrow$). 事象 A と B の独立性より,

$$\mathbb{P}(1_A = 1, 1_B = 1) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(1_A = 1)\mathbb{P}(1_B = 1).$$

次に，$A = (A \cap B) \cup (A \cap B^c)$ より,

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c).$$

よって

$$\mathbb{P}(1_A = 1, 1_B = 0) = \mathbb{P}(A \cap B^c) = \mathbb{P}(A)\mathbb{P}(B^c) = \mathbb{P}(1_A = 1)\mathbb{P}(1_B = 0).$$

A と B の役割を入れ替えることで

$$\mathbb{P}(1_A = 0, 1_B = 1) = \mathbb{P}(1_A = 0)\mathbb{P}(1_B = 1)$$

も従う．さらに，定理 2.1.4 (4) と A と B の独立性より

$$\begin{aligned}\mathbb{P}(A^c \cap B^c) &= \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B) \\ &= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) \\ &= \mathbb{P}(A^c)\mathbb{P}(B^c).\end{aligned}$$

よって

$$\mathbb{P}(1_A = 0, 1_B = 0) = \mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)\mathbb{P}(B^c) = \mathbb{P}(1_A = 0)\mathbb{P}(1_B = 0).$$

($\Leftarrow$). この主張は

$$\mathbb{P}(A \cap B) = \mathbb{P}(1_A = 1, 1_B = 1) = \mathbb{P}(1_A = 1)\mathbb{P}(1_B = 1) = \mathbb{P}(A)\mathbb{P}(B)$$

より従う.

(3). $E \subset \Omega^{f(X)}$, $F \subset \Omega^{g(Y)}$ を任意にとる．命題 2.2.5 より，$\Omega^{f(X)} = f(\Omega^X)$, $\Omega^{g(Y)} = g(\Omega^Y)$ であるから，$f^{-1}(E) \subset \Omega^X$, $g^{-1}(F) \subset \Omega^Y$ である．よって本定理 (1) の結果より

$$\begin{aligned}\mathbb{P}(f(X) \in E, g(Y) \in F) &= \mathbb{P}(X \in f^{-1}(E), Y \in g^{-1}(F)) = \mathbb{P}(X \in f^{-1}(E))\mathbb{P}(Y \in g^{-1}(F)) \\ &= \mathbb{P}(f(X) \in E)\mathbb{P}(g(Y) \in F).\end{aligned}$$

(4). 定理 2.3.1 (3) の証明と同様にして,

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{\substack{x \in \Omega^X \\ y \in \Omega^Y}} \mathbb{E}[XY, \{X = x, Y = y\}] = \sum_{\substack{x \in \Omega^X \\ y \in \Omega^Y}} \sum_{\omega \in \{X = x, Y = y\}} X(\omega)Y(\omega)\mathbb{P}\{\omega\} \\ &= \sum_{\substack{x \in \Omega^X \\ y \in \Omega^Y}} xy\mathbb{P}(X = x, Y = y) = \sum_{x \in \Omega^X} x\mathbb{P}(X = x) \sum_{y \in \Omega^Y} y\mathbb{P}(Y = y) \\ &= \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

□

> **例題 2.4.6** サイコロを $m+n$ 回振るとき，始めの m 回に出た目の和と後の n 回に出た目の積は独立であることを示せ．

解答. X_i を i 回目の出目とし, $X=(X_1,\dots,X_m)$, $Y=(X_{m+1},\dots,X_{m+n})$ とおく. このとき, $f(X)$ と $g(Y)$ が独立であることを示せばよい．ただし，関数 f,g はそれぞれ $f(x_1,\dots,x_m)=\sum_{i=1}^m x_i$, $g(y_1,\dots,y_n)=\prod_{i=1}^n y_i$ により定義する．

まず，$X_1,\dots,X_{m+n}$ が独立であるから，二つの確率ベクトル X,Y も独立である．これを確かめよう．独立性の定義で $\{1,\dots,m\}\subset\{1,\dots,m+n\}$ を考えれば，$x_i\in\Omega^{X_i}$, $i=1,\dots,m$, に対して $\mathbb{P}(X_1=x_1,\dots,X_m=x_m)=\mathbb{P}(X_1=x_1)\cdots\mathbb{P}(X_m=x_m)$．同様に，$\mathbb{P}(X_{m+1}=x_{m+1},\dots,X_{m+n}=x_{m+n})=\mathbb{P}(X_{m+1}=x_{m+1})\cdots\mathbb{P}(X_{m+n}=x_{m+n})$, $\mathbb{P}(X_1=x_1,\dots,X_{m+n}=x_{m+n})=\mathbb{P}(X_1=x_1)\cdots\mathbb{P}(X_{m+n}=x_{m+n})$ が任意の $x_i\in\Omega^{X_i}$, $i=m+1,\dots,m+n$, に対して成り立つ．よって，

$$
\begin{aligned}
&\mathbb{P}(X=(x_1,\dots,x_m),Y=(x_{m+1},\dots,x_{m+n}))\\
&=\mathbb{P}(X_1=x_1,\dots,X_m=x_m,X_{m+1}=x_{m+1},\dots,X_{m+n}=x_{m+n})\\
&=\prod_{i=1}^{m+n}\mathbb{P}(X_i=x_i)\\
&=\mathbb{P}(X_1=x_1,\dots,X_m=x_m)\mathbb{P}(X_{m+1}=x_{m+1},\dots,X_{m+n}=x_{m+n})\\
&=\mathbb{P}(X=(x_1,\dots,x_m))\mathbb{P}(Y=(x_{m+1},\dots,x_{m+n})).
\end{aligned}
$$

従って，定理 2.4.5 (3) より $f(X)$ と $g(Y)$ は独立である．

2.5　条件付き確率

引き続き，$(\Omega,\mathbb{P})$ を試行 $\mathbf{T}$ に対応する確率空間とする．

$\mathbb{P}(A)>0$ なる事象 A と事象 B とに対し，A の下での B の**条件付き確率** $\mathbb{P}(B\,|\,A)$ を

$$\mathbb{P}(B\,|\,A)=\frac{\mathbb{P}(B\cap A)}{\mathbb{P}(A)}$$

により定義する．これは A が起こる前提で B が起こる確率と解釈できる．

> **例題 2.5.1** People v. Collins 事件（例 1.2.1）を確率空間を用いて記述し，等式 (1.2.1) を正当化せよ．

解答. この場合，試行は計 n 組のカップルから 1 組を無作為抽出することである．対応する確率空間は $\Omega=\{1,\dots,n\}$, $\mathbb{P}\{i\}=1/n$, $i\in\Omega$, である．「ブロンドヘアをポニーテールに束ねた若い女性とあごひげを生やした黒人のカップルで，黄色い車に乗って黒人が運転している」という特徴を $(*)$ とおく．$A_i\subset\Omega$ をカップル i が特徴 $(*)$ に合致する事象とすると，仮定より $\mathbb{P}(A_i)=p$ が各 $i\in\Omega$ に

対して成り立つ．事象 A,B,C を

$$\begin{aligned} A &= \{\text{最低1組が特徴}(*)\text{に合致}\},\\ B &= \{\text{最低2組が特徴}(*)\text{に合致}\},\\ C &= \{\text{ちょうど1組だけ特徴}(*)\text{に合致}\} \end{aligned}$$

と定めると，求める確率は $\mathbb{P}(B\,|\,A)$ であり，C は排反事象の和として

$$C = (A_1 \cap A_2^c \cap A_3^c \cap \cdots \cap A_n^c) \cup (A_1^c \cap A_2 \cap A_3^c \cap \cdots \cap A_n^c) \cup \cdots \cup (A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n)$$

と表される．よって，カップル間の独立性より

$$\begin{aligned} \mathbb{P}(C) &= \mathbb{P}(A_1 \cap A_2^c \cap A_3^c \cap \cdots \cap A_n^c) + \mathbb{P}(A_1^c \cap A_2 \cap A_3^c \cap \cdots \cap A_n^c)\\ &\quad + \cdots + \mathbb{P}(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n)\\ &= \mathbb{P}(A_1)\mathbb{P}(A_2^c)\cdots\mathbb{P}(A_n^c) + \mathbb{P}(A_1^c)\mathbb{P}(A_2)\mathbb{P}(A_3^c)\cdots\mathbb{P}(A_n^c) + \cdots + \mathbb{P}(A_1^c)\cdots\mathbb{P}(A_{n-1}^c)\mathbb{P}(A_n)\\ &= np(1-p)^{n-1}. \end{aligned}$$

他方，$A = \bigcup_{i=1}^n A_i$ と A_i たちの独立性より，

$$\mathbb{P}(A^c) = \mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right) = \prod_{i=1}^n \mathbb{P}(A_i^c) = (1-p)^n.$$

よって定理 2.1.4 より，

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - (1-p)^n.$$

さらに，

$$B = A \cap C^c, \quad A \supset C$$

と定理 2.1.4 より

$$\mathbb{P}(B) = \mathbb{P}(A) - \mathbb{P}(C) = 1 - (1-p)^n - np(1-p)^{n-1}.$$

以上より，

$$\mathbb{P}(B\,|\,A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{1-(1-p)^n - np(1-p)^{n-1}}{1-(1-p)^n}.$$

条件付き確率は確率測度を定義する．

定理 2.5.2 $\mathbb{P}(A) > 0$ とする．このとき，$\mathbb{P}_A(B) := \mathbb{P}(B\,|\,A)$, $B \subset \Omega$, は Ω 上の確率である．

証明．定義 2.1.3 の 3 条件を確かめればよい．任意の $B \subset \Omega$ に対して，明らかに

$$\mathbb{P}_A(B) = \mathbb{P}(B\,|\,A) \geq 0.$$

$B \cap C = \emptyset$ なる任意の $B, C \subset \Omega$ に対して，

$$\mathbb{P}_A(B \cup C) = \frac{\mathbb{P}((B \cap A) \cup (C \cap A))}{\mathbb{P}(A)} = \frac{\mathbb{P}(B \cap A) + \mathbb{P}(C \cap A)}{\mathbb{P}(A)} = \mathbb{P}_A(B) + \mathbb{P}_A(C).$$

さらに，

$$\mathbb{P}_A(\Omega) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)} = 1.$$

ゆえに定理の主張が従う． □

ベイズの公式あるいは**ベイズの定理**の名で呼ばれる次の結果は，与えられた「結果」A の下で，「原因」B の起こる確率を計算する方法であり，応用上特に有用である．

定理 2.5.3 (**ベイズの公式**) $\mathbb{P}(A\cap B)>0$ のとき，

$$\mathbb{P}(B\,|\,A)=\frac{\mathbb{P}(A\,|\,B)\mathbb{P}(B)}{\mathbb{P}(A)} \tag{2.5.1}$$

証明. $A\cap B\subset A$ と定理 2.1.4 (2) より，$\mathbb{P}(A)\geq\mathbb{P}(A\cap B)>0$. 同様に $\mathbb{P}(B)>0$. 従って，(2.5.1) の両辺は意味を持ち，

$$\frac{\mathbb{P}(A\,|\,B)\mathbb{P}(B)}{\mathbb{P}(A)}=\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}\cdot\frac{\mathbb{P}(B)}{\mathbb{P}(A)}=\mathbb{P}(B\,|\,A).$$

□

例題 2.5.4 あなたは次のようなゲームに参加する．箱 $1,\ldots,N$ の一つに 100 万円が入っており，残りは「たわし」である．あなたは箱を一つ選ぶことができる．司会者はどれが当たりか知っている．例えば，あなたは箱 i を選ぶとする．この時点で当たりの確率は $1/N$ である．司会者は箱 i の中身を見せる前に箱 i 以外の箱を開ける．そこにはたわしが入っている（司会者は当たりの箱は開けない）．さて，この時点での当たりの確率について，正しいのはどちらか？

(1) 外れの箱が一つ減ったから，当たりの確率は $1/(N-1)$ に上がった．
(2) 司会者は外れの箱しか開けないから，当たりの確率は変わらない．

解答. 箱 1 が当たりで，選ばれ方は等確率で決まるとして解答を考えよう．A_i をあなたが箱 i を選ぶ事象，B_j を司会者が $j\in\{2,\ldots,N\}$ を選ぶ事象とする．このとき，$\mathbb{P}(A_i)=1/N$ であり，

$$\mathbb{P}(B_j\,|\,A_i)=\begin{cases}\dfrac{1}{N-1}, & (i=1),\\ \dfrac{1}{N-2}, & (i\neq 1)\end{cases}$$

である．ベイズの公式（定理 2.5.3）より，

$$\begin{aligned}\mathbb{P}(A_1\,|\,B_j)&=\frac{\mathbb{P}(B_j\,|\,A_1)\mathbb{P}(A_1)}{\mathbb{P}(B_j)}=\frac{\mathbb{P}(B_j\,|\,A_1)\mathbb{P}(A_1)}{\sum_{i\neq j}\mathbb{P}(B_j\,|\,A_i)\mathbb{P}(A_i)}\\&=\frac{\frac{1}{N-1}\cdot\frac{1}{N}}{\frac{1}{N-1}\cdot\frac{1}{N}+\sum_{\substack{i\neq j\\ i\neq 1}}\frac{1}{N-2}\cdot\frac{1}{N}}=\frac{\frac{1}{N-1}\cdot\frac{1}{N}}{\frac{N}{N-1}\cdot\frac{1}{N}}\\&=\frac{1}{N}.\end{aligned}$$

ゆえに，答えは (2) で，当たりの確率は変わらない．

2.6　大数の法則

ここでは $(\Omega,\mathbb{P})$ を n 回コイン投げ試行の確率空間とする．$\{X_i\}_{i=1}^n$ を**独立同分布**（Independently and Identically Distributed, IID）で，$\mathbb{P}(X_1=1)=\mathbb{P}(X_1=0)=1/2$ を満たすとする．すなわち，

確率変数 $X_1, \ldots, X_n$ は独立で，各 X_i の分布は $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = 1/2$ により与えられるとする．そして，X_i たちの算術平均

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

を考える．

定理 2.6.1 (ベルヌーイ（Bernoulli）の大数の法則) 任意の $\varepsilon > 0$ に対して，

$$\lim_{n\to\infty} \mathbb{P}\left(\left|S_n - \frac{1}{2}\right| > \varepsilon\right) = 0.$$

$\mathbb{E}X_1 = 1/2$ であるから，この定理は S_n が $\mathbb{E}X_1$ と「ずれる」確率が小さいということを述べている．平たく述べると，n が十分大きいとき，コインの表と裏は（ほぼ）同じ割合で出現するという，常識的に当然と思われることが数学的にも正しいということを示している．

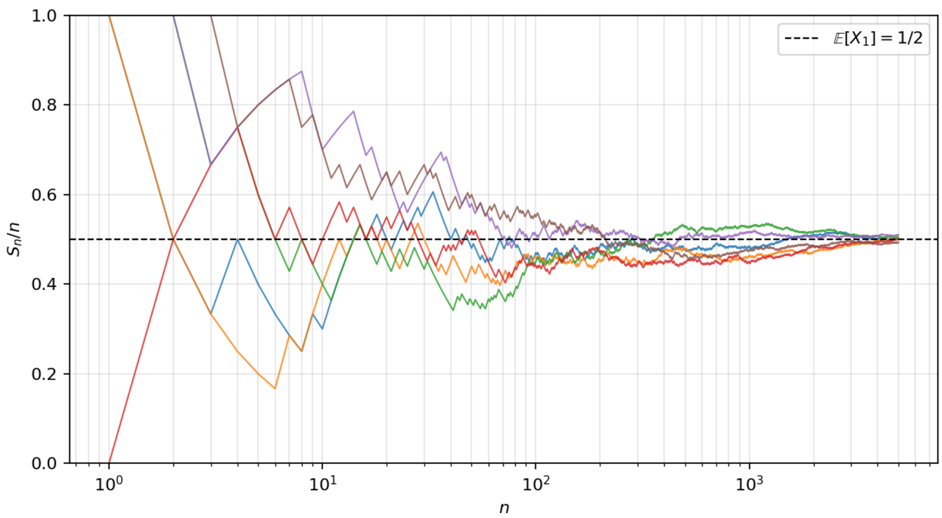


図 2.1　コイン投げにおける S_n/n の標本軌道 (6 本)．n が大きくなるにつれ S_n/n が $\mathbb{E}X_1 = 1/2$ に近づく様子．横軸は対数スケール．

定理 2.6.1 の証明．$A_n = \{|S_n - \mathbb{E}X_1| > \varepsilon\}$ とおく．$\omega \in A_n$ のとき，$1 < |S_n(\omega) - \mathbb{E}X_1|/\varepsilon$ だから，

$$\begin{aligned}
\mathbb{P}(A_n) &= \sum_{\omega\in A_n} \mathbb{P}\{\omega\} \le \sum_{\omega\in A_n} \frac{|S_n(\omega) - \mathbb{E}X_1|^2}{\varepsilon^2}\mathbb{P}\{\omega\} \le \sum_{\omega\in\Omega} \frac{|S_n(\omega) - \mathbb{E}X_1|^2}{\varepsilon^2}\mathbb{P}\{\omega\} \\
&= \frac{1}{\varepsilon^2}\mathbb{E}|S_n - \mathbb{E}X_1|^2 = \frac{1}{\varepsilon^2}\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}X_i)\right)^2\right] \\
&= \frac{1}{\varepsilon^2 n^2}\left\{\sum_{i=1}^n \mathbb{E}|X_i - \mathbb{E}X_i|^2 + \sum_{i\ne j} \mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)]\right\}.
\end{aligned}$$

ここで $\mathbb{V}[X_i] := \mathbb{E}|X_i - \mathbb{E}X_i|^2 = 1/4$ であり，定理 2.4.5 (3), (4) より，$i \neq j$ に対して

$$\mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)] = \mathbb{E}[X_i - \mathbb{E}X_i]\mathbb{E}[X_j - \mathbb{E}X_j] = 0.$$

従って，$n \to \infty$ のとき，

$$\mathbb{P}(A_n) \leq \frac{\mathbb{V}[X_1]}{\varepsilon^2 n} \to 0.$$

□

章末問題

問題 2.1 (1) $(\Omega, \mathbb{P})$ をある有限試行 $\mathbf{T}$ に対応する確率空間とする．$A_1, A_2, A_3 \subset \Omega$ に対して

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$$

が成り立つことを証明せよ．

(2) 300 人の社会人のうち，185 人は英会話クラス，90 人はプログラミング講座，160 人はビジネスマナー講座を受講している．英会話とプログラミングの両方を受講しているのは 48 人，プログラミングとビジネスマナーの両方は 42 人，英会話とビジネスマナーの両方は 103 人，3 講座すべてを受講しているのは 27 人である．
今，300 人から 1 人を無作為に選ぶ．その人が少なくとも 1 つの講座を受講している確率はいくらか？

問題 2.2 n 頭出馬の競馬を考える．馬 k のオッズは $1 + \pi_k$ である．すなわち，馬 k に A 円賭けた場合の払い戻し金は

$$\begin{cases} (1+\pi_k)A & (k\text{が 1 位}), \\ 0 & (k\text{が 2 位以下}). \end{cases}$$

今，

$$\sum_{k=1}^{n} \frac{1}{1+\pi_k} < 1$$

が成り立っていると仮定する．このとき，あなたならどう賭けるか？

$(\Omega, \mathbb{P})$ を有限試行の確率空間とし，任意の $\omega \in \Omega$ に対して $\mathbb{P}\{\omega\} > 0$ を仮定する．空でない事象 $A \subset \Omega$ と確率変数 X に対し，

$$\mathbb{E}[X \mid A] = \frac{\mathbb{E}[X, A]}{\mathbb{P}(A)}$$

を事象 A の下での X の**条件付き期待値**という．さらに，確率変数 X, Y に対し，

$$\mathbb{E}[X \mid Y] = \sum_{y \in \Omega^Y} \mathbb{E}[X \mid Y = y] 1_{\{Y=y\}}$$

を Y が与えられた下での X の条件付き期待値という．

問題 2.3 $(\Omega, \mathbb{P})$ を有限試行の確率空間とし，任意の $\omega \in \Omega$ に対して $\mathbb{P}\{\omega\} > 0$ を仮定する．X を実確率変数，Y は確率ベクトルとし，X と Y は独立とする．このとき，$f : \Omega^Y \to \mathbb{R}$ に対し，

$$\mathbb{E}[f(Y)X \mid Y] = f(Y)\mathbb{E}[X]$$

が成り立つことを示せ．

問題 2.4 N 回コイン投げ試行を考える．各コイン投げは独立とし，表の出る確率を $p \in (0, 1)$ とする．この試行に対応する確率空間を $(\Omega, \mathbb{P})$ とし，その上の確率変数列 $S_0, S_1, \ldots, S_N$ を次のように定義する:

$$S_n = S_{n-1} \times \begin{cases} u & (n\text{回目が表}), \\ d & (n\text{回目が裏}). \end{cases}$$

ただし，$0 < d < 1 < u$ で，S_0 は定数とする．

(1) $\mathbb{E}[S_3 \mid S_1]$ を求めよ．

(2) 任意の $n = 1, \ldots, N$ に対して $\mathbb{E}[S_n \mid S_{n-1}] = S_{n-1}$ が成り立つための必要十分条件を求めよ．

第 3 章

確率論の基本的枠組み

$(\Omega_n, \mathbb{P}_n)$ を n 回コイン投げ試行に関する確率空間とする．S_n を表の出る割合とすると，定理 2.6.1（ベルヌーイの大数の法則）より $\lim_{n\to\infty} \mathbb{P}_n(|S_n - 1/2| > \varepsilon) = 0$ であった．直感的には，もっと強く，「ω ごとに」$\lim_{n\to\infty} S_n(\omega) = 1/2$ が成立しそうである．しかし，数学的には，Ω_n は n 回までしか定義されていないので，n 無限大のときの標本は考えられない．この問題を解決するため，$\Omega_\infty = \{\omega_1\omega_2\cdots \mid \omega_i = 0\text{または}1\}$ を考えればよいのかもしれない．このとき，例えば，全て表が出る確率は $(1/2)^\infty = 0$ となる．後で示すように，この Ω_∞ は非可算無限集合なので，前章の枠組みでは，無限個の標本を持つ事象 A に対して，A が起こる確率 $= |A|/2^\infty = \infty \times 0$ となり意味がわからない．$\mathbb{P}(\Omega_\infty) = 1$, $\mathbb{P}\{\omega\} = 0$ であり，$\mathbb{P}(n\text{回目まで全て裏}) = 1/2^n$ 等，直感に合っていて，かつ $\mathbb{P}(\lim_{n\to\infty} S_n = 1/2)$ を考えられるような確率測度を定義したい．前章より一般的な理論を構築することでこの問いに答える．

3.1 無限集合おさらい

A を任意の集合とする．

- A が**有限集合**であるとは，A に属する元の個数 $|A|$ が有限個のときにいう．
- 空集合も有限集合と考える．
- 有限集合でない集合を**無限集合**という．
- A が**可算集合**であるとは，A から自然数全体 $\mathbb{N}$ への全単射が存在するときにいう．
- A が**高々可算集合**であるとは，A が有限集合または可算集合のときにいう．
- A が**非可算集合**であるとは，A が高々可算集合でないときにいう．

例題 3.1.1 整数全体の集合 $\mathbb{Z}$ は可算集合であることを確かめよ．

解答．$\mathbb{N}$ から $\mathbb{Z}$ への全単射が存在することを示せばよい．例えば，

$$f(n) = \begin{cases} 0 & (n = 1), \\ -k & (n = 2k,\ k \geq 1), \\ k & (n = 2k+1,\ k \geq 1) \end{cases}$$

と定義すると，$f : \mathbb{N} \to \mathbb{Z}$ は全単射である．

> **補題 3.1.2** (1) A が高々可算集合，$B \subset A$ ならば，B も高々可算集合
> (2) A, B が高々可算集合ならば，$A \times B$ も高々可算集合
> (3) 各 $n \in \mathbb{N}$ に対して A_n が高々可算集合ならば，$\bigcup_{n=1}^{\infty} A_n$ も高々可算集合

証明．(1)．A が有限集合のときは明らか．A が可算集合とすると，A は適当な全単射 $f : A \to \mathbb{N}$ の逆像なので，$A = \{a_n : n \in \mathbb{N}\}$ の形で表される．$a_n \in B$ なる n だけ抜き出し，$\mathbb{N}$ の部分列 $n(1) < n(2) < \cdots$ を作ると，

$$\{n(k) : k \in \mathbb{N}\} = \{n \in \mathbb{N} : a_n \in B\}$$

である．よって $B = \{a_{n(k)} : k \in \mathbb{N}\}$ と表されるので，これは可算集合である．A が有限集合の場合と併せて，結局，B は高々可算集合である．

(2)．$A = \{a_n : n \in \mathbb{N}\}$, $B = \{b_n : n \in \mathbb{N}\}$ とするとき，

$$A \times B = \{(a_k, b_\ell) : k, \ell \in \mathbb{N}\}$$

と書ける．$f : A \times B \to \mathbb{N}$ を

$$f(a_k, b_\ell) = \frac{1}{2}((k + \ell - 2)(k + \ell - 1) + 2\ell) \in \mathbb{N}, \quad (a_k, b_\ell) \in A \times B,$$

により定義すると，これは全単射である．実際，いくつか計算してみると，

$$f(a_1, b_1) = 1,\ \ f(a_2, b_1) = 2,\ \ f(a_1, b_2) = 3,$$
$$f(a_3, b_1) = 4,\ \ f(a_2, b_2) = 5,\ \ f(a_1, b_3) = 6.$$

一般の k, ℓ に対して，$k + \ell = n$ は固定し，$k > 1$ から $k = 1$ になるまで k を 1 つずつ減らしていくとき，$f(a_k, b_\ell) = ((n-2)(n-1) + 2(n-k))/2 = N$ とおけば，$k > 1$ のとき，

$$f(a_{k-1}, b_{\ell+1}) = \frac{1}{2}((n-2)(n-1) + 2(n-k+1)) = N + 1$$

となる．$k = 1$ まで来たら $k + \ell$ を 1 つ増やし，$k = n$ とする．すると，

$$f(a_n, b_1) = \frac{1}{2}((n+1-2)(n+1-1) + 2(n+1-n)) = \frac{1}{2}((n-2)(n-1) + 2(n-1) + 2) = N + 1.$$

以上により全射性が分かる．次に単射性を確認しよう．$T(m) := (m-2)(m-1)/2$ とおくと，$\ell = n - k$, $\ell' = n' - k'$ を用いて $f(a_k, b_\ell) = T(n) + \ell$ と書ける．$f(a_k, b_\ell) = f(a_{k'}, b_{\ell'})$，すなわち

$$T(n) + \ell = T(n') + \ell', \quad k, \ell, k', \ell' \geq 1, \tag{3.1.1}$$

が成り立つと仮定する．T は $m \geq 2$ 上で狭義単調増加で $T(m+1) - T(m) = m - 1$ を満たす．$n = n'$ なら $T(n) = T(n')$ より $\ell = \ell'$，よって $k = k'$．$n \neq n'$ と仮定して矛盾を導く．対称性より $n > n'$，すなわち $n \geq n' + 1$ としてよい．このとき

$$T(n) - T(n') \geq T(n'+1) - T(n') = n' - 1.$$

他方，(3.1.1) より $T(n) - T(n') = \ell' - \ell \leq \ell' - 1 \leq n' - 2$（最後は $k' \geq 1$ より $\ell' \leq n' - 1$）．以上を併せると $n' - 1 \leq n' - 2$ となり矛盾．よって $n = n'$, $k = k'$.

(3). 必要なら有限集合を除外すればよいので，A_n たちは全て可算集合と仮定してよい．集合列

$$B_1 := A_1, \quad B_n := A_n \setminus \bigcup_{j=1}^{n-1} A_j, \quad n \geq 2,$$

を考える．このとき，$\{B_n\}$ は互いに排反で，各 B_n が高々可算集合である．さらに，

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

各 B_n を $B_n = \{b_{n,k} : k \in \mathbb{N}\}$ と表すとき，$f(b_{n,k}) := (n,k)$ は $\bigcup_{n=1}^{\infty} B_n$ から $\mathbb{N} \times \mathbb{N}$ への単射である．他方，$f(\bigcup_{n=1}^{\infty} B_n) \subset \mathbb{N} \times \mathbb{N}$ であるから，本補題 (1) より，$f(\bigcup_{n=1}^{\infty} B_n)$ は高々可算であり，f は単射なので，$\bigcup_{n=1}^{\infty} B_n$ も高々可算集合である． □

命題 3.1.3 有理数全体 $\mathbb{Q}$ は可算集合である．

証明. $\mathbb{Q} = \{n/m : m, n \in \mathbb{N}\}$ と表される．ただし n/m は既約分数とする．$f : \mathbb{Q} \to \mathbb{N} \times \mathbb{N}$ を $f(n/m) = (m,n)$ により定義すると，$f(\mathbb{Q}) \subset \mathbb{N} \times \mathbb{N}$ であり，f は単射であるから，補題 3.1.2 より，$\mathbb{Q}$ は高々可算集合である．$\mathbb{Q}$ は明らかに有限集合ではないので，可算集合である． □

命題 3.1.4 無限回コイン投げ試行の標本空間 $\Omega = \{\omega_1\omega_2\cdots \mid \omega_n \in \{0,1\}, n \in \mathbb{N}\}$ は非可算無限集合である．

証明. Ω が有限集合でないのは明らか．Ω が可算集合と仮定すると，$\Omega = \{\omega^{(1)}, \omega^{(2)}, \cdots\}$ と書ける．各 $\omega^{(i)}$ を

$$\omega^{(i)} = \omega_1^{(i)}\omega_2^{(i)}\omega_3^{(i)}\cdots$$

と表すことにする．このとき，$\hat{\omega} = \hat{\omega}_1\hat{\omega}_2\hat{\omega}_3\cdots$ を

$$\hat{\omega}_i = \begin{cases} 0 & (\omega_i^{(i)} = 1 \text{のとき}), \\ 1 & (\omega_i^{(i)} = 0 \text{のとき}), \end{cases}$$

と定義すると，$\hat{\omega}$ は $\omega^{(1)}, \omega^{(2)}, \cdots$ の何とも一致しない．すなわち，$\hat{\omega} \notin \Omega$．これは Ω が可算集合であることに矛盾する． □

3.2　確率が意味をもつ集合

後述するように，無限回コイン投げのための標本空間 Ω に適切な確率測度を定義することができる．しかし Ω の全ての部分集合に対して確率を定義することは実はできない．一般には，$\{\text{確率が意味をもつ集合}\} \subsetneq 2^{\Omega}$ である．

以下，Ω を任意の空でない集合とする．

定義 3.2.1 Ω の部分集合族 $\mathcal{F}$ が σ-**集合体**（σ-field）または σ-**加法族**（σ-algebra）とは次の条件が満たされるときにいう．

(1) $\Omega \in \mathcal{F}$.
(2) $A \in \mathcal{F}$ ならば $A^c \in \mathcal{F}$.
(3) $A_1, A_2, \ldots \in \mathcal{F}$ ならば $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$.

- 2^Ω, $\{\emptyset, \Omega\}$ は共に Ω 上の σ-集合体である．

例題 3.2.2 $A \subset \Omega$ とする．このとき，$\sigma(A) := \{\emptyset, A, A^c, \Omega\}$ は Ω 上の σ-集合体であることを確かめよ．

解答．略.

定理 3.2.3 $\mathcal{A}$ を Ω の任意の部分集合族とする．このとき，$\mathcal{A}$ を含む最小の σ-集合体 $\sigma[\mathcal{A}]$ が一意的に存在する．

- $\sigma[\mathcal{A}]$ を $\mathcal{A}$ から生成される Ω 上の σ-集合体と呼ぶ.

定理 3.2.3 の証明．次の (A1), (A2), (A3) を満たす σ-集合体 $\sigma[\mathcal{A}]$ が一意に存在することを示す．

(A1) $\mathcal{A} \subset \sigma[\mathcal{A}]$.
(A2) $\sigma[\mathcal{A}]$ は Ω 上の σ-集合体である．
(A3) $\mathcal{A}$ を含む Ω 上の σ-集合体 $\mathcal{B}$ に対し，$\sigma[\mathcal{A}] \subset \mathcal{B}$ である．

この目的のため，

$$\mathcal{H} = \{\mathcal{B} \subset 2^\Omega \mid \mathcal{B} \supset \mathcal{A},\ \mathcal{B}\text{は } \sigma\text{-集合体}\}$$

を考える．2^Ω は $\mathcal{A}$ を含む σ-集合体であるから，$2^\Omega \in \mathcal{H}$. 特に，$\mathcal{H} \neq \emptyset$. $\mathcal{H}$ を $\mathcal{H} = \{\mathcal{B}_\lambda\}_{\lambda \in \Lambda}$ と表すことにして，

$$\sigma[\mathcal{A}] := \bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda$$

を考える．これが求めるものであることを示そう．

まず，定義より，任意の $\lambda \in \Lambda$ に対して，$\mathcal{A} \subset \mathcal{B}_\lambda$ である．よって，$\mathcal{A} \subset \bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda = \sigma[\mathcal{A}]$. すなわち (A1) は満たされる．

次に，$\mathcal{B}$ を $\mathcal{A}$ を含む σ-集合体とする．定義より，$\sigma[\mathcal{A}]$ はそのような σ-集合体の共通部分であるから，$\sigma[\mathcal{A}] \subset \mathcal{B}$. すなわち (A3) が成り立つ．

(A2) を示すために $\sigma[\mathcal{A}]$ が定義 3.2.1 の 3 条件を満たすか確認しよう. 各 $\mathcal{B}_\lambda$ は σ-集合体なので, 任意の $\lambda \in \Lambda$ に対して $\Omega \in \mathcal{B}_\lambda$. よって $\Omega \in \bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda = \sigma[\mathcal{A}]$ である．$A \in \sigma[\mathcal{A}]$ ならば，任意の $\lambda \in \Lambda$ に対して $A \in \mathcal{B}_\lambda$ である．$\mathcal{B}_\lambda$ は σ-集合体であるから，$A^c \in \mathcal{B}_\lambda$. よって $A^c \in \bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda = \sigma[\mathcal{A}]$ である．$A_n \in \sigma[\mathcal{A}]$, $n \in \mathbb{N}$, とするとき，任意の $n \in \mathbb{N}$, $\lambda \in \Lambda$ に対して $A_n \in \mathcal{B}_\lambda$ である．再び各 $\mathcal{B}_\lambda$ が σ-集合体であることを使うと，各 λ に対して $\bigcup_{n=1}^\infty A_n \in \mathcal{B}_\lambda$. ゆえに $\bigcup_{n=1}^\infty A_n \in \bigcap_{\lambda \in \Lambda} \mathcal{B}_\lambda = \sigma[\mathcal{A}]$.

最後に一意性を示そう．$\mathcal{C}$ を $\mathcal{A}$ を含む最小の σ-集合体とする．すなわち，$\mathcal{C}$ は (A1), (A2), (A3) を満たすとする．このとき，$\sigma[\mathcal{A}]$ の最小性と $\mathcal{C} \supset \mathcal{A}$ より，$\sigma[\mathcal{A}] \subset \mathcal{C}$ である．$\mathcal{C}$ と $\sigma[\mathcal{A}]$ の役割を入れ替えて $\mathcal{C} \subset \sigma[\mathcal{A}]$ を得る．従って，$\mathcal{C} = \sigma[\mathcal{A}]$. □

次に，$\mathbb{R}$ 上の区間の集合族

$$\mathcal{I} := \{(a,b] \mid -\infty \le a \le b \le +\infty\}$$

を導入する．ただし，$(a,+\infty]$ は $(a,+\infty)$ を表すものとする．

- $\mathcal{B}(\mathbb{R}) := \sigma[\mathcal{I}]$ を（1 次元）**ボレル集合体**という．
- $\mathcal{B}(\mathbb{R})$ に属する集合を（1 次元）**ボレル集合**という．

任意の $a, b \in \mathbb{R}$ に対して，$[a,b], (a,b), [a,b) \in \mathcal{B}(\mathbb{R})$ である．実際,

$$(a,b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right]$$

と $\mathcal{B}(\mathbb{R})$ が σ-集合体であることを使えば分かる．他も同様．従って，$\mathcal{B}(\mathbb{R})$ は高々可算個の区間から作られる集合族と解釈してよい．

例題 3.2.4 $\{\text{無理数}\} \in \mathcal{B}(\mathbb{R})$ を示せ．

解答．任意の $x \in \mathbb{R}$ と $n \in \mathbb{N}$ に対して，$(x - 1/n, x] \in \mathcal{B}(\mathbb{R})$ である．よって，$\{x\} = \bigcap_{n=1}^{\infty}(x - 1/n, x] \in \mathcal{B}(\mathbb{R})$．すなわち，1 点集合はボレル集合である．命題 3.1.3 より $\mathbb{Q}$ は可算集合であるから $\mathbb{Q} = \{q_1, q_2, \cdots\}$ と書ける．ゆえに $\mathbb{Q} = \bigcup_{n=1}^{\infty}\{q_n\} \in \mathcal{B}(\mathbb{R})$ であり，従って，$\mathbb{Q}^c \in \mathcal{B}(\mathbb{R})$.

多次元のボレル集合を定義しておこう．1 次元の場合と同様に $\mathbb{R}^d$ の集合族

$$\mathcal{I}_d := \{(a_1,b_1] \times (a_2,b_2] \times \cdots \times (a_d,b_d] \mid -\infty \le a_i \le b_i \le +\infty,\ i = 1,\ldots,d\}$$

を導入する．

- $\mathcal{B}(\mathbb{R}^d) := \sigma[\mathcal{I}_d]$ を d 次元ボレル集合体という．
- $\mathcal{B}(\mathbb{R}^d)$ に属する集合を d 次元ボレル集合という．
- $S \subset \mathbb{R}^d$ に対して，$\mathcal{B}(S) := \mathcal{B}(\mathbb{R}^d) \cap S$ は S 上の σ-集合体になる．

3.3 確率測度・確率分布

Ω を空でない集合，$\mathcal{F}$ を Ω 上の σ-集合体とする．

定義 3.3.1 $\mathcal{F}$ 上の集合関数 $\mathbb{P}$ で次の条件を満たすものを $(\Omega, \mathcal{F})$ 上の**確率測度**または単に**確率**という．

(1) 任意の $A \in \mathcal{F}$ に対して $0 \le \mathbb{P}(A) \le 1$.

(2) $\mathbb{P}(\Omega) = 1$.

(3) σ-加法性: $A_i \in \mathcal{F}$, $A_i \cap A_j = \emptyset$ $(i \ne j)$, $i, j \in \mathbb{N}$ のとき，

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

- 有限試行のときの加法性（定義 2.1.3）が「σ-加法性」に置き換わっていることに注意せよ．
- Ω が有限集合のとき，確率の加法性と σ-加法性は同値である．
- 3 つ組 $(\Omega, \mathcal{F}, \mathbb{P})$ を**確率空間**という．
- 確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ を考えているとき，Ω を**標本空間**，$A \in \mathcal{F}$ を**事象**と呼ぶ．

確率測度の例をいくつか挙げておこう．確率変数の状態空間上で定義される確率測度を**確率分布**と呼ぶことが多い．

例 3.3.2 (離散一様分布) $N \ge 2$ を自然数とする．$S := \{1, 2, \ldots, N\}$ を考え，2^S 上の集合関数 μ を

$$\mu(A) = \frac{|A|}{N}, \quad A \in 2^S,$$

により定義する．このとき，$(S, 2^S, \mu)$ は確率空間である．μ を**離散一様分布**と呼ぶ．$N = 2$ のとき，（公平な）コイン投げの試行に対応し，$N = 6$ のときは（公平な）サイコロ投げに対応する．

例 3.3.3 (ベルヌーイ分布) $0 \le p \le 1$ とする．$2^{\{0,1\}}$ 上の集合関数 μ を

$$\mu(\emptyset) = 0, \ \ \mu\{0\} = 1 - p, \ \ \mu\{1\} = p, \ \ \mu(\{0, 1\}) = 1,$$

により定義すると，$(\{0, 1\}, 2^{\{0,1\}}, \mu)$ は確率空間であり，これは成功／失敗の試行に対応する．μ を**ベルヌーイ（二項）分布**と呼ぶ．

例 3.3.4 (ポアソン分布) $S := \mathbb{N} \cup \{0\} = \{0, 1, 2, \ldots\}$ を考え，各 $n \in S$ に対して $p_n = e^{-cn} c^n / (n!)$ とおく．ただし c は正定数とする．2^S 上の集合関数 μ を

$$\mu(A) := \sum_{n \in A} p_n, \quad A \in 2^S,$$

により定義する．このとき $(S, 2^S, \mu)$ は確率空間である．μ を**ポアソン分布**と呼ぶ．ポアソン分布は件数や人数，品数などをモデル化するときによく用いられる．

例 3.3.5 (一様分布) $a, b \in \mathbb{R}$ は所与で，$a < b$ を満たすとする．このとき，$([a, b], \mathcal{B}[a, b])$ 上の確率測度 μ で，

$$\mu((x, y]) = \frac{y - x}{b - a}, \quad a \le x \le y \le b,$$

を満たすものが一意に存在する（証明略）．μ を $[a, b]$ 上の**一様分布**と呼ぶ．$[a, b] = [0, 1]$ のときの μ を特に**ルベーグ測度**という．

例 3.3.6 (指数分布) 所与のパラメータ $\lambda > 0$ に対し，$((0,\infty), \mathcal{B}(0,\infty))$ 上の確率測度 μ で，

$$\mu((a,b]) = \int_a^b \lambda e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}, \quad 0 \le a < b,$$

を満たすものが一意に存在する（証明略）．μ を**指数分布**と呼ぶ．関数

$$\rho(x) := \lambda e^{-\lambda x}, \quad x > 0,$$

を指数分布の**密度**または**密度関数**という．指数分布はイベントから次のイベントまでの時間間隔の分布として用いられることが多い．

例 3.3.7 (正規分布) 所与の $m \in \mathbb{R}$, $v > 0$ に対し，$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度 μ で，

$$\mu((a,b]) = \int_a^b \frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}} dx, \quad -\infty < a < b < \infty,$$

を満たすものが一意に存在する（証明略）．μ を平均 m，分散 v の**正規分布**あるいは**ガウス分布**と呼ぶ．関数

$$\rho(x) := \frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}}, \quad x \in \mathbb{R},$$

を正規分布の**密度**または**密度関数**という．$m = 0$, $v = 1$ のときの μ を特に**標準正規分布**と呼ぶ．

次に確率測度の基本的性質をみていこう．

命題 3.3.8 $(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間とするとき，以下が成り立つ．

(1) $\mathbb{P}(\emptyset) = 0$.

(2) $A_i \in \mathcal{F}$, $A_i \cap A_j = \emptyset$ $(i \ne j)$, $i, j = 1, \ldots, n$, ならば

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i).$$

(3) $A, B \in \mathcal{F}$ が $A \subset B$ を満たすとき，

$$\mathbb{P}(A) \le \mathbb{P}(B).$$

(4) $A, B \in \mathcal{F}$ に対し

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(5) $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, に対し，

$$\mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) \le \sum_{n=1}^\infty \mathbb{P}(A_n).$$

証明．(1)．$A_1 = \Omega$ とし，$n \ge 2$ に対し $A_n = \emptyset$ とおいて，$\mathbb{P}(\Omega) = 1$ と $\mathbb{P}$ の σ-加法性（定義 3.3.1）を使うと，

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mathbb{P}(A_n) = 1 + \sum_{n=2}^\infty \mathbb{P}(\emptyset).$$

従って，$\mathbb{P}(\emptyset)=0$.

(2). $i \geq n+1$ に対し $A_i=\emptyset$ とすると，直前の (1) の結果から，$i \geq n+1$ に対し $\mathbb{P}(A_i)=0$. このことと確率の σ-加法性より，

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right)=\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty}\mathbb{P}(A_i)=\sum_{i=1}^{n}\mathbb{P}(A_i).$$

(3). $A_1=A$, $A_2=B\setminus A$ として本命題 (2) を使うと，

$$\mathbb{P}(B)=\mathbb{P}(A_1\cup A_2)=\mathbb{P}(A_1)+\mathbb{P}(A_2)=\mathbb{P}(A)+\mathbb{P}(B\setminus A). \tag{3.3.1}$$

定義 3.3.1 (1) より $\mathbb{P}(B\setminus A)\geq 0$. よって主張する等式が成り立つ.

(4). $A_1=A\setminus(A\cap B)$, $A_2=B\setminus(A\cap B)$, $A_3=A\cap B$ として本定理 (2) および (3.3.1), すなわち $C\subset D$ ならば $\mathbb{P}(D\setminus C)=\mathbb{P}(D)-\mathbb{P}(C)$ を使うと，

$$\begin{aligned}\mathbb{P}(A\cup B)&=\mathbb{P}(A_1\cup A_2\cup A_3)=\mathbb{P}(A_1)+\mathbb{P}(A_2)+\mathbb{P}(A_3)\\&=\mathbb{P}(A\setminus(A\cap B))+\mathbb{P}(B\setminus(A\cap B))+\mathbb{P}(A\cap B)\\&=\mathbb{P}(A)-\mathbb{P}(A\cap B)+\mathbb{P}(B)-\mathbb{P}(A\cap B)+\mathbb{P}(A\cap B)\\&=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B).\end{aligned}$$

(5). $B_1=A_1$, $B_n=A_n\setminus(A_1\cup\cdots\cup A_{n-1})$ $(n\geq 2)$ により事象列 $\{B_n\}_{n=1}^{\infty}$ を定義する．このとき，$B_i\cap B_j=\emptyset$ $(i\neq j)$ であるから，確率の σ-加法性と本命題 (3) より，

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)=\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right)=\sum_{n=1}^{\infty}\mathbb{P}(B_n)\leq\sum_{n=1}^{\infty}\mathbb{P}(A_n).$$

□

命題 3.3.9 (確率の連続性) $(\Omega,\mathcal{F},\mathbb{P})$ を確率空間とするとき，以下が成り立つ.

(1) $\{A_n\}_{n=1}^{\infty}\subset\mathcal{F}$ が単調増加のとき，すなわち $A_1\subset A_2\subset\cdots$ のとき，

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)=\lim_{n\to\infty}\mathbb{P}(A_n).$$

(2) $\{A_n\}_{n=1}^{\infty}\subset\mathcal{F}$ が単調減少のとき，すなわち $A_1\supset A_2\supset\cdots$ のとき，

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)=\lim_{n\to\infty}\mathbb{P}(A_n).$$

証明. (1). $B_1=A_1$, $B_n=A_n\setminus A_{n-1}$ $(n\geq 2)$ により事象列 $\{B_n\}_{n=1}^{\infty}$ を定義すると，$B_i\cap B_j=\emptyset$ $(i\neq j)$ であるから，確率の σ-加法性と命題 3.3.8 (2) より，

$$\begin{aligned}\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)&=\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right)=\sum_{n=1}^{\infty}\mathbb{P}(B_n)=\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{P}(B_k)\\&=\lim_{n\to\infty}\mathbb{P}\left(\bigcup_{k=1}^{n} B_k\right)=\lim_{n\to\infty}\mathbb{P}(A_n).\end{aligned}$$

(2). $B_n = A_n^c$ $(n \geq 1)$ により定義される事象列 $\{B_n\}_{n=1}^{\infty}$ は単調増加である．よって本命題 (1) を適用し，

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

他方，

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} B_n\right)^c \tag{3.3.2}$$

であるから，

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = 1 - \lim_{n\to\infty} \mathbb{P}(B_n) = 1 - \lim_{n\to\infty}(1 - \mathbb{P}(A_n)) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

□

例題 3.3.10 $(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間とする．各 $n \in \mathbb{N}$ に対し $A_n \in \mathcal{F}$ が $\mathbb{P}(A_n) = 1$ を満たすとき，

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$$

が成り立つことを示せ．

解答．$\{B_n\}_{n=1}^{\infty} \subset \mathcal{F}$ を $B_n = A_n^c$ により定義すると，$\mathbb{P}(B_n) = 1 - \mathbb{P}(A_n) = 0$ である．このことと命題 3.3.8 (5) より，

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(B_n) = 0.$$

(3.3.2) を用いて

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = 1$$

を得る．

3.4　確率変数

以下，$(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間とする．確率変数とは，端的には，Ω 上の関数のことである．例えば，明日午前 9 時における地点 A の気温（℃）は確率変数とみなせる．この場合 $X(\omega) = 30$ などと表される．数学的に議論するためには，X に関係する集合が $\mathbb{P}$ の定義域である $\mathcal{F}$ に属さなければいけない．具体的には，$\{X > 25\} \in \mathcal{F}$ でないと $\mathbb{P}(X > 25)$ を定義できない．

定義 3.4.1 $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ が**確率変数**であるとは，

$$\{X > a\} \in \mathcal{F}, \quad a \in \mathbb{R}, \tag{3.4.1}$$

が満たされるときにいう．

この定義から，確率変数に関連する様々な集合が $\mathcal{F}$ に属すことが分かる．ボレル集合を用いた性質 (4) については測度論の知識が必要になるので証明は略す．

命題 3.4.2 $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ が確率変数であるとき，以下が成り立つ．

(1) 任意の $a \in \mathbb{R}$ に対して $\{X \geq a\} \in \mathcal{F}$.
(2) 任意の $a, b \in \mathbb{R}$ に対して $\{a < X \leq b\} \in \mathcal{F}$.
(3) 任意の $a, b \in \mathbb{R}$ に対して $\{a \leq X \leq b\} \in \mathcal{F}$.
(4) 任意の $B \in \mathcal{B}(\mathbb{R})$ に対して $\{X \in B\} \in \mathcal{F}$.
(5) $\{X = \infty\} \in \mathcal{F}$.
(6) $\{X = -\infty\} \in \mathcal{F}$.

証明．(1)．X が確率変数のとき (3.4.1) が成り立つ．このとき任意の $a \in \mathbb{R}$ に対して，

$$\{X \geq a\} = \bigcap_{n=1}^{\infty} \left\{X > a - \frac{1}{n}\right\} \in \mathcal{F}.$$

よって (1) が成り立つ．

(2)．(3.4.1) より，任意の $a, b \in \mathbb{R}$ に対して，$\{X > a\}, \{X > b\} \in \mathcal{F}$．これより，$\{X \leq b\} = \{X > b\}^c \in \mathcal{F}$．よって

$$\{a < X \leq b\} = \{X > a\} \cap \{X \leq b\} \in \mathcal{F}.$$

(3)．本命題 (1) の結果より $\{X \geq a\} \in \mathcal{F}$．よって

$$\{a \leq X \leq b\} = \{X \geq a\} \cap \{X > b\}^c \in \mathcal{F}, \quad a, b \in \mathbb{R}.$$

(5) は $\{X = \infty\} = \bigcap_{n=1}^{\infty}\{X > n\}$, (6) は $\{X = -\infty\} = \bigcap_{n=1}^{\infty}\{X \leq -n\} = \bigcap_{n=1}^{\infty}\{X > -n\}^c$ より従う． □

注意 3.4.3 X が実数値のときは，(3.4.1) は命題 (1)–(4) のいずれとも同値である．実際，命題 3.4.2 (1) が成り立っているとすると，任意の $a \in \mathbb{R}$ に対して，

$$\{X > a\} = \bigcup_{n=1}^{\infty} \left\{X \geq a + \frac{1}{n}\right\} \in \mathcal{F}.$$

また，命題 3.4.2 (2) が成り立つとき，

$$\{X > a\} = \bigcup_{n=1}^{\infty} \{a < X \leq a + n\} \in \mathcal{F}, \quad a \in \mathbb{R}.$$

命題 3.4.2 (3) を使う場合は,

$$\{X > a\} = \bigcup_{n=1}^{\infty} \left\{a + \frac{1}{n} \leq X \leq a + n\right\} \in \mathcal{F}, \quad a \in \mathbb{R}.$$

命題 3.4.2 (4) が満たされる場合は，ボレル集合として $B = (a, \infty)$ を考えればよい．

命題 3.4.4 $\{X_n\}_{n=1}^\infty$ を実確率変数列とするとき，

$$\sup_{n\in\mathbb{N}} X_n,\quad \inf_{n\in\mathbb{N}} X_n,\quad \limsup_{n\to\infty} X_n,\quad \liminf_{n\to\infty} X_n$$

は全て確率変数である．

証明の前に，上極限 $\limsup$ と下極限 $\liminf$ の定義を確認しておこう．実数列 $\{a_n\}_{n=1}^\infty$ に対して，

$$\limsup_{n\to\infty} a_n := \lim_{k\to\infty} \sup_{n\ge k} a_n \in \mathbb{R}\cup\{\pm\infty\},$$
$$\liminf_{n\to\infty} a_n := \lim_{k\to\infty} \inf_{n\ge k} a_n \in \mathbb{R}\cup\{\pm\infty\}.$$

命題 3.4.4 により，標語的には「確率変数列の極限も確率変数である」と言える．

命題 3.4.4 の証明．任意の $a\in\mathbb{R}$ に対して，

$$\left\{\sup_{n\ge1} X_n > a\right\} = \bigcup_{n=1}^\infty \{X_n > a\},\quad \left\{\inf_{n\ge1} X_n < a\right\} = \bigcup_{n=1}^\infty \{X_n < a\} \in \mathcal{F}$$

であるから，$\sup_{n\ge1} X_n$ と $\inf_{n\ge1} X_n$ はともに確率変数である．この事実と

$$\limsup_{n\to\infty} X_n = \lim_{n\to\infty} \sup_{k\ge n} X_k = \inf_{n\ge1} \sup_{k\ge n} X_k,$$
$$\liminf_{n\to\infty} X_n = \lim_{n\to\infty} \inf_{k\ge n} X_k = \sup_{n\ge1} \inf_{k\ge n} X_k$$

より $\limsup_{n\to\infty} X_n$ と $\liminf_{n\to\infty} X_n$ もともに確率変数であることも分かる．　□

命題 3.4.5 X を実確率変数，$f:\mathbb{R}\to\mathbb{R}$ を連続とするとき，$f(X)$ は確率変数である．

証明．$a\in\mathbb{R}$ とする．連続写像の定義より，$f^{-1}((a,\infty))$ は開集合である．ここで $\mathbb{R}$ の任意の開集合 B に対して，$B=\bigcup_{n=1}^\infty (a_n,b_n)$ なる $a_n,b_n\in\mathbb{R}$, $a_n<b_n$, $n\in\mathbb{N}$, が存在する（これについては例えば，[10] の 15 章をみよ）．この事実を $B=f^{-1}((a,\infty))$ に対して適用すると，

$$\{f(X)>a\} = \{X\in f^{-1}((a,\infty))\} = \bigcup_{n=1}^\infty \{X\in(a_n,b_n)\} \in \mathcal{F}.$$

□

$f:\mathbb{R}\to\mathbb{R}$ が**ボレル関数**であるとは，

$$\{f>\alpha\} = \{x\in\mathbb{R} \mid f(x)>\alpha\} \in \mathcal{B}(\mathbb{R}),\quad \alpha\in\mathbb{R},$$

が満たされるときにいう．

- 連続関数，階段関数はボレル関数である．
- 命題 3.4.5 の一般化として，次が成り立つ．X が実確率変数のとき，任意のボレル関数 f に対して $f(X)$ も実確率変数である．

多次元版のボレル関数を導入しておく．$f : \mathbb{R}^d \to \mathbb{R}$ が d 次元ボレル関数であるとは，

$$\{f > \alpha\} = \{x \in \mathbb{R}^d \mid f(x) > \alpha\} \in \mathcal{B}(\mathbb{R}^d), \quad \alpha \in \mathbb{R},$$

が満たされるときにいう．

- $X_1, \ldots, X_d$ を実確率変数，f を d 次元ボレル関数とするとき，$f(X_1, \ldots, X_d)$ も実確率変数である．
- $\{\mathbb{R}^d$ 上の連続関数$\} \subset \{d$ 次元ボレル関数$\}$.
- X, Y を実確率変数とするとき，任意の $\alpha, \beta \in \mathbb{R}$ に対して，$\alpha X + \beta Y$, XY も実確率変数である．

次に一般の確率空間における確率変数の分布を考えよう．

X を実確率変数とする．このとき，

$$\mu_X(B) := \mathbb{P}(X \in B), \quad B \in \mathcal{B}(\mathbb{R})$$

は $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度を定義する（読者はこれを確かめよ）．

- μ_X を X の**分布**または**確率法則**と呼ぶ.
- X が S-値（$S \subset \mathbb{R}$）のとき，μ_X は $(S, \mathcal{B}(S))$ 上の確率測度である．

次の定理が示すように，実確率変数 X の分布は $\mathbb{P}(a < X \leq b)$, $a \leq b$, によって決定される．証明については例えば [10, 定理 3.3] をみよ．

定理 3.4.6 二つの実確率変数 X, Y について，

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a < Y \leq b), \quad -\infty < a < b < \infty,$$

が成り立つとき，任意のボレル集合 $B \in \mathcal{B}(\mathbb{R})$ に対して $\mu_X(B) = \mu_Y(B)$. すなわち X と Y の分布は等しい．

- 実確率変数の分布は $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度ではあるが，これを決定するためには $\mathcal{B}(\mathbb{R})$ を生成する半開区間の族 $\mathcal{I} = \{(a, b] \mid -\infty \leq a < b \leq +\infty\}$ だけ考えれば十分だと定理 3.4.6 は述べている．
- 確率変数 X と Y の分布が等しいとき，$X \overset{\mathrm{d}}{=} Y$ と書くこともある．
- 確率変数 X に対し，

$$\mathbb{P}(a < X \leq b) = \int_a^b \rho(x)dx, \quad a < b, \ \ a, b \in \mathbb{R}, \tag{3.4.2}$$

  を満たす関数 ρ が存在するとき，ρ を X の**確率密度関数**あるいは単に**密度関数**という．
- 実用上現れる密度関数の多くは $\mathbb{R}$ 上の区分的連続関数である．そのような場合，(3.4.2) の右辺の積分は（広義）リーマン積分として解釈される．以下，特に断らない限り，本授業で現れる被積分関数は（広義）リーマン積分可能な関数と仮定して議論を進める．ルベーグ積分の理論を知っている読者はルベーグ積分として解釈すればよい．多次元の場合でも同様.

例 3.4.7 $m \in \mathbb{R}, v > 0$ を所与とする．X の密度関数 ρ が

$$\rho(x) = \frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}}, \quad x \in \mathbb{R},$$

により与えられるとき, X は平均 m, 分散 v の**正規分布**または**ガウス分布**に従うといい, $X \sim N(m, v)$ と書く．$m = 0$, $v = 1$ のときの正規分布を特に**標準正規分布**と呼ぶ．$Z \sim N(0, 1)$ ならば $X := m + \sqrt{v}Z \sim N(m, v)$ である．小さい $\varepsilon > 0$ に対して $Y \sim N(0, \varepsilon^2)$ であるような確率変数 Y を現象の観測ノイズ，あるいは推定誤差とみなすことが多い．

例 3.4.8 $a < b$ なる $a, b \in \mathbb{R}$ を所与とする．X の密度関数 ρ が

$$\rho(x) = \begin{cases} \dfrac{1}{b-a} & (a \le x \le b), \\ 0 & (\text{それ以外}), \end{cases}$$

により与えられるとき，X は $[a, b]$ 上の**一様分布**に従うといい，$X \sim U[a, b]$ と書く．

実確率変数 X に対して，

$$F(x) := \mathbb{P}(X \le x), \quad x \in \mathbb{R},$$

を X の**分布関数**という．定理 3.4.6 を踏まえると，X の分布関数により X の分布（確率法則）は完全に決定されることが分かる．

命題 3.4.9 確率変数 X の分布関数 F は次の性質を満たす．

(1) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.
(2) (単調性) $x \le y \Longrightarrow F(x) \le F(y)$.
(3) (右連続性) $\lim_{y \searrow x} F(y) = F(x)$.

証明．$\{x_n\}$ を $x_n \to -\infty$ なる任意の実数列とする．$y_n := \sup_{k \ge n} x_k$ とおくと，$\{y_n\}$ は単調非増加で，$y_n \to -\infty$．よって，命題 3.3.8 より $0 \le \mathbb{P}(X \le x_n) \le \mathbb{P}(X \le y_n)$ であり，確率の連続性（命題 3.3.9）より，

$$\lim_{n \to \infty} \mathbb{P}(X \le y_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \{X \le y_n\}\right) = \mathbb{P}(\emptyset) = 0.$$

$\{x_n\}$ は任意だったので，$\lim_{x \to -\infty} F(x) = 0$ が従う．

次に $\{\tilde{x}_n\}$ を $\tilde{x}_n \to \infty$ なる任意の実数列とする．$\tilde{y}_n := \inf_{k \ge n} \tilde{x}_k$ とおくと，$\{\tilde{y}_n\}$ は単調非減少で，$\tilde{y}_n \to \infty$．よって，$\mathbb{P}(X \le \tilde{y}_n) \le \mathbb{P}(X \le \tilde{x}_n) \le 1$ であり，再び確率の連続性より，

$$\lim_{n \to \infty} \mathbb{P}(X \le \tilde{y}_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{X \le \tilde{y}_n\}\right) = \mathbb{P}(\Omega) = 1.$$

ゆえに $\lim_{n \to \infty} \mathbb{P}(X \le \tilde{x}_n) = 1$ であり，$\{\tilde{x}_n\}$ は任意だったので，$\lim_{x \to \infty} F(x) = 1$ を得る．

次に単調性を確認する．$x \le y$ ならば，$\{X \le x\} \subset \{X \le y\}$ であるから，命題 3.3.8 (3) より $F(x) = \mathbb{P}(X \le x) \le \mathbb{P}(X \le y) = F(y)$.

最後に右連続性を示す．$x \in \mathbb{R}$ に対し，$x_n > x$, $\lim_{n\to\infty} x_n = x$ を満たす実数列 $\{x_n\}$ を任意にとる．このとき，単調性より，$F(x_n) \geq F(x)$ である．他方，$y_n = \sup_{k\geq n} x_k$ とおくと，$\{y_n\}$ は単調非増加で，$y_n \to x$．よって $F(x) \leq F(x_n) \leq F(y_n)$ であり，これと確率の単調性より，

$$F(x) \leq \lim_{n\to\infty} \mathbb{P}(X \leq y_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\{X \leq y_n\}\right) \leq \mathbb{P}(X \leq x) = F(x).$$

ゆえに $\lim_{y\searrow x} F(y) = F(x)$. □

- 命題 3.4.9 の 3 条件を満たす $F : \mathbb{R} \to [0,1]$ に対し，ある確率測度 μ が存在し，

$$F(x) = \mu((-\infty, x]), \quad x \in \mathbb{R}.$$

- X の分布関数 F が連続のとき，X は**連続分布**を持つという．

例題 3.4.10 $X \sim N(m, v)$，$0 < \alpha < 1$ のとき，$F^{-1}(\alpha)$ を求めよ．

解答．関数 $\Phi : \mathbb{R} \to \mathbb{R}$ を

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$$

により定める．これは標準正規分布に従う確率変数の分布関数である．このとき，$Z \sim N(0,1)$ とすると，

$$\mathbb{P}(X \leq x) = \mathbb{P}(m + \sqrt{v}Z \leq x) = \mathbb{P}\left(Z \leq \frac{x-m}{\sqrt{v}}\right) = \Phi\left(\frac{x-m}{\sqrt{v}}\right) \tag{3.4.3}$$

となる．Φ は連続かつ狭義単調増加で $\lim_{x\to\infty} \Phi(x) = 1$, $\lim_{x\to-\infty} \Phi(x) = 0$ を満たすから連続な逆関数 Φ^{-1} が存在する．従って，任意の $\alpha \in (0,1)$ に対し

$$\frac{x-m}{\sqrt{v}} = \Phi^{-1}(\alpha)$$

を満たす $x \in \mathbb{R}$ が一意に定まる．X の分布関数 F も同様に連続な逆関数 F^{-1} が存在するから，(3.4.3) より

$$F^{-1}(\alpha) = m + \sqrt{v}\Phi^{-1}(\alpha).$$

- 確率変数 X は投資による金融資産の損失額を表すと仮定し，$\alpha = 0.01$ とするとき，$F^{-1}(\alpha)$ は 1% で起こる最大の損失とみなせる．

定理 3.4.11 確率変数 X の分布関数 F は $\mathbb{R}$ 上連続で狭義単調増加とし，$U \sim U[0,1]$ とする．このとき $X \stackrel{\mathrm{d}}{=} F^{-1}(U)$.

証明．例題 3.4.10 の解答で示したように，F の連続な逆関数 $F^{-1} : \mathbb{R} \to \mathbb{R}$ が存在する．任意の $a < b$ に対し，F の狭義単調性より

$$\mathbb{P}(a < F^{-1}(U) \leq b) = \mathbb{P}(F(a) < U \leq F(b)) = F(b) - F(a) = \mathbb{P}(a < X \leq b).$$

従って，定理 3.4.6 より X と $F^{-1}(U)$ は同分布である． □

- 定理 3.4.11 は X の**サンプリング**の方法を示している．すなわち，X の擬似乱数を発生させるためには，U の擬似乱数を発生させ，$F^{-1}(U)$ を計算すればよい．

例題 3.4.12 (1 次元最適輸送) 実数 $a_1 \le a_2 \le \cdots \le a_n$，$b_1 \le b_2 \le \cdots \le b_n$ が与えられているとする．$\{1, \ldots, n\}$ 上の任意の順列 σ に対し，

$$C(\sigma) := \sum_{i=1}^{n} (a_i - b_{\sigma(i)})^2$$

を考える．これを最小化する σ を求めよ．

解答. $C(\sigma) = \sum_i a_i^2 + \sum_i b_i^2 - 2\sum_i a_i b_{\sigma(i)}$ と書けるので，第 1 項・第 2 項は σ によらず，$\sum_i a_i b_{\sigma(i)}$ を最大化する σ を求めればよい．

σ が恒等順列でないと仮定すると，ある $i < j$ で $\sigma(i) > \sigma(j)$ なるものが存在する．$\sigma(i)$ と $\sigma(j)$ を入れ替えた順列を σ' とおくと，

$$\begin{aligned}\sum_k a_k b_{\sigma'(k)} - \sum_k a_k b_{\sigma(k)} &= a_i b_{\sigma(j)} + a_j b_{\sigma(i)} - a_i b_{\sigma(i)} - a_j b_{\sigma(j)} \\ &= (a_j - a_i)(b_{\sigma(i)} - b_{\sigma(j)}) \ge 0.\end{aligned}$$

最後の不等号は $a_j \ge a_i$ と $b_{\sigma(i)} \ge b_{\sigma(j)}$（$\sigma(i) > \sigma(j)$ と b の単調性）から従う．すなわち，入れ替えにより和は減らない．これを繰り返すと恒等順列（全ての i に対し $\sigma(i) = i$ となる順列，記号で id と書く）に到達するので，最大値は $\sigma = \mathrm{id}$ で達成される．よって最小値も $\sigma = \mathrm{id}$ で達成される．

- 例題 3.4.12 は **1 次元最適輸送問題**の離散版である．有限集合 $\{a_1, \ldots, a_n\}$, $\{b_1, \ldots, b_n\}$ 上の経験分布 $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{a_i}$, $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{b_i}$ を考え，$\mathbb{P}(X = a_i, Y = b_{\sigma(i)}) = 1/n$ により (X, Y) の同時分布を定めると，$X \sim \mu, Y \sim \nu$ かつ

$$\mathbb{E}[(X - Y)^2] = \frac{1}{n} C(\sigma).$$

  すなわち例題 3.4.12 は，「周辺分布を μ, ν に固定したまま (X, Y) の同時分布を選び，$\mathbb{E}[(X - Y)^2]$ を最小化する」問題と読める．$\sigma = \mathrm{id}$ は，両者を順位どうしで結ぶ**単調カップリング**（monotone coupling）に対応する．
- 連続分布版についても同様の結果が成り立つ．X と Y がそれぞれ連続な狭義単調増加分布関数 F, G をもち，$\mathbb{E}|X|^2 < \infty, \mathbb{E}|Y|^2 < \infty$ とする．$U \sim U[0, 1]$ とおくとき，定理 3.4.11 より $F^{-1}(U) \stackrel{\mathrm{d}}{=} X$，$G^{-1}(U) \stackrel{\mathrm{d}}{=} Y$ である（$(F^{-1}(U), G^{-1}(U))$ を X と Y の単調カップリングと呼ぶ）．このとき，$X' \stackrel{\mathrm{d}}{=} X, Y' \stackrel{\mathrm{d}}{=} Y$ を満たす任意の同時分布 (X', Y') に対し，

$$\mathbb{E}[(X' - Y')^2] \ge \mathbb{E}[(F^{-1}(U) - G^{-1}(U))^2]$$

  が成り立つ（証明は離散版の議論を極限に飛ばすものになる．詳細は [2, Theorem 2.9] をみよ）．
- 右辺 $W_2(\mu, \nu)^2 := \mathbb{E}[(F^{-1}(U) - G^{-1}(U))^2] = \int_0^1 (F^{-1}(u) - G^{-1}(u))^2 du$ は分布 μ と ν の間の Wasserstein-2 **距離**の 2 乗と呼ばれ，分布間の距離を測る指標として広く用いられる．

定義 3.4.13 実確率変数 $X_1, X_2, \dots, X_n$ が**独立**であるとは,

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{a_i < X_i \le b_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(a_i < X_i \le b_i), \quad a_i, b_i \in \mathbb{R},\ \ i = 1, \dots, n,$$

が成り立つときにいう.

- $X_1, \dots, X_n$ が独立で，X_i の分布が全て同じとき，**独立同分布**（Independent and Identically Distributed）であるといい，IID と書く（i.i.d., iid とも書く）.

定義 3.4.14 実確率変数 $X_1, X_2, \dots$ が**独立**であるとは,

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{a_i < X_i \le b_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(a_i < X_i \le b_i), \quad a_i, b_i \in \mathbb{R},\ \ i \in \mathbb{N},$$

が成り立つときにいう.

- $\{X_n\}_{n=1}^{\infty}$ が独立で，X_n の分布が全て同じとき，**独立同分布列**であるという.

事象 $A_1, A_2, \dots$ が独立であるとは，任意の $\{i_1, \dots, i_n\} \subset \mathbb{N}$ に対して

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n}\mathbb{P}(A_{i_k})$$

が成り立つときにいう.

- 事象 $A \in \mathcal{F}$ は $\mathbb{P}(A) > 0$ を満たすとする．第 2 章と同様に，A の下での $B \in \mathcal{F}$ の条件付き確率を

$$\mathbb{P}(B \,|\, A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

により定義する.
- 条件付き確率を用いると，A と B が独立 $\iff$ $\mathbb{P}(B \,|\, A) = \mathbb{P}(B)$.

定理 2.4.5 と同様に，次の結果が得られる.

定理 3.4.15 事象 $A_1, A_2, \dots$ が独立 $\iff$ $1_{A_1}, 1_{A_2}, \dots$ が独立

また，次の定理を参考までに述べておこう.

定理 3.4.16 実確率変数 $X_1, \dots, X_n$ が独立であるための必要十分条件は

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \in B_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(X_i \in B_i), \quad B_i \in \mathcal{B}(\mathbb{R}),\ \ i = 1, \dots, n.$$

例 3.4.17 (**二項分布**) $p \in (0,1)$ とする．$\{X_i\}_{i=1}^{n}$ は $\mathbb{P}(X_1 = 1) = p = 1 - \mathbb{P}(X_1 = 0)$ を満たす IID とする．このとき，確率変数

$$S := X_1 + \cdots + X_n$$

をサイズ n，パラメータ p の**二項確率変数**といい，S の分布を**二項分布**という．S は n 回投げたコインの中で表が出た枚数を表しており，その分布は

$$\mathbb{P}(S = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \quad 0 \le k \le n,$$

により与えられるのであった．

例 3.4.18 (**幾何分布**) $p \in [0,1]$ とする．$\{X_n\}_{n=1}^{\infty}$ は IID で，共通の分布は $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = 0)$ を満たすとする．このとき，確率変数

$$T := \inf\{n \in \mathbb{N} : X_n = 1\}$$

はパラメータ p の**幾何分布**に従うという．ただし，$\inf \emptyset = +\infty$. T は $\mathbb{N} \cup \{+\infty\}$-値確率変数であり，コインが表になる最初の n を表している．$p \in (0,1]$ のとき，幾何分布は

$$\mathbb{P}(T = n) = p(1-p)^{n-1}, \quad n \in \mathbb{N},$$

により与えられる．特に，$\mathbb{P}(T = \infty) = 0$ であり，$p = 1$ のとき，$\mathbb{P}(T = 1) = 1$ である．また，$p = 0$ のときは $\mathbb{P}(T = \infty) = 1$ である．

例題 3.4.19 パラメータ $p \in [0,1)$ の幾何分布をもつ確率変数 T は次の**無記憶性**

$$\mathbb{P}(T \ge n + k \,|\, T > n) = \mathbb{P}(T \ge k), \quad n, k \in \mathbb{N}, \tag{3.4.4}$$

を満たす．

解答. $p = 0$ のとき (3.4.4) は両辺 0 で成立する．$p \in (0,1)$ のとき，条件付き確率の定義と例 3.4.18 より，

$$\begin{aligned}
\mathbb{P}(T \ge n + k \,|\, T > n) &= \frac{\mathbb{P}(T \ge n + k)}{\mathbb{P}(T > n)} = \frac{\mathbb{P}(T > n + k - 1)}{\mathbb{P}(T > n)} \\
&= \frac{(1-p)^{n+k-1}}{(1-p)^n} = (1-p)^{k-1} = \mathbb{P}(T \ge k).
\end{aligned}$$

- 無記憶性の意味は「これまでは失敗続きだった．これから成功するまでに要する回数は，これまでの失敗回数とは関係ない．」

例題 3.4.20 $Y \sim N(0,1)$, H は $\mathbb{P}(H = 1) = \mathbb{P}(H = -1) = 1/2$ を満たし，Y と H は独立とする．このとき，$X := H + Y$ の確率密度関数を求めよ．

解答. ϕ を標準正規分布の密度関数とする．このとき，Y と H の独立性より，

$$\begin{aligned}
\mathbb{P}(a < X \le x) &= \mathbb{P}(a < X \le x, H = 1) + \mathbb{P}(a < X \le x, H = -1) \\
&= \mathbb{P}(a - 1 < Y \le x - 1, H = 1) + \mathbb{P}(a + 1 < Y \le x + 1, H = -1) \\
&= \mathbb{P}(a - 1 < Y \le x - 1)\mathbb{P}(H = 1) + \mathbb{P}(a + 1 < Y \le x + 1)\mathbb{P}(H = -1) \\
&= \frac{1}{2} \int_{a-1}^{x-1} \phi(y) dy + \frac{1}{2} \int_{a+1}^{x+1} \phi(y) dy.
\end{aligned}$$

よって，X の密度関数 ρ は

$$\rho(x) = \frac{d}{dx}\mathbb{P}(a < X \leq x) = \frac{1}{2}\left(\phi(x-1) + \phi(x+1)\right), \quad x \in \mathbb{R},$$

により与えられる．

- この例題の X の分布を**混合正規分布**という．

ここで**少数の法則**を紹介しておこう．

> **定理** 3.4.21 (**少数の法則**) S_n はパラメータ (n,p) の二項分布とする．すなわち，$\{X_i\}$ は IID で各 X_i は $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$ を満たすとき，$S_n \stackrel{\mathrm{d}}{=} X_1 + \cdots + X_n$ と仮定する．$p = p_n$ が $\lim_{n\to\infty} np_n = c > 0$ を満たすならば，
>
> $$\lim_{n\to\infty} \mathbb{P}(S_n = k) = \frac{e^{-c}c^k}{k!}, \quad k = 0, 1, 2, \ldots$$

証明.（第一段階）．一般に，実数列 $\{a_n\}$, $\{b_n\}$ が $a_n \to 0$, $b_n \to \infty$, $a_n b_n \to c$ を満たすなら，$\lim_{n\to\infty}(1 + a_n)^{b_n} = e^c$ が成り立つ．実際，$\log(1+x)$ を $x = 0$ 近傍でテーラー展開すると

$$\log(1+x) = x + O(x^2), \quad x \to 0.$$

これより，

$$b_n \log(1 + a_n) = a_n b_n + O(a_n^2 b_n) \to c, \quad n \to \infty.$$

（第二段階）．仮定より，$k \geq 2$ に対して

$$n(n-1)\cdots(n-k+1)p^k = \left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)(np)^k \to c^k$$

であり，第一段階の結果より，

$$(1-p)^{n-k} = (1-p)^n(1-p)^{-k} \to e^{-c}.$$

ゆえに，任意の $k \geq 0$ に対して，

$$\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \to \frac{e^{-c}c^k}{k!}, \quad n \to \infty.$$

$\square$

少数の法則により，ポアソン分布を使って人数や件数，品数の分布をモデル化することが正当化される．例として，1 時間の間に銀行に来る人数の分布について考えよう．n が十分大きいとき，時間区間 $[(i-1)/n, i/n]$ の間に来る人数 X_i は互いに独立で，$X_i = 0$ または 1 とみなせる．さらに $X_i = 1$ の確率は n に反比例すると考えてよい．このとき，各 X_i はベルヌーイ分布となり，$\mathbb{P}(X_i = 1) = c/n$ を満たす正定数 c が存在する．定理 3.4.21 より，1 時間の間の到着人数 $X_1 + \cdots + X_n$ は近似的にパラメータ c のポアソン分布に従う．表 3.4.1 はパラメータ $(100, 0.03)$ の二項分布とパラメータ 3 のポアソン分布を比較したものである．

X と Y を実確率変数とする．

表 3.4.1　二項分布 $B(100, 0.03)$ とポアソン分布 Poi(3) の比較

k	$\mathbb{P}(X = k)$ (二項分布)	$\mathbb{P}(Y = k)$ (ポアソン分布)	絶対誤差
0	0.0476	0.0498	0.0022
1	0.1471	0.1494	0.0023
2	0.2252	0.2240	0.0011
3	0.2275	0.2240	0.0034
4	0.1706	0.1680	0.0026
5	0.1013	0.1008	0.0005
6	0.0496	0.0504	0.0008
7	0.0210	0.0216	0.0006

- 2 次元確率ベクトル (X, Y) の分布（確率法則）のことを X と Y の**同時分布**または**結合分布**と呼ぶ．
- 各成分 X, Y の分布を**周辺分布**という．
- 等式

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d \rho(x, y) dx dy, \quad a, b, c, d \in \mathbb{R}, \;\; a < b, \;\; c < d,$$

  を満たす関数 ρ が存在するとき，$\rho = \rho_{X,Y}$ を X と Y の**同時確率密度関数**または**結合確率密度関数**と呼ぶ．もう少し短く，**同時密度関数**または**結合密度関数**と呼ぶことも多い．
- 各成分 X, Y の確率密度関数 ρ_X, ρ_Y が存在するとき，それらを**周辺確率密度関数**あるいは短く**周辺密度関数**という．
- (X, Y) の結合密度関数 ρ が存在するとき，

$$\rho_X(x) = \int_{-\infty}^{\infty} \rho_{X,Y}(x, y) dy, \quad x \in \mathbb{R},$$

  は X の密度関数である．Y についても同様である．

命題 3.4.22 2 次元確率ベクトル (X, Y) は結合密度 $\rho_{X,Y}$ を持つと仮定する．さらに，X の密度関数 ρ_X と Y の密度関数 ρ_Y は $\mathbb{R}$ 上の連続関数であり，$\rho_{X,Y}$ は $\mathbb{R}^2$ 上の連続関数とする．このとき，

$$X \text{と} Y \text{が独立} \quad \Longleftrightarrow \quad \rho_{X,Y}(x, y) = \rho_X(x)\rho_Y(y), \quad x, y \in \mathbb{R}.$$

証明．$(\Rightarrow)$．任意の $a < x$, $b < y$ に対し，

$$\begin{aligned}
\int_a^x \int_b^y \rho_{X,Y}(x', y') dx' dy' &= \mathbb{P}(a < X \leq x, b < Y \leq y) \\
&= \mathbb{P}(a < X \leq x)\mathbb{P}(b < Y \leq y) = \int_a^x \rho_X(x') dx' \int_b^y \rho_Y(y') dy'.
\end{aligned}$$

両辺 x, y で偏微分し，$\rho_{X,Y}(x, y) = \rho_X(x)\rho_Y(y)$ を得る．

$(\Leftarrow)$. 重積分を累次積分とみなすと，任意の $a<c$, $b<d$ に対し，

$$\begin{aligned}\mathbb{P}(a<X\le c, b<Y\le d) &= \int_a^c\int_b^d \rho_{X,Y}(x',y')dx'dy' \\ &= \int_a^c \rho_X(x')dx'\int_b^d \rho_Y(y')dy' = \mathbb{P}(a<X\le c)\mathbb{P}(b<Y\le d).\end{aligned}$$

□

例題 3.4.23 確率ベクトル (X,Y) の同時密度関数 ρ が

$$\rho(x,y)=\begin{cases}\frac{3}{2}y^2 & (0\le x\le 2, 0\le y\le 1),\\ 0 & (\text{その他})\end{cases}$$

により与えられているとする．このとき以下の問いに答えよ．

(1) (X,Y) の周辺密度関数を求めよ．

(2) X と Y は独立か？

解答．まず，

$$\int_{\mathbb{R}^2}\rho(x,y)dxdy=\int_0^2\int_0^1\frac{3}{2}y^2dydx=1$$

より ρ は確かに密度関数である．X と Y の密度をそれぞれ ρ_X, ρ_Y とすると，$0\le x\le 2$ のとき，

$$\rho_X(x)=\int_{-\infty}^{\infty}\rho(x,y)dy=\frac{3}{2}\int_0^1 y^2dy=\frac{1}{2}$$

であり，$x<0$ または $2<x$ のとき $\rho_X(x)=0$. すなわち X は $[0,2]$ 上の一様分布に従う．さらに，$0\le y\le 1$ に対し，

$$\rho_Y(y)=\int_{-\infty}^{\infty}\rho(x,y)dx=3y^2$$

であり，$y<0$ または $1<y$ のとき $\rho_Y(y)=0$.

また，$\rho(x,y)=\rho_X(x)\rho_Y(y)$ が任意の $x,y\in\mathbb{R}$ に対して成り立つから命題 3.4.22 より X と Y は独立である．

結合密度関数が周辺密度関数の積と一致しない例を挙げておこう．

例 3.4.24 (X,Y) の結合密度関数 ρ が

$$\rho(x,y)=\begin{cases}\frac{21}{4}x^2y & (x^2\le y\le 1),\\ 0 & (\text{その他})\end{cases}$$

により与えられているとする．このとき，$-1\le x\le 1$ に対し

$$\rho_X(x)=\int_{-\infty}^{\infty}\rho(x,y)dy=\frac{21x^2}{4}\int_{x^2}^1 ydy=\frac{21}{8}x^2(1-x^2)$$

であり，$x < -1$ または $1 < x$ のとき $\rho_X(x) = 0$．さらに，$0 \le y \le 1$ に対し，

$$\rho_Y(y) = \int_{-\infty}^{\infty} \rho(x,y)dx = \frac{21y}{4}\int_0^{\sqrt{y}} x^2 dx = \frac{7}{4}y^{5/2}$$

であり，$y < 0$ または $1 < y$ のとき $\rho_Y(y) = 0$．よってもちろん一般には $\rho(x,y) \neq \rho_X(x)\rho_Y(y)$．

3.5　期待値

前節に引き続き，$(\Omega, \mathcal{F}, \mathbb{P})$ を任意の確率空間とする．特に，Ω は有限集合／無限集合どちらの場合もあり得る．有限試行の確率空間においては，確率変数 X の期待値は $\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}\{\omega\}$ により定義されるのであった．このように定義されるのは Ω が有限集合だから X が必然的に離散的に値をとるが故である．

Ω が無限集合の場合でも，X が**離散的**の場合は同様に期待値を定義できる．すなわち，適当な実数 $a_1, a_2, \ldots$ が存在して

$$X(\omega) = \sum_{n=1}^{\infty} a_n 1_{\{X=a_n\}}(\omega)$$

と表される場合に，その期待値 $\mathbb{E}[X]$ を

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} a_n \mathbb{P}(X = a_n) \tag{3.5.1}$$

と定義するのは自然なことであろう．

他方，X が**連続的**な場合，すなわち X が密度関数 p をもつ場合，その期待値 $\mathbb{E}[X]$ を

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xp(x)dx \tag{3.5.2}$$

と定義することも自然である．

では，X が離散的とも連続的とも言えない場合はどうだろうか？ 例えば，H は $\mathbb{P}(H = 0) = \mathbb{P}(H = 1) = 1/2$ を満たす確率変数，$Z \sim N(0,1)$ とし，H と Z は独立とする．このとき $X := HZ$ を考える．$Z = 0$ の確率は 0 だから

$$\mathbb{P}(X = 0) = \mathbb{P}(\{H = 0\} \cup \{Z = 0\}) = \mathbb{P}(H = 0) = \frac{1}{2}.$$

よって X は連続的とは言えないし，任意の $b > 0$ に対し

$$\mathbb{P}(0 < X < b) = \mathbb{P}(H = 1, 0 < Z < b) = \mathbb{P}(H = 1)\mathbb{P}(0 < Z < b) = \frac{1}{2}\int_0^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx$$

と表されるので，離散的でもない．このような確率変数は決して病的なものでないことに気付くだろう．この場合は期待値を (3.5.1), (3.5.2) どちらの方法でも定義することはできない．

以下では，まず任意の確率変数に対して適用可能な形で期待値を定義する．そしてその定義の帰結として，離散的な確率変数に対しては (3.5.1) が成立し，連続的な確率変数に対しては (3.5.2) が成立することをみる．

単純確率変数に対する定義

X が**単純確率変数**であるとは，X が適当な $a_i \in \mathbb{R}$, $C_i \in \mathcal{F}$, $i = 1, \ldots, n$, を用いて $X = \sum_{i=1}^{n} a_i 1_{C_i}$ と表されるときにいう．ただし $C_i \cap C_j = \emptyset$ $(i \neq j)$．このとき，X の**期待値** $\mathbb{E}[X]$ を

$$\mathbb{E}[X] = \sum_{i=1}^{n} a_i \mathbb{P}(C_i)$$

により定義する．

非負確率変数に対する定義

非負の確率変数 X に対して，単純確率変数の単調非減少列 $\{X_n\}_{n=1}^{\infty}$ が存在して，

$$X(\omega) = \lim_{n\to\infty} X_n(\omega), \quad \omega \in \Omega,$$

とできる．例えば，

$$X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\}} + n1_{\{X \geq n\}}$$

とすればよい．このとき，

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n] \in [0, \infty]$$

と定義する．この定義は近似単純確率変数列の取り方に依らない（証明略）．

一般の確率変数に対する定義

一般の $\mathbb{R} \cup \{\pm\infty\}$-値確率変数 X に対して，$X^+ := \max\{X, 0\}$, $X^- := \max\{-X, 0\}$ とおく．
$\mathbb{E}[X^+] < \infty$ または $\mathbb{E}[X^-] < \infty$ のとき，

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$$

と定義する．

- $\mathbb{E}[X]$ を X の**期待値**または**平均値**と呼ぶ.
- $\mathbb{E}[X]$ は $\pm\infty$ をとり得る.
- $\mathbb{E}[X^+] < \infty$ かつ $\mathbb{E}[X^-] < \infty$ のとき，X は**可積分**または**平均可能**という.

期待値は離散的な確率変数の場合は和で，連続的な場合は積分で表される．

命題 3.5.1 X が $\mathbb{N} \cup \{0\}$-値確率変数のとき，

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} n\mathbb{P}(X = n) \in [0, \infty].$$

証明．$X_n := \sum_{k=1}^{n} k1_{\{X=k\}}$ を考える．$\{X_n\}_{n=1}^{\infty}$ は単調非減少な確率変数列で，$\lim_{n\to\infty} X_n = X$ である．すなわち，$\{X_n\}$ は X の近似単純確率変数列である．よって非負確率変数に対する期待値

の定義より，

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} \sum_{k=1}^{n} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty} k\mathbb{P}(X = k).$$

この等式は右辺が発散する場合も含む. □

命題 3.5.2 実確率変数 X が密度関数 p を持つとする．このとき，

$$\int_{-\infty}^{\infty} |x|p(x)dx < \infty$$

ならば X は可積分で，

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xp(x)dx.$$

証明. 各 $n \geq 1$ に対して非負確率変数 X_n^+ を

$$X_n^+ = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{\{\frac{k-1}{2^n} \leq X^+ < \frac{k}{2^n}\}} + n1_{\{X^+ \geq n\}}$$

と定義すると，$\{X_n^+\}_{n=1}^{\infty}$ は X^+ の近似単純確率変数列である．よって期待値の定義より

$$\mathbb{E}[X^+] = \lim_{n\to\infty} \mathbb{E}[X_n^+].$$

$\alpha > 0$ のとき $\{X^+ \geq \alpha\} = \{X \geq \alpha\}$ だから

$$X_n^+ = \sum_{k=2}^{n2^n} \frac{k-1}{2^n} 1_{\{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\}} + n1_{\{X \geq n\}}.$$

これより

$$\begin{aligned}\mathbb{E}[X_n^+] &= \sum_{k=2}^{n2^n} \frac{k-1}{2^n} \mathbb{P}\left(\frac{k-1}{2^n} \leq X < \frac{k}{2^n}\right) + n\mathbb{P}(X \geq n) \\ &= \sum_{k=2}^{n2^n} \frac{k-1}{2^n} \int_{(k-1)/2^n}^{k/2^n} p(x)dx + n\int_n^{\infty} p(x)dx.\end{aligned}$$

このことと

$$\int_0^{\infty} xp(x)dx = \sum_{k=1}^{n2^n} \int_{(k-1)/2^n}^{k/2^n} xp(x)dx + \int_n^{\infty} xp(x)dx$$

より，

$$\begin{aligned}0 &\leq \int_0^{\infty} xp(x)dx - \mathbb{E}X_n^+ \\ &= \int_0^{1/(2^n)} xp(x)dx + \sum_{k=2}^{n2^n} \int_{(k-1)/2^n}^{k/2^n} \left(x - \frac{k-1}{2^n}\right) p(x)dx + \int_n^{\infty} (x-n)\, p(x)dx \\ &\leq \frac{1}{2^n} \int_0^{1/(2^n)} p(x)dx + \frac{1}{2^n} \sum_{k=2}^{n2^n} \int_{(k-1)/2^n}^{k/2^n} p(x)dx + \int_n^{\infty} xp(x)dx \\ &= \frac{1}{2^n} \int_0^n p(x)dx + \int_n^{\infty} xp(x)dx.\end{aligned}$$

右辺は $n \to \infty$ のとき 0 に収束するので，

$$\mathbb{E}[X^+] = \int_0^\infty xp(x)dx$$

が従う．同様にして，

$$\mathbb{E}[X^-] = \int_{-\infty}^0 (-x)p(x)dx$$

を得る．ゆえに X は可積分であり，

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-] = \int_{-\infty}^\infty xp(x)dx.$$

□

次の二つの命題はそれぞれ命題 3.5.1, 3.5.2 の一般化であり有用である．証明は略す．

命題 3.5.3 X を $\mathbb{Z}^d$-値確率変数，$f : \mathbb{Z}^d \to \mathbb{R}$ とする．このとき，

$$\mathbb{E}[f(X)] = \sum_{n\in\mathbb{Z}^d} f(n)\mathbb{P}(X = n).$$

ただし，等号は右辺，左辺どちらかが有限値であれば他辺も有限値で，その値が一致することを意味する．f が非負なら，等号は両辺ともに $+\infty$ となる場合を含む．

命題 3.5.4 X は d 次元確率変数で密度関数 p を持つとする．このとき，$\mathbb{R}^d$ 上の連続関数 f に対し，

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)p(x)dx.$$

ただし，等号は右辺，左辺どちらかが有限値であれば他辺も有限値で，その値が一致することを意味する．f が非負なら，等号は両辺ともに $+\infty$ となる場合を含む．

$\mathbb{E}|X|^2 < \infty$ のとき，

$$\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

を X の**分散**といい，

$$\mathrm{Cov}(X,Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

を X と Y の**共分散**という．また，$\mathrm{Cov}(X,Y) = 0$ のとき，X と Y は**無相関**であるという．

例 3.5.5 $Z \sim N(0,1)$ のとき，$xe^{-x^2/2}$ が奇関数であることから

$$\mathbb{E}[Z] = \int_{-\infty}^\infty x\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = 0$$

であり，部分積分により

$$
\begin{aligned}
\mathbb{E}[Z^2] &= \int_{-\infty}^{\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} x \left(-\frac{e^{-x^2/2}}{\sqrt{2\pi}} \right)' dx \\
&= \left[x \left(-\frac{e^{-x^2/2}}{\sqrt{2\pi}} \right) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \left(-\frac{e^{-x^2/2}}{\sqrt{2\pi}} \right) dx \\
&= 1.
\end{aligned}
$$

よって $\mathbb{V}[Z] = 1$.

$X \sim N(m, v)$ のとき，

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}} dx = \int_{-\infty}^{\infty} (m + \sqrt{v} y) \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\
&= m + \sqrt{v} \int_{-\infty}^{\infty} y \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\
&= m.
\end{aligned}
$$

また，

$$
\begin{aligned}
\mathbb{E}[(X - \mathbb{E}[X])^2] &= \int_{-\infty}^{\infty} (x-m)^2 \frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}} dx = \int_{-\infty}^{\infty} (\sqrt{v} y)^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\
&= v \int_{-\infty}^{\infty} y^2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = v.
\end{aligned}
$$

よって $\mathbb{V}[X] = v$.

例 3.5.6 X がパラメータ λ の指数分布に従うとき，命題 3.5.4 より

$$
\begin{aligned}
\mathbb{E}[X] = \mathbb{E}[X^+] &= \int_0^{\infty} x \lambda e^{-\lambda x} dx = \int_0^{\infty} x (-e^{-\lambda x})' dx = \int_0^{\infty} e^{-\lambda x} dx \\
&= \frac{1}{\lambda}.
\end{aligned}
$$

例題 3.5.7 ポアソン分布に従う確率変数の平均と分散を求めよ．

解答. X がパラメータ c のポアソン分布に従うとすると（このことを $X \sim \mathrm{Poi}(c)$ と書くこともある），

$$
\mathbb{P}(X = n) = e^{-c} \frac{c^n}{n!}, \quad n = 0, 1, 2, \ldots
$$

である．これより，

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{n=0}^{\infty} n \mathbb{P}(X = n) = c e^{-c} \sum_{n=1}^{\infty} \frac{c^{n-1}}{(n-1)!} = c e^{-c} \sum_{n=0}^{\infty} \frac{c^n}{n!} \\
&= c.
\end{aligned}
$$

また，

$$\begin{aligned}\mathbb{E}[(X-c)^2] &= \sum_{n=0}^{\infty}(n-c)^2\mathbb{P}(X=n) = \sum_{n=1}^{\infty} n^2\frac{e^{-c}c^n}{n!} - 2c\sum_{n=0}^{\infty} n\frac{e^{-c}c^n}{n!} + c^2\sum_{n=0}^{\infty}\frac{e^{-c}c^n}{n!} \\ &= c\sum_{n=1}^{\infty} n\frac{e^{-c}c^{n-1}}{(n-1)!} - 2c^2 + c^2 = c\sum_{n=0}^{\infty}(n+1)\frac{e^{-c}c^n}{n!} - c^2 \\ &= c(c+1) - c^2 \\ &= c.\end{aligned}$$

よって $\mathbb{V}[X] = c$.

> **例題 3.5.8** $X \sim N(m, v)$ のとき，$\mathbb{P}(e^X \le \mathbb{E}[e^X]) > \dfrac{1}{2}$ を確かめよ．

解答．命題 3.5.4 より，

$$\mathbb{E}[e^X] = \int_{-\infty}^{\infty} e^x \frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}}\,dx$$

であり，

$$x - \frac{(x-m)^2}{2v} = -\frac{1}{2v}(x-(m+v))^2 + m + \frac{v}{2}$$

であるから，

$$\mathbb{E}[e^X] = e^{m+v/2}\int_{-\infty}^{\infty}\frac{e^{-(x-(m+v))^2/(2v)}}{\sqrt{2\pi v}}\,dx = e^{m+v/2}.$$

よって，

$$\begin{aligned}\mathbb{P}(e^X \le \mathbb{E}e^X) = \mathbb{P}\left(X \le m + \frac{v}{2}\right) &= \int_{-\infty}^{m+v/2}\frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}}\,dx \\ &= \frac{1}{2} + \int_{m}^{m+v/2}\frac{e^{-(x-m)^2/(2v)}}{\sqrt{2\pi v}}\,dx > \frac{1}{2}.\end{aligned}$$

- $Y = e^X$ の分布を**対数正規分布**と呼ぶ．
- 世帯年収を対数正規分布でモデル化するとき，上の結果はラフには「平均世帯年収以下の世帯数は半数を超える」ことを意味している．

> **例題 3.5.9** $X \sim U[-2, 2]$, $Y := X^6$ について，X と Y は無相関だが，独立ではないことを示せ．

解答．命題 3.5.4 より，

$$\mathbb{E}[X] = \frac{1}{4}\int_{-2}^{2} x\,dx = 0, \quad \mathbb{E}[Y] = \frac{1}{4}\int_{-2}^{2} x^6 dx = \frac{64}{7}$$

であり，定数 c に対し，

$$\mathbb{E}[X(Y-c)] = \mathbb{E}[X^7 - cX] = \frac{1}{4}\int_{-2}^{2}(x^7 - cx)\,dx = 0.$$

よって

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[X\left(Y - \frac{64}{7}\right)\right] = 0.$$

すなわち X と Y は無相関である．

一方，

$$\mathbb{P}(X \geq 1, Y \leq 1) = \mathbb{P}(X \geq 1, X^6 \leq 1) = \mathbb{P}(X = 1) = 0,$$
$$\mathbb{P}(X \geq 1) = \frac{1}{4}, \quad \mathbb{P}(Y \leq 1) = \mathbb{P}(|X| \leq 1) = \frac{1}{2}.$$

よって $0 = \mathbb{P}(X \geq 1, Y \leq 1) < \mathbb{P}(X \geq 1)\mathbb{P}(Y \leq 1)$ であるから X と Y は独立ではない．

期待値の基本的性質をみていこう．

定理 3.5.10 X, Y を実確率変数とするとき，以下が成り立つ．

(1) $\mathbb{E}|X| < \infty$，$\mathbb{E}|Y| < \infty$ のとき，任意の $\alpha, \beta \in \mathbb{R}$ に対して

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y].$$

(2) $X \leq Y$ で，$\mathbb{E}[X]$ と $\mathbb{E}[Y]$ は共に存在するとき，

$$\mathbb{E}[X] \leq \mathbb{E}[Y].$$

(3) $\mathbb{E}|X| < \infty$ のとき，

$$|\mathbb{E}[X]| \leq \mathbb{E}|X|.$$

証明．(1)．確率ベクトル (X, Y) の結合密度関数 $\rho_{X,Y}$ が存在する場合に示そう．一般の場合の証明は [10] や [8] など測度論的確率論のテキストを参照のこと．この場合，X と Y の密度関数 ρ_X, ρ_Y が存在し，それぞれ

$$\rho_X(x) = \int_{-\infty}^{\infty} \rho_{X,Y}(x,y)dy, \quad \rho_Y(y) = \int_{-\infty}^{\infty} \rho_{X,Y}(x,y)dx$$

により与えられる．このことと，命題 3.5.4，および積分の線形性より

$$\begin{aligned}\mathbb{E}[\alpha X + \beta Y] &= \int_{\mathbb{R}^2} (\alpha x + \beta y)\rho(x,y)dxdy = \alpha\int_{\mathbb{R}^2} x\rho(x,y)dxdy + \beta\int_{\mathbb{R}^2} y\rho(x,y)dxdy \\ &= \alpha\int_{\mathbb{R}} x\rho_X(x)dx + \beta\int_{\mathbb{R}} y\rho_Y(y)dy = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y].\end{aligned}$$

(2)．$Z = Y - X$ とおくと，Z は非負確率変数だから期待値の定義より

$$\mathbb{E}[Z] = \lim_{n\to\infty}\left\{\sum_{k=1}^{n2^n} \frac{k-1}{2^n}\mathbb{P}\left(\frac{k-1}{2^n} \leq Z < \frac{k}{2^n}\right) + n\mathbb{P}(Z \geq n)\right\} \geq 0.$$

これと本定理 (1) の結果より $\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[Z] \geq 0$.

(3)．$|X| = X^+ + X^-$ と書ける．X^+, X^- は共に非負だから本定理 (2) の結果より，$\mathbb{E}[X^+] \geq 0$, $\mathbb{E}[X^-] \geq 0$. これらと期待値の定義，線形性より，

$$|\mathbb{E}[X]| = |\mathbb{E}[X^+] - \mathbb{E}[X^-]| \leq \mathbb{E}[X^+] + \mathbb{E}[X^-] = \mathbb{E}|X|.$$

□

- 定理 3.5.10 より，$\mathbb{E}[X], \mathbb{E}[X^2]$ がともに存在するとき，

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{3.5.3}$$

後で示すコーシー・シュワルツの不等式（系 3.5.19）を使うと，$\mathbb{E}X^2 < \infty$ のとき，$\mathbb{E}[X]$ も存在するので，(3.5.3) が成り立つ．同様に，$\mathbb{E}[X], \mathbb{E}[Y], \mathbb{E}[XY]$ がすべて存在するとき，

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \tag{3.5.4}$$

これについてもコーシー・シュワルツの不等式（系 3.5.19）を使うと，$\mathbb{E}X^2 < \infty, \mathbb{E}[Y^2] < \infty$ のとき，$\mathbb{E}[X], \mathbb{E}[Y], \mathbb{E}[XY]$ がすべて存在するので (3.5.4) が成り立つ．

- $\mathbb{E}[X, A] := \mathbb{E}[X 1_A]$, $A \in \mathcal{F}$.
- $\mathbb{P}(X = Y) = 1$ のとき，X と Y は**ほとんど確実に**（almost surely）等しいといい，

$$X = Y, \quad \text{a.s.}$$

と書く．

命題 3.5.11 X, Y を平均可能確率変数とする．このとき，

$$\mathbb{E}[X, A] = \mathbb{E}[Y, A], \quad A \in \mathcal{F},$$

ならば

$$X = Y \quad \text{a.s.}$$

証明．$Z = X - Y$ とおく．各 $n \in \mathbb{N}$ に対し，事象 $A_n^+ := \{Z > 1/n\}$, $A_n^- := \{Z < -1/n\}$ を考える．このとき，

$$0 = \mathbb{E}[Z, A_n^+] \geq \frac{1}{n}\mathbb{P}(A_n^+),$$
$$0 = \mathbb{E}[Z, A_n^-] \leq -\frac{1}{n}\mathbb{P}(A_n^-)$$

が成り立つ．よって任意の $n \in \mathbb{N}$ に対して

$$\mathbb{P}(A_n^+) = \mathbb{P}(A_n^-) = 0.$$

従って，確率の連続性（命題 3.3.9）より，

$$\begin{aligned}\mathbb{P}(Z \neq 0) = \mathbb{P}(Z > 0) + \mathbb{P}(Z < 0) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^+\right) + \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^-\right) \\ &= \lim_{n\to\infty} \mathbb{P}(A_n^+) + \lim_{n\to\infty} \mathbb{P}(A_n^-) \\ &= 0.\end{aligned}$$

□

命題 3.5.12 X は非負確率変数で $\mathbb{E}[X]=0$ を満たすとする．このとき，

$$X=0 \text{ a.s.}$$

証明．各 $n\in\mathbb{N}$ に対して事象 $A_n=\{X>1/n\}$ を考える．このとき，

$$0=\mathbb{E}[X]\geq\mathbb{E}[X,A_n]\geq\frac{1}{n}\mathbb{P}(A_n).$$

よって任意の $n\geq 1$ に対して $\mathbb{P}(A_n)=0$. これと確率の連続性より

$$\mathbb{P}(X>0)=\mathbb{P}\left(\bigcup_{n=1}^{\infty}A_n\right)=\lim_{n\to\infty}\mathbb{P}(A_n)=0.$$

□

命題 3.5.13 X が平均可能確率変数ならば，$|X|<\infty$ a.s.

- 逆は成り立たない．例えば，$Z\sim N(0,1)$ として $X:=e^{Z^2}$ を考えよ．

命題 3.5.13 の証明．$A=\{|X|=+\infty\}\in\mathcal{F}$ とおく．このとき，任意の $\omega\in A$ と $N\geq 1$ に対して $|X(\omega)|\geq N$ である．これより，

$$\mathbb{P}(A)\leq\mathbb{E}\left[\frac{|X|}{N},A\right]\leq\frac{\mathbb{E}|X|}{N}.$$

$N\to\infty$ として $\mathbb{P}(A)=0$ を得る． □

命題 3.5.14 平均可能確率変数 X に対し，

$$\mathbb{E}[|X|]=\int_0^{\infty}\mathbb{P}(|X|>x)dx.$$

証明の概略．フビニの定理により非負確率変数に対する期待値と積分の順序変更が正当化される．フビニの定理については [10] や [8] など測度論的確率論のテキストあるいはルベーグ積分のテキストをみよ．これを用いると，

$$\begin{aligned}\mathbb{E}|X|&=\mathbb{E}\left[\int_0^{|X|}dx\right]=\mathbb{E}\int_0^{\infty}1_{\{|X|>x\}}dx=\int_0^{\infty}\mathbb{E}[1_{\{|X|>x\}}]dx\\&=\int_0^{\infty}\mathbb{P}(|X|>x)dx.\end{aligned}$$

□

確率変数 X が $\mathbb{N}\cup\{0\}$ に値をとるとき，$n\leq x<n+1$ ならば $\mathbb{P}(X>x)=\mathbb{P}(X>n)$ である．よって

$$\int_0^{\infty}\mathbb{P}(X>x)dx=\sum_{n=0}^{\infty}\int_n^{n+1}\mathbb{P}(X>x)dx=\sum_{n=0}^{\infty}\mathbb{P}(X>n).$$

ゆえに命題 3.5.14 の系として次が得られる．

系 3.5.15 $\mathbb{N} \cup \{0\}$-値確率変数 X に対して，

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n).$$

注意 3.5.16 系 3.5.15 は次のようにして直接導くこともできる．X が $\mathbb{N} \cup \{0\}$-値確率変数で，$p_n = \mathbb{P}(X = n)$ とおくとき，

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{n=1}^{\infty} np_n = p_1 + 2p_2 + 3p_3 + \cdots \\
&= \sum_{n=1}^{\infty} p_n + \sum_{n=2}^{\infty} p_n + \sum_{n=3}^{\infty} p_n + \cdots \\
&= \mathbb{P}(X \geq 1) + \mathbb{P}(X \geq 2) + \mathbb{P}(X \geq 3) + \cdots \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X > n).
\end{aligned}$$

独立性と期待値の関係については次が基本的である．

定理 3.5.17 X と Y はともに平均可能確率変数で，互いに独立とする．このとき，XY も平均可能で，

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- 定理 3.5.17 より，X と Y が独立で**二乗可積分**（i.e., $\mathbb{E}|X|^2 < \infty$, $\mathbb{E}|Y|^2 < \infty$）ならば，

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y].$$

定理 3.5.17 の証明．（第一段階）．まず，$X, Y \geq 0$ を仮定する．確率変数列 $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$ を

$$\begin{aligned}
X_n &= \sum_{k=1}^{n2^n} (k-1)2^{-n} 1_{\{(k-1)2^{-n} \leq X < k2^{-n}\}}, \\
Y_n &= \sum_{k=1}^{n2^n} (k-1)2^{-n} 1_{\{(k-1)2^{-n} \leq Y < k2^{-n}\}}
\end{aligned}$$

により定義すると，

$$0 \leq X_1Y_1 \leq X_2Y_2 \leq \cdots \leq XY$$

かつ

$$XY = \lim_{n\to\infty} X_nY_n$$

が成り立つ．よって $\{X_nY_n\}_{n=1}^{\infty}$ は XY の近似単純確率変数列である．ゆえに期待値の定義より

$$\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_nY_n].$$

他方，独立性より，

$$
\begin{aligned}
&\mathbb{E}[X_nY_n]\\
&=\mathbb{E}\left[\sum_{k,j=1}^{n2^n}(k-1)(j-1)2^{-2n}1_{\{(k-1)2^{-n}\le X<k2^{-n},(j-1)2^{-n}\le Y<j2^{-n}\}}\right]\\
&=\sum_{k,j=1}^{n2^n}(k-1)(j-1)2^{-2n}\mathbb{P}((k-1)2^{-n}\le X<k2^{-n},(j-1)2^{-n}\le Y<j2^{-n})\\
&=\sum_{k=1}^{n2^n}(k-1)2^{-n}\mathbb{P}((k-1)2^{-n}\le X<k2^{-n})\sum_{j=1}^{n2^n}(j-1)2^{-n}\mathbb{P}((j-1)2^{-n}\le Y<j2^{-n})\\
&=\mathbb{E}[X_n]\mathbb{E}[Y_n].
\end{aligned}
$$

従って，

$$
\mathbb{E}[XY]=\lim_{n\to\infty}\mathbb{E}[X_n]\mathbb{E}[Y_n]=\mathbb{E}[X]\mathbb{E}[Y].
$$

（第二段階）．X,Y が一般の場合に示す．第一段階の結果より，$\mathbb{E}|XY|=\mathbb{E}|X|\mathbb{E}|Y|<\infty$．よって XY は可積分である．さらに，第一段階の結果を X^+, X^-, Y^+, Y^- たちに適用して

$$
\begin{aligned}
\mathbb{E}[XY]&=\mathbb{E}[(X^+-X^-)(Y^+-Y^-)]=\mathbb{E}[X^+Y^+]-\mathbb{E}[X^+Y^-]-\mathbb{E}[X^-Y^+]+\mathbb{E}[X^-Y^-]\\
&=\mathbb{E}[X^+]\mathbb{E}[Y^+]-\mathbb{E}[X^+]\mathbb{E}[Y^-]-\mathbb{E}[X^-]\mathbb{E}[Y^+]+\mathbb{E}[X^-]\mathbb{E}[Y^-]\\
&=\mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}
$$

を得る．　□

次に，期待値を評価するために有用な不等式をいくつか紹介する．

定理 3.5.18 (ヘルダーの不等式) X,Y を実確率変数，$p,q\in\mathbb{R}$ は $1<p,q<\infty$ かつ $(1/p)+(1/q)=1$ を満たすとする．このとき，

$$
\mathbb{E}|XY|\le(\mathbb{E}[|X|^p])^{1/p}(\mathbb{E}[|Y|^q])^{1/q}.
$$

証明. $\mathbb{E}|X|^p=0$ のとき，命題 3.5.12 より $X=0$ a.s. よってこの場合，命題の不等式は両辺 0 で成立する．$\mathbb{E}|Y|^q=0$ のときも同様である．

$\mathbb{E}|X|^p>0$, $\mathbb{E}|Y|^q>0$ を仮定する．$a,b\ge0$ を任意にとって固定する．関数

$$
f(t)=\frac{t^p}{p}+\frac{1}{q}-t,\quad t\ge0,
$$

は $t=1$ のとき最小値 0 をとる．よって任意の $t\ge0$ に対して $f(t)\ge0$. これを書き換えると，

$$
t\le\frac{t^p}{p}+\frac{1}{q}
$$

であり，両辺 b^q を掛けて

$$
b^qt\le\frac{b^qt^p}{p}+\frac{b^q}{q},\quad t\ge0,
$$

を得る．$t=ab/b^q$ として

$$
ab\le\frac{a^p}{p}+\frac{b^q}{q}.
$$

$a = |X|(\mathbb{E}|X|^p)^{-1/p}$, $b = |Y|(\mathbb{E}|Y|^q)^{-1/q}$ を代入し，両辺期待値をとると，

$$\frac{\mathbb{E}|XY|}{\mathbb{E}[|X|^p]^{1/p}\mathbb{E}[|Y|^q]^{1/q}} \le \frac{1}{p}\mathbb{E}\left[\frac{|X|^p}{\mathbb{E}|X|^p}\right] + \frac{1}{q}\mathbb{E}\left[\frac{|Y|^q}{\mathbb{E}|Y|^q}\right] = \frac{1}{p} + \frac{1}{q} = 1.$$

これより命題が従う．□

系 3.5.19 (コーシー・シュワルツの不等式) X, Y を実確率変数とするとき，

$$\mathbb{E}|XY| \le (\mathbb{E}[|X|^2])^{1/2}(\mathbb{E}[|Y|^2])^{1/2}.$$

- コーシー・シュワルツの不等式より，$\mathbb{E}|X|^2 < \infty$ のとき，X は平均可能である．このことと期待値の線形性より

  $$\mathbb{V}[X] = \mathbb{E}|X|^2 - (\mathbb{E}X)^2.$$

  さらに，$\mathbb{E}|Y|^2 < \infty$ ならば，定理 3.5.10 (3) とコーシー・シュワルツの不等式より

  $$\begin{aligned}|\mathrm{Cov}(X,Y)| &= |\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]| \le \mathbb{E}|(X-\mathbb{E}X)(Y-\mathbb{E}Y)| \\ &\le \sqrt{\mathbb{E}|X-\mathbb{E}X|^2}\sqrt{\mathbb{E}|Y-\mathbb{E}Y|^2} = \sqrt{\mathbb{V}[X]}\sqrt{\mathbb{V}[Y]} < \infty.\end{aligned} \tag{3.5.5}$$

  Y も可積分であるから，結局，期待値の線形性より，

  $$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

- (3.5.5) より，**相関係数**

  $$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbb{V}[X]}\sqrt{\mathbb{V}[Y]}}$$

  は

  $$-1 \le \rho(X,Y) \le 1$$

  を満たしている．今，$\langle U, V\rangle = \mathbb{E}[UV]$ を確率変数 U, V の「内積」，$\|U\|_2 = \sqrt{\mathbb{E}U^2}$ を U の「ノルム」とみなすと（ここでは厳密さは気にしなくてもよい），

  $$\rho(X,Y) = \frac{\langle \tilde{X}, \tilde{Y}\rangle}{\|\tilde{X}\|_2\|\tilde{Y}\|_2}$$

  と表すことができる．ただし $\tilde{X} = X - \mathbb{E}X$, $\tilde{Y} = Y - \mathbb{E}Y$．これは 2 つのベクトルのなす角の余弦（コサイン）の形をしている．すなわち相関係数は，**偏差** $\tilde{X}, \tilde{Y}$ を「ベクトル」とみなしたときの「なす角の余弦」と解釈できる．特に，$\rho(X,Y) = \pm 1$ は $\tilde{X}$ と $\tilde{Y}$ が「平行」であること（互いに定数倍の関係）に，$\rho(X,Y) = 0$ は $\tilde{X}$ と $\tilde{Y}$ が「直交」していること（無相関）に対応する．

系 3.5.20 実確率変数 X と $0 < p < q < \infty$ を満たす $p, q \in \mathbb{R}$ に対し，

$$(\mathbb{E}[|X|^p])^{1/p} \le (\mathbb{E}[|X|^q])^{1/q}.$$

証明．$X' = |X|^p, Y' = 1$, $p' = q/p$, $q' = q/(q-p)$ に対してヘルダーの不等式（定理 3.5.18）を適用すると，

$$\mathbb{E}|X|^p \leq \left(\mathbb{E}|X|^{p \cdot q/p}\right)^{p/q} \mathbb{E}[1^{q/(q-p)}]^{(q-p)/q} = (\mathbb{E}|X|^q)^{p/q}.$$

これより系が得られる． □

定理 3.5.21 (**マルコフの不等式**) X を非負確率変数とする．このとき，任意の $a > 0$ に対し，

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

証明．$1_{\{X \geq a\}} \leq (X/a)1_{\{X \geq a\}}$ より，

$$\mathbb{P}(X \geq a) = \mathbb{E}[1_{\{X \geq a\}}] \leq \mathbb{E}\left[\frac{X}{a} 1_{\{X \geq a\}}\right] \leq \frac{1}{a}\mathbb{E}X.$$

□

- マルコフの不等式の右辺の $1/a$ は**最良定数**である．実際，$\mathbb{P}(X = 0) = 24/25$, $\mathbb{P}(X = 5) = 1/25$ のとき，$\mathbb{P}(X \geq 5) = 1/25 = (1/5)\mathbb{E}[X]$.

定理 3.5.21 は次のように一般化できる．証明はほとんど同じなので割愛する．

定理 3.5.22 (**一般化マルコフの不等式**) X を実確率変数，$\varphi : [0, \infty) \to (0, \infty)$ を単調非減少関数とする．このとき，

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[\varphi(|X|)]}{\varphi(a)}.$$

次の**チェビシェフの不等式**も有用である．これは一般化マルコフの不等式により容易に導くことができる．

定理 3.5.23 (**チェビシェフの不等式**) X は確率変数で $\mathbb{E}|X|^2 < \infty$ を満たすとする．このとき，任意の $\varepsilon > 0$ に対し，

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathbb{V}[X]}{\varepsilon^2}.$$

最後に**イエンセンの不等式**を紹介して不等式シリーズを閉めよう．

定理 3.5.24 (**イエンセンの不等式**) f を区間 J 上に定義された連続凸関数とする．X は平均可能確率変数で $f(X)$ も平均可能と仮定する．このとき，

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

3.6 期待値の収束

確率変数列 $\{X_n\}_{n=1}^{\infty}$ が確率変数 X に**概収束**するとは，$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$, すなわち $\lim_{n\to\infty} X_n = X$ a.s. が成り立つときにいう．

例 3.6.1 $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}[0,1], \mathrm{Leb})$ において，$X_n(\omega) = n\omega^n$ は 0 に概収束する．

$\{X_n\}$ を非負確率変数 X の近似単純確率変数列とするとき，期待値の定義の際にみたように，$\{X_n\}$ は X に概収束，かつ $\lim_{n\to\infty}\mathbb{E}[X_n]=\mathbb{E}[X]$ が成り立つ．すなわち，この場合は

$$\lim_{n\to\infty}\mathbb{E}[X_n]=\mathbb{E}\left[\lim_{n\to\infty}X_n\right] \tag{3.6.1}$$

が成り立っている．

一般の確率変数列に対して期待値と極限の交換 (3.6.1) が成り立つための十分条件を調べよう．

定理 3.6.2 (**単調収束定理**) $X_1, X_2, \ldots$ は非負確率変数列で，

$$0\le X_1\le X_2\le\cdots\le X_n\le\cdots \quad \text{a.s.}$$

とする．このとき，

$$\lim_{n\to\infty}\mathbb{E}[X_n]=\mathbb{E}\left[\lim_{n\to\infty}X_n\right].$$

- 定理 3.6.2 において単調性の仮定は本質的である．例えば $(\Omega,\mathcal{F},\mathbb{P})=([0,1],\mathcal{B}[0,1],\mathrm{Leb})$ において，$X_n(\omega)=n\omega^{n-1}$ は $\lim_{n\to\infty}X_n=0$, a.s. を満たすが $\lim_{n\to\infty}\mathbb{E}[X_n]=1$.

定理 3.6.2 の証明．仮定より

$$A:=\{\omega\in\Omega \mid 0\le X_1(\omega)\le X_2(\omega)\le\cdots\}=\bigcap_{n=1}^{\infty}\{X_n\le X_{n+1}\}\in\mathcal{F}$$

は $\mathbb{P}(A)=1$ を満たす．$\omega\in A$ に対し $X(\omega):=\lim_{n\to\infty}X_n(\omega)$ と定め，$\omega\in A^c$ に対しては $X_n(\omega)=0$ と再定義し，$X(\omega):=0$ と定める．この定義の下，任意の $\omega\in\Omega$ に対して

$$0\le X_1(\omega)\le X_2(\omega)\le\cdots$$

となる．このように X_n を再定義しても

$$\mathbb{E}[X_n,A]=\mathbb{E}[X_n]-\mathbb{E}[X_n,A^c]=\mathbb{E}[X_n]$$

であるから期待値に影響を与えない．すなわち修正後の $\{X_n\}$ に対して定理を示せばよい．

（第一段階）．まず $\mathbb{E}[X]<\infty$ の場合に示す．$X_n\le X_{n+1}$ より

$$\lim_{n\to\infty}\mathbb{E}[X_n]\le\mathbb{E}[X] \tag{3.6.2}$$

は明らか．期待値の定義より，単調非減少単純確率変数列 $\{\tilde{X}_n\}$ が存在し，$\mathbb{E}[X]=\lim_{n\to\infty}\mathbb{E}[\tilde{X}_n]$. よって，任意の $\varepsilon>0$ に対し $k\in\mathbb{N}$ が存在し，

$$\mathbb{E}[X]<\mathbb{E}[\tilde{X}_k]+\varepsilon.$$

$Y=\tilde{X}_k$ とおき，事象 $A_n:=\{X_n+\varepsilon>Y\}$ を考える．このとき，

$$Y1_{A_n}\le(X_n+\varepsilon)1_{A_n}\le X_n+\varepsilon.$$

各点で X_n は単調非減少で X に収束することと，$Y\le X$ より，

$$A_n\subset A_{n+1},\ \ n\ge1,\quad \bigcup_{n=1}^{\infty}A_n=\Omega.$$

さらに，Y は $\sum_i c_i 1_{B_i}$ の形で表されるので，確率の連続性より，$n \to \infty$ の極限で，

$$\mathbb{E}[Y1_{A_n}] = \sum_i c_i \mathbb{P}(B \cap A_n) \to \sum_i c_i \mathbb{P}\left(\bigcup_{N=1}^{\infty} A_N \cap B\right) = \mathbb{E}[Y].$$

よって，

$$\mathbb{E}[X] - \varepsilon < \mathbb{E}[Y] = \lim_{n\to\infty} \mathbb{E}[Y1_{A_n}] \le \lim_{n\to\infty} \mathbb{E}[X_n] + \varepsilon.$$

ε は任意だったから

$$\mathbb{E}[X] \le \lim_{n\to\infty} \mathbb{E}[X_n].$$

これと (3.6.2) を併せて定理の主張が従う．

（第二段階）．$\mathbb{E}[X] = +\infty$ のとき，単調非減少非負単純確率変数列 $\{\tilde{X}_n\}$ が存在し，

$$\lim_{n\to\infty} \tilde{X}_n = X, \quad \lim_{n\to\infty} \mathbb{E}[\tilde{X}_n] = +\infty.$$

よって，任意の $N \ge 1$ に対し $Y := \tilde{X}_k$ が存在し，$\mathbb{E}[Y] > N$．$A_n = \{X_n \ge Y\}$ とおいて第一段階と同じ議論を行うことにより，

$$N < \mathbb{E}[Y] = \lim_{n\to\infty} \mathbb{E}[Y1_{A_n}] \le \lim_{n\to\infty} \mathbb{E}[X_n 1_{A_n}] \le \lim_{n\to\infty} \mathbb{E}[X_n].$$

$N \to \infty$ として，

$$\lim_{n\to\infty} \mathbb{E}[X_n] = +\infty = \mathbb{E}[X]$$

を得る．□

単調収束定理の系として次が得られる．

系 3.6.3 $X, X_1, X_2, \ldots$ を非負確率変数とする．このとき以下が成り立つ．

(1) 期待値が発散する場合も含めて

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[X_n].$$

(2) $A_n \in \mathcal{F}$, $A_i \cap A_j = \emptyset$ $(i \ne j)$ ならば

$$\mathbb{E}\left[X, \bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[X, A_n].$$

証明．(1)．$Y_n := \sum_{k=1}^{n} X_k$ は定理 3.6.2 の条件を満たすので，

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \mathbb{E}\left[\lim_{n\to\infty} Y_n\right] = \lim_{n\to\infty} \mathbb{E}[Y_n] = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}[X_k] = \sum_{n=1}^{\infty} \mathbb{E}[X_n].$$

(2)．直前の (1) の結果より，

$$\mathbb{E}\left[X, \bigcup_{n=1}^{\infty} A_n\right] = \mathbb{E}\left[\sum_{n=1}^{\infty} X1_{A_n}\right] = \sum_{n=1}^{\infty} \mathbb{E}[X1_{A_n}].$$

□

定理 3.6.4 (ルベーグの収束定理（優収束定理）) $\{X_n\}$ は概収束する確率変数列で，ある平均可能非負確率変数 Y が存在し

$$|X_n| \leq Y, \quad n = 1, 2, \ldots$$

を満たすと仮定する．このとき，

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right].$$

証明. X を $\{X_n\}$ の概収束極限とする．すなわち，$X = \lim_{n\to\infty} X_n$, a.s. このとき，極限が存在しているから，$\{X_n\}$ の上極限と下極限は確率 1 で一致する．言い換えると，

$$X = \lim_{n\to\infty} \inf_{k\geq n} X_k = \lim_{n\to\infty} \sup_{k\geq n} X_k, \quad \text{a.s.}$$

各 $n \in \mathbb{N}$ に対して，$Z_n = \inf_{k\geq n}(Y - X_k)$, $Z'_n = \inf_{k\geq n}(Y + X_k)$ とおくと，$\{Z_n\}$ と $\{Z'_n\}$ はともに非負単調非減少であり，

$$\lim_{n\to\infty} Z_n = Y - \lim_{n\to\infty} \sup_{k\geq n} X_k = Y - X,$$
$$\lim_{n\to\infty} Z'_n = Y + \lim_{n\to\infty} \inf_{k\geq n} X_k = Y + X.$$

よって単調収束定理（定理 3.6.2）より，

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}[Z_n] &= \mathbb{E}[Y] - \mathbb{E}[X], \\ \lim_{n\to\infty} \mathbb{E}[Z'_n] &= \mathbb{E}[Y] + \mathbb{E}[X]. \end{aligned} \tag{3.6.3}$$

他方，Z_n, Z'_n の定義より，$Z'_n - Y \leq X_n \leq Y - Z_n$. よって

$$\mathbb{E}[Z'_n] - \mathbb{E}[Y] \leq \mathbb{E}[X_n] \leq \mathbb{E}[Y] - \mathbb{E}[Z_n], \quad n \in \mathbb{N}. \tag{3.6.4}$$

従って，(3.6.3) と (3.6.4) より，はさみうちの原理を適用し，

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n].$$

□

- 定理 3.6.4 における確率変数 Y が定数でとれるとき，ルベーグの収束定理を**有界収束定理**と呼ぶこともある．

微分と期待値の交換は次の命題によって正当化される．

命題 3.6.5 X を $\mathbb{R}^d$-値確率変数，I を $\mathbb{R}$ の開区間とする．関数 $f : I \times \mathbb{R}^d \to \mathbb{R}$ は，任意の $(t, x) \in I \times \mathbb{R}^d$ に対して偏微分 $(\partial/\partial t)f(t, x)$ が存在し，$f(t, X)$ が任意の $t \in I$ に対して確率変数となるものとする．さらに，

$$\sup_{t\in I} \left|\frac{\partial f}{\partial t}(t, X)\right|$$

は平均可能と仮定する．このとき，

$$\frac{d}{dt}\mathbb{E}[f(t, X)] = \mathbb{E}\left[\frac{\partial f}{\partial t}(t, X)\right].$$

証明．$t \in I$ を任意にとり固定する．$\{h_n\}$ を 0 に収束する任意の数列とする．このとき，

$$\frac{\partial}{\partial t} f(t,X) = \lim_{n\to\infty} \frac{f(t+h_n,X) - f(t,X)}{h_n}.$$

確率変数の極限として表されるので，命題 3.4.4 より，$(\partial/\partial t)f(t,X)$ は確率変数である．平均値の定理より，$s_n(\omega) \in I$ が存在して，

$$\left| \frac{f(t+h_n,X) - f(t,X)}{h_n} \right| = \left| \frac{\partial f}{\partial t}(s_n,X) \right| \leq \sup_{s\in I} \left| \frac{\partial f}{\partial t}(s,X) \right|.$$

ここで

$$X_n = \frac{f(t+h_n,X) - f(t,X)}{h_n}, \quad Y = \sup_{s\in I} \left| \frac{\partial f}{\partial t}(s,X) \right|$$

とおくと，命題 3.4.4 と仮定より Y は可積分確率変数である．よって，ルベーグの収束定理（定理 3.6.4）より，

$$\frac{d}{dt}\mathbb{E}[f(t,X)] = \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right] = \mathbb{E}\left[\frac{\partial f}{\partial t}(t,X)\right].$$

□

X を実確率変数，$I \subset \mathbb{R}$ を開区間とする．

$$G(t) := \mathbb{E}[e^{tX}], \quad t \in I,$$

が存在するとき，G を X の**モーメント母関数**と呼ぶ.

定理 3.6.6 X のモーメント母関数 G が $I = (-\varepsilon, \varepsilon)$ 上存在するような適当な $\varepsilon > 0$ がとれると仮定する．このとき，任意の $n \in \mathbb{N}$ に対し，

$$\mathbb{E}[X^n] = G^{(n)}(0).$$

証明．$t \in I$ を任意にとり固定する．このとき，$-\varepsilon < t-\delta < t+\delta < \varepsilon$ を満たすような $\delta > 0$ が存在する．$\delta > 0$ に対し実数 $x \geq 0$ で $x \leq \frac{1}{e\delta}e^{\delta x}$ が成り立つ（$x \mapsto e^{\delta x}/(e\delta) - x$ の最小値を $x = 1/\delta$ でとり，値は 0）．これより，

$$\begin{aligned}
\mathbb{E}\left[|X| \sup_{|t|\leq\varepsilon/2} e^{tX}\right] &= \mathbb{E}\left[X 1_{\{X>0\}} \sup_{|t|\leq\varepsilon/2} e^{tX} + (-X)1_{\{X\leq 0\}} \sup_{|t|\leq\varepsilon/2} e^{tX}\right] \\
&\leq \frac{1}{e\delta}\mathbb{E}\left[1_{\{X>0\}} \sup_{|t|\leq\varepsilon/2} e^{(t+\delta)X} + 1_{\{X\leq 0\}} \sup_{|t|\leq\varepsilon/2} e^{(t-\delta)X}\right] \\
&= \frac{1}{e\delta}\mathbb{E}\left[1_{\{X>0\}} e^{(\varepsilon/2+\delta)X} + 1_{\{X\leq 0\}} e^{(-\varepsilon/2-\delta)X}\right] \\
&\leq \frac{1}{e\delta}\left(\mathbb{E}\left[e^{(\varepsilon/2+\delta)X}\right] + \mathbb{E}\left[e^{(-\varepsilon/2-\delta)X}\right]\right) < \infty.
\end{aligned}$$

従って，命題 3.6.5 が適用でき，

$$\frac{d}{dt}\mathbb{E}[e^{tX}] = \mathbb{E}\left[Xe^{tX}\right], \quad |t| < \frac{\varepsilon}{2}.$$

$t = 0$ として $G'(0) = \mathbb{E}[X]$ を得る．同様の議論を繰り返すことで，一般の $n \in \mathbb{N}$ に対して $G^{(n)}(0) = \mathbb{E}[X^n]$ が成り立つことも示せる． □

例題 3.6.7 $X \sim N(0,1)$ のとき，モーメント母関数を用いて

$$\mathbb{E}[X^n] = \begin{cases} (n-1)!! & (n\text{が偶数}), \\ 0 & (n\text{が奇数}) \end{cases}$$

を確かめよ．

解答．まず，任意の $t \in \mathbb{R}$ に対して，

$$G(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} dx = e^{t^2/2} = \sum_{j=0}^{\infty} \frac{t^{2j}}{2^j j!}.$$

これより，$k \geq 1$ に対し，

$$G^{(2k-1)}(t) = \sum_{j=k}^{\infty} \frac{(2j)(2j-1)\cdots(2j-(2k-2))}{2^j j!} t^{2j-(2k-1)},$$
$$G^{(2k)}(t) = \sum_{j=k}^{\infty} \frac{(2j)(2j-1)\cdots(2j-(2k-1))}{2^j j!} t^{2j-2k}.$$

よって，

$$G^{(2k-1)}(0) = 0, \quad G^{(2k)}(0) = \frac{(2k)!}{2^k k!} = (2k-1)!!.$$

3.7 無限回コイン投げ空間

無限回コイン投げ標本空間 $\Omega = \{\omega = (\omega_1, \omega_2, \ldots) \mid \omega_n = 0, 1,\ \ n \in \mathbb{N}\}$ を考える．各 $i_m = 0, 1$ に対し，

$$\Omega_{i_1 i_2 \cdots i_k} := \{\omega = (i_1, i_2, \ldots, i_k, \omega_{k+1}, \omega_{k+2}, \ldots) \mid \omega_n = 0, 1,\ \ n \geq k+1\}$$

と定める．すなわち，$\Omega_{i_1 i_2 \cdots i_k} = \{1 \text{ 回目} = i_1, \ldots, k \text{ 回目} = i_k\}$ である．従って，この事象 $\Omega_{i_1 i_2 \cdots i_k}$ の確率 $P(\Omega_{i_1 i_2 \cdots i_k})$ を $P(\Omega_{i_1 i_2 \cdots i_k}) := 2^{-k}$ により定めるのは自然である．さらに，$\Omega_{i_1 i_2 \cdots i_k}$ たちから作られる集合族 $\mathcal{A}$ を以下のように定義する．

$$\mathcal{A} := \left\{ \bigcup_{i=1}^{n} I_i \ \middle|\ I_i \in \mathcal{I},\ I_i \cap I_j = \emptyset\ (i \neq j) \right\}.$$

ただし，

$$\mathcal{I} := \{\Omega_{i_1 i_2 \cdots i_k} \mid i_\ell = 0, 1\ (\ell = 1, \ldots, k),\ k \geq 1\}.$$

$\mathcal{A}$ の事象については

$$P(A) := \sum_{i=1}^{n} P(I_i), \quad A = \bigcup_{i=1}^{n} I_i \in \mathcal{A},$$

により確率を付与する．

命題 3.7.1 P は $\mathcal{A}$ 上で σ-加法的である．これより，$\mathcal{F} := \sigma[\mathcal{A}]$ 上の確率測度 $\mathbb{P}$ が一意的に存在し，

$$\mathbb{P}(A) = P(A), \quad A \in \mathcal{A}.$$

- この $(\Omega, \mathcal{F}, \mathbb{P})$ が求める無限回コイン投げ空間である．
- 証明については [8] の定理 2.4, 例題 2.2 をみよ．

次の命題は $\mathbb{P}$ が定義できない $A \subset \Omega$ の存在を意味する．

命題 3.7.2 無限回コイン投げ空間 $(\Omega, \mathcal{F}, \mathbb{P})$ において，$\mathbb{P}$ の 2^Ω 上への拡張は存在しない．

証明の概略．そのような拡張が存在すると仮定し，$\{X_n\}$ を IID で $\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = 1/2$ とする．このとき，$X := \sum_{n=1}^\infty 2^{-n} X_n \sim U[-1, 1]$ である（これについては後述の例題 5.1.11 をみよ）．ここで，E を $[0, 1] \subset E \subset [-1, 2]$ で $\mathrm{Leb}(E)$ が定義されない集合とすると（このような E はルベーグ非可測集合と呼ばれている．構成については [9] 参照)，$\{X \in E\} \in 2^\Omega$ だが $\mathbb{P}(X \in E) = \mathrm{Leb}(E)$ は定義されない． □

章末問題

問題 3.1 $\Omega = [0, 1] \times [0, 1] \subset \mathbb{R}^2$, A を点 $(0, 0)$, $(1, 0)$, $(0, 1)$ を結ぶ三角形，B を点 $(0, 0)$, $(1, 0)$, $(1, 1)$ を結ぶ三角形とする．このとき，$\sigma[\{A, B\}]$ に属する集合で $\emptyset$ 以外のものを全て図示せよ．

問題 3.2 かき氷屋 A が 1 日に売るかき氷の氷の量を X（10kg 当たり）とする．過去の傾向から X の分布をモデル化し，仕入れの量を再検討したい．そこで，X の確率密度関数 ρ が

$$\rho(x) = \begin{cases} \dfrac{a}{(1 + bx)^2}, & (x > 0), \\ 0, & (x \le 0) \end{cases}$$

により表されると仮定する．ただし，a, b は正定数である．過去の実績から，おおよそ $\mathbb{P}(X < 3) = 0.9$ である．これを満たすような a, b の値を求めよ．

有限試行の確率と同様に，一般の確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ においても，条件付き確率を

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

と定義する．ただし，$A, B \in \mathcal{F}$ で $\mathbb{P}(A) > 0$.

問題 3.3 確率変数 X はパラメータ $\lambda > 0$ の指数分布に従うとする．以下の問いに答えよ．

(1) $\lim_{\Delta x \to 0} \dfrac{\mathbb{P}(x < X \le x + \Delta x \mid X > x)}{\Delta x} = \lambda$ を示せ．

(2) 無記憶性: $\mathbb{P}(X > x + y \mid X > x) = \mathbb{P}(X > y)$, $x, y > 0$, を示せ．

問題 3.4 確率空間 $([0, 1], \mathcal{B}[0, 1], \mathrm{Leb})$ を考える．ただし，Leb は $[0, 1]$ 上のルベーグ測度である．このとき，

$$\mathrm{Leb}(\mathbb{Q}) = 0$$

を示せ．

問題 3.5 A さんが加入している損害保険は，損害額（保険請求）に対する保険金の上限が 2000 万円，自己負担分（免責分）が 10 万円である．推定される損害額から自己負担分を引いた額が保険金として支払われる．損害額が自己負担分以下なら支払いなし，上限以上なら上限から自己負担分を引いた額が支払いとなる．例えば,

- 損害額が 5 万円なら支払いゼロ,
- 損害額が 120 万円なら 110 万円の支払い,
- 損害額が 2005 万円なら 1990 万円の支払い

となる．このとき次の問いに答えよ.

(1) 非負確率変数 X により損害額を表すことにする．支払われる保険金 Y を X の関数として表せ.
(2) Y の分布関数を求め，グラフの概形を描け.

問題 3.6 ある交差点では，週末に発生する交通事故の件数はポアソン分布に従うと考えられている．過去 5 週の週末に記録された事故件数は

$$1,\ 0,\ 2,\ 1,\ 3$$

であった.

(1) このデータから 1 週末あたりの事故件数の平均を求め，それによりポアソン分布のパラメータを推定せよ.
(2) 今週末に 2 件以上の事故が起こる確率を求めよ.

問題 3.7 (**再生性**) X, Y は実確率変数で，互いに独立とする.

(1) X と Y が共にポアソン分布に従うとき，$X+Y$ もポアソン分布に従うことを示せ.
(2) X と Y が共に標準正規分布に従うとき，$X+Y$ は平均 0，分散 2 の正規分布に従うことを示せ.

問題 3.8 ギャンブルを繰り返し行う．確率変数列 $\{X_n\}_{n=1}^{\infty}$ を

$$X_n = \begin{cases} 1 & (n\text{回目が勝ち}), \\ 0 & (n\text{回目が負け}) \end{cases}$$

と定義する．各回の勝敗は独立に決まるとし，各 n に対し $\mathbb{P}(X_n = 1) = p \in (0,1)$ と仮定する.

(1) 初めて勝つまでに要する賭け数

$$T_1 := \inf\{n \geq 1 \mid X_n = 1\}$$

の期待値を求めよ.

(2) T_2 を 2 勝目までに要する賭け数とする．すなわち,

$$T_2 = \inf\{n > T_1 \mid X_n = 1\}$$

とする．このとき，確率変数 $T_2 - T_1$ はパラメータ p の幾何分布に従うことを示せ.

(3) 2 回勝つまでに必要な平均賭け数を求めよ．

問題 3.9 (Buffon **の針**) ある用紙の上に距離 d の間隔で平行な直線が規則的に書かれている．長さ ℓ の針を用紙の上にランダムに落とす．このとき，針がいずれかの直線と交わる確率を求めよう．

そのために，まず $\ell < d$ を仮定する（同時に 2 本の平行線と交わらないようにするため）．X を落とした針の中点から最も近い平行線までの垂直距離，Θ を針の軸と平行線がなす鋭角とする．次の問いに答えよ．

(1) 落とした針がある平行線と交わるための必要十分条件が

$$X \leq \frac{\ell}{2}\sin\Theta$$

であることを説明せよ．

(2) X は $[0, d/2]$ 上の一様分布，Θ は $[0, \pi/2]$ 上の一様分布にそれぞれ従うとし，X と Θ は独立と仮定する．この仮定の下で，落とした針がいずれかの平行線と交わる確率を求めよ．

問題 3.10 X, Y は $\mathbb{N} \cup \{0\}$-値確率変数で，独立同分布とする．各 $n \geq 0$ に対して $p_n = \mathbb{P}(X = n) = \mathbb{P}(Y = n)$ とおき，さらに $r_n = \sum_{k=n}^{\infty} p_k$ とおく．このとき，系 3.5.15 を使って

$$\mathbb{E}[\min(X, Y)] = \sum_{n=1}^{\infty} r_n^2$$

を確かめよ．

問題 3.11 (**複合分布**) $\{X_n\}_{n=1}^{\infty}$ は IID で，X_1 がパラメータ $\lambda > 0$ の指数分布に従うとする．確率変数 N はパラメータ c のポアソン分布に従い，$\{X_n\}$ と独立とする（すなわち，$N, X_1, X_2, \ldots$ は独立）．このとき，$Z := \sum_{n=0}^{N} X_n$ の期待値と分散を求めよ．ただし，$\sum_{n=1}^{0} = 0$ とする．

問題 3.12 実数列 $\{a_i\}_{i=1}^{n}$, $\{p_i\}_{i=1}^{n}$ は $a_i, p_i > 0$, $\sum_{i=1}^{n} p_i = 1$ を満たすとする．このとき，イェンセンの不等式を用いて，不等式

$$a_1 p_1 + \cdots + a_n p_n \geq a_1^{p_1} \cdots a_n^{p_n}$$

を証明せよ（ヒント: $\mathbb{P}(X = \log a_i) = p_i$ を満たす確率変数 X と $f(x) = e^x$ を考える）．

問題 3.13 確率変数 X の分散が 0 ならば，X はほとんど確実に定数であることを示せ（ヒント: 命題 3.5.12 を用いる）．

問題 3.14 (**モンティ・ホール問題**) あなたは次のようなゲームに参加する: 箱 $1, \ldots, N$ の一つに 100 万円，残りはたわし．箱を一つ選ぶ．司会者はどれが当たりか知っている．あなたは箱 i を選ぶとする．この時点で当たりの確率は $1/N$．司会者は箱 i の中身を見せる前に箱 i 以外の箱（例えば箱 j）を開ける．そこにはたわしが入っている（司会者は当たりの箱は開けない）．このとき，あなたは次の 2 つの選択肢のうち，一方を選べる．

(a) あくまで最初に選んだ箱 i を最終選択とする．

(b) 改めて i, j 以外から箱 k を選び，箱 k を最終選択とする．

両者で当たりの確率は異なるか？
（ヒント：箱 1 が実際に当たりとする．A_i をあなたが箱 i を選ぶ事象，B_j を司会者が箱 $j \in \{2,\ldots,N\} \setminus \{i\}$ を選ぶ事象とすると，授業中に示した通り，$\mathbb{P}(A_i) = 1/N$ で，$i = 1$ のとき $\mathbb{P}(B_j \mid A_i) = 1/(N-1)$, $i \neq 1$ のとき $\mathbb{P}(B_j \mid A_i) = 1/(N-2)$ である．C_1 を改めて i, j 以外から選んだ箱が当たりである事象として，$\mathbb{P}(C_1 \,|\, B_j)$ と $\mathbb{P}(A_1 \,|\, B_j)$ を比較すればよい）

問題 3.15 **(包除公式)** $A_1, \ldots, A_n$ を事象とし，$A = \bigcup_{i=1}^n A_i$ とおく．

(1) $1_A = 1 - \prod_{i=1}^n (1 - 1_{A_i})$ を示せ.
(2) (1) の右辺を展開し，両辺期待値をとることにより

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1}\mathbb{P}\left(\bigcap_{i=1}^n A_i\right)$$

を示せ.

問題 3.16 確率変数 X, Y は同時密度関数 $\rho_{X,Y}(x,y)$ を持つとする．このとき，

$$\rho_{X|Y}(x \,|\, y) = \begin{cases} \dfrac{\rho_{X,Y}(x,y)}{\rho_Y(y)} & (\rho_Y(y) \neq 0), \\ 0 & (\rho_Y(y) = 0) \end{cases}$$

と定め，$\rho_{X|Y}(x \,|\, y)$ を $Y = y$ の下での X の**条件付き密度**と呼ぶ．ここで，ρ_Y は Y の密度関数を表す．以下の問いにおいて，現れる積分や期待値は全て存在すると仮定する.

(1) $m(y) := \int_{-\infty}^{\infty} x\rho_{X|Y}(x \,|\, y)dx$ とおくとき，

$$\mathbb{E}[X, \{a < Y \leq b\}] = \mathbb{E}[m(Y), \{a < Y \leq b\}]$$

を示せ.

(2) 確率密度に関するベイズの公式

$$\rho_{Y|X}(y \,|\, x) = \frac{\rho_{X|Y}(x \,|\, y)\rho_Y(y)}{\rho_X(x)}$$

を示せ．ここで，ρ_X は X の密度関数.

(3) $\mathbb{E}[f(X,Y) \,|\, Y] := \int_{-\infty}^{\infty} f(x,Y)\rho_{X|Y}(x \,|\, Y)dx$ とおく．このとき，次の 2 性質

$$\mathbb{E}[Xg(Y) \,|\, Y] = g(Y)\mathbb{E}[X \,|\, Y],$$
$$\mathbb{E}[\mathbb{E}[f(X,Y) \,|\, Y]] = \mathbb{E}[f(X,Y)]$$

を示せ.

(4)

$$\min_g \mathbb{E}[|X - g(Y)|^2] = \mathbb{E}[|X - \mathbb{E}[X \mid Y]|^2]$$

を示せ．ただし，最小化は g はボレル関数で $g(Y)$ が二乗可積分となるもの全体で考える．
（ヒント：$|X - g(Y)|^2 = |X - g(Y) + \mathbb{E}[X \,|\, Y] - \mathbb{E}[X \,|\, Y]|^2 = (X - \mathbb{E}[X \,|\, Y])^2 + 2(X - \mathbb{E}[X \,|\, Y])(\mathbb{E}[X \,|\, Y] - g(Y)) + (\mathbb{E}[X \,|\, Y] - g(Y))^2$ と展開し，期待値を計算してみよ.）

第 4 章

大数の法則

第 3 章のはじめに，ベルヌーイの大数の法則（定理 2.6.1）は ω ごとに成立しそうだと述べた．この章ではいよいよその法則を示す．そのために必要な枠組みは第 3 章で全て整えた．以下，$(\Omega, \mathcal{F}, \mathbb{P})$ を一般の確率空間として，現れる確率変数は全てこの確率空間上で定義されているとする．

4.1 ボレル・カンテリの補題

ボレル・カンテリの補題は概収束を証明するための強力な道具である．結果を述べる前に集合列の上極限の定義を確認しておこう．$\{A_n\} \subset \mathcal{F}$ に対し，

$$\begin{aligned}\limsup_{n\to\infty} A_n &:= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \\ &= \{\omega \in \Omega \mid \omega \in A_n \text{となる } n \text{ が無限個ある}\}.\end{aligned}$$

- $\limsup_{n\to\infty} A_n = \{A_n,\ \text{i.o.}\}$ (infinitely often の略) と書く．

定理 4.1.1 (ボレル・カンテリの第一補題) $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$ とする．このとき，

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0.$$

証明．確率の連続性と劣加法性（命題 3.3.9 (2), 命題 3.3.8 (5)）より，

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0.$$

□

- 定理 4.1.1 の逆は成立しない．反例: $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}[0,1], \text{Leb})$, $A_n = (0, 1/n]$.

例題 4.1.2 $\{X_n\}_{n=1}^{\infty}$ は独立で，$\mathbb{P}(X_n = 1) = p_n = 1 - \mathbb{P}(X_n = 0)$ とする．このとき，$\sum_{n=1}^{\infty} p_n < \infty$ ならば，$X_n \to 0$, a.s. であることを証明せよ．

解答．$A_n = \{X_n = 1\}$ とおき，ボレル・カンテリの第一補題を適用すると，$\mathbb{P}(A_n,\ \text{i.o.}) = 0$. すなわち, $B := (\limsup_{n\to\infty} A_n)^c = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c$ は $\mathbb{P}(B) = 1$ を満たす. $\omega \in B$ ならば, ある $n = n(\omega)$ が存在し，$k \geq n$ に対し，$X_k(\omega) = 0$. よって，$\lim_{n\to\infty} X_n(\omega) = 0$. ゆえに $\mathbb{P}(X_n \to 0) = 1$.

定理 4.1.3 (**ボレル・カンテリの第二補題**) $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$ を独立事象の列とし，$\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \infty$ を仮定する．このとき,

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

証明. $\mathbb{P}$ の連続性と $\{A_n\}$ の独立性，および不等式 $1 - x \leq e^{-x}$ $(x \geq 0)$ を使うと,

$$\begin{aligned}
\mathbb{P}\left(\left(\limsup_{n\to\infty} A_n\right)^c\right) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n\to\infty}\lim_{m\to\infty}\mathbb{P}\left(\bigcap_{k=n}^{m} A_k^c\right)\\
&= \lim_{n\to\infty}\lim_{m\to\infty}\prod_{k=n}^{m}\mathbb{P}(A_k^c) = \lim_{n\to\infty}\lim_{m\to\infty}\prod_{k=n}^{m}(1 - \mathbb{P}(A_k))\\
&\leq \lim_{n\to\infty}\lim_{m\to\infty}\exp\left(-\sum_{k=n}^{m}\mathbb{P}(A_k)\right) = 0.
\end{aligned}$$

□

例題 4.1.4 100 個のキーを持つ PC のキーボードを考える．あるサルが一様にランダムにキータイプを繰り返すとする．このとき，14 文字 `W. Shakespeare`（これらは 100 個のキーに含まれるとする）がほとんど確実に無限回現れることを示せ.

解答．X_n をサルが n 回目にタイプした文字とし,

$$A_n = \{X_{14(n-1)+1}\cdots X_{14(n-1)+14} = \texttt{W. Shakespeare}\}$$

とおく．このとき，$\{A_n\}$ は独立で，$\mathbb{P}(A_n) = 100^{-14}$. B を 14 文字が無限回現れる事象とすると，$\{A_n,\ \text{i.o.}\} \subset B$ である．よってボレル・カンテリの第二補題より，$\mathbb{P}(B) \geq \mathbb{P}(A_n,\ \text{i.o.}) = 1$.

4.2　大数の弱法則

- 確率変数列 $\{X_n\}$ が確率変数 X に**確率収束**するとは，任意の $\varepsilon > 0$ に対して

$$\lim_{n\to\infty}\mathbb{P}(|X_n - X| > \varepsilon) = 0$$

  が成り立つときにいう.
- X_n が X に確率収束することを $X_n \xrightarrow{\text{P}} X$ と書く.

定理 4.2.1 $X, X_1, X_2, \ldots$ を確率変数とする．このとき，以下が成り立つ.

(1) $X_n \to X$, a.s. $\Longrightarrow X_n \xrightarrow{\text{P}} X$.
(2) $X_n \xrightarrow{\text{P}} X \Longrightarrow X_{n_k} \to X$, a.s. となる部分列 $\{n_k\}$ が存在する.

証明. (1). $X_n \to X$, a.s. とする．このとき，

$$
\begin{aligned}
0 = \mathbb{P}(X_n\text{が}X\text{に収束しない}) &= \mathbb{P}\left(\bigcup_{p=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\left\{|X - X_k| > \frac{1}{p}\right\}\right) \\
&= \lim_{p\to\infty}\mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\left\{|X - X_k| > \frac{1}{p}\right\}\right).
\end{aligned}
$$

任意の $\varepsilon > 0$ に対して，十分大きな $p \geq 1$ で $\varepsilon > 1/p$ なるものがとれるので，

$$
\begin{aligned}
\lim_{p\to\infty}\mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\left\{|X - X_k| > \frac{1}{p}\right\}\right) &\geq \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X - X_k| > \varepsilon\}\right) \\
&= \lim_{n\to\infty}\mathbb{P}\left(\bigcup_{k=n}^{\infty}\{|X - X_k| > \varepsilon\}\right) \\
&\geq \limsup_{n\to\infty}\mathbb{P}(|X - X_n| > \varepsilon).
\end{aligned}
$$

従って，

$$
\lim_{n\to\infty}\mathbb{P}(|X - X_n| > \varepsilon) = 0, \quad \varepsilon > 0.
$$

すなわち $\{X_n\}$ は X に確率収束する．

(2). $X_n \xrightarrow{\mathrm{P}} X$ とする．このとき，部分列 $n_1 < n_2 < \cdots$ を

$$
\mathbb{P}\left(|X_{n_k} - X| > \frac{1}{2^k}\right) < \frac{1}{2^k}
$$

を満たすようにとれる．$A_k = \{|X_{n_k} - X| > 2^{-k}\}$ とおくと，$\sum_{k=1}^{\infty}\mathbb{P}(A_k) < \infty$. よって，ボレル・カンテリの第一補題より，$\mathbb{P}(A_n,\ \text{i.o.}) = 0$ である．$\omega \in (\limsup_{n\to\infty} A_n)^c = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c$ ならば，ある $n = n(\omega)$ が存在し，任意の $k \geq n$ に対して $|X(\omega) - X_{n_k}(\omega)| \leq 2^{-k}$. よって $\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega)$. ゆえに $X_n \to X$, a.s. □

定理 4.2.1 (1) の逆は一般に成立しない．

例 4.2.2 $\{X_n\}_{n=1}^{\infty}$ は 0 か 1 に値をとる独立確率変数列で，$p_n = \mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0)$ とおくとき，

$$
\lim_{n\to\infty} p_n = 0, \quad \sum_{n=1}^{\infty} p_n = \infty
$$

を満たしていると仮定する. このとき, 任意の $\varepsilon > 0$ に対して $\mathbb{P}(X_n > \varepsilon) \leq p_n \to 0$. よって $X_n \xrightarrow{\mathrm{P}} 0$. 一方，ボレル・カンテリの第二補題より，$\mathbb{P}(\text{無限個の}n\text{で}X_n = 1) = 1$. よって $\mathbb{P}(X_n \to 0) = 0$.

ベルヌーイの大数の法則（定理 2.6.1）の一般化が次の**大数の弱法則**である．

定理 4.2.3 (**大数の弱法則**) $X_1, X_2, \ldots$ は無相関で，$m = \mathbb{E}[X_n]$, $\sup_{n\geq 1}\mathbb{V}[X_n] \leq C < \infty$ とする．すなわち，期待値は一定で分散は有界とする．このとき，

$$
\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\mathrm{P}} m.
$$

証明．$S_n = \sum_{i=1}^n X_i$ とおく．このとき，

$$\begin{aligned}\mathbb{E}\left|\frac{S_n}{n} - m\right|^2 &= \mathbb{E}\left|\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right|^2 = \frac{1}{n^2}\sum_{i,j=1}^n \mathrm{Cov}(X_i, X_j) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}[X_i] \\ &\leq \frac{C}{n}.\end{aligned}$$

チェビシェフの不等式（定理 3.5.23）より，任意の $\varepsilon > 0$ に対し，

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left|\frac{S_n}{n} - m\right|^2 \to 0, \quad n \to \infty.$$

ゆえに $S_n/n \xrightarrow{\mathrm{P}} m$. □

- 定理 4.2.3 の証明で示しているように，$\lim_{n\to\infty} \mathbb{E}|S_n/n - m|^2 = 0$ が成り立っている．一般に，$\lim_{n\to\infty} \mathbb{E}|X_n - X|^2 = 0$ が成り立つとき，$\{X_n\}$ は X に L^2 **収束**するという．

例題 4.2.4 (Bernstein の多項式近似) $[0, 1]$ 上の任意の連続関数 f は多項式

$$f_n(x) := \sum_{k=0}^n f\left(\frac{k}{n}\right)\frac{n!}{k!(n-k)!}x^k(1-x)^{n-k}, \quad x \in [0, 1]$$

により一様に近似できること，すなわち，

$$\sup_{0\leq x\leq 1} |f(x) - f_n(x)| \to 0, \quad n \to \infty, \tag{4.2.1}$$

が成り立つことを示せ．

解答．$f_n(0) = f(0)$, $f_n(1) = f(1)$ なので，$0 < x < 1$ の場合を考える．$\{X_n\}$ を $\{0, 1\}$-値の IID で $\mathbb{P}(X_n = 1) = x$ とする．このとき，$\mathbb{E}[X_n] = x$ である．$S_n = X_1 + \cdots + X_n$ とおくと，大数の弱法則より，$S_n/n \xrightarrow{\mathrm{P}} x$．より強く，定理 4.2.3 の証明中に示したように，

$$\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| > c\right) \leq \frac{\mathbb{V}[X_1]}{nc^2} \tag{4.2.2}$$

が任意の $c > 0$, $n \in \mathbb{N}$ に対して成り立つ．

他方，S_n は二項分布に従うから

$$\mathbb{P}(S_n = k) = \frac{n!}{k!(n-k)!}x^k(1-x)^{n-k}, \quad k = 0, 1, \ldots, n,$$

であり，これより，

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^n f\left(\frac{k}{n}\right)\mathbb{P}(S_n = k) = \sum_{k=0}^n f\left(\frac{k}{n}\right)\frac{n!}{k!(n-k)!}x^k(1-x)^{n-k} = f_n(x).$$

f は $[0, 1]$ 上一様連続なので，$\varepsilon > 0$ に対し，$\delta > 0$ が存在し，$|y - y'| \leq \delta$ ならば $|f(y) - f(y')| \leq \varepsilon$.

よって，$M = \max_{0\le y\le 1}|f(y)|$ とおけば，(4.2.2) より，

$$
\begin{aligned}
&|f(x) - f_n(x)| \\
&= \left| f(x) - \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]\right| \le \mathbb{E}\left| f(x) - f\left(\frac{S_n}{n}\right)\right| \\
&= \mathbb{E}\left[\left| f(x) - f\left(\frac{S_n}{n}\right)\right|, \left\{\left|\frac{S_n}{n} - x\right| > \delta\right\}\right] + \mathbb{E}\left[\left| f(x) - f\left(\frac{S_n}{n}\right)\right|, \left\{\left|\frac{S_n}{n} - x\right| \le \delta\right\}\right] \\
&\le 2M\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| > \delta\right) + \varepsilon\mathbb{P}\left(\left|\frac{S_n}{n} - x\right| \le \delta\right) \\
&\le \frac{2M}{n\delta^2}\mathbb{V}[X_1] + \varepsilon.
\end{aligned}
$$

$\mathbb{V}[X_1] = x(1-x) \le 1$ より，

$$
\sup_{0\le x\le 1} |f(x) - f_n(x)| \le \frac{2M}{n\delta^2} + \varepsilon
$$

を得る．上式で $n \to \infty$ としてから $\varepsilon \to 0$ とすることで (4.2.1) が従う．

例題 4.2.5 (高次元超立方体はほとんど球の境界) $\varepsilon \in (0,1)$ を固定し，

$$
A_n = \left\{ x \in \mathbb{R}^n ; \frac{n}{3}(1-\varepsilon) \le |x|^2 \le \frac{n}{3}(1+\varepsilon)\right\}
$$

を考える．このとき，大数の弱法則を用いて

$$
\lim_{n\to\infty} \frac{1}{2^n}\,|A_n \cap (-1,1)^n| = 1
$$

を示せ．ただし，$|B|$ は B の体積（ルベーグ測度）である．

- この結果は，n が十分大きいとき，超立方体 $(-1,1)^n$ の体積のほとんどが半径 $\sqrt{n/3}$ の球の境界付近 A_n の体積で占められていることを意味する．

解答．$\{X_n\}$ を IID で，各 X_n が $(-1,1)$ 上の一様分布に従うとする．このとき，

$$
\mathbb{E}[X_n^2] = \frac{1}{2}\int_{-1}^{1} x^2 dx = \frac{1}{3}, \quad \mathbb{E}[|X_n|^4] = \frac{1}{2}\int_{-1}^{1} x^4 dx = \frac{1}{5} < \infty.
$$

$\{X_n^2\}$ も IID なので，大数の弱法則より，

$$
\mathbb{P}\left(\left|\frac{1}{n}(X_1^2 + \cdots + X_n^2) - \frac{1}{3}\right| > \varepsilon\right) \to 0.
$$

ゆえに，

$$
\lim_{n\to\infty} \frac{1}{2^n}|A_n \cap (-1,1)^n| = \lim_{n\to\infty} \mathbb{P}((X_1,\ldots,X_n) \in A_n \cap (-1,1)^n) = 1.
$$

4.3　大数の強法則

大数の弱法則は独立同分布列の標本平均が期待値に確率収束することであった．次の**大数の強法則**は概収束を保証するものである．これにより，n 回のコイン投げで表が出た割合は ω ごとに（確率 1 で）1/2 に収束することが示される．

定理 4.3.1 $\{X_n\}_{n=1}^{\infty}$ は IID で，各 X_n は平均可能とする．このとき，

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i=\mathbb{E}[X_1],\quad \text{a.s.}$$

証明．$M:=\mathbb{E}[|X_1|^4]<\infty$ を仮定して証明する．一般の場合の証明については [8] や [10] などをみよ．

一般性を失うことなく $\mathbb{E}[X_1]=0$ を仮定できる．系 3.5.20 より，

$$\left(\mathbb{E}|X_1|^2\right)^{1/2}\le\left(\mathbb{E}|X_1|^3\right)^{1/3}\le\left(\mathbb{E}|X_1|^4\right)^{1/4}=M^{1/4}.$$

これより各 X_n の 2 次モーメント $\mathbb{E}[X_n^2]$, 3 次モーメント $\mathbb{E}[X_n^3]$ が存在することに注意せよ．$S_n=X_1+\cdots+X_n$ とおく．このとき，多項定理より

$$S_n^4=\sum_i X_i^4+4\sum_{\substack{i,j\\ i\ne j}}X_i^3X_j+6\sum_{i<j}X_i^2X_j^2+12\sum_{\substack{i,j,k\\ \text{相異なる}}}X_i^2X_jX_k+24\sum_{i<j<k<\ell}X_iX_jX_kX_\ell$$

と表される．定理 3.5.17 と $\mathbb{E}[X_i]=0$ より，相異なる i,j,k,ℓ に対して $\mathbb{E}[X_i^3X_j]=\mathbb{E}[X_i^2X_jX_k]=\mathbb{E}[X_iX_jX_kX_\ell]=0$. 従って，

$$\mathbb{E}[S_n^4]=\sum_i\mathbb{E}[X_i^4]+6\sum_{i<j}\mathbb{E}[X_i^2]\mathbb{E}[X_j^2]=nM+3n(n-1)\left(\mathbb{E}[X_1^2]\right)^2\le 3Mn^2.$$

よって，マルコフの不等式を適用し，

$$\mathbb{P}\left(\frac{1}{n}|S_n|\ge\frac{1}{n^{1/8}}\right)=\mathbb{P}(|S_n|^4\ge n^{7/2})\le n^{-7/2}\mathbb{E}|S_n|^4\le 3Mn^{-3/2}.$$

よって $\sum_{n=1}^{\infty}\mathbb{P}(|S_n/n|\ge n^{-1/8})<\infty$ であるからボレル・カンテリの第一補題より $\mathbb{P}(\Omega_0)=1$. ただし $\Omega_0=\bigcup_{N\ge1}\bigcap_{n\ge N}\{|S_n/n|<n^{-1/8}\}$. これより Ω_0 上では $\lim_{n\to\infty}S_n/n=0$. ゆえに定理の主張が従う． □

例題 4.3.2 $X_1,X_2,\ldots,X_n$ を無作為抽出された標本で，共通の（未知の）分布関数 F を持つと仮定する．このとき，**経験分布関数** $F_n(x):=(1/n)\sum_{i=1}^{n}1_{\{X_i\le x\}}$, $x\in\mathbb{R}$, は

$$\lim_{n\to\infty}F_n(x)=F(x),\quad x\in\mathbb{R},\tag{4.3.1}$$

を満たす．これを確認せよ．

解答．任意の $x\in\mathbb{R}$ に対し，確率変数列 $1_{\{X_n\le x\}}$, $n\ge1$, は IID で，共通の期待値は $F(x)=\mathbb{E}[1_{\{X_n\le x\}}]$ により与えられる．よって定理 4.3.1 より (4.3.1) が従う．

章末問題

問題 4.1 確率変数列 $\{X_n\}_{n=0}^{\infty}$ を次のように帰納的に定義する．$X_0=1$ とし，X_{n+1} は区間 $[-|X_n|,|X_n|]$ からランダムに選ばれる．すなわち，$X_{n+1}/|X_n|$ は $[-1,1]$ 上の一様分布に従い，

$X_1, \dots, X_n$ と独立である．このとき，$n^{-1} \log |X_n|$ がある定数 c に概収束することを示せ．さらに，この定数 c を求めよ．

問題 4.2 $X_1, \dots, X_n$ は確率変数 X からサンプリングされたものとする（i.e., $\{X_j\}_{j=1}^n$ は IID で，$X_j \overset{\mathrm{d}}{=} X$）．このとき，

$$\frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X})^2$$

を**不偏分散**という．ここで，$\overline{X} = n^{-1}\sum_{j=1}^n X_j$. 条件 $\mathbb{E}|X|^2 < \infty$ の下，次が成り立つことを示せ．

(1) $\mathbb{E}\left[\dfrac{1}{n-1}\displaystyle\sum_{j=1}^n (X_j - \overline{X})^2\right] = \mathbb{V}[X]$.

(2) $\displaystyle\lim_{n\to\infty} \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X})^2 = \mathbb{V}[X]$, a.s.

第 5 章

中心極限定理

5.1 特性関数

i を虚数単位とする（$i^2 = -1$）．実確率変数 X に対し

$$\varphi(t) := \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)], \quad t \in \mathbb{R},$$

を X の**特性関数**という．

- 複素数値確率変数 Z に対し，$\mathbb{E}[Z] := \mathbb{E}[\mathrm{Re}(Z)] + i\mathbb{E}[\mathrm{Im}(Z)]$.
- $\mathrm{Re}(a+ib) = a$, $\mathrm{Im}(a+ib) = b$.
- 複素共役: $\overline{a+ib} = a - ib$.
- $|a+ib| = ((a+ib)\overline{(a+ib)})^{1/2} = (a^2+b^2)^{1/2}$.

これからみていくように，特性関数には以下の特徴がある．

- 分布の情報を全て含む．
- 独立性の確認に役立つ．
- 確率変数の分布が一致するか，収束するかどうか確認するときに使う．

定理 5.1.1 $\varphi(t)$ を X の特性関数とするとき，以下が成立する．

(1) $\varphi(0) = 1$.
(2) $\varphi(-t) = \overline{\varphi(t)}$.
(3) $|\varphi(t)| \le 1$.
(4) $\varphi(t)$ は $\mathbb{R}$ 上の一様連続関数である．

証明．(1) は明らか．

(2)．$t \in \mathbb{R}$ に対し，

$$\begin{aligned}\varphi(-t) &= \mathbb{E}[e^{-itX}] = \mathbb{E}[\cos(tX)] - i\mathbb{E}[\sin(tX)] \\ &= \overline{\mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)]} = \overline{\varphi(t)}.\end{aligned}$$

(3). コーシー・シュワルツの不等式より，$t \in \mathbb{R}$ に対し，

$$|\varphi(t)|^2 = \mathbb{E}[\cos(tX)]^2 + \mathbb{E}[\sin(tX)]^2 \le \mathbb{E}[\cos^2(tX)] + \mathbb{E}[\sin^2(tX)] = 1.$$

(4). 任意の $s, t \in \mathbb{R}$ に対し，

$$\begin{aligned}|\varphi(t) - \varphi(s)| &= |\mathbb{E}[e^{itX} - e^{isX}]| = |\mathbb{E}[\cos(tX) - \cos(sX)] + i\mathbb{E}[\sin(tX) - \sin(sX)]| \\ &\le \mathbb{E}[(\cos(tX) - \cos(sX))^2 + (\sin(tX) - \sin(sX))^2]^{1/2} \\ &= \mathbb{E}[2 - 2\cos(tX)\cos(sX) - 2\sin(tX)\sin(sX)]^{1/2} \\ &= \mathbb{E}[2 - 2\cos((t-s)X)]^{1/2}.\end{aligned}$$

$c_n \to 0$ なる任意の実数列 $\{c_n\}$ に対して，$|\cos(c_nX)| \le 1$ だからルベーグの収束定理を適用し，

$$\lim_{n\to\infty} \mathbb{E}[2 - 2\cos(c_nX)] = \mathbb{E}\left[2 - 2\lim_{n\to\infty} \cos(c_nX)\right] = 0.$$

従って，$\lim_{h\to 0} \mathbb{E}[2 - 2\cos(hX)] = 0$. ゆえに，任意の $\varepsilon > 0$ に対し，$\delta > 0$ が存在して，

$$\mathbb{E}[2 - 2\cos(uX)] \le \varepsilon^2, \quad |u| \le \delta.$$

これより，$|t - s| \le \delta$ ならば，$|\varphi(t) - \varphi(s)| \le \varepsilon$. これは φ の一様連続性を意味している． □

例 5.1.2 X がパラメータ c のポアソン分布に従うとき，

$$\varphi(t) = \exp(c(e^{it} - 1)), \quad t \in \mathbb{R}.$$

実際，定義通りに計算すると，

$$\varphi(t) = \sum_{n=0}^{\infty} e^{itn} \frac{e^{-c}c^n}{n!} = e^{-c} \sum_{n=0}^{\infty} \frac{(e^{it}c)^n}{n!} = \exp(c(e^{it} - 1)).$$

例 5.1.3 X がパラメータ λ の指数分布に従うとき，

$$\varphi(t) = \frac{\lambda}{\lambda - it}.$$

例 5.1.4 $X \sim N(0,1)$ のとき，

$$\varphi(t) = e^{-t^2/2}, \quad t \in \mathbb{R}. \tag{5.1.1}$$

以下，これを確かめよう．sin 関数は奇関数なので，

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} \cos(tx) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

$(\partial/\partial t)\cos(tx) = -x\sin(tx)$ であり，$\mathbb{E}[\sup_{t\in\mathbb{R}} |-X\sin(tX)|] \le \mathbb{E}|X| < \infty$ であるから，命題 3.6.5 より微分と期待値の交換が可能で，

$$\varphi'(t) = \int_{-\infty}^{\infty} (-x\sin(tx)) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

が成り立つ．部分積分により，

$$\varphi'(t) = -\int_{-\infty}^{\infty} t\cos(tx) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = -t\varphi(t).$$

この微分方程式を解くことで (5.1.1) を得る．実際,

$$\frac{d}{dt}\left(\varphi(t)e^{t^2/2}\right)=\varphi'(t)e^{t^2/2}+\varphi(t)te^{t^2/2}=0$$

より $\varphi(t)e^{t^2/2}=\varphi(0)=1$.

例 5.1.5 $X\sim N(m,v)$ のとき,

$$\varphi(t)=e^{imt-vt^2/2},\quad t\in\mathbb{R}. \tag{5.1.2}$$

これを確かめるために，$Z\sim N(0,1)$ ならば $X\sim m+\sqrt{v}Z$ であることを思い出そう．(5.1.1) を使えば,

$$\varphi(t)=\mathbb{E}[e^{it(m+\sqrt{v}Z)}]=e^{itm}e^{-(t\sqrt{v})^2/2}$$

であるから (5.1.2) を得る．

この節のはじめに特性関数は分布の情報を全て含むと述べた．このことは次の Lévy の**反転公式**により分かる．

定理 5.1.6 (Lévy **の反転公式**) $\mathbb{P}(X=a)=\mathbb{P}(X=b)=0$ のとき,

$$\mathbb{P}(a<X<b)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ibt}-e^{-iat}}{-it}\mathbb{E}[e^{itX}]dt.$$

定理 5.1.6 の証明のために補題を一つ準備する．

補題 5.1.7 任意の $T>0$ に対し,

$$\left|\int_0^T\frac{\sin t}{t}dt-\frac{\pi}{2}\right|\le\frac{2}{T}.$$

証明．まず,

$$\int_0^T\frac{\sin t}{t}dt=\int_0^T\sin t\int_0^\infty e^{-tx}dxdt=\int_0^\infty\int_0^T e^{-tx}\sin tdtdx$$

と表されることに注目し,

$$J_T(x)=\int_0^T e^{-tx}\sin tdt$$

とおく．部分積分により,

$$\begin{aligned}J_T(x)&=\int_0^T e^{-tx}(-\cos t)'dt=[e^{-tx}(-\cos t)]_0^T-\int_0^T(-x)e^{-tx}(-\cos t)dt\\&=1-e^{-Tx}\cos T-x\int_0^T e^{-tx}\cos tdt.\end{aligned}$$

もう一度部分積分し,

$$\int_0^T e^{-tx}\cos tdt=[e^{-tx}\sin t]_0^T+x\int_0^T e^{-tx}\sin tdt=e^{-Tx}\sin T+xJ_T(x).$$

よって，

$$J_T(x) = \frac{1}{1+x^2}\left(1 - e^{-Tx}(x\sin T + \cos T)\right).$$

$x = \tan\theta$ の変換により

$$\int_0^\infty \frac{dx}{1+x^2} = \int_0^{\pi/2} d\theta = \frac{\pi}{2}$$

であるから，$\sup_{x\geq 0}(1+x)/(1+x^2) \leq 2$ より

$$\begin{aligned}\left|\int_0^T \frac{\sin t}{t}dt - \frac{\pi}{2}\right| &= \left|\int_0^\infty \frac{e^{-Tx}}{1+x^2}(x\sin T + \cos T)dx\right| \leq \int_0^\infty e^{-Tx}\frac{x+1}{1+x^2}dx \\ &\leq 2\int_0^\infty e^{-Tx}dx = \frac{2}{T}.\end{aligned}$$

□

定理 5.1.6 の証明．$a < b$ を定理の仮定を満たすものとする．$T > 0$ に対し，

$$I_T := \int_{-T}^T \frac{e^{-ibt} - e^{-iat}}{-it}\mathbb{E}[e^{itX}]dt$$

を考える．被積分関数が有界なので，期待値と積分の順序交換が可能（フビニの定理．[8] や [10] を参照）である．これより

$$\begin{aligned}I_T &= \int_{-T}^T \int_a^b e^{-itx}dx\mathbb{E}[e^{itX}]dt = \int_{-T}^T\int_a^b \mathbb{E}[e^{it(X-x)}]dxdt = \int_{-T}^T \mathbb{E}\left[\int_a^b e^{it(X-x)}dx\right]dt \\ &= \int_{-T}^T\left\{\mathbb{E}\left[\frac{\sin(t(X-a)) - \sin(t(X-b))}{t}\right] + i\mathbb{E}\left[\frac{\cos(t(X-a)) - \cos(t(X-b))}{t}\right]\right\}dt \\ &= \mathbb{E}\left[\int_{-T}^T \frac{\sin(t(X-a)) - \sin(t(X-b))}{t}dt\right] + i\mathbb{E}\left[\int_{-T}^T \frac{\cos(t(X-a)) - \cos(t(X-b))}{t}dt\right] \\ &= \mathbb{E}\left[\int_{-T}^T \frac{\sin(t(X-a)) - \sin(t(X-b))}{t}dt\right].\end{aligned}$$

最後の等式は t に関する積分の被積分関数が奇関数であることから従う．

次に，$\theta \in \mathbb{R}$ に対し

$$R(\theta, T) := \int_{-T}^T \frac{\sin(\theta t)}{t}dt$$

を考える．$\theta > 0$ のときは

$$R(\theta, T) = 2\int_0^{\theta T}\frac{\sin t}{t}dt$$

と書け，これより，一般の $\theta \in \mathbb{R}$ については

$$R(\theta, T) = 2\mathrm{sgn}(\theta)\int_0^{|\theta|T}\frac{\sin t}{t}dt$$

と表される．ここで，

$$\mathrm{sgn}(\theta) = \begin{cases} 1 & (\theta > 0), \\ 0 & (\theta = 0), \\ -1 & (\theta < 0). \end{cases}$$

よって,

$$Y_T = 2\mathrm{sgn}(X-a)\int_0^{T|X-a|}\frac{\sin t}{t}dt - 2\mathrm{sgn}(X-b)\int_0^{T|X-b|}\frac{\sin t}{t}dt$$

とおけば $I_T = \mathbb{E}[Y_T]$. 補題 5.1.7 より,

$$\lim_{T\to\infty} Y_T = \begin{cases} 2\pi & (a < X < b), \\ \pi & (X = a \text{ または } X = b), \\ 0 & (X < a \text{ または } X > b). \end{cases}$$

さらに, $\sup_{x\in\mathbb{R}}|\sin x/x| \le 1$ と再び補題 5.1.7 より,

$$|Y_T| \le 4\sup_{s\ge 0}\left|\int_0^s \frac{\sin t}{t}dt\right| \le 4\int_0^1 \frac{\sin t}{t}dt + 4\sup_{s\ge 1}\left|\int_0^s \frac{\sin t}{t}dt\right| \le 4 + 4\left(\frac{\pi}{2}+2\right).$$

ゆえに, ルベーグの収束定理が適用可能であり, $\mathbb{P}(X=a) = \mathbb{P}(X=b) = 0$ を考慮すれば

$$\lim_{T\to\infty} I_T = \mathbb{E}\left[\lim_{T\to\infty} Y_T\right] = 2\pi\mathbb{P}(a < X < b)$$

を得る. □

X の密度関数が存在する場合, その密度は特性関数の**フーリエ反転公式**により与えられる.

命題 5.1.8 確率変数 X の確率密度関数 ρ が存在すると仮定する. このとき, $\int_{-\infty}^{\infty}|\mathbb{E}[e^{itX}]|dt < \infty$ ならば,

$$\rho(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\mathbb{E}[e^{itX}]dt. \tag{5.1.3}$$

証明. $a < b$ なる $a, b \in \mathbb{R}$ を任意にとる. X の密度関数が存在するので, $\mathbb{P}(X=a) = \mathbb{P}(X=b) = 0$ である. $a < x < b$ に対し,

$$F_T(x) = \int_{-T}^{T} e^{-itx}\mathbb{E}[e^{itX}]dt$$

とおく. このとき, U を (a,b) 上の一様分布に従う確率変数とすれば,

$$\int_{-T}^{T}\int_a^b e^{-itx}dx\mathbb{E}[e^{itX}]dt = \int_a^b F_T(x)dx = (b-a)\mathbb{E}[F_T(U)].$$

仮定より,

$$|F_T(U)| \le \int_{-T}^{T}|\mathbb{E}[e^{itX}]|dt \le \int_{-\infty}^{\infty}|\mathbb{E}[e^{itX}]|dt < \infty.$$

また,

$$\lim_{T\to\infty} F_T(U) = \int_{-\infty}^{\infty} e^{-itU}\mathbb{E}[e^{itX}]dt$$

であるから, ルベーグの収束定理と Lévy の反転公式より,

$$\begin{aligned}\mathbb{P}(a < X < b) &= \lim_{T\to\infty}\frac{b-a}{2\pi}\mathbb{E}[F_T(U)] = \frac{b-a}{2\pi}\mathbb{E}\left[\int_{-\infty}^{\infty} e^{-itU}\mathbb{E}[e^{itX}]dt\right] \\ &= \frac{1}{2\pi}\int_a^b\int_{-\infty}^{\infty} e^{-itx}\mathbb{E}[e^{itX}]dt.\end{aligned}$$

a, b は任意だったから, 上式は密度関数 ρ が (5.1.3) により与えれれることを意味する. □

離散的な確率変数に適用できる反転公式は以下で与えられる．

命題 5.1.9 任意の確率変数 X, 任意の $a \in \mathbb{R}$ に対して,

$$\mathbb{P}(X = a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} e^{-ita}\mathbb{E}[e^{itX}]dt.$$

証明. 任意の $T > 0$ に対して,

$$\begin{aligned}\int_{-T}^{T} e^{-ita}\mathbb{E}[e^{itX}]dt &= \mathbb{E}\left[\int_{-T}^{T} \cos(t(X-a))dt\right] + i\mathbb{E}\left[\int_{-T}^{T} \sin(t(X-a))dt\right] \\ &= \mathbb{E}\left[\int_{0}^{2T} \cos(t(X-a))dt\right] \\ &= \mathbb{E}\left[2T1_{\{X=a\}} + \frac{1}{|X-a|}\sin(2T|X-a|)1_{\{X\neq a\}}\right].\end{aligned}$$

従って，$\sup_{x\in\mathbb{R}} |\sin x/x| \leq 1$ とルベーグの収束定理より，$T \to \infty$ の極限で,

$$\begin{aligned}\frac{1}{2T}\int_{-T}^{T} e^{-ita}\mathbb{E}[e^{itX}]dt &= \mathbb{P}(X=a) + \mathbb{E}\left[\frac{1}{2T|X-a|}\sin(2T|X-a|)1_{\{X\neq a\}}\right] \\ &\to \mathbb{P}(X=a).\end{aligned}$$

□

二つの確率変数が同分布であるための必要十分条件は対応する特性関数が一致することである．

定理 5.1.10 二つの確率変数 X, Y に対して,

$$X \overset{\mathrm{d}}{=} Y \quad \Longleftrightarrow \quad \mathbb{E}[e^{itX}] = \mathbb{E}[e^{itY}], \quad t \in \mathbb{R}.$$

部分的証明. 同分布ならば特性関数が等しいのは明らか．ここでは，X と Y がともに密度関数をもつときに逆が成り立つことを確認する．一般の場合の証明は [8] や [10] など，測度論的確率論のテキストをみよ．この仮定の下で，任意の $x \in \mathbb{R}$ に対して $\mathbb{P}(X = x) = \mathbb{P}(Y = x) = 0$ である．よって定理 5.1.6 より，任意の $a < b$ について $\mathbb{P}(a < X \leq b) = \mathbb{P}(a < Y \leq b)$. あとは定理 3.4.6 を適用すればよい． □

例題 5.1.11 $\{X_n\}_{n=1}^{\infty}$ は IID で，$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ を満たすとする．このとき,

$$X := \sum_{n=1}^{\infty} \frac{X_n}{2^n} \sim U[-1, 1]$$

を示せ.

解答. $t \in \mathbb{R}$ を任意にとる．$U \sim U[-1, 1]$ とすると,

$$\mathbb{E}[e^{itU}] = \frac{1}{2}\int_{-1}^{1} e^{itx}dx = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}.$$

$\lim_{x\to 0} \sin x/x = 1$ であるから，上式は $t = 0$ のときにも成立することに注意せよ.

他方，ルベーグの収束定理と独立性より，

$$
\begin{aligned}
\mathbb{E}[e^{itX}] &= \mathbb{E}\left[\lim_{N\to\infty} e^{it\sum_{n=1}^{N} X_n 2^{-n}}\right] = \lim_{N\to\infty} \mathbb{E}\left[e^{it\sum_{n=1}^{N} X_n 2^{-n}}\right] \\
&= \lim_{N\to\infty} \prod_{n=1}^{N} \mathbb{E}\left[e^{it2^{-n}X_n}\right] = \prod_{n=1}^{\infty} \left(\frac{1}{2}e^{it2^{-n}} + \frac{1}{2}e^{-it2^{-n}}\right) \\
&= \prod_{n=1}^{\infty} \cos\left(\frac{t}{2^n}\right) \\
&= \frac{\sin t}{t}.
\end{aligned}
$$

$t \in \mathbb{R}$ は任意だったから，定理 5.1.10 より $X \overset{\mathrm{d}}{=} U$.

d 次元確率変数 $X = (X_1, \ldots, X_d)$ に対する特性関数は

$$
\varphi(\mathbf{t}) := \mathbb{E}[e^{i\sum_{i=1}^{d} t_i X_i}], \quad \mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{R}^d
$$

により定義される．

Lévy の反転公式の多次元版は以下のようになる．証明は略す．

定理 5.1.12 (**多次元版反転公式**) $\mathbb{P}(X_k = a_k) = \mathbb{P}(X_k = b_k) = 0$ $(k = 1, \ldots, d)$ のとき，

$$
\begin{aligned}
&\mathbb{P}(X \in \prod_{k=1}^{d} (a_k, b_k)) \\
&= \lim_{T\to\infty} \frac{1}{(2\pi)^d} \int_{-T}^{T} \cdots \int_{-T}^{T} \prod_{k=1}^{d} \frac{e^{-ib_k t_k} - e^{-ia_k t_k}}{-it_k} \mathbb{E}[e^{i\sum_{k=1}^{d} t_k X_k}] dt_1 \cdots dt_d.
\end{aligned}
$$

定理 5.1.10 の多次元版は以下のようになる．こちらも証明は略す．

定理 5.1.13　d 次元確率ベクトル $X = (X_1, \ldots, X_d)$, $Y = (Y_1, \ldots, Y_d)$ に対し，

$$
X \overset{\mathrm{d}}{=} Y \iff \mathbb{E}[e^{i\sum_{k=1}^{d} t_k X_k}] = \mathbb{E}[e^{i\sum_{k=1}^{d} t_k Y_k}],\ \ t_1, \ldots, t_d \in \mathbb{R}.
$$

次の結果は特性関数による独立性の特徴付けである．

定理 5.1.14　確率変数 $X_1, \ldots, X_n$ に対して以下が成り立つ．

$$
X_1, \ldots, X_n \text{ が独立} \iff \mathbb{E}[e^{i\sum_{k=1}^{n} t_k X_k}] = \prod_{k=1}^{n} \mathbb{E}[e^{it_k X_k}], \quad t_1, \ldots, t_n \in \mathbb{R}.
$$

証明．$n = 2$ の場合の証明を与える．$X = X_1$, $Y = X_2$ とおく．

$(\Rightarrow)$. 任意の $s, t \in \mathbb{R}$ をとって固定する．このとき，独立性より

$$
\begin{aligned}
&\mathbb{E}[e^{itX+isY}] \\
&= \mathbb{E}[e^{itX}e^{isY}] = \mathbb{E}[(\cos(tX) + i\sin(tX))(\cos(sY) + i\sin(sY))] \\
&= \mathbb{E}[\cos(tX)\cos(sY) - \sin(tX)\sin(sY)] + i\mathbb{E}[\sin(tX)\cos(sY) + \cos(tX)\sin(sY)] \\
&= \mathbb{E}[\cos(tX)]\mathbb{E}[\cos(sY)] - \mathbb{E}[\sin(tX)]\mathbb{E}[\sin(sY)] \\
&\quad + i\mathbb{E}[\sin(tX)]\mathbb{E}[\cos(sY)] + i\mathbb{E}[\cos(tX)]\mathbb{E}[\sin(sY)] \\
&= \mathbb{E}[e^{itX}]\mathbb{E}[e^{isY}].
\end{aligned}
$$

$(\Leftarrow)$. 確率変数 $\tilde{Y}$ を Y と同分布で，(X, Y) とは独立なものとする．すなわち，$X, Y, \tilde{Y}$ は独立とする．上で示したことと，定理の仮定より，任意の $s, t \in \mathbb{R}$ に対し，

$$
\mathbb{E}[e^{itX+is\tilde{Y}}] = \mathbb{E}[e^{itX}]\mathbb{E}[e^{is\tilde{Y}}] = \mathbb{E}[e^{itX}]\mathbb{E}[e^{isY}] = \mathbb{E}[e^{itX+isY}].
$$

よって，定理 5.1.13 より $(X, \tilde{Y}) \overset{\mathrm{d}}{=} (X, Y)$．ゆえに，任意のボレル集合 A, B に対し，

$$
\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A, \tilde{Y} \in B) = \mathbb{P}(X \in A)\mathbb{P}(\tilde{Y} \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).
$$

従って X と Y は独立である． □

d 次元確率変数 $X = (X_1, \ldots, X_d)$ の分布が密度関数

$$
\rho(x) = \frac{1}{\sqrt{(2\pi)^d \det(V)}} \exp\left(-\frac{1}{2}(x-m)^\mathsf{T} V^{-1}(x-m)\right) \tag{5.1.4}
$$

を持つとき，X は**平均ベクトル** m, **分散共分散行列** V **の正規分布**に従うといい，$X \sim N(m, V)$ と書く．ここで，$m \in \mathbb{R}^d$, V は d 次元正定値対称行列，a^T は a の転置である．

(5.1.4) の ρ が確かに密度関数になっていること，すなわち $\mathbb{R}^d$ 上での積分が 1 に等しいことを確かめてみよう．V を対角化し，$V = PDP^{-1}$ と表す．ここで P は直交行列（i.e., $P^{-1} = P^\mathsf{T}$）で，

$$
D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \text{\huge 0} \\ & & \ddots & \\ \text{\huge 0} & & & \lambda_d \end{pmatrix}.
$$

$\lambda_k > 0$ は V の固有値である．$x = m + Py$ と変数変換すれば $|\det P| = 1$ だから

$$
\begin{aligned}
\int_{\mathbb{R}^d} \rho(x)dx &= \frac{1}{\sqrt{(2\pi)^d \det(V)}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} y^\mathsf{T} D^{-1} y\right) dy \\
&= \frac{1}{\sqrt{(2\pi)^d \det(V)}} \prod_{k=1}^{d} \int_{-\infty}^{\infty} e^{-|y_k|^2/(2\lambda_k)} dy_k = \frac{\sqrt{\det(D)}}{\sqrt{\det(V)}} = 1.
\end{aligned}
$$

多次元正規分布の特性関数は次の式で与えられる．

命題 5.1.15 $X \sim N(m, V)$ のとき，

$$
\mathbb{E}[e^{iz^\mathsf{T} X}] = \exp\left(im^\mathsf{T} z - \frac{1}{2} z^\mathsf{T} V z\right), \quad z \in \mathbb{R}^d.
$$

証明．上と同様に V を対角化し，$x = m + Py$ と変数変換し，$m = (m_k)$, $P = (P_{\ell k})$ と表すことにすれば，

$$\begin{aligned}\mathbb{E}[e^{iz^\top X}] &= \frac{1}{\sqrt{(2\pi)^d \det(V)}} \prod_{k=1}^d \int_{-\infty}^\infty \exp\left(i m_k z_k + i \sum_{\ell=1}^d z_\ell P_{\ell k} y_k - |y_k|^2/(2\lambda_k) \right) dy_k \\ &= \frac{1}{\sqrt{(2\pi)^d \det(V)}} \prod_{k=1}^d \exp\left(i m_k z_k - \lambda_k (\sum_{\ell=1}^d z_\ell P_{\ell k})^2/2 \right) \sqrt{2\pi\lambda_k}.\end{aligned}$$

ここで $\sum_{k=1}^d \lambda_k (\sum_{\ell=1}^d z_\ell P_{\ell k})^2/2 = z^\top V z$ だから結局，

$$\mathbb{E}[e^{iz^\top X}] = \frac{\sqrt{\det(D)}}{\sqrt{\det(V)}} \exp\left(i m^\top z - \frac{1}{2} z^\top V z \right) = \exp\left(i m^\top z - \frac{1}{2} z^\top V z \right).$$

□

例題 5.1.16 X と Y をそれぞれ標準正規分布に従う独立確率変数とする．このとき，任意の $0 \le \theta < 2\pi$ に対して，$V := X\cos\theta - Y\sin\theta$ と $W := X\sin\theta + Y\cos\theta$ も標準正規分布に従う独立確率変数であることを示せ．

解答．$s, t \in \mathbb{R}$ を任意にとる．定理 5.1.14 と例 5.1.4 より，

$$\begin{aligned}\mathbb{E}[e^{itV+isW}] &= \mathbb{E}\left[e^{i(t\cos\theta + s\sin\theta)X} e^{i(-t\sin\theta + s\cos\theta)Y} \right] \\ &= \mathbb{E}\left[e^{i(t\cos\theta + s\sin\theta)X} \right] \mathbb{E}\left[e^{i(-t\sin\theta + s\cos\theta)Y} \right] \\ &= e^{-(t\cos\theta + s\sin\theta)^2/2} e^{-(-t\sin\theta + s\cos\theta)^2/2} \\ &= e^{-(t^2+s^2)/2}.\end{aligned}$$

この等式で $t = 0$ として，

$$\mathbb{E}[e^{itV}] = e^{-t^2/2}$$

を得る．よって $V \sim N(0,1)$. 同様にして $\mathbb{E}[e^{isW}] = e^{-s^2/2}$. よって $W \sim N(0,1)$. 以上を合わせて

$$\mathbb{E}[e^{itV+isW}] = \mathbb{E}[e^{itV}]\mathbb{E}[e^{isW}]$$

が任意の $s, t \in \mathbb{R}$ に対して成り立つことが分かる．ゆえに定理 5.1.14 より V と W は独立である．

モーメント母関数が存在する場合は，これによる同分布性や独立性による特徴付けが可能である．

定理 5.1.17 X, Y を確率変数とする．ある $r > 0$ が存在し，任意の $t \in (-r, r)$ に対し，$\mathbb{E}[e^{tX}] < \infty$, $\mathbb{E}[e^{tY}] < \infty$ を仮定する．このとき，以下が成り立つ．

(1) $X \overset{\mathrm{d}}{=} Y \iff \mathbb{E}[e^{tX}] = \mathbb{E}[e^{tY}], \quad t \in (-r, r)$.
(2) X と Y が独立 $\iff \mathbb{E}[e^{uX+vY}] = \mathbb{E}[e^{uX}]\mathbb{E}[e^{vY}], \quad u, v \in (-r, r)$.

証明．(1). ($\Leftarrow$). 定理 3.6.6 より，任意の $n \in \mathbb{N}$ に対し $\mathbb{E}[X^n] = \mathbb{E}[Y^n]$. $t \in (-r, r)$ を任意にとって固定する．このとき，

$$\sum_{n=0}^\infty \left| \frac{(-1)^n}{(2n)!} (tX)^{2n} \right| = \cosh(tX) \le e^{tX} + e^{-tX}$$

であり，$\mathbb{E}[e^{tX}] < \infty, \mathbb{E}[e^{-tX}] < \infty$（$\pm t \in (-r, r)$）．よってルベーグの収束定理より

$$\begin{aligned}
\mathbb{E}[\cos(tX)] &= \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}(tX)^{2n}\right] = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n)!}\mathbb{E}[X^{2n}] \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n)!}\mathbb{E}[Y^{2n}] = \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}(tY)^{2n}\right] \\
&= \mathbb{E}[\cos(tY)].
\end{aligned}$$

さらに，

$$\mathbb{E}[e^{|tX|}] = \mathbb{E}[1_{\{X>0\}}e^{|t|X} + 1_{\{X\le 0\}}e^{-|t|X}] \le \mathbb{E}[e^{|t|X}] + \mathbb{E}[e^{-|t|X}] < \infty.$$

このことと

$$\sum_{n=0}^{\infty} \left| \frac{(-1)^n}{(2n+1)!}(tX)^{2n+1} \right| \le e^{|tX|}$$

から，再びルベーグの収束定理を適用し，

$$\begin{aligned}
\mathbb{E}[\sin(tX)] &= \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}(tX)^{2n+1}\right] = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n+1}}{(2n+1)!}\mathbb{E}[X^{2n+1}] \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n+1}}{(2n+1)!}\mathbb{E}[Y^{2n+1}] = \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}(tY)^{2n+1}\right] \\
&= \mathbb{E}[\sin(tY)].
\end{aligned}$$

よって任意の $t \in (-r, r)$ に対して $\mathbb{E}[e^{itX}] = \mathbb{E}[e^{itY}]$．

一般の $t \in \mathbb{R}$ に対しては，$t_0 \in (-r, r)$, $h \in \mathbb{R}$ を適当に選び，$t = t_0 + h$ と表す．X, Y の特性関数をそれぞれ φ_X, φ_Y と書くことにすると，

$$\begin{aligned}
\varphi_X(t) &= \mathbb{E}\left[e^{it_0X}\sum_{n=0}^{\infty} \frac{(ihX)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{h^n}{n!}\mathbb{E}[(iX)^n e^{it_0X}] = \sum_{n=0}^{\infty} \frac{h^n}{n!}\frac{d^n\varphi_X}{dt^n}(t_0) \\
&= \sum_{n=0}^{\infty} \frac{h^n}{n!}\frac{d^n\varphi_Y}{dt^n}(t_0) = \sum_{n=0}^{\infty} \frac{h^n}{n!}\mathbb{E}[(iY)^n e^{it_0Y}] = \mathbb{E}\left[e^{it_0Y}\sum_{n=0}^{\infty} \frac{(ihY)^n}{n!}\right] \\
&= \varphi_Y(t).
\end{aligned}$$

ゆえに $\mathbb{R}$ 上で $\varphi_X = \varphi_Y$．従って定理 5.1.13 より $X \overset{\mathrm{d}}{=} Y$．

(2)．定理 5.1.14 の証明と同じ議論を使えばよい． □

5.2　法則収束

確率変数列 $\{X_n\}_{n=1}^{\infty}$ が確率変数 X に**法則収束**するとは，$\mathbb{P}(X = x) = 0$ を満たす任意の $x \in \mathbb{R}$ に対し，

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x)$$

が成り立つときにいう．

- X_n が X に法則収束することを $X_n \xrightarrow{\mathrm{d}} X$ と書く．X_n が X に**分布収束**するともいう．
- 法則収束は「確率変数の収束」ではなく「確率変数の分布の収束」である．X_n や X の定義される確率空間は異なっていてもよい．

条件「$\mathbb{P}(X = x) = 0$」はなぜ必要か．$X_n = a + 1/n$, $X = a$ のとき，X_n は X に「収束」しているとみなしたいが，$\mathbb{P}(X = x) > 0$ なる点 x も含めて法則収束を考えると，$x = a$ のとき，

$$0 = \mathbb{P}(X_n \le x) \nrightarrow \mathbb{P}(X \le x) = 1$$

となってしまう．よって分布関数がジャンプする点を排除する方が「収束」を適切に記述できる．

離散型確率変数の場合，法則収束は各状態確率の収束と同値である．

命題 5.2.1 確率変数 $X, X_1, X_2, \ldots$ が全て $\mathbb{N}$-値確率変数のとき，

$$X_n \xrightarrow{\mathrm{d}} X \iff \mathbb{P}(X_n = k) \to \mathbb{P}(X = k), \ \ k \in \mathbb{N}.$$

- この命題により，少数の法則（定理 3.4.21）は二項分布がポアソン分布に法則収束していることと言い換えることができる．

命題 5.2.1 の証明. ($\Rightarrow$). $\varepsilon \in (0, 1)$ を任意にとる．このとき，任意の $k \in \mathbb{N}$ に対し $\mathbb{P}(X = k \pm \varepsilon) = 0$ であるから，分布収束の定義より，

$$\mathbb{P}(X_n = k) = \mathbb{P}(X_n \le k + \varepsilon) - \mathbb{P}(X_n \le k - \varepsilon) \to \mathbb{P}(X \le k + \varepsilon) - \mathbb{P}(X \le k - \varepsilon) = \mathbb{P}(X = k).$$

($\Leftarrow$). 任意の $x \ge 1$ に対して，$\lfloor x \rfloor$ で x を超えない最大の整数を表すことにすると，

$$\mathbb{P}(X_n \le x) = \sum_{k=1}^{\lfloor x \rfloor} \mathbb{P}(X_n = k) \to \sum_{k=1}^{\lfloor x \rfloor} \mathbb{P}(X = k) = \mathbb{P}(X \le x).$$

□

法則収束が密度関数の収束を意味するとは限らない．一般には，法則収束は密度関数の収束よりも弱い．

例 5.2.2 各 $n \in \mathbb{N}$ に対して，関数

$$\rho_n(x) = \begin{cases} 1 + \cos(2\pi n x) & (0 \le x \le 1), \\ 0 & (\text{その他}) \end{cases}$$

を考える．

$$\int_{-\infty}^{\infty} \rho_n(x) dx = \int_0^1 (1 + \cos(2\pi n x)) dx = 1$$

だから ρ_n は確率密度関数である．X_n を ρ_n を密度としてもつような確率変数とする．このとき，$x \in (0, 1)$ に対し，

$$\mathbb{P}(X_n \le x) = \int_0^x \rho_n(y) dy = x + \frac{\sin(2\pi n x)}{2\pi n x} \to x, \quad n \to \infty.$$

よって，X_n は $(0, 1)$ 上の一様分布に従う確率変数に法則収束する．他方，各 $x \in (0, 1)$ に対し，$\rho_n(x)$ は $n \to \infty$ で振動するので収束しない．

次の結果は法則収束を特徴付ける．高度な議論が必要なので証明は略す．詳細については測度論的確率論のテキストをみよ．

定理 5.2.3 確率変数列 $\{X_n\}_{n=1}^{\infty}$, 確率変数 X に対して以下は同値.

(1) $X_n \xrightarrow{\mathrm{d}} X$.
(2) $\mathbb{E}[e^{itX_n}] \to \mathbb{E}[e^{itX}]$, $t \in \mathbb{R}$.
(3) 任意の有界連続関数 $f : \mathbb{R}^d \to \mathbb{R}$ に対して，$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

5.3 中心極限定理

大数の法則は IID 列の算術平均が平均値に収束することであった．**中心極限定理**はこの収束の様子をもう少し精密に記述する．

- $Z \sim N(0,1)$ を**標準正規（ガウス）確率変数**と呼ぶ.

定理 5.3.1 (**中心極限定理**) $\{X_n\}_{n=1}^{\infty}$ は IID で $\mathbb{E}|X_1|^2 < \infty$ を満たすとする．m, v をそれぞれ共通の期待値，分散とする．すなわち，$m = \mathbb{E}[X_1]$, $v := \mathbb{V}[X_1]$. このとき，$\dfrac{1}{\sqrt{n}}\displaystyle\sum_{k=1}^{n}\dfrac{(X_k - m)}{\sqrt{v}}$ は標準正規確率変数に法則収束する.

法則収束の定義より，任意の $a < b$ に対し，

$$\lim_{n\to\infty} \mathbb{P}\left(a < \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{(X_k - m)}{\sqrt{v}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

粗く述べると，$Z \sim N(0,1)$ とするとき，十分大きな n で

$$\frac{1}{n}\sum_{k=1}^{n} X_k \approx m + \sqrt{\frac{v}{n}}Z.$$

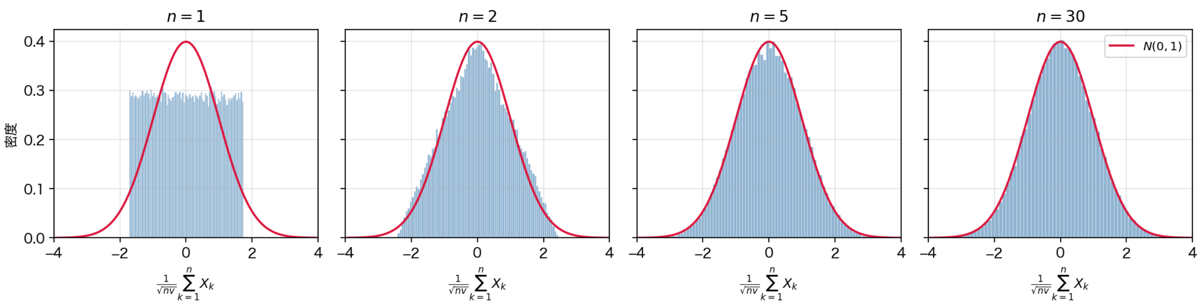


図 5.1　$X_k \sim U[-1,1]$（$v = 1/3$）の正規化和 $\frac{1}{\sqrt{nv}}\sum_{k=1}^{n} X_k$ のヒストグラム（10 万サンプル）. n が増えるにつれ，分布が標準正規分布の密度（赤線）へ近づく.

定理 5.3.1 の証明のために補題を一つ用意する.

補題 5.3.2 $x \in \mathbb{R}$ に対し，

$$e^{ix} = 1 + ix - \frac{x^2}{2} + x^2 \int_0^1 (1-s)(1-e^{ixs})ds. \tag{5.3.1}$$

証明．微積分学の基本定理，変数変換，積分順序の変更により，

$$\begin{aligned} e^{ix} &= 1 + \int_0^x ie^{iy}dy = 1 + ix\int_0^1 e^{ixt}dt = 1 + ix\int_0^1 \left(1 + ix\int_0^t e^{ixs}ds\right)dt \\ &= 1 + ix - x^2\int_0^1\int_s^1 e^{ixs}dtds = 1 + ix - x^2\int_0^1 (1-s)e^{ixs}ds \\ &= 1 + ix - \frac{x^2}{2} + x^2\int_0^1 (1-s)(1-e^{ixs})ds. \end{aligned}$$

□

定理 5.3.1 の証明．一般性を失うことなく，$m = 0, v = 1$ を仮定できる．(5.3.1) より，

$$e^{itX_1/\sqrt{n}} = 1 + \frac{it}{\sqrt{n}}X_1 - \frac{t^2}{2n}X_1^2 + \frac{t^2}{n}W_n$$

と書ける．ここで，

$$W_n = X_1^2\int_0^1 (1-s)(1-e^{istX_1/\sqrt{n}})ds.$$

各 ω に対して $\lim_{n\to\infty} W_n(\omega) = 0$ であり，

$$|W_n| \le |X_1|^2\int_0^1 (1-s)|1-e^{istX_1/\sqrt{n}}|ds \le |X_1|^2.$$

よってルベーグの収束定理を適用し，

$$\varepsilon_n := \mathbb{E}[W_n] \to 0.$$

この記法と X_k たちが IID であることから，

$$\mathbb{E}[e^{it(X_1+\cdots+X_n)/\sqrt{n}}] = \left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon_n\right)^n = \left(1 - \frac{t^2}{2n}\right)^n\left(1 + \frac{t^2\varepsilon_n/n}{1-t^2/(2n)}\right)^n.$$

二項定理より，$2n > t^2$ を満たす大きな n に対し，

$$\begin{aligned} \left(1 + \frac{t^2\varepsilon_n/n}{1-t^2/(2n)}\right)^n &= 1 + \sum_{k=1}^n \frac{n!}{k!(n-k)!}\left(\frac{2t^2\varepsilon_n}{2n-t^2}\right)^k \\ &= 1 + n\left(\frac{2t^2\varepsilon_n}{2n-t^2}\right)\sum_{k=1}^n \frac{(n-1)!}{k!(n-k)!}\left(\frac{2t^2\varepsilon_n}{2n-t^2}\right)^{k-1} \\ &= 1 + \left(\frac{2nt^2}{2n-t^2}\right)\varepsilon_n\sum_{k=0}^{n-1} \frac{(n-1)!}{(k+1)!(n-1-k)!}\left(\frac{2t^2\varepsilon_n}{2n-t^2}\right)^k \\ &\le 1 + 2t^2\varepsilon_n\left(1 + \frac{2t^2\varepsilon_n}{2n-t^2}\right)^{n-1} \\ &\le 1 + 2t^2\varepsilon_n\left(1 + \frac{t^2\varepsilon_n}{n-1}\right)^{n-1} \\ &\le 1 + 2t^2\varepsilon_n e^{t^2\varepsilon_n}. \end{aligned}$$

従って，

$$\left|\left(1+\frac{t^2\varepsilon_n/n}{1-t^2/(2n)}\right)^n-1\right|\le 2t^2\varepsilon_n e^{t^2\varepsilon_n}\to 0,\quad n\to\infty.$$

ゆえに

$$\lim_{n\to\infty}\mathbb{E}[e^{it\sum_{k=1}^n X_k/\sqrt{n}}]=\lim_{n\to\infty}\left(1-\frac{t^2}{2n}\right)^n=e^{-t^2/2}.$$

あとは定理 5.2.3 を適用すればよい． □

例題 5.3.3 コインを 10000 回投げるとき，約 95% の確率で，表の出る回数は 4900 回から 5100 回の間である．これを確かめよ．

解答. $\{X_n\}_{n=1}^\infty$ を IID で，$\mathbb{P}(X_1=0)=\mathbb{P}(X_1=1)=1/2$ を満たすものとする．このとき，$\mathbb{E}[X_1]=1/2$, $\mathbb{V}[X_1]=1/4$ であり，中心極限定理より，$S_n=X_1+\cdots+X_n$ とおくとき，

$$\frac{S_n-\frac{n}{2}}{\sqrt{\frac{n}{4}}}\xrightarrow{\mathrm{d}} Z\sim N(0,1).$$

$n=10000$ が十分大きいとすれば，

$$\begin{aligned}\mathbb{P}(\text{表が4900〜5100回出る})&=\mathbb{P}(S_n\in[4900,5100])=\mathbb{P}(S_n-n/2\in[-\sqrt{n},\sqrt{n}])\\&=\mathbb{P}\left(\left|\frac{S_n-\frac{n}{2}}{\sqrt{\frac{n}{4}}}\right|\le 2\right)\approx\mathbb{P}(|Z|\le 2)=1-2\mathbb{P}(Z>2).\end{aligned}$$

$\mathbb{P}(Z>\alpha)$ の値は正規分布表などから算出でき，$\mathbb{P}(Z>2)\approx 0.0228$ である．従って，

$$\mathbb{P}(\text{表が4900〜5100回出る})\approx 1-2\times 0.0228=0.9544.$$

例題 5.3.4 モンテカルロ法により $\int_0^1 f(x)dx$ を計算したい．$U_1,\dots,U_n$ を独立に発生させた $[0,1]$ 上の一様乱数として，$\sigma^2=\mathbb{V}[f(U_1)]$ とおく．近似誤差 $\left|\frac{1}{n}\sum_{k=1}^n f(U_k)-\int_0^1 f(x)dx\right|$ が 0.1σ 以下である確率を 99% 以上にするためには，何個くらいの乱数を発生させればよいか？

解答.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n f(U_k)-\int_0^1 f(x)dx\right|>0.1\sigma\right)<0.01 \tag{5.3.2}$$

となる n を求める．n として十分大きな数を想定して中心極限定理を適用すると，

$$\begin{aligned}(5.3.2)\text{の左辺}=\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{k=1}^n\frac{f(U_k)-\mathbb{E}[f(U_1)]}{\sqrt{\mathbb{V}[f(U_1)]}}\right|>\frac{\sqrt{n}}{10}\right)&\approx\mathbb{P}(|Z|>\sqrt{n}/10)\\&=2\mathbb{P}(Z>\sqrt{n}/10).\end{aligned}$$

ただし $Z\sim N(0,1)$.

$$\mathbb{P}(Z>2.58)\approx 0.0049<0.0051\approx\mathbb{P}(Z>2.57)$$

なので，$\sqrt{n}/10\approx 2.58$ を満たす n をとればよい．これより $n=666$ 個の乱数で十分ということがわかる．

例題 5.3.5 （ヨーロピアン・スタイルの）ルーレットゲームを考える．ルーレットには 37 のポケットがあり，数字は 1 から 36 までのうち 18 個は赤，残りの 18 個は黒，0 は緑である．ゲームの配当は 2 倍である．すなわち，収益は $\pm 1\times$ 賭け数である．
今，毎回赤にチップを 1 枚賭けるギャンブラーを考える．$1369 = 37^2$ 回ゲームするとき，平均で 37 枚チップを失うことになるが，負けない確率が約 16% あることを確かめよ．

解答．ギャンブラーの各回の収益を X_n で表すと，$\{X_n\}_{n=1}^{\infty}$ は IID で，$\mathbb{P}(X_n = 1) = 18/37$, $\mathbb{P}(X_n = -1) = 19/37$．これより，

$$\begin{aligned}\mathbb{E}[X_n] &= \frac{18}{37} - \frac{19}{37} = -\frac{1}{37},\\ \mathbb{V}[X_n] &= 1 - \frac{1}{37^2} \approx 0.9992 \approx 1.\end{aligned}$$

よって，n 回目までの収益を S_n で表すと，$\mathbb{E}[S_n] = -n/37$ であり，従って，$\mathbb{E}[S_{1369}] = -37$.
　一方，中心極限定理を適用し，

$$\begin{aligned}\mathbb{P}(S_n \geq 0) &= \mathbb{P}\left(\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{\mathbb{V}[X_1]n}} \geq -\frac{n\mathbb{E}[X_1]}{\sqrt{\mathbb{V}[X_1]n}}\right)\\ &\approx \mathbb{P}(Z \geq 1)\\ &\approx 0.1587.\end{aligned}$$

章末問題

　非負整数値確率変数 X に対して，

$$g_X(z) := \mathbb{E}[z^X], \quad z \in \mathbb{C},\ |z| \leq 1$$

を X の**確率母関数**という．

問題 5.1 非負整数値確率変数列 $\{X_n\}_{n=1}^{\infty}$ は IID で共通の確率母関数 $g_X(z)$ をもつとする．T を非負整数値確率変数で，$\{X_n\}$ と独立とする．このとき，$Y := \sum_{n=1}^{T} X_n$ は

$$g_Y(z) = g_T(g_X(z))$$

を満たすことを示せ．ただし，$\sum_{n=1}^{0} = 0$ とする．

問題 5.2 $X_1, \ldots, X_n$ は独立で，各 k について $X_k \sim N(m_k, \sigma_k^2)$ であるとする．このとき，特性関数を用いて，$X_1 + \cdots + X_n \sim N(\bar{m}, \bar{\sigma}^2)$ を示せ．ただし，$\bar{m} = m_1 + \cdots + m_n$, $\bar{\sigma} = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$.

問題 5.3 $m \in \mathbb{R}^d$, $V \in \mathbb{R}^{d\times d}$ は正定値対称行列とし，$X \sim N(m, V)$ とする．このとき，$Y := B^{-1}(X - m) \sim N(0, I_d)$ であることを示せ．ここで B は正則な対称行列で $B^2 = V$ を満たすもので，I_d は d 次の単位行列とする．

問題 5.4 f を $[0, \infty)$ 上の有界連続関数とする．$x \in (0, \infty)$ を任意にとる．$\{X_n\}_{n=1}^{\infty}$ は IID で，共通の分布はパラメータ x のポアソン分布であるとする．

(1) $S_n := X_1 + \cdots + X_n$ がパラメータ nx のポアソン分布に従うことを示せ．

(2) 大数の強法則とルベーグの収束定理を用いて，

$$f(x) = \lim_{n\to\infty} \sum_{\ell=0}^{\infty} f\left(\frac{\ell}{n}\right) e^{-nx} \frac{(nx)^\ell}{\ell!}$$

を示せ．（ヒント: $\mathbb{E}[f(S_n/n)]$ を計算してみよ．）

問題 5.5 $\{X_n\}_{n=1}^{\infty}$ は IID で，$X_1 \sim U([0,1])$ とする．このとき，$Y_n := n \cdot \min(X_1, \ldots, X_n)$ はパラメータ 1 の指数分布に従う確率変数に法則収束することを示せ（ヒント: $\mathbb{P}(Y_n > x)$ を計算する）．

問題 5.6 （偏りのない）コインを 900 回投げるとする．中心極限定理を用いて，495 回以上表が出る確率の近似値を求めよ．ただし，$Z \sim N(0,1)$ とするとき，$\mathbb{P}(Z \geq 3) = 0.0013$ であることを用いてもよい．

問題 5.7 中心極限定理を用いて，

$$\lim_{n\to\infty} e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!} = \frac{1}{2}$$

を示せ．

参考文献

[1] M. H. De Groot, M. J. Schervish, 椿広計 監訳, 大野忠士 監訳, 領家美奈 監訳, 浅野美代子 等訳. デグルート＆シャービッシュ 確率と統計. 共立出版, 2022. 原著第 4 版の翻訳.

[2] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Birkhäuser/Springer, Cham, 2015.

[3] W. フェラー, 国沢清典 監訳, 羽鳥裕久, 大平坦 訳. 確率論とその応用. 第 2 上. 紀伊國屋書店, 1969.

[4] W. フェラー, 河田竜夫 監訳, ト部舜一 等訳. 確率論とその応用. 第 1 上. 紀伊國屋書店, 1960.

[5] W. フェラー, 河田竜夫 監訳, ト部舜一 等訳. 確率論とその応用. 第 1 下. 紀伊國屋書店, 1961.

[6] W. フェラー 著, 国沢清典 監訳, 羽鳥裕久, 大平坦 訳. 確率論とその応用. 第 2 下. 紀伊國屋書店, 1970.

[7] P. ブレモー 著, 向井久 訳. モデルで学ぶ確率入門. 丸善出版, 新装版, 2012.

[8] 伊藤清. 確率論. 岩波書店, 1991.

[9] 伊藤清三. ルベーグ積分入門. 裳華房, 1963.

[10] 佐藤担. 測度から確率へ. 共立出版, 1994.

[11] 吉田伸生. 確率の基礎から統計へ. 日本評論社, 新装版, 2021.