講義ノート

確率論

確率空間と極限定理

村田 昇

version: 2020 年 8 月 21 日

早稲田大学 先進理工学部
電気・情報生命工学科

はじめに

工学においては「不確定性を伴なって観測されるデータから対象とする現象の数学的なモデルを構築する」といった問題に直面することが多々あります．例えば実験データの解析や，ノイズを伴なって観測されるシステムの入出力関係の同定といった問題です．典型的には

- 「データの生成機構は不確定性を持っている」
- 「不確定性を記述するための (パラメトリック) モデルを作る」
- 「我々が現実の場面で利用できるのは，この生成機構から出てくる有限個のサンプルである」
- 「この有限個のサンプルを使ってモデルのパラメタを推測する」

といった条件で数学的に妥当なモデルを考えることになります．このような場面では不確定性や曖昧さを記述する手法として確率モデルが重要な役割を果します．

この講義は，信号処理・情報理論・制御理論・学習理論といった分野で必要とされる確率モデルを扱うために，確率論と数理統計学の基本的な枠組を学んでもらうことを目的とします．

確率・統計は対にして語られることが多いですが，その考え方は対照的です．確率は演繹的 (deductive) であり，ある確率的な構造をもつシステムから観測されるデータの性質を論じることを目的とします．一方統計は帰納的 (inductive) であり，観測されたデータから，それを生成したシステムの確率的な構造を推測することが目的となり，どちらも重要な工学の基礎概念となります．

本講義で扱う「確率」は場合の数を数え上げる組合せ論的なものではなく，測度論を基礎とする確率の現代的な枠組みです．確率空間・測度・密度・確率変数といった概念を学んでもらい，工学の様々な場面で必要となる確率論の基本的知識を身につけてもらうことを目標とします．特に

- 確率空間
- 様々な確率分布とそれに関わる量
- 独立な確率変数の和に関する定理

といった項目について学びます．

一方「統計」では数理統計学の基本となる考え方を紹介します．この枠組は学習や制御あるいは信号処理といった分野と関連が深く，特に線形モデルを主体とした統計的手法は本質的に非線形となるモデルを扱わなくてはいけない現実的な場面においても大事な基礎となります．主として

- 推定 (不偏推定，最尤推定)
- 検定

を取り上げ，統計的な考え方を学んでもらいます．また時間が許せば

- ベイズ統計
- 多変量解析 (回帰分析，主成分分析，判別分析など)

といった項目についても簡単に説明します．

教科書は特に指定しませんが，末尾にいくつか参考書を挙げてあるので，必要に応じて参照して下さい．

目次

有限試行の確率空間 1

1.1 試行と見本空間

確率論は，不確定性のある現象における物事の起こり易さ・起こり難さを記述するための数学的な枠組である．確率を考えるためには，対象となる現象を客観的に調べる実験の方法と，その実験の結果観測される事柄を規定しておく必要がある．まずはじめに，これらを記述するために必要ないくつかの用語を定義する．

1.1.1 試行と見本点

ある現象の背後に潜む仕組みを調べることを考える．この「仕組み」とは確率論の言葉で記述される数学的な構造である．このためには通常何らかの「実験」を行い，現象の全体，あるいはその一部分を「観測」する必要がある．この「実験」のことを**試行** (trial) と呼ぶ．また，試行の結果観測される事柄を**見本点** (sample point) あるいは**標本**という．統計などの別の分野では見本点を**観測値** (observation)，あるいは**実現値** (realization) と呼ぶこともある．

例 1.1 (コイン投げ)．「コインを投げ，表か裏か調べる」という実験を考える．この場合「コインを投げる」ことが試行に対応する．また「表」と「裏」が観測される見本点となる．

例 1.2 (骰子振り)．「骰子 (サイコロ) を振って，どの目が出易いか調べる」という実験を考える．この場合「骰子を振る」ことが試行に対応する．またこの試行では「1, 2, 3, 4, 5, 6 のいずれかの目」が出る可能性があるので，これら全てが見本点となる．

例 1.3 (ルーレット回し)．賭博に用いるルーレットではなく，思考実験のために以下のような特別なルーレットを用意する．円周の長さが 1m の円盤を用意し，円周上の一点を起点として周長に従って 0 から 1 の目盛りを付ける．ただし，実際に細かく目盛りが付けられる訳ではないので仮想的に付けるものとする．また 0 と 1 は円周上では一致していることに注意する．起点であるこの点には “0” と “1” という別の名前が付いてしまうが，以下では 0 は用いずに 1 のみを用いることにする．さて，この円盤の中心に回転するように釘を打ち，円周の外側の適当な位置に印 (円周の位置を示すための矢印) を付ける．これが以降で考えるルーレットである．そこで「ルーレットを回して止まったときに印が指している目盛りを読む」という実験を考える．この場合「ルーレットを回す」ことが試行に対応し，試行の結果得られる見本点は「$(0,1]$ の間のいずれかの実数値」となる．ただし，$[a,b]$ は閉区間，(a,b) は開区間を表すものとし，$(a,b]$ とした場合は a は区間に含まれないが b は区間に含まれることを表す．

1.1.2 見本空間

見本点全体の集合，すなわち試行を数限りなく繰り返したとき観測される全ての見本点を集めた集合は**見本空間** (sample space) あるいは**標本空間**と呼ばれる．慣習的に見本空間は Ω (ギリシャ文字のオメガの大文字) のように大文字で表される．またその要素である 1 つ 1 つの見本点は ω (オメガの小文字) のように小文字で表される．

例 1.4 (コイン投げ)**.** 試行 $T =$「コインを投げる」とすると見本空間は $\Omega = \{$表, 裏$\}$ と表される．

例 1.5 (骰子振り)**.** 試行 $T =$「骰子を振る」の見本空間は $\Omega = \{1, 2, 3, 4, 5, 6\}$ と表され，試行 T の結果 2 の目が出た場合は「2 が観測された」，あるいは「$\omega = 2$ が観測された」というように記述される．

例 1.6 (ルーレット回し)**.** 試行 $T =$「ルーレットを回す」の見本空間は $\Omega = (0, 1]$ である．

定義からわかるように，見本空間は試行によって規定される．このとき見本空間が有限集合 (要素の数が有限個) である試行を**有限試行**と呼ぶ．これに対して見本空間が無限集合 (要素の数が無限個) となる試行を**無限試行**と呼ぶ．

例 1.7. 「コイン投げ」「骰子振り」の場合，見本点はそれぞれ 2 つ，6 つしかないので明らかに有限試行である．一方「ルーレット回し」の場合，見本点の数は数え切れないので無限試行である．

例 1.8. 「1 が出るまで骰子を振り続ける」という試行を考える．この場合，例えば出た目を順に並べた数字の列 $\langle 4, 2, 5, 6, 1\rangle$ が 1 つの見本点になる．このとき見本点としていくらでも長い系列が存在することに注意しよう．見本空間を直接書き下すことはできないが，いくつか例示すると $\Omega = \{\langle 1\rangle, \langle 2, 1\rangle, \langle 3, 1\rangle, \dots, \langle 4, 2, 5, 6, 1\rangle, \dots\}$ となり，その要素の数は無限となる．したがってこの試行は無限試行である．

このように同じ骰子を使った実験でも，考察の対象とする試行によって見本点の個数は有限にも無限にもなり，見本空間のもつ性質が本質的に異なる場合があることに注意しなくてはならない．

1.2 事象とその表現

確率を考える対象としては，単一の見本点だけではなく，複数の見本点の集まりを考えることができる．ここでは見本点の集合である事象とその記述の仕方，および基本的な演算についてまとめておく．

1.2.1 事象

見本点の集合，すなわち見本空間の部分集合は**事象** (event) と呼ばれる．ある試行 T の見本空間を Ω とし，その部分集合を A としたとき，試行 T の結果として部分集合 A に属する見本点が出現することを「**事象** A **が起こる**」と表現する．1 つの見本点だけからな

る集合も事象となるが，これを**根元事象** (elementary event) と呼ぶことがある．また，特別な事象として見本空間全体 Ω を**全事象** (full event, whole event)，空集合 $\emptyset$ を**空事象** (null event, empty event) と呼ぶ．

例 1.9 (骰子振り)**.** 事象 A として「偶数の目」$A = \{2, 4, 6\}$ を考えると，骰子を振って 4 の目が出た場合は 4 は A に属している ($4 \in A$) ので「事象 A が起こった」ことになり，5 の目が出た場合は $5 \notin A$ なので「事象 A は起こらなかった」ことになる．

例 1.10 (ルーレット回し)**.** 事象 A を区間 $(0, 0.5]$ とする．ルーレットを回して止まったとき印が 0.5 を指していたら $0.5 \in A$ なので「事象 A が起こった」ことになり，$\pi/4 = 0.785\cdots$ を指していたら $\pi/4 \notin A$ なので「事象 A は起こらなかった」ことになる．

1.2.2 事象の演算

事象を用いた演算については以下の 4 つが重要である．また，事象は集合なので，事象の演算は集合論での演算に対応することに注意する．

- **和事象** (sum event)– A または B が起こることを表す．A, B の和集合 (union) $A \cup B$ で表記する．
- **交事象** (product event)– A と B が同時に起こることを表す．A, B の交集合 (intersection) $A \cap B$ で表記する．
- **差事象** (difference event)– A が起こり B が起こらないことを表す．A, B の差 (difference) $A \backslash B$ で表記する．
- **余事象** (complementary event)– A が起こらないことを表す．A の補集合 (complement) A^c，または $\bar{A}$ で表記する．

これらに関連して特殊な概念および演算として以下の 3 つがある．

- **排反事象** (exclusive event)– 事象 A, B が同時に起こらないことを表す．A, B が**互いに素** (disjoint) ともいう．この関係は空集合 (空事象)$\emptyset$ を用いて $A \cap B = \emptyset$ で表現される．
- **直和** (direct sum, disjoint union)– 排反事象の和 $A \cup B$ を $A + B$ と書く．逆に $A + B$ と書いた場合 A と B が排反であることが暗黙の了解となる．3 つ以上の事象についても同様である．
- **固有差** – $A \supset B$ のとき，すなわち事象 B が起これば必ず事象 A が起こるとき，A と B の差事象 $A \backslash B$ を $A - B$ と書く．

なお，差事象は余事象を用いると $A \backslash B = A \cap B^c$ と書くこともできる．また，余事象は固有差を用いると $A^c = \Omega - A$ と書くことができる．

例 1.11 (骰子振り)**.** 3 つの事象

事象 A として「偶数の目」$A = \{2, 4, 6\}$,
事象 B として「奇数の目」$B = \{1, 3, 5\}$,
事象 C として「素数の目」$C = \{2, 3, 5\}$

を考えると，例えば

$$\begin{aligned}&A\text{の余事象は } B = A^c,\\ &B\text{と}C\text{の交事象は } B\cap C=\{3,5\},\\ &A\text{と}C\text{の和事象は } A\cup C=\{2,3,4,5,6\},\\ &A\text{と}B\text{の直和は全事象 } A+B=\Omega=\{1,2,3,4,5,6\}\end{aligned}$$

のようになる．

練習問題 (1) (2)

1.2.3 条件による事象の表現

事象は見本点 ω に関する**条件** (condition) として表すこともある．この場合は条件を表す式 $\alpha(\omega)$ を用いて

$$A = \{\omega |\ \alpha(\omega)\}$$

のように書く．場合によっては条件 α を事象 A と同一視して，事象 α ということもあるので注意する．

例 1.12 (骰子振り)**.** 「偶数の目」という事象 A は

$$A = \{\omega |\ \omega\text{が偶数}\} = \{2,4,6\}$$

のように表される．

例 1.13 (ルーレット回し)**.** 「区間 $(0,0.5]$ の数」という事象 A は

$$A = \{\omega |\ 0 < \omega \le 0.5\}$$

のように表される．

同様に和事象，交事象，余事象をそれぞれ条件に関する論理式を用いて

- α または β : $\alpha \vee \beta$ (和事象)

  $$\{\omega |\ \alpha(\omega)\vee\beta(\omega)\} = \{\omega |\ \alpha(\omega)\}\cup\{\omega |\ \beta(\omega)\}$$

- α かつ β : $\alpha \wedge \beta$ (交事象)

  $$\{\omega |\ \alpha(\omega)\wedge\beta(\omega)\} = \{\omega |\ \alpha(\omega)\}\cap\{\omega |\ \beta(\omega)\}$$

- α の否定 : $\alpha^{\neg}$ (余事象)

  $$\{\omega |\ \alpha(\omega)^{\neg}\} = \Omega - \{\omega |\ \alpha(\omega)\}$$

のように表すこともある．

例 1.14 (骰子振り)**.** それぞれの事象を条件で書くと

$$\begin{aligned}&\{\omega |\ (\omega\text{が偶数})^{\neg}\} = \{\omega |\ \omega\text{が奇数}\} \quad (\text{余事象}),\\ &\{\omega |\ (\omega\text{が奇数})\wedge(\omega\text{が素数})\} = \{3,5\} \quad (\text{交事象}),\\ &\{\omega |\ (\omega\text{が偶数})\vee(\omega\text{が素数})\} = \{2,3,4,5,6\} \quad (\text{和事象}),\\ &\{\omega |\ (\omega\text{が偶数})\vee(\omega\text{が奇数})\} = \{1,2,3,4,5,6\} \quad (\text{和事象})\end{aligned}$$

となる．

練習問題 (3) (4)

1.3 有限試行の確率空間

これまでの準備をもとに，各事象の起こり易さである確率を計算する関数を定義し，確率を考える空間を規定する．この節では有限試行の場合について考える．

1.3.1 有限試行の確率測度

まず見本空間上で事象の起こる確率を表す関数としての確率測度を定義する．

定義 1.15 (確率測度)**.** 見本空間 Ω と任意の事象 $A, B \subset \Omega$ に対して以下の性質をもつ実数値集合関数 P(Ω の部分集合に作用して実数を出力する関数) を**確率測度** (probability measure) または**確率分布** (probability distribution) あるいは単に**分布** (distribution) という．

(P.1) $P(A) \geq 0,$ (**正値性**; positivity)

(P.2) $P(A+B) = P(A) + P(B),$ (**加法性**; additivity)

(P.3) $P(\Omega) = 1$ (**全確率**は 1)

確率測度とはある事象が起こる頻度を測るものさしである．上の条件は確率に要請される自然な性質に対応し，それぞれ

(P.1) ある事象の起こる確率は 0 または正の値を取る

(P.2) 排反な事象の和の確率はそれぞれの事象の確率の和となる

(P.3) ある試行を行ったとき見本空間の中のどれか 1 つの見本点は必ず観測される

ことを意味する．

例 1.16 (骰子振り)**.** いかさまのない骰子を 1 回振る試行 T の確率測度 P とは，

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

$$P(\text{素数の目が出る}) = P(\{2,3,5\}) = \frac{1}{2}$$

のように，適当な事象を入れるとその事象の起こる確率を返してくれる関数 P である．

上に挙げた性質 (P.1–3) が確率を定義するために必要であることは「骰子振り」におけるさまざまな状況を考えれば明らかであろう．

練習問題 (5) (6)

1.3.2 確率空間

ある試行の確率を考える枠組は，その見本空間と確率測度によって次のように規定される．

定義 1.17 (確率空間)**.** 見本空間 Ω と事象の集合 (集合族)$\mathcal{F}$ と確率測度 P の組 $(\Omega, \mathcal{F}, P)$ を，**確率空間** (probability space) と呼ぶ．

確率測度 P は「事象」の起こる確率を計算する関数であるから，事象の集合 (集合族)$\mathcal{F}$ は P の定義域であり，確率を考えるべき「事象」=「Ω の部分集合」の集合である．通常は空集合と見本空間全体を含む Ω のあらゆる部分集合の集合 (冪集合 2^{Ω}; 部分集合は，各要素について見ると含むか含まないかの 2 通り，全体では Ω の要素数乗個ある) を考えればよい．なお，「集合の集合」という言い方は数学的には正しい言い方ではなく，後に集合族という概念を説明する．

試行 T が定まると，その見本空間 Ω，考えるべき事象の集合 $\mathcal{F}$，およびそのの確率法則 P を考えることができる．この確率空間 $(\Omega, \mathcal{F}, P)$ を，試行 T を明示して**試行 T の確率空間**と呼ぶことがある．確率空間 $(\Omega, \mathcal{F}, P)$ は試行 T の数学的側面を完全に表していることになる．

試行 T の結果，どの見本点がどのくらいの頻度で起こるかを表す**確率法則** (probability low)P は，試行 T の見本空間 Ω 上の確率測度 P によって記述されるので，通常は確率法則と確率測度を同一視する．本稿では P のことを確率法則と呼んだり確率測度，あるいは確率分布と呼んだりすることがあるが，基本的には同じものを指し示していると思えばよい．

定理 1.18. 事象の演算に関しては以下の関係が成り立つ．[1]

(1) $$P\Big(\sum_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i)$$

(2) $$P(A - B) = P(A) - P(B)$$

(3) $$P(A^c) = 1 - P(A)$$

(4) $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(5) $$P(A) = \sum_{\omega \in A} P\{\omega\}$$

[1] $\sum_{i=1}^{n} A_i = A_1 + A_2 + \cdots + A_n$ は直和を意味しており，A_i, $i = 1, 2, \ldots$ は排反 (互いに素) であることを，また $A - B$ については $A \supset B$ であることを暗黙のうちに表わしていることに注意する．なお以降では記法を簡便にするため，1 点 ω からなる集合 $\{\omega\}$ の確率測度を $P\{\omega\}$ と表すことがあるので注意する．

練習問題 (7)

当たり前ではあるが有限試行の場合，(5) の性質よりあらゆる見本点 ω についてその起こる確率 $P\{\omega\}$ が決まれば確率法則 P は完全に定まることがわかる．

例 1.19 (骰子振り)**.** いかさまのない骰子を 1 回振る試行 T において素数の目が出る確率は

$$\begin{aligned} P(\text{素数の目}) &= P(\{2, 3, 5\}) \\ &= P\{2\} + P\{3\} + P\{5\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \end{aligned}$$

となる．

例 1.20 (コイン投げ)**.** 2 枚のコインを投げる試行 T において 2 つとも表が出る確率を考える．$\langle$表，表$\rangle$, $\langle$表，裏$\rangle$, $\langle$裏，表$\rangle$, $\langle$裏，裏$\rangle$ の 4 通りが等しい確率で起こるので

$$P\left(\langle \text{表，表} \rangle\right) = \frac{1}{4}$$

である．

練習問題 (8) (9) (10)

1.4 試行が有限でない場合の問題点

次に一般の試行 T と，それに付随する確率空間 $(\Omega, \mathcal{F}, P)$ がどのようなものになるかを考えてみよう．確率測度が持つべき性質は有限試行の場合とほぼ同じであるが，Ω の要素の数が有限でないということが本質的な問題を引き起こすことがある．

まず以下のような無限試行で確率の値を求める問題を考えてみよう．

例 1.21 (骰子振り)．「1 が出るまで骰子を振り続ける」という試行で $\langle 6,5,4,3,2,1\rangle$ という系列が得られる確率は，6 回骰子を振った場合の 6^6 通りの中のただ 1 通りであり，どの系列が出るかは等しい確率なので

$$P\left(\langle 6,5,4,3,2,1\rangle\right) = \frac{1}{6^6}$$

となる．また，ちょうど 6 回で 1 が出て終わる確率は，最初の 5 回は 1 以外の目が出て，最後に 1 の目が出る 5^5 通りの場合があるので

$$P\left(\langle *,*,*,*,*,1\rangle\right) = \frac{5^5}{6^6}$$

である．ただし $*$ は 1 以外の目を表すものとする．[2]

[2] 確率の積でも計算できるが，まだ独立性を定義していないので，ここでは場合の数で説明している．

この試行は無限試行であるが，特定の見本点の確率を求めるには有限試行と同じように計算することができる．

一方，次の無限試行の例を考えてみよう．

例 1.22. 「ルーレット回し」の試行で，ちょうど 0.5 の値が出る確率はいくつか？

この問題を有限試行と同じように扱うと以下のような矛盾が生じる．

まず，事象 A を適当な数の集合として，これを $A = \{a, b, c, \dots\}$ と書くことにする．前出の加法性に従うなら事象 $A = \{a, b, c, \dots\}$ に対して，1 回の試行で例えば見本点 a と b が同時に観測されることはないので，

$$P(A) = P\{a\} + P\{b\} + P\{c\} + \cdots$$

として良さそうである．仮に各要素の出現確率が同じ $\epsilon > 0$ という値である場合

$$P\{a\} = P\{b\} = P\{c\} = \cdots = \epsilon \quad (\neq 0)$$

を考えてみよう．「ルーレット回し」の場合には，円盤が精確に作られているのであればこの条件が成り立つと考えて良いだろう．無限試行の場合には，事象 A が無限個の要素を持っている場合もあることに注意しよう．このときにはどんな小さな ϵ であっても，足し合わせるうちにいつか 1 を越え，最終的には

$$P(A) = \epsilon + \epsilon + \epsilon + \cdots \quad \to \infty$$

となってしまう．これは確率の値が 1 を越えてしまうので，明らかに矛盾である．このことから，事象 A に含まれる見本点が無限個あり，各見本点が同様に起こり易い場合には，

$$P\{a\} = P\{b\} = P\{c\} = \cdots = 0$$

でなくてはならないことになる．一方，任意の見本点についてその確率が 0 なら，いくら足しても

$$P(A) = P\{a\} + P\{b\} + P\{c\} + \cdots = 0$$

である．さらに見本空間全体についてこの議論を行えば $P(\Omega) = 0$ となってしまうので，全事象の確率がうまく定義できなくなってしまい，やはり矛盾が生じることになる．

この問題をうまく解決するには，加法性の概念を考え直す必要がある．この時，事象を構成する要素の数の「無限の程度」が，数えられるものであるかどうかということが実は重要な問題となる．

まとめ

- 確率論の基本用語: 試行，見本点，見本空間，事象
- 事象は見本空間の部分集合
- 事象の演算は集合の演算と等価
- 有限試行の確率測度の基本的な性質 (正値性，加法性，全確率)
- 確率空間は見本空間，事象の集合 (定義域)，確率測度の三つ組

練習問題

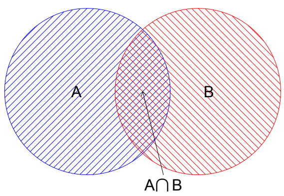

図 1.1: ベン図の例．交事象 $A \cap B$ が示す領域を表している．

(1) 和事象，交事象，差事象，余事象，排反事象，固有差，直和をベン図 (Venn diagram) を描いて説明せよ (図 1.1 参照)．

(2) よく切ったトランプから 1 枚カードを引く試行を考える．このとき適当な事象と，事象の演算を考えてみよ．

(3) よく切ったトランプから 1 枚カードを引く試行を考える．このとき適当な事象を条件で記述し，それらの演算を考えてみよ．

(4) ド・モルガンの法則 (de Morgan's laws) について説明せよ．

(5) よく切ったトランプから 1 枚カードを引く試行を考える．この試行の確率測度がどのような関数になるか説明せよ．

(6) 2 つの骰子を投げる試行を考える．この試行の確率測度がどのような関数になるか説明せよ．

(7) 確率測度の定義にもとづいて，定理 1.18 の各式を証明せよ．

(8) 自分を含めて教室に 20 人の生徒がいるとして，以下の問に答えよ．

a) 自分と同じ誕生日の人がいる確率はいくつか？

b) 誰かと誰かが同じ誕生日である確率はいくつか？

ただし，どの日に生れるかは，一様にばらついている (ある特定の日に生まれた人が多いということはない) とし，閏年は考えないことにする．

また，計算上必要ならば次の近似式を用いてもよい．

$$n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \qquad \text{(Stirling の公式)}$$

(9) 1 から 13 までのスペードとハートのカードを別々に良く切っておく．教室にいる 13 人の生徒にスペードを 1 枚，ハートを 1 枚ずつ配ったとき，同じ数字のカードを誰かが持っている確率はいくつか？

(10) M 個の箱に無作為に N 個の玉を投げ入れたとき，同じ箱に 2 つ以上の玉が入る確率を求めよ．

ヒント

(7) 例えば (1) は

$$\begin{aligned}P(A_1+A_2+\cdots+A_n) &= P\left(A_1+(A_2+\cdots+A_n)\right)\\ &= P(A_1)+P(A_2+\cdots+A_n)\end{aligned}$$

となるので，加法性を順次適用すればよい．また (2) は

$$A=(A-B)+B$$

となることより容易に示すことができる．

(8) どちらも余事象を考えればよい．

a) ある人の誕生日が自分の誕生日と違う確率は $1-1/365$ である．

b) 異なる N 人の誕生日と $N+1$ 人目の誕生日がさらに異なる確率は $1-N/365$ である．

(9) スペードとハートの 2 種類のカードが配られているが，スペードを 1 から 13 の順に並べ直してハートの 1 から 13 の並べ替えだけを考えれば本質的には十分である．これは，例えば 13 人がプレゼントを持ってきて，無作為に交換したときに自分のプレゼントを受け取らない確率を求める問題と同じである．

さて，13 枚の場合を直接求めることは難しいが，N 枚の場合に一般化して以下のように考えると漸化式を求めることができる．

N 枚のカードを良く切って並べたとき，スペードとハートの数字が全て一致していなかったとする．この場合の数を C_N で表すことにする．例えば

♠	1	2	3	⋯	N
♡	3	1	N	⋯	2

♠	1	2	3	⋯	N
♡	3	N	1	⋯	2

のように並んでいるとしよう．ここで，ハートの N のカードと，スペードの N の位置にあるハートのカードを入れ替えたとする．上の場合

♠	1	2	3	⋯	N
♡	3	1	$\boxed{2}$	⋯	$\boxed{\text{N}}$

♠	1	2	3	⋯	N
♡	3	$\boxed{2}$	1	⋯	$\boxed{\text{N}}$

となる．このとき 1 から $N-1$ のカードの並びで起こりうる事象は次の 2 通りである．

- $N-1$ 枚のカードの並びが全て異なる (左の場合)．
- 入れ換えた場所でカードが一致し，それ以外の $N-2$ 枚のカードの並びが全て異なる (右の場合)．

もともと考えていた N 枚全てが一致していない並びで，ハートの N のカードが来る位置はスペードの 1 から $N-1$ の位置に対応する $N-1$ 通りであるので，

$$C_N=(N-1)(C_{N-1}+C_{N-2})$$

となる．C_1 と C_2 の場合の数を数え上げれば任意の N について場合の数を求めることができる．あとは，並べ替えの総数と余事象を考えればよい．

(10) これは (8b) の問題を一般化したものである ($M = 365$, $N = 20$)．事象 A を

$$A = \{1\text{ つの箱に }2\text{ つ以上の玉が入る}\}$$

とすると，その余事象の確率は場合の数を数え上げて

$$P(A^c) = \frac{(1\text{ つの箱に }2\text{ つ以上の玉が入らない場合の数})}{(M\text{ 個の箱に }N\text{ 個の玉を入れる場合の数})}$$

として計算すれば良いので，

$$\begin{aligned} P(A^c) &= \frac{\overbrace{M \times (M-1) \times \cdots \times (M-N+1)}^{M\text{ 個の別々の箱に }N\text{ 個の玉が入る}}}{\underbrace{M \times M \times \cdots \times M}_{M\text{ 個のいずれかの箱に }N\text{ 個の玉が入る}}} \\ &= \frac{M!}{M^N (M-N)!} \end{aligned}$$

である．ここで Stirling の公式を使うと A の余事象の確率は

$$P(A^c) \simeq \frac{1}{M^N} \cdot \sqrt{2\pi M} \left(\frac{M}{e}\right)^M \cdot \frac{1}{\sqrt{2\pi(M-N)}} \left(\frac{e}{M-N}\right)^{M-N}$$

$$= \frac{1}{e^N} \left(\frac{M}{M-N}\right)^{M-N+\frac{1}{2}}$$

と近似できる．その対数を考えると

$$\begin{aligned} \log P(A^c) &\simeq -N + \left(M - N + \frac{1}{2}\right) \log \frac{M}{M-N} \\ &= -N - \left(M - N + \frac{1}{2}\right) \log \left(1 - \frac{N}{M}\right) \end{aligned}$$

と計算されるので，$\log(1-x)$ は x が小さければ

$$\log(1-x) \simeq -x - \frac{x^2}{2}$$

と Taylor 展開により近似できるので，N が M に比べて十分に小さい場合には

$$\begin{aligned} \log P(A^c) &\simeq -N + \left(M - N + \frac{1}{2}\right) \left(\frac{N}{M} + \frac{1}{2}\left(\frac{N}{M}\right)^2\right) \\ &= -\frac{N^2}{2M} + \frac{N}{2M} + \frac{N^2}{4M^2} - \frac{N^3}{2M^2} \\ &= -\frac{N^2}{2M} + (\text{影響の小さい項}) \end{aligned}$$

として良い．ここで影響の小さい第 2 項以下を無視して N について解くと

$$N = \sqrt{M} \cdot \sqrt{-2\log P(A^c)} = \sqrt{M} \cdot \sqrt{-2\log(1 - P(A))}$$

となる．この式は事象 A が起こる確率をある値とするために必要な玉の数を箱の数から求める式となっている．$P(A)$ のいくつかの場合について $\sqrt{-2\log P(A^c)}$ を計算すると

$$\begin{array}{lll}
P(A) = 0.5 & P(A^c) = 0.5 & \sqrt{-2\log P(A^c)} = 1.18 \\
P(A) = 0.8 & P(A^c) = 0.2 & \sqrt{-2\log P(A^c)} = 1.79 \\
P(A) = 0.9 & P(A^c) = 0.1 & \sqrt{-2\log P(A^c)} = 2.15 \\
P(A) = 0.95 & P(A^c) = 0.05 & \sqrt{-2\log P(A^c)} = 2.45 \\
P(A) = 0.99 & P(A^c) = 0.01 & \sqrt{-2\log P(A^c)} = 3.03
\end{array}$$

となるので，例えば誕生日の問題 ($M = 365$[日]) では，

$$N = 1.18 \times \sqrt{365} = 22.5 \simeq 23[\text{人}]$$

の中に 0.5 の確率で

$$N = 3.03 \times \sqrt{365} = 57.9 \simeq 58[\text{人}]$$

の中に 0.99 の確率で同じ誕生日の人がいると概算できる．

一般の確率空間 2

2.1 集合の濃度

集合をその性質に従って分類する場合，含まれる要素の数によって分類することがある．この場合無限個の要素からなる集合でも，それが「数えられる無限個」なのか「数えられない無限個」なのかによって集合の性質が異なる．この性質の違いは「集合の濃度 (濃さ)」が違うと表現される．一般の確率測度を考える場合，見本点の数が無限個ある見本空間を対象としなくてはならないが，このとき無限個の質の違いを捉えておく必要がある．

確率を論じる時に注意すべき集合の濃度として以下のものがある．

有限の濃度 有限個の要素からなる集合

可算の濃度 自然数，有理数など

連続の濃度 無理数，実数など

より詳しくは位相と集合に関する成書を参照して欲しいが，以下では要素を無限個含む集合である有理数と無理数を取り上げて，可算と連続の性質の違いを考えてみることにする．

2.1.1 有理数の可算性

ある集合が**可算** (**可付番**) (enumerable, countable) であるとは，集合の全ての要素に自然数で順番に番号が与えられることである．集合の要素が有限個であれば可算であることは当たり前の話であるが，要素数が無限であっても番号付けができる集合がある．当然のことながら自然数は可算集合である．

練習問題 (1)

例 2.1. 区間 $(0,1)$ に含まれる全ての有理数が可算であるかどうか調べるには，有理数と自然数に 1 対 1 の対応 (全単射)，すなわち 1 つの自然数を与えるとそれに対応する有理数が，逆に有理数を与えるとそれに対応する自然数が求められる規則があるかどうかを調べればよい．次のような表を考えてみよう．

	1	2	3	4	5	6	$\cdots$
1		1/2	1/3	1/4	1/5	1/6	$\cdots$
2			2/3	*	2/5	*	$\cdots$
3				3/4	3/5	*	$\cdots$
4					4/5	*	$\cdots$
5						5/6	$\cdots$
$\vdots$							$\ddots$

表 2.1: 有理数の表.

表 2.1 は区間 $(0,1)$ の間の有理数を，分母を横に，分子を縦に，それぞれ小さいものから順に並べたものである．表中の * は約分できて既に同じ分数が存在しているものを表す．また，表 2.2 は

表 2.2: 番号付けの表.

	1	2	3	4	5	6	$\cdots$
1		1	2	4	6	10	$\cdots$
2			3		7		$\cdots$
3				5	8		$\cdots$
4					9		$\cdots$
5						11	$\cdots$
$\vdots$							$\ddots$

既約な分数に順番に番号を振っていったものである．必要に応じてこの表を大きくしていけば，どんな有理数もある 1 つの自然数に対応づけられることがわかる．ただし，表からは約分できるものを取り除かなくてはいけないので，この対応関係の規則を陽に表す計算式を作るのは意外と難しいことがわかる．

実は 1 対 1 の対応を示す代わりに，自然数全体から有理数の部分集合への 1 対 1 対応 (自然数から有理数への単射) と，有理数全体から自然数の部分集合への 1 対 1 対応 (有理数から自然数への単射) があることを言えば，有理数が可算であることが示される (*Bernstein* **の定理**，図 2.1 参照)．どちらの集合も要素の数は無限個なので，本来その数を比べることはできないが，直感的には以下のように解釈すればよい．まず自然数全体から有理数の一部へ 1 対 1 の対応関係があれば，有理数の方に余りがあるので有理数の要素数が自然数の要素数以上である

$$(\text{自然数の個数}) \leq (\text{有理数の個数})$$

といえる．また逆に有理数全体から自然数の一部へ 1 対 1 の対応関係があれば，自然数の方に余りがあるので自然数の要素数が有理数の要素数以上である

$$(\text{自然数の個数}) \geq (\text{有理数の個数})$$

といえる．したがって 2 つの不等式が成り立つためには等号が成立し，結局自然数と有理数ではその集合の要素の個数が一致していると考えることができる．

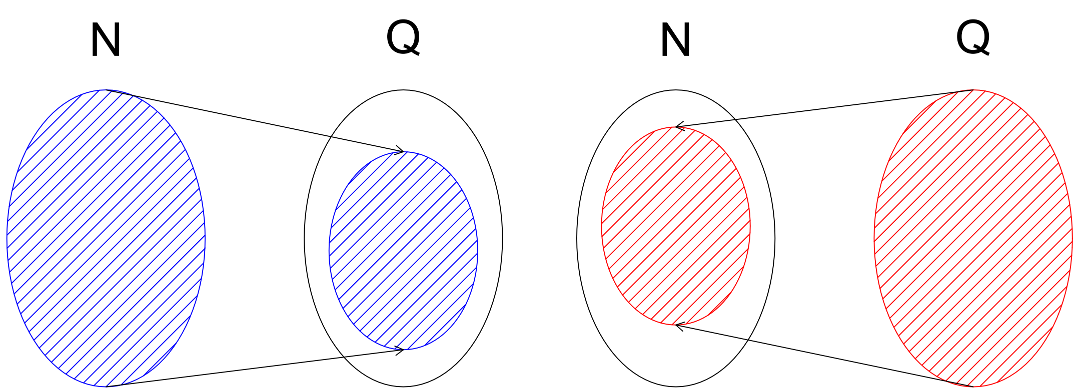

図 2.1: 自然数 ($\mathbb{N}$) と有理数 ($\mathbb{Q}$) の対応関係を考える Bernstein の定理の概念図.

実際このような対応関係は以下のようにして構成することができる．

まず自然数 $1, 2, 3, \ldots, n, \ldots$ に対して区間 $(0, 1)$ にある有理数のうち $1/2, 1/3, \ldots, 1/(n+1), \ldots$ を対応させる，すなわち

$$\text{自然数: } n \mapsto \text{有理数: } \frac{1}{n+1}$$

である．これによって全ての自然数は区間 $(0, 1)$ の中のどれか 1 つの有理数に対応づけられたことになる．この対応関係においては $2/3$ のように分子が 1 でない有理数は用いていないので，有理数の方にまだ余りがある．すなわち図 2.1 の左の状況で，直感的には

$$(\text{自然数の個数}) \leq (\text{有理数の個数})$$

であるといえる．

次に自然数と有理数を並べた 2 つの表を考える．

	1	2	3	4	5	6	⋯
1		1	2	4	7	11	⋯
2			3	5	8	12	⋯
3				6	9	13	⋯
4					10	14	⋯
5						15	⋯
⋮							⋱

表 2.3: 自然数の表．

	1	2	3	4	5	6	⋯
1		1/2	1/3	1/4	1/5	1/6	⋯
2			2/3	2/4	2/5	2/6	⋯
3				3/4	3/5	3/6	⋯
4					4/5	4/6	⋯
5						5/6	⋯
⋮							⋱

表 2.4: 重複ありの有理数の表．

表 2.4 の中には約分していない有理数，例えば $2/4 = 1/2$ や $2/6 = 1/3$ などが存在しているが，この表の中には全ての有理数が含まれている．表 2.3 の自然数の中には既約でなく重複している有理数の番号もあることになるが，全ての有理数に 1 つの自然数が規則的に割り当てられている．したがって，有理数の既約な表現を p/q とすれば，任意の有理数から自然数への対応関係は

$$\text{有理数: } \frac{p}{q} \mapsto \text{自然数: } \sum_{k=1}^{q-1}(k-1) + p = \frac{(q-1)(q-2)}{2} + p$$

と書くことができる．この場合，既約でない有理数の番号となる自然数は使われていないことになるので，

$$(\text{自然数の個数}) \geq (\text{有理数の個数})$$

となり，図 2.1 の右の状況であることがわかる．

以上二つの不等式から，結局

$$(\text{自然数の個数}) = (\text{有理数の個数})$$

であることがわかる．ただし，先にも書いたが要素数は無限個なので，この書き方はあくまで直感的なものであり，厳密には集合の濃度の定義を用いて示される．いずれにせよ上記のような 2 つの対応関係が構成できることから Bernstein の定理により区間 $(0,1)$ に含まれる有理数が可算であることが示される．

例 2.2. 整数が可算であることは，自然数と整数の間の非常に簡単な 1 対 1 の対応づけを用いて示してもよいが，Bernstein の定理により以下のような対応を考えて示すこともできる．自然数 n から整数 z への対応を単純な

$$n \mapsto z = n$$

とすれば，任意の n に対して異なる z が対応する．ただし，定義から非正の整数は使われずに残っているので，自然数全体から整数の部分集合への 1 対 1 対応となっている．一方，整数 z から自然数 n への対応を例えば

$$\begin{aligned} z \mapsto n &= \begin{cases} 5^{z} & z \geq 0 \\ 3^{-z} & z < 0 \end{cases} \\ &= (4 + \mathrm{sign}(z))^{|z|} \end{aligned}$$

とすれば，3 と 5 は互いに素なので任意の z に対して異なる n が対応する．ただし，$\mathrm{sign}(z)$ は z の符号が正なら 1，負なら -1 を返す関数である．ここで，自然数として使われているのは 3 と 5 の冪乗のみで，明らかに全ての自然数が使われているわけではない．また n の素因数分解を考えれば元の z がわかることに注意すれば，整数全体から自然数の部分集合への 1 対 1 対応となっていることがわかる．

練習問題 (2)

2.1.2 無理数の非可算性

一方区間 $(0,1)$ に含まれる無理数は可算ではない．可算でないことを**非可算** (unenumerable, uncountable; あるいは**不可算**) という．無理数の非可算性を証明するには背理法を用いるが，この証明法は Cantor の対角線論法と呼ばれる．

証明．区間 $(0,1)$ に含まれる無理数の集合が可算であると仮定する．有理数の集合は上で示したように可算であるから，仮定より無理数と有理数の和である実数も可算であり，自然数の番号付けができることになる．例えば有理数に付けられた番号を 2 倍して 1 を引き奇数の自然数を割り当て，一方無理数に付けられた番号を 2 倍して偶数の自然数を割り当てればよい．この番号付けに従って区間 $(0,1)$ に含まれる全ての実数を以下のように順番に並べたとする．

$$\begin{array}{ll} 1: & 0.\,d_1^1\,d_2^1\,d_3^1\,d_4^1\cdots \\ 2: & 0.\,d_1^2\,d_2^2\,d_3^2\,d_4^2\cdots \\ 3: & 0.\,d_1^3\,d_2^3\,d_3^3\,d_4^3\cdots \\ 4: & 0.\,d_1^4\,d_2^4\,d_3^4\,d_4^4\cdots \\ \vdots & \quad\vdots \end{array}$$

ここで d_j^i は第 i 番の実数の第 j 桁の数字 (0〜9) を表す．ただし，無理数は無限小数なので何桁目でも数字が存在するが，有理数は循環小数でなければ途中の桁で止まってしまうものもあるので，有限桁の場合は最後の桁以降を 0 で埋めることにする．このようにすると任意の番号 i の有理数に対して，その番号に対応する小数の第 i 桁の数字 d_i^i を定義することができる．そこで次のようにして 1 つの数を作る．まず自然数で付けた番号が 1 番の数の小数点以下 1 桁目の数字 d_1^1 を取り出し，それと異なる数字 $\tilde{d}_1^1(\neq d_1^1)$ を 1 つ選ぶ．次に 2 番の数の小数点以下 2 桁目の数字 d_2^2 を取り出し，それと異なる数字 $\tilde{d}_2^2$ を 1 つ選ぶ．3 番以降も同様にして順に数字を選ぶ．つまり上の表の対角線にあたる場所の数字を左上から順に取り出し，その数字と異なる数字を 1 つずつ選んでいき，これらを並べて作られる実数

$$0.\,\tilde{d}_1^1\,\tilde{d}_2^2\,\tilde{d}_3^3\,\tilde{d}_4^4\cdots$$

を考える．例えば規則的に $\tilde{d}_n^n = d_n^n + 1 \pmod{10}$ (1 を加えて 10 で割った余りを考える) などとしてもよい．この実数は区間 $(0,1)$ に含まれる数であるが，その作り方からわかる通り表に並べられた第 n 番の数とは第 n 桁目が違っている．すなわちこの数はもとの表にはない，番号が振られていない数ということになり，区間 $(0,1)$ の実数が可算であることに矛盾する．

以上より無理数が可算であるとした最初の仮定が間違っていたことがわかる． □

練習問題 (3)

2.2 一般の確率測度

有限試行の場合と同様に確率測度 P は見本空間 Ω の部分集合に作用する．しかしながら，その部分集合，すなわち事象は無条件に何でも使ってよい訳ではなく，適当な条件を満たすものを考えないと確率測度 P を矛盾なく定義することはできない．以下では確率測度 P がどんな性質をもたなくてはいけないかを簡単に説明する．

2.2.1 可測集合と σ-加法族

定義が前後するが確率測度 P が与えられたときに，確率測度 P で確率の値を測ることができる集合を**可測集合** (measurable set)，あるいは**可測事象** (measurable event) と呼ぶ．また，集合 A を確率測度 P で測ることができるとき，A は**可測** (measurable) であるという．言い換えると見本空間の部分集合の中で事象として考えてよいもの (確率を考える対象としてよいもの) が可測集合である．

以下では確率測度 P が評価する対象全体を $\mathcal{F}$ と書く．すなわち $\mathcal{F}$ は P の定義域である．確率測度 P は実数値集合関数 (集合に作用して実数を出力する関数) であるから，定義域 $\mathcal{F}$ は，確率測度 $P(A)$ がきちんと定義される事象の全体，すなわち見本空間の部分集合 $A \subset \Omega$ (可測集合) の全体を表す，いわば集合の集合であり，これを**集合族** (family of sets) という．

確率測度として P をうまく定義するためには，その定義域 $\mathcal{F}$ は以下の三つの条件を満たさなくてはならない．

$$(\sigma.1) \qquad \Omega \in \mathcal{F}$$
$$(\sigma.2) \qquad A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$$
$$(\sigma.3) \qquad A_n \in \mathcal{F},\ (n = 1, 2, \dots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$$

それぞれの意味するところは

(σ.1) 見本空間 (全事象) は可測である (P で測ることができる)．

(σ.2) ある事象が可測なら，その余事象も可測である．

(σ.3) 可測集合の可算 (自然数で番号付けできる) 無限和も可測である．

となる．特に大事なものは (σ.3) である．

定義 2.3 (σ-加法族)**.** (σ.1)-(σ.3) の 3 つの条件を満たす集合族を σ-**加法族** (σ-algebra) と呼ぶ．

また確率測度 P の取る値については以下の条件が必要となる．

$$(\text{P.1}) \qquad P(A) \geq 0,\ A \in \mathcal{F}$$
$$(\text{P.2}) \qquad P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n),\ A_n \in \mathcal{F}$$

(σ-**加法性**; σ-additivity)

$$(\text{P.3}) \qquad P(\Omega) = 1$$

(P.2) は，可算無限個の排反事象の和の確率がそれぞれの事象の確率の和となることを意味しており，それ以外の条件の意味するところは有限事象と同じである．加法性を可算無限個の和に拡張したところに重要な意味がある．

定義 2.4 (確率測度)**.** 集合関数 P は条件 (P.1), (P.2) を満たすとき**測度** (measure) と呼ばれ，さらに (P.3) まで満たすとき**確率測度** (probability measure) と呼ばれる．また $P(A)$ を A の**測度**という．

この確率測度を用いて一般の見本空間における確率空間は定義される．

定義 2.5 (確率空間)**.** 見本空間 Ω と確率測度 P，および P の定義域である σ-加法族 $\mathcal{F}$ の組 $(\Omega, \mathcal{F}, P)$ を**確率空間** (probability spcae) という．

2.2.2 Lebesgue 測度

次に応用上最も重要な確率測度を定義するために，以下の例を考えてみよう．

例 2.6. 試行 T を「区間 $(0,1]$ から無作為に一点抜き出す」こととする (無作為はでたらめと言い換えてもよい). このとき見本空間は

$$\Omega = (0,1]$$

であり，Ω に含まれる見本点は無数にあるので無限試行となる. 確率測度は事象 A が区間 $[a,b]$, (a,b), $[a,b)$ または $(a,b]$ といった簡単な集合 (確率測度が開区間か閉区間かによらないことは後ほどわかる) であれば

$$P(A) = |A| = b-a, \ (|A| \text{ は区間の長さを表す})$$

となる. 例えば $P([0.5,0.6])$ は

> 「区間 $(0,1]$ から無作為に抜き出した点が区間 $[0.5,0.6]$ の間に入っている確率」

であるが，Ω のどの点が抜き出されるかはでたらめで同じ可能性があるのでこれは区間全体の長さに対して注目している区間の長さの比で評価され

$$P([0.5,0.6]) = \frac{(\text{注目している区間の長さ})}{(\text{区間全体の長さ})} = \frac{|0.6-0.5|}{|1-0|} = 0.1$$

となる. 同様にして $P((0.1,0.11)) = 0.01$ などとなる.

容易にわかるようにこの「区間 $(0,1]$ から無作為に一点抜き出す」試行は「ルーレット回し」の試行と等価である.

さて，上の例で考えた事象 A はもっと複雑なものを考えることもでき，可算個の適当な区間を組み合わせてつくられるどのような集合であってもよい. 試行 T においては，こうした集合演算で作られるいろいろな集合を集めたものが確率測度 P の定義域 $\mathcal{F}$ となる. 直感的には

$$\mathcal{F} = \{\text{可算個の任意の区間の和，差，交，余集合から作られる集合}\}$$

である. こうした集合の測度 (確率) は (P.2) の性質を用いることによって，排反な区間の長さの和から計算できる.

定義 2.7 (Lebesgue 測度)**.** 「区間 $(0,1]$ から無作為に一点抜き出す」試行 T によって考えられる確率測度 P を $(0,1]$ 上の *Lebesgue* **測度** (Lebesgue measure) という. また Lebesgue 測度の定義域となる σ-加法族を $\mathbb{R}$ の *Borel* **集合族** (Borel field) という. [3]

[3] Lebesgue 可測でない集合 (Borel 集合族に含まれない集合) というのは，非常に特殊であるが存在する. しかしながら，応用上現れる集合のほとんどは可測と考えて差し支えない.

Borel 集合や Lebesgue 測度の厳密な定義については，Lebesgue 積分の教科書などを参考にしてほしい.

例 2.8. 「区間 $(0,1]$ から無作為に一点抜き出す」試行，すなわち Lebesgue 測度において，$P\{a\}$(一点 a が抜き出される確率) は 0 となる. 何故なら σ-加法性より，この値が少しでも 0 より大きいと全確率が 1 を越えてしまうからである. 例えば $\{1, 1/2, 1/3, \dots\}$ という可算集合を考えると

$$\begin{aligned} &P\left\{\frac{1}{n}\right\} \geq \varepsilon > 0 \\ &\quad \Rightarrow P\left(\left\{1, \frac{1}{2}, \frac{1}{3}, \dots\right\}\right) = \sum_{n=1}^{\infty} P\left\{\frac{1}{n}\right\} \geq \sum_{n=1}^{\infty} \varepsilon \to \infty \end{aligned}$$

となり，(P.3) を満たさない．同様にして可算個の点からなる事象の確率は Lebesgue 測度においてはいずれも 0 となる．例えば

$$P\left(\left\{1, \frac{1}{2}, \frac{1}{2^2}, \dots, \frac{1}{2^n}, \dots\right\}\right) = 0$$

である．[4]

例 2.9. 「区間 $(0,1]$ から無作為に一点抜き出す」試行において，抜き出した点が有理数である確率を考える．以下では

$$\begin{aligned}\mathbb{R}_{(0,1]} &= \{\text{区間 } (0,1] \text{ 上の実数全体}\}(= \Omega)\\ \mathbb{Q}_{(0,1]} &= \{\text{区間 } (0,1] \text{ 上の有理数全体}\}\end{aligned}$$

と書くことにする．有理数は可算であるからその要素に番号が付けられ，

$$\mathbb{Q}_{(0,1]} = \{q_1, q_2, q_3, \dots, q_n, \dots\}$$

のように書ける．確率測度の σ-加法性 ((P.2) の性質) によれば

$$\begin{aligned}P(\mathbb{Q}_{(0,1]}) &= P\left(\{q_1, q_2, q_3, \dots, q_n, \dots\}\right)\\ &= P\{q_1\} + P\{q_2\} + P\{q_3\} + \cdots + P\{q_n\} + \cdots\\ &= 0 + 0 + 0 + \cdots + 0 + \cdots\\ &= 0\end{aligned}$$

となる．一方，無理数である確率は，

$$(\text{区間 } (0,1] \text{ 上の無理数全体}) = \mathbb{R}_{(0,1]} - \mathbb{Q}_{(0,1]}$$

を用いて

$$P(\mathbb{R}_{(0,1]} - \mathbb{Q}_{(0,1]}) = P(\mathbb{R}_{(0,1]}) - P(\mathbb{Q}_{(0,1]}) = 1 - 0 = 1$$

と求めることができる．無理数全体は可算でないため有理数のように成分毎の可算無限和では書けず，したがって (P.2) の性質を使って計算することはできないことに注意する．

例 2.10. 「区間 $(0,5]$ から無作為に一点抜き出す」試行を考えたとき，その見本空間は $\Omega = (0,5]$ となる．この試行の確率測度は Lebesgue 測度を定数倍 (正規化) し，

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{\mu(A)}{5}, \quad A \in \mathcal{F}$$

とすることによって構成できる．

例 2.11 (Buffon の針; Buffon's needle)**.** 図 2.2 のように 2 次元平面上に等間隔 d で平行線が引いてある．長さ l の針をこの平面上にランダム (でたらめ) に落としたとき，平行線と交わる確率を求めることを考える．ただし $l < d$ とする．針の中心位置と一番近い平行線を原点とし，水平方向の座標を x とする．x は区間 $[-d/2, d/2]$ で一様に分布する (どの点が得られるかは無作為) と考えられる．また針に向きがあるとして，針が図の水平方向となす角度を θ と

[4] Lebesgue 測度はもともと確率測度として定義されたものではなく，実数上の任意の区間の集まりの長さの総和を表す一般の測度として考えられたものである．その定義は上で考えた区間 $(0,1]$ 上の Lebesgue 測度をそのまま実数全体に広げたものと思えばよい．Lebesgue 測度は通常 μ で表され，例えば $\mu((0,3]) = 3$ のように考える区間によっては 1 を越える．見本空間を $\Omega = (0,1]$ としたときには $\mu(\Omega) = 1$ となるので，この場合は確率測度になっている．

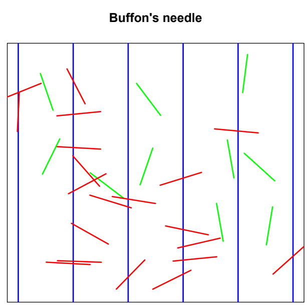

図 2.2: 平行線と交わった針を赤，それ以外を緑で表している．

する．θ も同様に区間 $[0, 2\pi]$ で一様に分布すると考えられる．この試行の見本点は (x, θ) で表され，その見本空間は

$$\Omega = [-d/2, d/2] \times [0, 2\pi] \subset \mathbb{R}^2$$

となる．確率測度とその定義域となる σ-加法族は，Lebesgue 測度を 2 次元に自然に拡張したものを考えればよい．すなわち，$\mathbb{R}^2$ の Borel 集合族は 1 次元の場合の区間を 2 次元の矩形領域に置き換え

$$\mathcal{F} = \{\text{可算個の矩形領域の和，差，交，余集合から作られる集合}\}$$

によって定義し，確率測度は長さを面積に置き換え

$$P(A) = \frac{(A\text{ の面積})}{(\Omega\text{ の面積})} = \frac{(A\text{ の面積})}{2\pi d},\ A \in \mathcal{F}$$

として計算すればよい．以上より，この試行の確率空間 $(\Omega, \mathcal{F}, P)$ が構成される．

さて，針が平行線と交わる条件は，針の両端の座標の符号が異なること

$$\left(x + \frac{l}{2}\cos\theta\right)\left(x - \frac{l}{2}\cos\theta\right) < 0$$

であり，この条件を満たす (x, θ) の領域を図示すると図 2.3 のようになる．

したがって

$$P(\text{針が平行線と交わる}) = \frac{4l}{2\pi d} = \frac{2l}{\pi d}$$

となる．

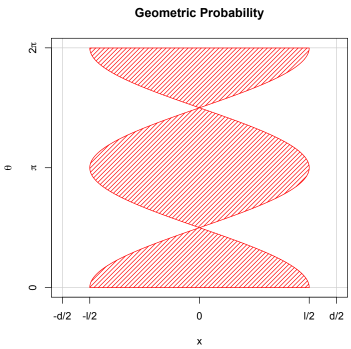

図 2.3: 見本空間と条件 (平行線と交わる) を満たす領域．

練習問題 (4) (5) (6) (7) (8)

2.2.3 零集合

例にあげたいくつかの事象のように，事象 A の確率が 0 であるからといって，事象 A に含まれる見本点が全く起こり得ない訳ではない．確率 0 であるのに起こらない訳ではないというのは一見矛盾しているように思えるが，これは確率の定義を場合の数を数え上げられる有限，あるいは可算濃度の見本空間から，連続濃度の見本空間上に拡大したためである．区間 $(0, 1]$ からのでたらめな抜き出しの場合どの点も同様に抜き出されるので，ある一点を引く確率は他のどれかの点を引く確率と同じである．例えば $1/2$ と $1/\sqrt{2}$ を比べた場合，どちらかが抜き出される確率は同じで，有理数だから無理数より抜き出され難いなどということはない．確率 0 となるのはある意味で比較の問題であり，ある一点が抜き出される確率はそれ以外のどれかの点が抜き出される確率に比べて遥かに小さく，比率としては $0 : 1$ となると考えることができると言っているのである．これは有理数あるいは無理数がそれぞれ抜き出される事象を比べた場合も同様で，直感的には無理数の個数が有理数の個数に比べて圧倒的に多いということを表している．

このように一般の試行を考えた場合，空集合でない集合 A でその事象が起こる確率が $0(P(A)=0)$ となるものがある．こうした集合を**零集合** (null set) と呼ぶ．

また，確率・統計では以下のような特殊な言い回しをすることがある．事象 $A=\{\omega|\alpha(\omega)\}$ の確率が 1 であるとき，「**ほとんど確実に** (almost surely) 条件 $\alpha(\omega)$ **が成り立つ**」，あるいは「条件 $\alpha(\omega)$ **が確率** 1 **で成り立つ**」といい，

$$\alpha(\omega) \text{ a.s.}$$

と書く．これは条件 $\alpha(\omega)$ の余事象 A^c が零集合となると言い換えてもよい．条件 $\alpha(\omega)$ は成り立たないこともある (A^c は空でない) が，それの起こる確率が 0 であり，直感的には，成り立たない場合は成り立つ場合に比べて圧倒的に少ないことを意味している．

まとめ

- 有限集合と無限集合
- 集合の濃度
- 可算集合: 自然数，整数，有理数
- 連続集合: 無理数，実数
- 一般の確率測度と σ-加法性
- Lebesgue 測度と Borel 集合族
- 「ほとんど確実に成り立つ」

練習問題

(1) 整数が可算であることを示せ．

(2) 正の有理数全体が可算であるかどうか論じよ．

(3) 以下の集合 V のうち，可算集合はどれか？
- 区間 $(0,1)$ に含まれる有理数．
- 区間 $(0,1)$ に含まれる無理数．
- 区間 $(0,1)$ に含まれる実数で 1/整数 となるもの．
- 正の有理数．
- 実数 $\mathbb{R}$(区間 $(-\infty,\infty)$)
- 実数 $\mathbb{R}$(区間 $(-\infty,\infty)$) 上の有理数．
- 2 つの実数の組 $(x,y)\in\mathbb{R}^2$
- 2 次元空間において xy 座標がともに整数となる点の集合．
- 1 の目が出るまでサイコロを投げ続ける試行の見本空間．
- 表が出るまでコインを投げ続ける試行の見本空間．

(4) 図 2.4 のように二次元平面上に等間隔 d で平行線が引いてある．長さ l の 2 本の針が，その中点で直角に交わった十字形を，この平面上に位置，角度ともにランダムに落とす．ただし $l < d$ とする．

a) 十字形が平行線と交わる確率を求めよ．

b) 十字形が平行線と一点だけで交わる確率を求めよ．

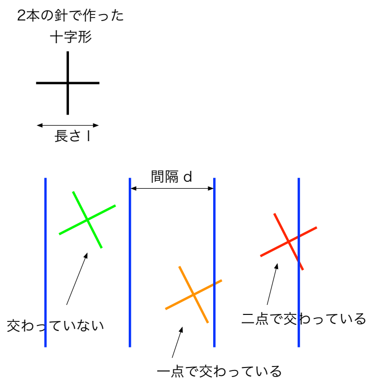

図 2.4: Buffon の針 (十字針の場合).

(5) 二次元平面上に等間隔 d の平行線で作られた白黒の市松模様がある．長さ l の針をこの平面上に位置，角度ともにランダムに落とす．ただし $l < d$ とする．このとき針が黒と交わる確率を求めよ．

(6) 正方形のタイルを敷き詰めた床面に，1 個のコインを投げる試行を考える．図 2.5 は投げたコインが 3 枚のタイルにまたがって含まれている状態を表している．タイルは床面の十分大きな領域に敷き詰められており，コインは必ずタイルの上に落ちるとする．また，コインの落ちる位置はタイル上で無作為 (ランダム) と考えてよいものとする．タイルの一辺を l[cm]，コインの直径を r[cm]，また $l > r$ であるとして以下の問に答えなさい．

a) コイン全体が 1 枚のタイルの中に完全に含まれる確率はいくつか？

b) コインが 2 枚のタイルにまたがって含まれる確率はいくつか？

c) コインが 3 枚のタイルにまたがって含まれる確率はいくつか？

d) コインが 4 枚のタイルにまたがって含まれる確率はいくつか？

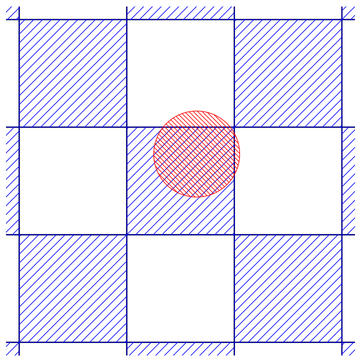

図 2.5: コイン投げの問題 (3 枚のタイルに含まれる場合).

(7) 床に 4cm×8cm のタイルを隙間なく敷き詰め，直径 2cm のコインを無作為に投げる．以下の問いに答えよ．

a) コインが 1 つのタイルに完全に含まれる確率を求めよ．

b) コインが 2 つのタイルのみに跨 (また) がる確率を求めよ．

c) コインが 4 つのタイルに跨がる確率を求めよ．

(8) 長さが d[m] で電荷 $\pm Q$[C] を両端に持つ双極子が，xy 平面上にランダム (双極子の中心座標が xy 平面で一様に分布) に並んだ 2 次元の誘電体 (双極子の向きを表す $-Q$ から $+Q$ に向かうベクトルが xy 平面に含まれている) を考える．電界がない状態では双極子の向きは xy 平面内でランダム (x 軸となす角度が一様に分布) であるが，電界が加えられると双極子の中心を軸として xy 平面を回転して一斉に電界に沿って並ぶものとする．

電界のない状態から x 軸に平行に左から右向きに電界を加えたとする．このとき以下の問いに答えなさい．

a) 電界を加えたとき，中心の x 座標が a $(-d/2 \le a \le d/2)$ である双極子の正電荷が y 軸を左から右に横切る確率を求めなさい．

b) 電界を加えたとき，中心の x 座標が区間 $[-d/2, d/2]$ にある双極子の正電荷が y 軸を左から右に横切る確率を求めなさい．

c) 単位面積あたりの双極子の数を ρ[個/m^2] とする．電界を加えたとき，y 軸の単位長さあたりを横切る電荷 (正も負もあることに注意) を求めなさい．

ヒント

(4) 一本の針の問題を参考に，1点または2点で交わる場合を分けて図示すればよい．

(5) 縦・横・回転の3つの自由度を考えて確率空間を構成すればよい．

(6) コインの中心位置を適当に場合分けして確率空間を構成すればよい．

(8) 実在の双極子は3次元空間で考えなくてはいけないので繁雑であるが，2次元の場合は比較的簡単に確率空間を作ることができる．

確率密度 3

3.1 確率測度と確率密度

実データを扱う工学的な応用においては，見本空間は抽象的なものではなく，実数上に値を取るものを考えることが一般的である．このとき確率測度を具体的に表現するためには，本章で述べる確率密度による積分表現を用いることになる．

3.1.1 積分による確率測度の表現

似たような事象であっても観測される確率が異なる試行として，以下の例を考えてみよう．

例 3.1 (歪んだ目盛りのルーレット)**.** 起点 0 から時計回りに周に沿って進んだ長さ l の位置の目盛りの値を

$$x = l^2$$

としてルーレットを作る．このルーレットでは，例えば目盛りの指す 2 つの区間 $[0.1, 0.2]$ と $[0.8, 0.9]$ に対応する弧の長さが異なる．このため，それぞれの区間でルーレットが止まる確率も異なり，弧の長さを考えると

$$P([0.1, 0.2]) > P([0.8, 0.9])$$

となる．これから，これまで扱ってきた単純なルーレットとは異なる確率法則に従うことがわかる．歪んだ目盛りの指す区間 $[a, b]$ は，起点からの周長で表した区間 (これまでのルーレットの目盛り) では $[\sqrt{a}, \sqrt{b}]$ に対応するので，

$$P([a, b]) = \sqrt{b} - \sqrt{a}$$

となる．

この問題では，確率測度を具体的に計算することができるが，もう少し単純な例を考えて確率測度を表現する別の方法を考えてみることにする．

例 3.2. 「$\Omega = [0, 1]$ の点を異なる確率で抜き出す」という試行を考える．[5] 異なる確率とは，例えば区間 $[0.5, 1]$ から点が抜き出される事象と，区間 $[0, 0.5)$ から点が抜き出される事象を比較すると，前者の方が後者より 3 倍起こりやすいといった状況を考えてみる．「ルーレット回し」の試行で同様なことを考えるとすれば，ルーレットの目盛りの振り方を均等ではなく，1/4 の区間に $(0, 0.5]$ を，残りの区間に $(0.5, 1]$ を割り当てるなどすればよい．このとき

$$f(x) = \begin{cases} 0.5, & (0 \leq x < 0.5) \\ 1.5, & (0.5 \leq x \leq 1) \end{cases}$$

[5] この問題では 1 点 0 を抜き出す確率は $P\{0\} = 0$ なので，「ルーレット回し」の試行における見本空間 $\Omega = (0, 1]$ とほぼ同様に見本空間 $\Omega = [0, 1]$ の確率法則を考えればよい．

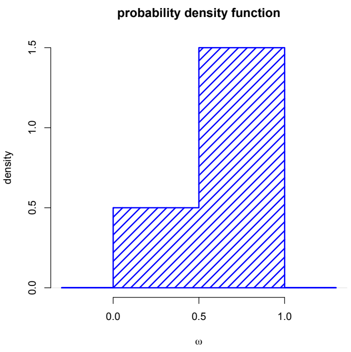

図 3.1: 重みの付いた抜き出しの確率密度.

とすると (図 3.1 参照)，この試行において抜き出される点が $A \subset \Omega$ に含まれる確率測度は

$$P(A) = \int_A f(\omega) d\omega$$

と書くことができる.

実際，このように定義した積分が確率測度の条件を満たしているかどうかは，以下のように簡単に確かめることができる.

(P.1) $f(x)$ の正値性から明らか,

(P.2) 積分区間を分割しても積分値が変わらないという積分の性質から明らか,

(P.3) 区間 $[0, 1]$ 上の積分

$$P([0,1]) = \int_0^1 f(\omega) d\omega = 0.5 \times 0.5 + 0.5 \times 1.5 = 1$$

より全確率が 1 となっている.

また同じ長さの区間でも 0.5 より小さい場合と大きい場合では，例えば

$$P([0.1, 0.2]) = \int_{0.1}^{0.2} f(\omega) d\omega = [0.5\omega]_{0.1}^{0.2} = 0.05$$
$$P([0.7, 0.8]) = \int_{0.7}^{0.8} f(\omega) d\omega = [1.5\omega]_{0.7}^{0.8} = 0.15$$

のように確率測度は 3 倍になっている．以上より，積分によってこの試行の確率測度が表現できることがわかる.

定義 3.3 (確率密度)**.** 確率測度がある関数の積分

$$P(A) = \int_A f(\omega) d\omega$$

で書かれているとき，その被積分関数を**確率密度** (probability density) または**確率密度関数** (probability density function) という. また確率密度を持つ確率測度を**絶対連続** (absolutely continuous) な分布と呼ぶ. [6]

[6] 別の記述の仕方では

$$P(A) = \int_A f(\omega) d\omega = \int_A P(d\omega)$$

のように書くことがある．この記法は後述する平均値などの計算でも用いることがある.

例 3.4 (歪んだ目盛りのルーレット)**.** 最初に挙げた特殊なルーレットの問題における確率密度は，任意の a, b $(0 < a < b \leq 1)$ に対して

$$P([a,b]) = \int_a^b f(\omega) d\omega = \sqrt{b} - \sqrt{a}$$

となる関数 f なので,

$$f(\omega) = \begin{cases} \frac{1}{2\sqrt{\omega}}, & (0 < \omega \leq 1) \\ 0, & (\text{それ以外}) \end{cases}$$

となる.

3.1.2 Riemann 積分と Lebesgue 積分

上の例では積分については特別な注意を払わなかったが，この積分は正確には Lebesgue 積分というものである．通常の積分は Riemann 積分と呼ばれるもので，これらはそれぞれ以下のように定義される．

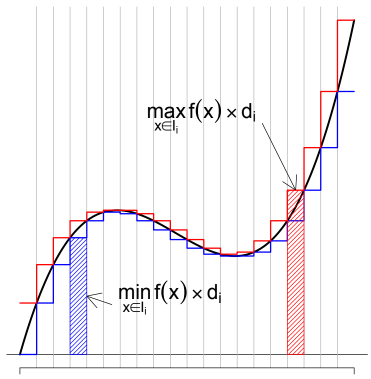

図 3.2: Riemann 積分の概念図.

定義 3.5 (Riemann 積分). 有界な積分領域 K を n 個の区間 I_i に分割 (等分でも，そうでなくても良い) する．区間 I_i の幅を d_i とし，各区間における関数 $f(x)$ の最小値と最大値を用いて有限和 s_n,S_n を次のように定義する (図 3.2 参照).

$$s_n = \sum_{i=1}^{n} \min_{x \in I_i} f(x) \times d_i$$
$$S_n = \sum_{i=1}^{n} \max_{x \in I_i} f(x) \times d_i$$

各区間の大小関係から

$$s_n \leq S_n$$

が成り立っているが，分割を細かくしたとき，それぞれの極限が存在して，なおかつ一致するならば，これを $f(x)$ の *Riemann* **積分** (Riemann integral)

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} S_n = \int_K f(x)dx$$

という．

直観的には上側と下側をそれぞれ階段関数で近似したとき，上下からから挟み込んでいくらでも精度よく表すことができる滑らかな関数に対して定義されるのが Riemann 積分である．なお，定義域の中に階段状の有界な飛びがあっても，積分区間をこの飛びで分割したとき各区間で積分値が確定すれば全体の積分も定義できるので，区分的に滑らかであれば十分である．

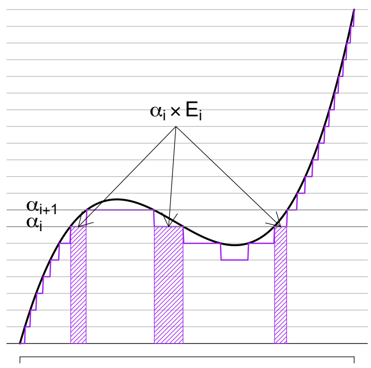

図 3.3: Lebesgue 積分の概念図.

定義 3.6 (Lebesgue 積分). 区間 K 上で定義された関数 $f(x)$ の値域を n 個の区間に分割 (等分でも，そうでなくても良い) する．区間の端点を α_i とし，$f(x)$ が区間 $[\alpha_i, \alpha_{i+1})$ で値を取る x の領域を $E_i \subset K$ で表すこととし，各 E_i 上で値が α_i となる階段関数 $f_n(x)$(**単関数**という) を考える (図 3.3 参照)．このとき μ を Lebesgue 測度とし，単関数 $f_n(x)$ の Lebesgue 積分を

$$\int_K f_n(x)dx = \sum_{i=1}^{n} \alpha_i \mu(E_i)$$

で定義する．分割を細かくしたとき $f_n(x)$ は $f(x)$ に近づいていくが，この極限が存在するならば，これを $f(x)$ の *Lebesgue* **積分** (Lebesgue integral)

$$\lim_{n\to\infty} \int_K f_n(x)dx = \int_K f(x)dx$$

という．

なお，二つの積分を明示的に区別するために，Lebesgue 積分を

$$\int_K f(x)\mu(dx), \quad (\mu \text{ は Lebesgue 測度})$$

と書くこともある．
上の定義に現れる集合 E_i が区間の可算回の演算 (この場合は和) で表すことができるのであれば Lebesgue 測度で測ることができる事象 (Lebesgue 可測) となるので，Lebesgue 積分可能かどうかは関数値が同じ値となる点の集合 E_i の性質による．
2 つの積分の違いをみるために次のような例を考えてみる．

例 3.7. $[0,1]$ 上で次のような関数

$$f(x) = \begin{cases} 1, & x \text{ が有理数} \\ 0, & x \text{ が無理数} \end{cases} \tag{3.1}$$

を考える．[7] どんなに短い区間 I_i の中にも有理数と無理数は混在するので，

$$\min_{x \in I_i} f(x) = 0, \quad \max_{x \in I_i} f(x) = 1$$

したがって

$$s_n = \sum_{i=0}^{n} 0 \times d_i = 0$$
$$S_n = \sum_{i=0}^{n} 1 \times d_i = 1$$

となり，n をいくら大きくしてもその極限は一致しない．したがってこの関数の Riemann 積分は存在しない．一方適当な値域の分割において

$$\alpha_i \leq 0 < \alpha_{i+1}, \quad \alpha_j \leq 1 < \alpha_{j+1}$$

となるのであれば，定義より

$$\int f_n(x)dx = \alpha_i \times \mu(\text{無理数}) + \alpha_j \times \mu(\text{有理数}) = \alpha_i \times 1 + \alpha_j \times 0 = \alpha_i$$

となるが，$n \to \infty$ のとき $\alpha_i \to 0$, $\alpha_j \to 1$ であるので

$$\int f(x)dx = 0 \times \mu(\text{無理数}) + 1 \times \mu(\text{有理数}) = 0$$

となり Lebesgue 積分可能である．

[7] この関数は Dirichlet の関数と呼ばれ

$$f(x) = \lim_{n\to\infty} \lim_{k\to\infty} \cos^{2k}(n!\pi x)$$

と書けることが知られている．

練習問題 (1)

上の例でわかるように，Riemann 積分と Lebesgue 積分は一般には一致しない．しかしながら適当な条件を満たせば二つの積分は一致する．二つが一致する典型的な条件としては以下のものが重要である．

定理 3.8. *有界閉区間 K で f が連続なら，K 上の f の Lebesgue 積分の値は Riemann 積分の値と一致する．*

したがって応用上重要な問題に現れる「あまり特殊でない関数」ではほとんど一致すると考えてよく，普通の意味での積分 (Riemann 積分) の方法をそのまま使って計算することができる．ただし，定義域が有界でない場合は積分値の取り扱いを注意しなくてはいけない場合もある．より詳しくは微積分や Lebesgue 積分の成書を参照されたい．

3.2 確率密度関数の例

Lebesgue 積分を用いて確率密度を定義したが，よく用いられる確率測度の密度関数を以下に例示する．

練習問題 (2) (3) (4) (5)

3.2.1 正規分布

平均 μ，分散 σ^2 の**正規分布** (Gauss 分布; normal distribution, Gaussian distribution) の密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
$$(-\infty < x < \infty;\ -\infty < \mu < \infty,\ \sigma > 0)$$

で表される (図 3.4 参照)．見本空間は $(-\infty, \infty)$ である．特に $\mu = 0, \sigma = 1$ のとき**標準正規分布** (標準 Gauss 分布; standard normal distribution, standard Gaussian distribution) と呼ぶ．

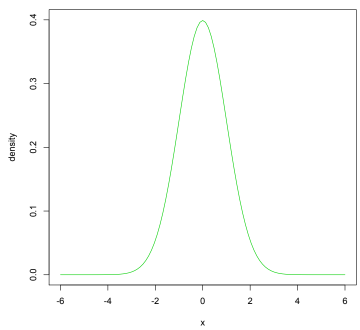

図 3.4: 確率密度関数: 正規分布 (平均 0, 分散 1).

3.2.2 一様分布

区間 $[a, b]$ を見本空間とする**一様分布** (uniform distribution) の密度関数は

$$\frac{1}{b-a} \quad (b \le x \le a)$$

で表される (図 3.5 参照)．なお，定義域に端点を含まない場合も同様に定義される．

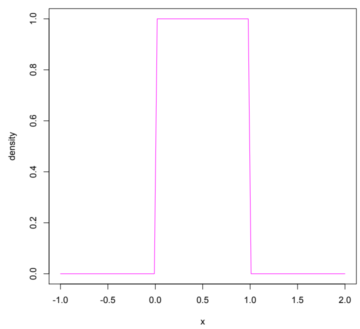

図 3.5: 確率密度関数: 一様分布 (区間 $[0, 1]$).

3.2.3 Cauchy 分布

Cauchy **分布** (Cauchy distribution) の密度関数は

$$f(x) = \frac{1}{\pi\sigma}\left(1 + \frac{(x-\mu)^2}{\sigma^2}\right)^{-1}$$
$$(-\infty < x < \infty;\ -\infty < \mu < \infty,\ \sigma > 0)$$

で表される (図 3.6 参照)．2 つの母数 (パラメタ) を持ち，μ は位置母数 (location parameter)，σ は尺度母数 (scale parameter) と呼ばれ，それぞれ分布の中央値 (それより大きい値と小さい値の出る確率がそれぞれ 1/2 である点) と拡がりを規定する．見本空間は $(-\infty, \infty)$ であり，裾の重い分布の典型として用いられることが多い．

2 つの量 Y, Z が標準正規分布に従うとき，そのの比 $X = Y/Z$ は $\mu = 0,\ \sigma = 1$ の Cauchy 分布に従う．

Cauchy 分布は平均値や分散が定義されない分布の例としても有名である．なお，平均値，分散については後節で定義する．

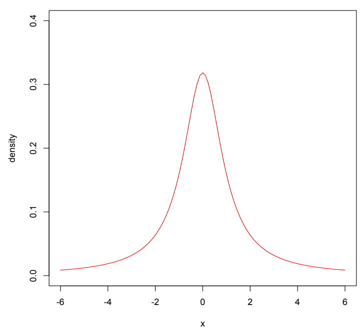

図 3.6: 確率密度関数: Cauchy 分布 (中央値 0, スケール 1).

3.2.4 χ^2-分布

ν 個の量 $X_1, X_2, \ldots, X_\nu$ が標準正規分布に従うとき，その 2 乗和

$$Y = X_1^2 + X_2^2 + \cdots + X_\nu^2$$

の分布を自由度 ν の χ^2-**分布** (χ^2 distribution) といい，その密度関数は見本空間 $[0, \infty)$ 上で

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} x^{\nu/2-1} e^{-x/2} \qquad (0 \le x < \infty)$$

と表される (図 3.7 参照)．ただし $\Gamma(z)$ は Γ 関数で

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$$

で定義される．この分布は適合度や分散などの検定において利用される．

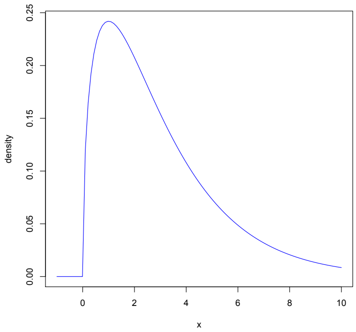

図 3.7: 確率密度関数: χ^2-分布 (自由度 3).

3.2.5 t-分布

標準正規分布に従う 1 つの量 Z と，それとは独立な ν 個の標準正規分布に従う量 $X_1, X_2, \ldots, X_\nu$ の 2 乗和

$$Y = X_1^2 + X_2^2 + \cdots + X_\nu^2$$

で定義される量

$$X = \frac{Z}{\sqrt{Y/\nu}}$$

の分布を自由度 ν の t-**分布** (Student's t distribution) といい，その密度関数は見本空間 $(-\infty, \infty)$ 上で

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)} \qquad (-\infty < x < \infty)$$

と表される (図 3.8 参照)．この分布は平均値などの検定において利用される．

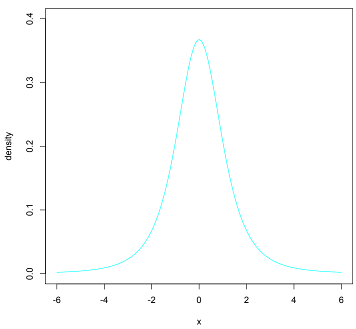

図 3.8: 確率密度関数: t-分布 (自由度 3).

3.2.6 指数分布

指数分布 (exponential distribution) は見本空間が $[0, \infty)$ 上で定義される確率分布で，比率 (rate) とよばれる 1 つの母数 (パラメタ)λ を持ち，その密度関数は

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \quad (0 \le x < \infty;\ \lambda > 0)$$

で表される (図 3.9 参照)．タクシーなどの待ち時間をモデル化する基本的な分布として利用されることが多い．

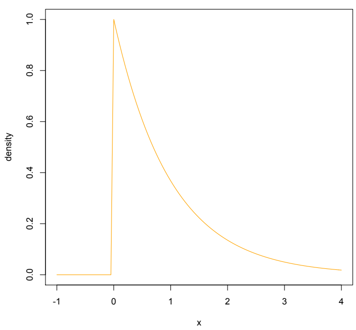

図 3.9: 確率密度関数: 指数分布 (比率 1).

3.2.7 確率密度がない分布の例

もちろん全ての確率測度に確率密度が存在する訳ではないので，以下にいくつか密度がない例も挙げておく．

例 3.9 (δ 分布)**.** 確率測度を

$$P(A) = \delta_a(A) = \begin{cases} 1, & a \in A \\ 0, & a \notin A \end{cases}$$

で定義する．[8] これは

$$(a \text{ が抜き出される確率}) = 1$$
$$(\Omega - a \text{ の点が抜き出される確率}) = 0$$

であるから，Ω から一点抜き出すとそれはほとんど確実に a であるような抜き出しを考えていることになる．

[8] 便宜的に δ 分布の密度関数を Dirac の δ 関数 (Dirac's delta function) を用いて表すことがある．しかしながら δ 関数は関数値が陽に書ける関数ではなく，超関数と呼ばれる特殊な関数であることに注意しなくてはならない．

例 3.10. 有理数 $\mathbb{Q}_{[0,1]}$ 全体に番号をふり，これを $q_1, q_2, \ldots$ で表す．

$$P(A) = \sum_{i: q_i \in A} 2^{-i}$$

という確率測度を考える．例えば

$$P(\mathbb{Q}_{[0,1]}) = 1, \quad P(\Omega - \mathbb{Q}_{[0,1]}) = 0$$

であるが，

$$P(\mathbb{Q}_{[0,1]}) = \int_{\mathbb{Q}_{[0,1]}} f(x)\mu(dx) = 1$$

となるような密度関数 f を考えることはできない．なぜなら 1 点 q_i の起こる確率は

$$P\{q_i\} = 2^{-i}$$

であるので

$$P\{q_i\} = \int_{\{q_i\}} f(x)\mu(dx) = f(q_i) \times \mu(q_i) = 2^{-i}$$

となるような密度関数 f を考えなくてはいけないが，1 点の Lebesgue 測度は 0 なので

$$f(q_i) \times \mu(q_i) = f(q_i) \times 0 = 0$$

となり，条件を満たす f を構成することは不可能であることがわかる．

例 3.11. コインを無限回投げるという試行を考える．コインの表を 1，裏を 0 で表すことにすると見本空間は

$$\Omega = \{\omega = (\omega_1, \omega_2, \ldots) |\ \omega_n = 0, 1\ (n = 1, 2, \ldots)\}$$

のように表される無限試行である．例えば最初に表が出るか裏が出るかで 2 つの排反事象

$$A_1 = \{\omega = (1, \omega_2, \omega_3, \dots) |\ \omega_n = 0, 1\ (n = 2, 3, \dots)\}$$
$$A_0 = \{\omega = (0, \omega_2, \omega_3, \dots) |\ \omega_n = 0, 1\ (n = 2, 3, \dots)\}$$

を考えると

$$A_1 + A_0 = \Omega, \quad \text{かつ} \quad P(A_1) = P(A_0)$$

なので，

$$P(A_1) = \frac{1}{2}, \quad P(A_0) = \frac{1}{2}$$

となる．また，k 回目までの表裏の出る順番を指定した事象

$$A_{i_1, i_2, \dots, i_k} = \{\omega = (i_1, i_2, \dots, i_k, \omega_{k+1}, \dots) |\ \omega_n = 0, 1\ (n \geq k+1)\}$$

を考えて $i_1, i_2, \dots, i_k$ を適当に 0,1 に固定すると，その事象の起こる確率は

$$P(A_{i_1, i_2, \dots, i_k}) = 2^{-k}$$

となる．したがってこれらの事象を集めた集合を

$$\mathcal{I} = \{A_{i_1, i_2, \dots, i_k} |\ i_\nu = 0, 1\ (\nu = 1, 2, \dots, k), k = 1, 2, \dots\}$$

とすると，任意の $A \in \mathcal{I}$ について確率 $P(A)$ は普通の意味で定義される．しかし Ω が無限次元の空間なので，積分そのものをうまく定義することができない．

この例で取り上げた P が確率測度になることを厳密に示すためには確率測度の拡張定理なるものを用いる必要があるが，講義ではそこまで深く立ち入らない．直観的には ω に対して区間 $[0, 1]$ の 2 進数 $0.\ \omega_1 \omega_2 \cdots$ を考えることにすれば，見本空間 Ω と区間 $[0, 1]$ を同一視できるので，確率測度として Lebesgue 測度を対応させればよい．

練習問題 (6) (7)

まとめ

- 確率密度の積分による確率測度の表現
- 確率密度の条件
- Riemann 積分と Lebesgue 積分
- いろいろな確率分布

練習問題

(1) 関数 (3.1) が有理数で 1，無理数で 0 となることを確かめよ．

(2) 正規分布の確率密度関数が $(-\infty, \infty)$ 上で積分すると 1 となることを確かめよ．

(3) Cauchy 分布において，平均値や分散が定義されないことを示せ．

(4) z が自然数のとき，Γ 関数の値を求めよ．

(5) 次の関数 f が確率密度関数となるように k を定めよ．

a) その 1

$$f(x) = k \cdot e^{-\lambda|x|} \quad (-\infty < x < \infty;\ \lambda > 0)$$

b) その 2

$$f(x) = \begin{cases} k(x+b), & -b \leq x < 0 \\ -k(x-b), & 0 \leq x \leq b \\ 0, & \text{それ以外} \end{cases}$$

c) その 3

$$f(x) = \begin{cases} 0 & (x < 0) \\ \exp(kx) & (x \geq 0) \end{cases}$$

d) その 4

$$f(x) = \begin{cases} 0 & (|x| > 1) \\ k\cos(\pi x/2) & (|x| \leq 1) \end{cases}$$

e) その 5

$$f(x) = k\exp(-4x^2) \quad (-\infty < x < \infty)$$

f) その 6

$$f(x) = \begin{cases} 0, & |x| > 1, \\ k(x^2 - 1) & |x| \leq 1. \end{cases}$$

(6) 長さ 1m の一本の紐を無作為に一箇所で切り (一様分布に従って切る場所を選ぶと考えてよい)，二本の紐を作るとする．以下の問に答えよ．

a) 短い方の紐の長さを L[m] とするとき，L の確率密度関数を求めよ．

b) A 君と B 君がそれぞれ別の紐を切断する．得られた短い方の紐の長さをそれぞれ L_A[m], L_B[m] とし，L_A と L_B の合計を L_1 とする．$L_1 < a$ (a は実数) となる確率 $P(L_1 < a)$ を a を用いて表せ．

c) L_1 の確率密度関数を求めよ．

d) L_A と L_B のうち長い方を L_2 とする．$L_2 < a$ (a は実数) となる確率 $P(L_2 < a)$ を a を用いて表せ．

e) L_2 の確率密度関数を求めよ．

f) A 君と B 君と C 君がそれぞれ別の紐を切断して得られた短い方の紐の長さをそれぞれ L_A, L_B, L_C とする．L_A, L_B, L_C の中で最も短いものを L_3 とするとき，L_3 の確率密度関数を求めよ．

(7) 周長 1m で，起点から周に沿って進んだ長さ lm の位置の目盛りを

$$x = \sqrt{l}$$

とするルーレットを回す試行の確率密度関数 $f(x)$ を求めよ．

ヒント

(2) 2つの正規分布の積を考え，その重積分を極座標を用いて変数変換すればよい．

(3) 平均が定まるためには，どの区間に対しても積分値が確定しなくてはいけない．

(4) $z = 1$ の場合を計算し，z に関する漸化式を作ればよい．

(5) 確率密度関数の全確率の条件を考えればよい．

(6) L_A と L_B が互いに無関係であることに注意すれば，確率空間は2次元空間内に比較的簡単に書くことができる．また L_A と L_B と L_C の関係は3次元空間内に表さなくてはいけないが，2次元空間から類推すればよい．

(7) 確率測度 $P(x < a)$ は簡単に求められるので，これを確率密度関数の積分として表すことを考えればよい．

確率変数 4

4.1 1次元の確率変数

この章では確率空間の上の関数として定義される確率変数の性質について纏める．確率を扱う多くの実問題では，数値的な評価を行うために様々な量を計算する必要があるが，確率変数は具体的な量を扱うための最も重要な概念である．

4.1.1 確率変数の定義

まず確率変数の定義を述べておく．

定義 4.1 (確率変数)**.** ある試行 T の確率空間を $(\Omega, \mathcal{F}, P)$ とする．Ω から実数 $\mathbb{R}$ への写像

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

として定義される実数値関数 $X(\omega)$ を確率空間 $(\Omega, \mathcal{F}, P)$ 上の**確率変数** (random variable) と呼ぶ．あるいは実数であることを明示するために**実確率変数**と呼ぶこともある．

確率変数は，試行 T の結果として見本点 ω が出現したとき，その見本点に対応して $X(\omega)$ という値を取る偶然量を表している．一般の見本空間は，骰子の目の系列や工場の製品など，数値とは限らない要素の集合を考えているため，そのままでは後に定義する「平均値」や「分散」といった数値と対応づけることが難しい．例えば「1 が出るまで骰子を振り続ける」試行の見本点として得られる長さの異なる系列に対して「見本点の平均」を定義することはできないだろう．一方，確率変数は重さや長さ，あるいは回数など実数に値を取る具体的な数値を考えているので，「平均値」や「分散」といった量を考えることができる．

例 4.2 (骰子振りの確率変数)**.** 骰子を振ると (出た目 $\times 100$) 円の賞金が貰えるとする．骰子の目を ω で表すと賞金 X は

$$X(\omega) = 100\omega$$

という確率変数であると考えることができる．

例 4.3 (二回の骰子振りの確率変数)**.** 骰子を二回振ったときの出た目の和を考える．このとき見本空間は

$$\Omega = \{(\omega_1, \omega_2) |\ \omega_1, \omega_2 = 1, 2, \dots, 6\} = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$$

となり，出た目の和 X は

$$X(\omega_1, \omega_2) = \omega_1 + \omega_2$$

という確率変数である．また 1 回目と 2 回目に出た目の大小で

$$Y(\omega_1, \omega_2) = \begin{cases} 1, & \omega_1 > \omega_2 \\ 0, & \omega_1 = \omega_2 \\ -1, & \omega_1 < \omega_2 \end{cases}$$

といった確率変数を定義することもできる．

例 4.4 (ルーレット回しの確率変数)．「ルーレット回し」の試行において，ルーレットの目盛り ω が有理数なら 1 点，無理数なら 0 点の得点が得られるとする．このとき点数

$$X(\omega) = \begin{cases} 1, & \omega \text{ が有理数} \\ 0, & \omega \text{ が無理数} \end{cases}$$

は $\Omega = (0, 1]$ 上の確率変数と考えることができる．同様に出た目盛りが 0 にどれだけ近いかに応じて点数がもらえるとする．この点数を

$$Y(\omega) = \log \frac{1}{\omega}$$

で定めるとすれば，Y も $\Omega = (0, 1]$ 上の確率変数と考えることができる．

例 4.5 (骰子振り)．「1 が出るまで骰子を振り続ける」試行で得られる見本点を ω とする．このとき，系列 ω の長さ

$$X(\omega) = \text{系列 } \omega \text{ の長さ}$$

は自然数 $\mathbb{N}$ 上の確率変数と考えることができる．

例 4.6. ある工場で生産されるエンジンの抜き取り検査をして，最大出力と燃費を測定したとする．この場合は見本点 ω は生産されるエンジンの個体を表し，最大出力 $X(\omega)$ と燃費 $Y(\omega)$ は個体 ω ごとに異なる確率変数であると考えることができる．

4.1.2 確率変数の確率空間

確率変数 $X(\omega)$ はある決まった値ではなく，ω が従う確率法則によってばらつく確率的な量であるため，確率変数 $X(\omega)$ の確率空間を考えることができる．確率変数 $X(\omega)$ の取り得る全ての値の集合を確率変数 X の**見本空間** (sample space) と呼び，

$$\Omega^X = X(\Omega) \qquad (\Omega \text{ の各点を } X \text{ で写像した像 (点) の集合})$$

と書く．

また，その確率法則は以下のようにして考えることができる．例えば，見本空間 Ω^X の任意の部分集合 B に対して「X の値が B に入る」$(X \in B)$ という事象を定義し，その確率を考える．この事象に含まれる元の見本空間 Ω の見本点は

$$A = \{\omega |\ X(\omega) \text{ が } B \text{ に入る}\} = \{\omega |\ X(\omega) \in B\} \subset \Omega$$

であり，A は元の確率空間で確率を考えることができるので，

$$P(X\text{ の値が }B\text{ に入る}) = P(A)$$

となる．ここで X を Ω の部分集合から $\mathbb{R}$ の部分集合へ対応させる関数

$$B = \{X(\omega) |\ \omega \in A\} = X(A)$$

とみなして，その逆関数を

$$X^{-1}(B) = \{\omega |\ X(\omega) \in B\} = A$$

によって定義すると，X の値が B に入る確率は

$$P(X^{-1}(B)) = P\{\omega |\ X(\omega) \in B\}$$

と表される．これを B の関数とみて

$$P^X(B) = P(X^{-1}(B))$$

と書くと，P^X は Ω^X 上の X の確率分布 (あるいは確率測度と言ってもよい) を表す．Ω^X は $\mathbb{R}$ の部分集合なので，Ω^X 上の Borel 集合族を $\mathcal{F}^X$ と書くことにすれば，$(\Omega^X, \mathcal{F}^X, P^X)$ も確率空間になる．実際

$$\begin{aligned}
&X^{-1}(B_1 + B_2) = X^{-1}(B_1) + X^{-1}(B_2)\\
&X^{-1}(\Omega^X) = \Omega
\end{aligned}$$

であることを用いると，P^X が確率測度の定義を満たしていることが確認できる (図 4.1 参照).

$$\begin{aligned}
P^X(B) &= P(X^{-1}(B)) \geq 0 && \text{(正値性)}\\
P^X(B_1 + B_2) &= P(X^{-1}(B_1 + B_2))\\
&= P(X^{-1}(B_1) + X^{-1}(B_2))\\
&= P(X^{-1}(B_1)) + P(X^{-1}(B_2))\\
&= P^X(B_1) + P^X(B_2) && \text{(加法性)}\\
P^X(\Omega^X) &= P(X^{-1}(\Omega^X))\\
&= P(\Omega) = 1 && \text{(全確率)}
\end{aligned}$$

加法性は 2 つの集合について示したが，σ-加法性も同様に示すことができる．

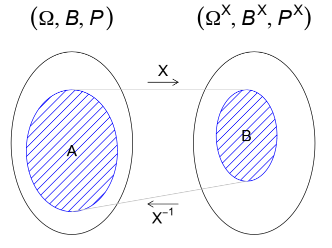

図 4.1: 元の確率空間と確率変数 X の確率空間の関係.

例 4.7 (二回の骰子振りの確率変数)．X の見本空間は

$$\Omega^X = \{2, 3, 4, \ldots, 11, 12\}$$

である．またその確率法則は

$$
\begin{aligned}
P^X(X=2) &= P\left((\omega_1,\omega_2)\in\{(1,1)\}\right) = \frac{1}{36},\\
P^X(X=3) &= P\left((\omega_1,\omega_2)\in\{(1,2),(2,1)\}\right) = 2\cdot\frac{1}{36} = \frac{1}{18},\\
P^X(X=4) &= P\left((\omega_1,\omega_2)\in\{(1,3),(2,2),(3,1)\}\right)\\
&= 3\cdot\frac{1}{36} = \frac{1}{12},\\
&\vdots\\
P^X(X=11) &= P\left((\omega_1,\omega_2)\in\{(5,6),(6,5)\}\right) = 2\cdot\frac{1}{36} = \frac{1}{18},\\
P^X(X=12) &= P\left((\omega_1,\omega_2)\in\{(6,6)\}\right) = \frac{1}{36}
\end{aligned}
$$

となる．同様に Y の見本空間は

$$\Omega^Y = \{1,0,-1\}$$

である．またその確率法則は

$$
\begin{aligned}
P^Y(Y=1) &= P\left((\omega_1,\omega_2)\in\{(2,1),(3,1),\dots,(6,5)\}\right)\\
&= 15\cdot\frac{1}{36} = \frac{5}{12},\\
P^Y(Y=0) &= P\left((\omega_1,\omega_2)\in\{(1,1),(2,2),\dots,(6,6)\}\right)\\
&= 6\cdot\frac{1}{36} = \frac{1}{6},\\
P^Y(Y=-1) &= P\left((\omega_1,\omega_2)\in\{(1,2),(1,3),\dots,(5,6)\}\right)\\
&= 15\cdot\frac{1}{36} = \frac{5}{12}
\end{aligned}
$$

となる．

例 4.8 (ルーレット回しの確率変数)．X の見本空間は

$$\Omega^X = \{1,0\}$$

である．また，その確率法則は

$$
\begin{aligned}
P^X(X=1) &= P(\omega\text{ が有理数}) = 0\\
P^X(X=0) &= P(\omega\text{ が無理数}) = 1
\end{aligned}
$$

となる．同様に Y の見本空間は

$$\Omega^Y = \log\frac{1}{(0,1]} = -\log((0,1]) = [0,\infty)$$

である．また，その確率法則は

$$
\begin{aligned}
P^Y(a<Y<b) &= P\left(\{\omega|\ a < -\log\omega < b\}\right)\\
&= P\Big(\{\omega|\ e^{-a} > \omega > e^{-b}\}\Big)\\
&= e^{-a} - e^{-b} = -e^{-b} - (-e^{-a})
\end{aligned}
$$

となることから，Y の確率分布の密度関数は

$$f(y) = e^{-y}, \qquad P^Y(a < Y < b) = \int_a^b f(y)dy$$

で与えられることがわかる．

X は実数値を取るので，X の見本空間 Ω^X は $\mathbb{R}$ の部分集合である．また連続分布を考える場合，測度の定義域は特に断わらない限り Lebesgue 測度において自然な定義域である Borel 集合族を考える．なお Ω^X が可算集合であれば

$$P^X(B) = \sum_{b \in B} P^X\{b\}$$

となり，各点での確率 $P^X\{b\}$ によって確率法則は完全に決まる．

上の骰子やルーレットの例のように，同じ確率空間の上で異なる確率変数を考えることができる．前述したエンジンの例のように，実際の問題ではある現象を理解するために何種類かの異なる実験を行って具体的な物理量を測定する．現象の総体を表すのが元となる確率空間であり，確率変数はその一面を捉えるために行う実験の観測値に対応している．実験で観測される変数はいろいろあるが，背後にある原因は共通で，ω はその原因を表す何か抽象的なものを象徴していると考えればよい (図 4.2 参照)．

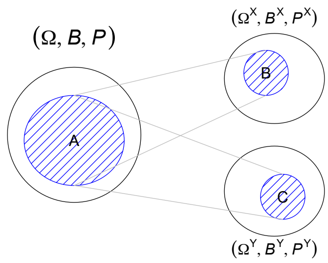

図 4.2: 異なる確率変数 X, Y と確率空間の関係．$B = X(A)$, $C = Y(A)$.

4.2 多次元の確率変数

前述したように，同じ確率空間の上にいくつもの確率変数を考えることができ，これを纏めてベクトルと見る場合もある．例えば 2 次元の値を取る関数

$$\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega))$$

を考えると，2 次元の確率変数を考えることができる．これを *2* **次元確率変数** (2-dimensional random variable) あるいは 2 次元の**確率ベクトル** (random vector) という．見本空間や確率法則は，1 次元の場合と同様にそれぞれ

$$\Omega^{\boldsymbol{X}} = \boldsymbol{X}(\Omega)$$

$$P^{\boldsymbol{X}}(B) = P\{\omega | \ \boldsymbol{X}(\omega) \in B\} = P(\boldsymbol{X}^{-1}(B)), \quad B \subset \Omega^{\boldsymbol{X}}$$

で定義される．$X_1(\omega), X_2(\omega)$ を $\boldsymbol{X}(\omega)$ の**成分変数**，$\boldsymbol{X}(\omega)$ を $X_1(\omega), X_2(\omega)$ の**結合変数**という．また，$\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega))$ の確率法則を表す $P^{\boldsymbol{X}}$ を $X_1(\omega), X_2(\omega)$ の**同時確率分布** (**結合確率分布**; joint probability distribution)，あるいは**同時確率法則** (**結合確率法則**; joint probability law) と呼ぶ．

ここで

$$\Omega^{\boldsymbol{X}} \subset \Omega^{X_1} \times \Omega^{X_2} \subset \mathbb{R}^2$$

となるが，結合変数の見本空間 $\Omega^{\boldsymbol{X}}$ と各成分変数の見本空間の積空間 $\Omega^{X_1} \times \Omega^{X_2}$ が一致するとは限らないことに注意する (図 4.3 参照)．

以上の議論は n **次元確率変数** (n-dimensional random variable) についても同様である．

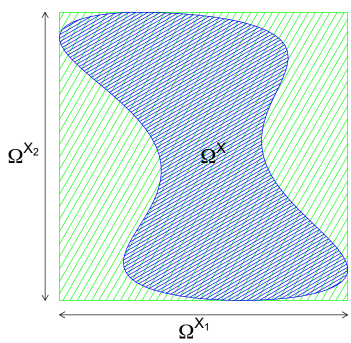

図 4.3: 結合変数の見本空間 $\Omega^{\boldsymbol{X}}$ と成分変数の積空間 $\Omega^{X_1} \times \Omega^{X_2}$ の関係．

4.3 確率変数の変換

実数値関数

$$
\begin{aligned}
\phi : \Omega^X &\to \mathbb{R} \\
X &\mapsto \phi(X)
\end{aligned}
$$

に対して

$$
Y(\omega) = \phi(X(\omega)) \qquad (Y = \phi \circ X \text{と書くこともある})
$$

とすると $Y(\omega)$ という新しい実数値確率変数ができる．

例 4.9 (二回の骰子振りの確率変数の変換)**.** 骰子を二回振ったときの出た目の和を考え，和が偶数なら 100 円貰え，奇数なら 50 円支払うというゲームを考える．骰子を二回振ったときの出た目をそれぞれ ω_1, ω_2 とする．また，目の和を

$$
X(\omega_1, \omega_2) = \omega_1 + \omega_2
$$

で表すとすると，賞金 (罰金)Y は

$$
Y(\omega_1, \omega_2) = Y(X) = \begin{cases} 100, & X \text{ が偶数} \\ -50, & X \text{ が奇数} \end{cases}
$$

という X が変換された確率変数であると考えることができる．

新しく作られた確率変数 Y の見本空間は，変換を順に追って

$$
\Omega^Y = \phi(X(\Omega)) = (\phi \circ X)(\Omega)
$$

で定義される (図 4.4 参照)．また，Ω^Y の適当な部分集合 C に対して，Ω^X で対応する集合を B，Ω で対応する集合を A とすると

$$
\begin{aligned}
C &= \phi(B) = \phi(X(A)) \\
B &= X(A) = \phi^{-1}(C) \\
A &= X^{-1}(B) = X^{-1}(\phi^{-1}(C))
\end{aligned}
$$

となるので，その確率法則は

$$
\begin{aligned}
P^Y(C) &= P^X(B) = P(A) \\
&= P(X^{-1}(\phi^{-1}(C))) = P\{\omega |\ \phi(X(\omega)) \in C\} \quad C \subset \Omega^Y
\end{aligned}
$$

で計算されることがわかる．

4.4 確率変数の平均値

一般の見本空間では抽象的な空間を考えることができるが，反面数量的な取り扱いが難しいことも多い．一方，確率変数はその見本空間が実数となり平均値や分散といった具体的な量を計算することができる．まず，確率変数の平均について定義するが，これは確率変数の定義されるもとの見本空間，あるいは確率変数そのものの見本空間の性質によって 2 つの場合がある．

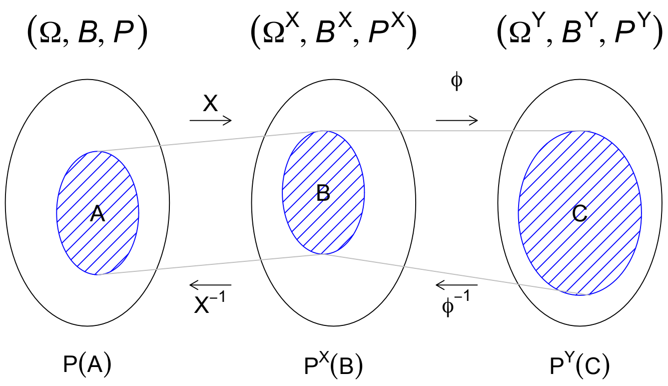

図 4.4: 確率変数の変換.

4.4.1 離散分布の平均値

定義 4.10 (平均値). 見本空間が可算の場合，その確率空間上で定義された確率変数 X の**平均値** (**期待値**; expectation) は

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

で定義される.

$X(\omega)$ は ω の関数であるが，全ての可能な ω について和を取っているので，$\mathbb{E}[X]$ は ω によらない量となっていることに注意する.
また，確率変数は関数であるので，$\mathbb{E}$ は関数 $X(\omega)$ に作用する作用素であると考えることができる.

例 4.11 (骰子振りの確率変数). 出た目の 100 倍の値である賞金 X の平均値 (期待値) は，骰子の目を ω で表すことにすれば $P\{\omega\} = \frac{1}{6}$ なので，

$$\begin{aligned}
\mathbb{E}[X] &= X(1) \cdot P\{1\} + X(2) \cdot P\{2\} + X(3) \cdot P\{3\} \\
&\quad + X(4) \cdot P\{4\} + X(5) \cdot P\{5\} + X(6) \cdot P\{6\} \\
&= 100 \cdot \frac{1}{6} + 200 \cdot \frac{1}{6} + 300 \cdot \frac{1}{6} + 400 \cdot \frac{1}{6} + 500 \cdot \frac{1}{6} + 600 \cdot \frac{1}{6} \\
&= 350
\end{aligned}$$

である.

例 4.12 (二回の骰子振りの確率変数). 出た目の和 X の平均値は，二つの目を ω_1, ω_2 で表すことにすれば $P\{(\omega_1, \omega_2)\} = \frac{1}{36}$ なので，

$$\begin{aligned}
\mathbb{E}[X] &= X(1,1) \cdot P\{(1,1)\} + X(1,2) \cdot P\{(1,2)\} + \\
&\quad \cdots + X(6,5) \cdot P\{(6,5)\} + X(6,6) \cdot P\{(6,6)\} \\
&= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \cdots + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} \\
&= 7
\end{aligned}$$

である．また二つ目の大小関係で $\{-1,0,1\}$ の値を取る確率変数 Y の平均値は

$$\begin{aligned}\mathbb{E}[Y] &= Y(1,1)\cdot P\{(1,1)\} + Y(1,2)\cdot P\{(1,2)\} + \\ &\qquad \cdots + Y(6,5)\cdot P\{(6,5)\} + Y(6,6)\cdot P\{(6,6)\} \\ &= 1\cdot\frac{15}{36} + 0\cdot\frac{6}{36} + (-1)\cdot\frac{15}{36} \\ &= 0\end{aligned}$$

である．

4.4.2 連続分布の平均値

定義 4.13 (平均値)**.** 確率分布 P に確率密度 f がある場合，その確率空間上で定義された確率変数 X の**平均値** (**期待値**; expectation) は

$$\mathbb{E}[X] = \int_\Omega X(\omega)P(d\omega) = \int_\Omega X(\omega)f(\omega)d\omega$$

で定義される．

連続分布の場合も $X(\omega)$ は ω の関数であるが，ω に関して積分しているので，平均値 $\mathbb{E}[X]$ は ω によらない量となっていることに注意する．

例 4.14 (ルーレット回しの確率変数)**.** ルーレットの目盛りが有理数か無理数かで $\{1,0\}$ の値を取る確率変数 X の平均値は

$$\mathbb{E}[X] = \int_0^1 X(\omega)\mu(d\omega) = 1\cdot\mu(\text{有理数}) + 0\cdot\mu(\text{無理数}) = 0$$

である．これは ω の見本空間を用いた計算として示してあるが，X の確率空間を考えるとその見本空間は $\Omega^X = \{0,1\}$ であり，確率法則は $P^X(X=0)=1, P^X(X=1)=0$ となるので，

$$\mathbb{E}[X] = 0\cdot P^X(X=0) + 1\cdot P^X(X=1) = 0\cdot 1 + 1\cdot 0 = 0$$

と計算してもよい．また目盛りの負の対数を取る確率変数 Y の平均値は

$$\begin{aligned}\mathbb{E}[Y] &= \int_0^1 (-\log\omega)\,\mu(d\omega) \\ &= [-\omega\log\omega + \omega]_0^1 = 1\end{aligned}$$

である．

見本空間を $A\subset\Omega$ に制限した場合の平均値を

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

$$\mathbb{E}[X,A] = \int_A X(\omega)P(d\omega) = \int_A X(\omega)f(\omega)d\omega$$

と書くことがある．これは次節で述べる条件付期待値とは異なることに注意する．

なお，$\boldsymbol{X}$ がベクトルの場合は，その平均値は各成分ごとに計算すればよい．

$$\mathbb{E}[\boldsymbol{X}] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_n]) \in R^n$$

4.4.3 平均値の性質

平均値について以下の性質が重要である．

定理 4.15. 平均値には以下のような性質がある．

(1) $\mathbb{E}[aX+bY]=a\,\mathbb{E}[X]+b\,\mathbb{E}[Y]$ （**線形性**; linearity）

(2) $\mathbb{E}[X,\sum_{i=1}^{n}A_i]=\sum_{i=1}^{n}\mathbb{E}[X,A_i]$

(3) A の上で恒等的に確率変数が定数，すなわち $X(\omega)=a$ ならば
$\mathbb{E}[X,A]=a\,P(A)$
特に $\mathbb{E}[a]=a$ (定数の平均値はその定数)

また，X の確率分布が考えられ，X の確率分布が P^X，あるいは密度関数が f^X で与えられる場合には以下が成り立つ．

(4) $\mathbb{E}[X]=\sum_{x\in\Omega^X}xP^X\{x\}$ (X が離散分布)
$\mathbb{E}[X]=\int_{x\in\Omega^X}xf^X(x)dx$ (X が連続分布)

(5) $Y(\omega)=\phi(X(\omega))$ ならば
$\mathbb{E}[Y]=\sum_{x\in\Omega^X}\phi(x)P^X\{x\}$ (X が離散分布)
$\mathbb{E}[Y]=\int_{x\in\Omega^X}\phi(x)f^X(x)dx$ (X が連続分布)

練習問題 (1)

例 4.16 (二回の骰子振りの確率変数)**.** 出た目の和 X の平均値は平均の和の関係を用いると

$$\begin{aligned}\mathbb{E}[X]&=\mathbb{E}[\omega_1+\omega_2]=\mathbb{E}[\omega_1]+\mathbb{E}[\omega_2]=2\mathbb{E}[\omega_1]\\&=2\left(1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}\right)\\&=7\end{aligned}$$

としても求まる．

例 4.17 (ルーレット回しの確率変数)**.** 目盛りの負の対数 Y の平均値は Y の確率密度を用いて

$$\begin{aligned}\mathbb{E}[Y]&=\int_0^\infty ye^{-y}dy\\&=\left[-ye^{-y}\right]_0^\infty+\int_0^\infty e^{-y}dy=\left[-e^{-y}\right]_0^\infty\\&=1\end{aligned}$$

としてもよい．

4.5 確率変数の分散

4.5.1 分散・共分散

平均値と同様に確率変数の性質を決める大事な量として分散がある．これは確率変数が平均値のまわりにどのくらいばらつくかを示す指標となる．

定義 4.18 (分散)．確率変数 X の**分散** (variance) は平均値を用いて

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

で定義される．また，分散の平方根を**標準偏差** (standard deviation) という．

定義 4.19 (共分散)．確率変数 X, Y の**共分散** (covariance) は

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

で定義される．

分散の計算も，確率空間 $(\Omega, \mathcal{F}, P)$ 上で平均値の定義に基づいて計算してもよいが，確率変数の確率法則が記述される場合にはそちらを用いてもよい．例えば X, Y が離散分布に従う場合には

$$\mathrm{Var}(X) = \sum_{x \in \Omega^X} (x - \mathbb{E}[X])^2 P^X\{x\}$$

$$\mathrm{Cov}(X, Y) = \sum_{x \in \Omega^X, y \in \Omega^Y} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) P^{(X,Y)}\{x, y\}$$

のように順次展開して計算していけばよい．また連続分布に従う場合には確率密度関数を用いて

$$\mathrm{Var}(X) = \int_{x \in \Omega^X} (x - \mathbb{E}[X])^2 f^X(x) dx$$

$$\mathrm{Cov}(X, Y) = \int_{x \in \Omega^X, y \in \Omega^Y} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) f^{(X,Y)}(x, y) dx dy$$

となる．

例 4.20 (二回の骰子振りの確率変数)．出た目の和 X の分散は

$$\begin{aligned}
&\mathbb{E}[(X - \mathbb{E}[X])^2] \\
&\qquad = (2-7)^2 \cdot \frac{1}{36} + (3-7)^2 \cdot \frac{2}{36} + \cdots + (12-7)^2 \cdot \frac{1}{36} \\
&\qquad = \frac{35}{6}
\end{aligned}$$

である．また二つの目の大小関係を表す Y の分散は

$$\begin{aligned}
&\mathbb{E}[(Y - \mathbb{E}[Y])^2] \\
&\qquad = (1-0)^2 \cdot \frac{5}{12} + (0-0)^2 \cdot \frac{1}{6} + (-1-0)^2 \cdot \frac{5}{12} \\
&\qquad = \frac{5}{6}
\end{aligned}$$

である．

例 4.21 (ルーレット回しの確率変数). 有理数・無理数を区別する X の分散は

$$\mathbb{E}[(X-\mathbb{E}[X])^2]=(1-0)^2\cdot 0+(0-0)^2\cdot 1=0$$

である．また負の対数 Y の分散は

$$\begin{aligned}\mathbb{E}[(Y-\mathbb{E}[Y])^2]&=\int_0^\infty (y-1)^2e^{-y}dy\\&=-\left[\left\{(y-1)^2+2(y-1)+2\right\}e^{-y}\right]_0^\infty\\&=1\end{aligned}$$

である．

4.5.2 分散の性質

分散・共分散については以下の性質が重要である．

定理 4.22. 分散・共分散には以下の性質がある．

(1) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$ (対称性)

(2) $\mathrm{Cov}(X,a)=0$

(3) $\mathrm{Cov}(X,X)=\mathrm{Var}(X)\geq 0$ (正値性)

(4) $\mathrm{Var}(aX+b)=a^2\,\mathrm{Var}(X)$

$\mathrm{Cov}(aX+b,cY+d)=ac\mathrm{Cov}(X,Y)$

(5) $\mathrm{Var}(aX+bY)=a^2\,\mathrm{Var}(X)+2ab\,\mathrm{Cov}(X,Y)+b^2\,\mathrm{Var}(Y)$

(6) $\mathrm{Var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2$

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

(7) $|\mathrm{Cov}(X,Y)|\leq\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$

(Schwarz の不等式の一つの表現)

特に (6) の関係は重要である．

練習問題 (2)

4.5.3 モーメント

分散をより一般化した量としてはモーメントがある．

定義 4.23 (モーメント). 確率変数 X の k **次モーメント** (積率; k-th moment) は，確率変数の k 乗の平均

$$m_k(X)=\mathbb{E}[X^k]$$

で定義される．分散と同じように平均値を中心として考えた**平均値のまわりの** k **次モーメント**は

$$\mu_k(X)=\mathbb{E}[(X-\mathbb{E}[X])^k]$$

で定義される．

また，確率変数の絶対値の k 乗のモーメント

$$\tilde{m}_k(X) = \mathbb{E}[|X|^k]$$

を k **次絶対モーメント**と呼ぶことがある．

なお，分散は平均値のまわりの 2 次モーメントである．

4.6 特性関数

平均，分散，モーメントは確率変数の一面を捉えるための 1 つの量であるが，確率変数が従う確率測度の特徴を表すものとしては以下の特性関数がある．

定義 4.24 (特性関数)**. 特性関数** (characteristic function) は指数関数を用いて変数変換された確率変数の平均として

$$\phi(z) = \mathbb{E}[e^{izX}]$$

で定義される．

特性関数は実数値ではなく複素数値をとることに注意する．

1 次元の連続分布 P に対しては

$$\phi(z) = \int_R e^{izx} P(dx)$$

であるが，密度関数 f が存在すれば

$$\phi(z) = \int_R e^{izx} f(x)dx$$

となり，一般の定義とは定数倍の違いはあるが確率変数 X の確率密度の Fourier 変換に他ならない．離散分布では

$$\phi(z) = \sum_{x \in \Omega} P\{x\} e^{izx},$$

により計算される．

Fourier 変換には逆 Fourier 変換が存在してもとの関数を再構成できるように，特性関数と確率測度 (確率密度) は一対一に対応している (*Lévy-Haviland* **の反転公式**)．すなわち 2 つの分布の特性関数が同じであれば，2 つの分布が同じであることが言える．後に中心極限定理の証明で述べるように特性関数は確率測度の性質を調べるために重要な働きをする．

まとめ

練習問題 (3) (4) (5) (6) (7) (8) (9) (10) (11) (12) (13) (14) (15) (16) (17) (18) (19) (20)

- 確率空間上の関数としての確率変数
- 確率変数の確率空間
- 複数の確率変数の同時確率 (結合確率)
- 確率変数の平均，分散，モーメント
- 確率変数の特性関数

練習問題

(1) 定理 4.15 のそれぞれの性質を証明せよ．

(2) 定理 4.22 のそれぞれの性質を証明せよ．

(3) 長さ 1m の棒をランダムに一箇所で切断する．このとき以下の問に答えよ．

a) 切断したときにできる短い方の棒の長さの従う確率分布の密度関数を求めよ．

b) 長い方の棒の長さの従う確率分布の密度関数を求めよ．

c) 長い方の棒と短い棒の長さの差の従う確率分布の密度関数を求めよ．

d) 短い方の棒の長さの平均値と分散を求めよ．

e) 長い方の棒の長さの平均値と分散を求めよ．

f) 長い方の棒と短い棒の長さの差の平均値と分散を求めよ．

(4) 長さ 1m の棒をランダムに二箇所で切断する．このとき以下の問に答えよ．

a) 一番短い棒，真中の長さの棒，一番長い棒の従う確率分布の密度関数を求めよ．

b) 一番短い棒，真中の長さの棒，一番長い棒のそれぞれの平均値を求めよ．

(5) A 君と B 君がそれぞれ長さ 1m の紐を切断し，切られた短い方の紐の長さをそれぞれ L_A[m], L_B[m] とする．このとき以下の問に答えよ．

a) L_A と L_B の合計を M とする．確率 $P(M < a)$ を a(a は実数) を用いて表せ．

b) L_A と L_B のうち長い方を N とする．確率 $P(N < a)$ を a(a は実数) を用いて表せ．

c) M と N のそれぞれの確率分布の密度関数を求めよ．

d) M と N のそれぞれの平均値を求めよ．

(6) 確率変数 X が平均 μ，分散 σ^2 の正規分布に従うとき，X の特性関数を求めよ．ただし，平均 μ，分散 σ^2 の正規分布の確率密度関数は次の式で表される．

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty).$$

(7) 確率変数 X が，確率密度関数

$$f(x) = \begin{cases} \frac{1}{2} & (-1 \leq x \leq 1), \\ 0 & \text{それ以外}, \end{cases}$$

で表される分布に従うとき，以下の問に答えよ．

a) X の特性関数を求めよ．

b) 実数 k を用いて新たな確率変数 $Y = kX$ を定義する．Y の特性関数を求めよ．

(8) 確率変数 N は平均 0，分散 b^2 の正規分布に従うとする．以下の問に答えよ．

a) 確率変数 N の特性関数を求めよ．

b) 新しい確率変数 $X = \cos(N)$ を定義する．X の平均と分散を求めよ．

c) 新しい確率変数 $Y = 2\sin(N)$ を定義する．Y の分散が 1 となるように b を定めよ．

(9) 確率変数 U は区間 $[0, 1]$ 上の一様分布に従うとする．以下で定義される確率変数 X の平均，分散，および確率密度関数を求めよ．

a) $X = 2U - 1$

b) $X = U^2$

c) $X = (2U - 1)^3$

d) $X = \sin(\pi(U - 0.5))$

(10) 確率変数 U は区間 $[-a, a]$ $(a > 0)$ 上の一様分布に従うとする．以下の問に答えよ．

a) 確率変数 U の分散が $\pi^2/12$ となるように a を定めよ．

b) 確率変数 U の特性関数を求めよ．

c) 新しい確率変数 $X = \cos(U)$ を定義する．X の平均と分散を求めよ．

(11) 確率変数 U_1, U_2 は独立に区間 $[-1, 1]$ 上の一様分布に従うとする．以下で定義される確率変数 X の平均，分散，確率密度関数および特性関数を求めよ．

a) $X = U_1 - U_2$

b) $X = U_1 + 2U_2$

(12) 確率変数 U_1, U_2 は独立に区間 $[0, 1]$ 上の一様分布に従うとする．以下の問に答えよ．

a) 新しい確率変数 $X = (U_1 - U_2)/2$ を定義する．X の分散を求めよ．

b) X の特性関数と確率密度関数を求めよ．

c) 新たな確率変数 Y を

$$Y = \sqrt{-2\log U_1}\cos(2\pi U_2)$$

で定義する．Y の平均と分散を求めよ．

d) 確率変数 Y の特性関数を求めよ．

e) 新たな確率変数 Z を

$$Z = \sqrt{-2\log U_1}\sin(2\pi U_2)$$

で定義する．Y と Z の共分散 $\mathrm{Cov}((Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z]))$ を求めよ．

f) 確率変数 Z の密度関数を求めよ．

(13) 確率変数 N_1, N_2 は独立に平均 0，分散 1 の正規分布に従うとする．以下の問に答えよ．

a) 確率変数 $\sigma_1 N_1$ の特性関数を求めよ．ただし σ_1 は正の実数とする．

b) 新たな確率変数 $X = \sigma_1 N_1 + \sigma_2 N_2$ を定義する．ただし σ_1, σ_2 は正の実数とする．X の分散を求めよ．

c) 確率変数 X の特性関数を求めよ．

d) 確率変数 X の確率密度関数を求めよ．

e) 新しい確率変数 $Y = N_1^2 + N_2^2$ を定義する．Y の特性関数を求めよ．

f) 確率変数 Y の従う分布の確率密度関数を求めよ．

(14) 確率変数 N は平均 1，分散 1 の正規分布に従うとする．以下の問に答えよ．

a) N の特性関数を求めよ．

b) N と同じ分布に従う 2 つの独立な確率変数 N_1, N_2 を用いて新しい確率変数 $X = N_1 - N_2$ を定義する．X の従う分布の確率密度関数を求めよ．

c) N と同じ分布に従う 2 つの独立な確率変数 N_3, N_4 を用いて新しい確率変数 $Y = 4N_3 - 3N_4$ を定義する．Y の従う分布の確率密度を求めよ．

(15) 確率変数 U が，区間 $[0, 1]$ 上の一様分布に従うとする．すなわち U の確率密度関数は

$$f(u) = \begin{cases} 1, & 0 \leq u \leq 1 \\ 0, & \text{それ以外} \end{cases}$$

で表される．新しい確率変数 X, Y を

$$X = \cos(\pi U), \quad Y = \sin(2\pi U)$$

で定義するとき，以下の問に答えよ．

a) X と Y の平均と分散を求めよ．

b) X と Y の共分散を求めよ．

c) 区間 I を以下で定める．

$$I = \left[\frac{1}{2}, \frac{1}{\sqrt{2}}\right]$$

$X \in I$ となる U の範囲を求めよ．また，$X \in I$ となる確率 $P(X \in I)$ を求めよ．

d) $Y \in I$ となる U の範囲を求めよ．また，$Y \in I$ となる確率 $P(Y \in I)$ を求めよ．

e) 前二問の結果を参考に，Y の従う分布の確率密度関数を求めよ．

(16) 確率変数 U は区間 $(0,1]$ 上の一様分布に従い，確率変数 N は平均 0，分散 1 の正規分布に従うとする．このとき，以下の問いに答えよ．

a) 確率変数 U を変換して，新しい確率変数

$$X = 2U - 1$$

を定義する．X の従う分布の確率密度関数を求めよ．

b) 確率変数 X の平均と分散を求めよ．

c) 確率変数 U を対数で変換して，新しい確率変数

$$Y = -\log U$$

を定義する．Y の従う分布の確率密度関数を求めよ．

d) 確率変数 Y の平均，分散，および特性関数を求めよ．

e) N と同じ分布に従う独立な確率変数 N_1，N_2 を用いて新しい確率変数

$$Z = \frac{N_1^2 + N_2^2}{2}$$

を定義する．Z の特性関数を求めよ．

f) 確率変数 Z の平均と分散を求めよ．

g) 上の計算を踏まえて，確率変数 N から確率変数 U を作る方法について論じよ．

(17) 次の関数 f が確率密度関数となるように k を定めたとする．確率変数 X が f で表される分布に従うとき，X の平均，分散，特性関数を求めよ．

a)

$$f(x) = \begin{cases} k(x+b) & (-b \leq x < 0), \\ -k(x-b) & (0 \leq x \leq b), \\ 0 & \text{それ以外}. \end{cases}$$

b)

$$f(x) = k \cdot e^{-\lambda|x|} \quad (-\infty < x < \infty).$$

(18) 2 つのバスケットボールチーム A, B の今シーズンのこれまでの試合を調べたところ，A チームの得点は平均 77 点，分散 12^2，B チームの得点は平均 67 点，分散 16^2 の正規分布に従うことがわかった．A チームと B チームが対戦する次の試合でも，各チームの得点はこの分布に従うと仮定してよい．以下の問に答えよ．

a) A チームと B チームの得点差の分布はどのようになるか説明しよ．

b) A チームが勝つ確率 (より多くの得点を取る確率) を求めよ．

c) B チームが 10 点差で勝つ確率を求めよ．

(19) 周長 1[m] の歪みのないルーレットを回し，起点から針の指した位置まで周に沿って進んだ長さ ω[m] を見本点 $\omega \in (0,1]$ とする試行を考える．ルーレットを 2 度回して得られた見本点 (互いに無関係である) をそれぞれ ω_1, ω_2 とするとき，以下の問いに答えよ．

a) 確率変数 X を

$$X = \omega_1 + \omega_2$$

とする．確率変数 X の値が $X < a$ となる確率 $P(X < a)$ を求めよ．

b) 確率変数 X の確率密度関数を求めよ．

c) 確率変数 X の平均値と分散を求めよ．

d) 新しい確率変数を $Y = 1/X$ とする．Y の平均値を求めよ．

(20) とある世界では，毎朝神様がある確率分布に従って世界中の全ての人に一つの数字を割当て，数字の大きさに従って不幸の種を散蒔くという．つまり数字が大きい程その日良くないことが起こるらしい．

人々はこの数字がどういう分布に従っているかは教えてもらえないし，数字の範囲も分布の形もどうやら毎日変わるので推測することは難しいらしい．

数字だけ眺めていてもなんだか居心地が悪いと考えたある人が，その日自分がどのくらい運が悪いかを明確にするために次のようなことを考えた．

> 「出逢う人ごとに数字を尋ねることにして，何人目で自分より悪い数字 (大きな数字) を持つ人に出逢ったかでその日の運の悪さを測ることしよう．」

つまり自分より大きな数字の人がいれば，取りも直さず自分より運が悪い人がいると安心できるので，安心するまでに掛った回数で不幸の度合を測ろうと考えた訳である．例えば最初に逢った人がいきなり自分より悪い数字だったらその日は-1 ポイント，20 人目だったら-20 ポイントというように不幸の点数を決めてしまおうということである．なおこの世界には無限に多くの人が住んでいるとして，自分より悪い人に出逢うまで何回でも尋ねるとしておこう．

さて，このアイデアで彼は本当に居心地が良くなるのだろうか？

ヒント

(2) 定義に従って展開すればよい．なお，Schwarz の不等式については，t を実数として確率変数 X, Y の合成変数 $tX + Y$ を考え，その分散が正となること，すなわち

$$\mathrm{Var}(tX+Y) = t^2\,\mathrm{Var}(X) + 2t\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y) \geq 0$$

を利用すれば判別式より容易に求められる．

(3) 切断する場所を見本点として確率空間を構成し，短い棒や長い棒の長さを表す確率変数を考えればよい．

(4) 切断する 2 点を見本点として 2 次元の見本空間を用いて確率空間を構成し，対応する確率変数を考えればよい．

(5) 切断するそれぞれの点を見本点として 2 次元の確率空間を構成し，指示された事象の確率を考えればよい．

(9) 確率密度関数を求めるには，適当な事象 ($X < a$ など) に対する確率 ($P(X < a)$ など) を考えて微分すればよい．

(17) 確率密度関数となる条件は前章のとおり．あとは定義に従って計算すればよい．

(18) 異なる正規分布の差の分布がどのような分布に従うかを考えてみよう．

(20) この人は平均何点となるか (平均何回の人に尋いて安心できるか) を考えてみよう．

条件付確率と独立性 5

5.1 条件付確率と Bayes の定理

条件付確率とは，一種の因果関係を記述するための確率論の枠組である．一般の因果律では原因と結果の間に自然な時間関係が存在するが，条件付確率は時間的な前後とは関係なく事象の間に定義される．また事象をうまく分解することによって導かれる Bayes の定理を利用して逆向きの関係を陽に計算することもできるため，確率的な推論には欠くことのできない重要な概念でもある．

5.1.1 条件付確率

定義 5.1. 確率空間 $(\Omega, \mathcal{F}, P)$ において 2 つの事象 $A, B \subset \Omega$ を考える．ただし，事象 A は空でない $A \neq \emptyset$ とする．事象 A のもとにおける事象 B の**条件付確率** (conditional probability) を $P(B|A)$ または $P_A(B)$ と表記し，

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

で定義する．

事象 A と B が同時に起こる確率 $P(A \cap B)$ は**同時確率** (joint probability) と言い，これは

「事象 A と事象 B が同時に起きている確率」

を表す．一方，条件付確率は

「事象 A が起きたときに事象 B が起きている確率」

を与えることになる (図 5.1 参照)．

なお，$P(A) = 0$ のときは $P(B|A)$ は意味がないが，適当に一点 $\omega_0 \in A$ を選び

$$P(B|A) = \begin{cases} 1, & \omega_0 \in B \\ 0, & \omega_0 \notin B \end{cases}$$

とすると都合がよいことが多い．

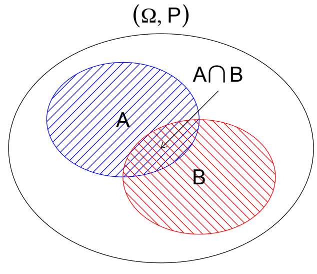

図 5.1: 条件付確率の意味．A に対する $A \cap B$ の割合．

例 5.2. 骰子を一回振る試行について，以下の事象を定義する．

$$A = \{偶数の目が出る\}$$
$$B = \{素数の目が出る\}$$

このとき

$$\begin{aligned}
P(B|A) &= P(\text{偶数の目が出たとき，それが素数である}) \\
&= \frac{P(A\cap B)}{P(A)} = \frac{P(\text{偶数かつ素数の目が出る})}{P(\text{偶数の目が出る})} \\
&= \frac{P\{2\}}{P(\{2,4,6\})} = \frac{1/6}{1/2} = \frac{1}{3} \\
P(B^c|A) &= P(\text{偶数の目が出たとき，それが素数でない}) \\
&= \frac{2/6}{1/2} = \frac{2}{3} = 1 - P(B|A)
\end{aligned}$$

となる．

条件付確率を考える際に良く使われるのは，確率変数を用いた表現である．例えば，確率変数が一つの値を取る場合を考えると，これは

$$\begin{aligned}
A = \{\omega|X(\omega) = x\} \Leftrightarrow A = X^{-1}(x)\ (x \in \Omega^X) \\
B = \{\omega|Y(\omega) = y\} \Leftrightarrow B = Y^{-1}(y)\ (y \in \Omega^Y)
\end{aligned}$$

のようにある事象に対応づけられる．これを用いて

$$P(Y\!=\!y|X\!=\!x) \quad \text{または} \quad P_{X=x}(Y\!=\!y)$$

といった条件付確率を考えることができる．

同様に確率変数がある集合に含まれる場合

$$\begin{aligned}
A = \{\omega|X(\omega) \in E\} \Leftrightarrow A = X^{-1}(E)\ (E \subset \Omega^X) \\
B = \{\omega|Y(\omega) \in F\} \Leftrightarrow B = Y^{-1}(F)\ (F \subset \Omega^Y)
\end{aligned}$$

となるが，この場合は

$$P(Y \in F|X \in E) \quad \text{または} \quad P_{X\in E}(Y \in F)$$

というように書かれる．

また，こうした確率変数を用いた表記では同時確率は単に変数を並べて

$$\begin{aligned}
P(X\!=\!x, Y\!=\!y) &= P(\{\omega|X(\omega) = x\} \cap \{\omega|Y(\omega) = y\}) \\
P(X\!\in E, Y\!\in F) &= P(\{\omega|X(\omega) \in E\} \cap \{\omega|Y(\omega) \in F\})
\end{aligned}$$

と表すことが多い．

さらに，確率変数 X,Y のいろいろな値に対する同時確率や条件付確率をまとめて確率法則 (確率分布) として表現する場合には，

$$P(X,Y), \quad P(Y|X)$$

といった記述を行うことが多いので，注意してほしい．

定義から明らかであるが，条件付確率の以下の性質は重要である．

定理 5.3. 条件付確率には次の性質がある．

(1) $P(A\cap B) = P(A)P(B|A)$

(2) $P(A|A) = 1$

(3) $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$

(4) $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

(2) より事象 A を固定したとき $P(\cdot|A)$ は A を見本空間とする確率法則になっていることがわかる．これを A のもとにおける**条件付確率分布** (conditional probability distribution) あるいは**条件付確率法則** (conditional probability law) という．またこの分布のもとでの平均値を

$$\mathbb{E}[Y \mid A] = \sum_{\omega \in A} Y(\omega)P(\omega|A) \quad (\text{離散分布の場合})$$

$$\mathbb{E}[Y \mid A] = \int_A Y(\omega)P(d\omega|A) \quad (\text{連続分布の場合})$$

で定義できるが，これを**条件付平均値** (conditional expectation) と呼ぶ．

5.1.2 Bayes の定理

また条件付確率では定理の (4) より導かれる次の等式が重要である．

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

これを用いると，$\Omega = A_1 + A_2 + \cdots + A_n$ のとき $A_k \cap B$ は互いに排反な事象となることに注意すれば

$$P(B) = P(\Omega \cap B) = P(\sum_{k=1}^{n} A_k \cap B) = \sum_{k=1}^{n} P(A_k \cap B)$$

と分解して記述されるので，A_i のもとにおける B の条件付確率法則から B のもとにおける A_i の条件付確率法則を求める以下の公式が導かれる．

定理 5.4 (Bayes の定理)**.** $\Omega = A_1 + A_2 + \cdots + A_n$ のとき

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)}$$

となる．

5.2 独立性

独立性とは 2 つの確率変数が確率法則の意味で無関係となることである．その定義の仕方はいくつかあるが，本質は条件付確率の性質によって理解できる．

5.2.1 確率変数の独立性

まず有限試行の場合を考える．

二つの確率変数 X, Y が互いに**独立** (independent) であるとは，任意の x, y に対して

$$P(X=x, Y=y) = P(X=x)P(Y=y)$$

となることであると定義される．左辺の $X = x$ と $Y = y$ が同時に起こる確率は条件付確率を用いると

$$P(X=x, Y=y) = P(X=x)P(Y=y|X=x)$$

であるから，一般に X の値に応じて Y の起こる確率は変化する．独立な場合は X,Y がそれぞれ起こる確率の積であり，Y の起こる確率は X に左右されないこと

$$P(Y=y|X=x) = P(Y=y)$$

を示している．

無限試行の場合の厳密な定義は少々面倒であるが，大雑把に言うと任意の事象 (集合)$E \subset \Omega^X$,$F \subset \Omega^Y$ に対して

$$P(X \in E, Y \in F) = P(X \in E)P(Y \in F)$$

が成り立つことと思えばよい．確率密度を持つ場合には X, Y の同時確率分布の密度関数 (**同時確率密度**; joint probability density) を $f(x, y)$，X, Y のそれぞれの確率分布の密度関数 (**周辺確率密度**; marginal probability density) を $g(x), h(y)$ とすると，確率変数 X, Y が互いに独立とは任意の x, y に対して

$$f(x, y) = g(x)h(y)$$

となることである．

一般の場合には特性関数を用いて

$$\mathbb{E}[e^{isX+itY}] = \mathbb{E}[e^{isX}]\mathbb{E}[e^{itY}]$$

が成り立つことと同値である (*Kac* **の定理**).

独立性に関連して重要な性質を以下に纏めておく．

定理 5.5. 独立性に関しては以下の性質が成り立つ．

(1) 確率変数 X, Y が独立 $\Leftrightarrow$ $P(Y|X) = P(Y)$ *a.s.*

(周辺分布 $P(Y)$ と条件付分布 $P(Y|X)$ が確率 1 で等しい)

(2) 確率変数 $X_1, X_2, \ldots, X_n$ が独立 $\Rightarrow$ 変数変換した確率変数

$$Y_1 = \phi_1(X_1), Y_2 = \phi_2(X_2), \ldots, Y_n = \phi_n(X_n)$$

は独立 (逆が成り立つとは限らないことに注意)

(3) 確率変数 $X_1, X_2, \ldots, X_n$ が独立 $\Rightarrow$

$$\mathbb{E}[X_1X_2\cdots X_n] = \mathbb{E}[X_1]\mathbb{E}[X_2]\cdots\mathbb{E}[X_n]$$

(平均値の乗法性)

(4) 確率変数 $X_1, X_2, \ldots, X_n$ の任意の *2* 変数が独立 $\Rightarrow$

$$\mathrm{Var}(X_1+X_2+\cdots+X_n) = \mathrm{Var}(X_1)+\mathrm{Var}(X_2)+\cdots+\mathrm{Var}(X_n)$$

(分散の加法性) [9]

5.2.2 事象の独立性

確率変数の独立に対して**事象の独立**という概念もある．集合 A の指示関数 (または定義関数ともいう) を

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

で定義する．指示関数は実数値関数であり，$Y = 1_A(\omega)$ とおくと Y は見本点 ω が事象 A に含まれたかどうかを表す確率変数と考えることができる．ここで確率変数の独立性の定義を用いて，事象 $A_1, A_2, \ldots, A_n$ のそれぞれの指示関数で定義される確率変数 $Y_1, Y_2, \ldots, Y_n$ が互いに独立である場合，事象 $A_1, A_2, \ldots, A_n$ は独立であるという．

[9] 平均値の加法性は独立の仮定とは無関係に成り立つ．

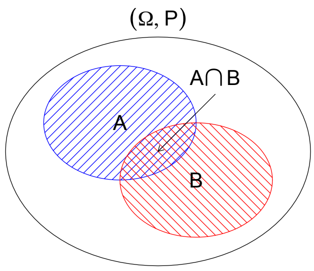

図 5.2: 独立性の直観的な意味．A の中に限っても，全体で考えても B の起こる頻度は変わらないとき，A と B は独立である．

定理 5.6. 以下の *3* つは同値である．

1. 事象 $A_1, A_2, \ldots, A_n$ が独立

2. $B_i = A_i$ または $B_i = A_i^c$ としたとき

$$P(B_1 \cap B_2 \cap \cdots \cap B_n) = P(B_1)P(B_2)\cdots P(B_n)$$

3. $\forall k = 2, 3, \ldots, n,\ 1 \leq \forall i_1 < \forall i_2 < \cdots \forall i_k \leq n$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$$

まとめ

- 同時確率，周辺確率，条件付確率
- 条件付確率分布，条件付確率密度，条件付平均値
- 条件付確率と Beyes の定理
- 確率変数と事象の独立性

練習問題 (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11)

練習問題

(1) K 先生はテニス観戦が趣味で，お気に入りの S 選手の成績で大会の翌日の機嫌が大きく変化する．詳しく調べた結果，次のことが判った．
- K 先生の機嫌は，良いか悪いかの二通りしかない．
- S 選手がベスト 4 に残る (1 位から 4 位のいずれかになる) と，0.6 の確率で機嫌が良い．
- S 選手がベスト 4 に残らないと，0.6 の確率で機嫌が悪い．
- S 選手が優勝 (1 位になる) すると，0.9 の確率で機嫌が良い．
- S 選手は，0.2 の確率でベスト 4 に残る．
- S 選手は，0.05 の確率で優勝する．

このとき，以下の問いに答えよ．

a) K 先生の機嫌が良いときに，前日の大会で S 選手がベスト 4 に残った確率はいくつか？

b) K 先生の機嫌が良いときに，前日の大会で S 選手が優勝した確率はいくつか？

c) S 選手がベスト 4 には残ったけれど優勝できなかった (2 位から 4 位のいずれかになった) ときに，K 先生の機嫌が良い確率はいくつか？

(2) 野球とサッカーの好きな W 先生は，お気に入りのチーム (野球の A チームとサーカーの B チーム) の成績で翌日の機嫌が大きく変化する．詳しく調べた結果，次のことが判った．

- W 先生の機嫌は，良いか悪いかの二通りしかない．
- A チームが勝つと 0.75 の確率で機嫌が良い．
- A チームが負けると 0.7 の確率で機嫌が悪い．
- B チームが負けると 0.6 の確率で機嫌が悪い．
- A チーム・B チームがともに負けると 0.9 の確率で機嫌が悪い．
- A チームの勝率は 0.6 である．
- B チームの勝率は 0.5 である．
- A チームと B チームの勝ち負けは独立と考えてよい．

このとき，以下の問いに答えよ．

a) A チーム・B チームがともに勝つとき、W 先生の機嫌が良い確率はいくつか?

b) W 先生の機嫌が悪いときに、前日 B チームが負けた確率はいくつか?

c) W 先生の機嫌が良いときに、前日 A チームが負けて B チームが勝った確率はいくつか?

(3) アイスクリームと水着の売上と夏の気温の関係を調べたところ，以下のことがわかった．

- それぞれの売上は，冷夏かそうでないかに関係する．
- 冷夏の場合，0.8 の確率でアイスクリームの売上は例年より小さい．
- 冷夏の場合，アイスクリームと水着の売上がともに例年より大きくなる確率は 0.1 である．
- 冷夏でない場合，アイスクリームと水着の売上の大小は互いに独立である．
- 冷夏でない場合，アイスクリームと水着の売上がともに例年より大きくなる確率は 0.6 である．
- 冷夏である確率は 0.3 である．

なお，売上の大小は例年の平均値と比べて，大きいか小さいかの 2 通りだけを考えるものとする．

さらに，確率変数 X,Y によってアイスクリームと水着の売上をそれぞれ表し，売上が例年より大きいとき $+1$ の値を，小さいとき -1 の値をとるものとすると次のことがわかった．

- X の平均値は 0.17 であり，Y の平均値は 0.24 である．

これらを踏まえて，以下の問に答えなさい．

a) 冷夏のとき，アイスクリームと水着の売上がともに例年より小さくなる確率はいくつか？

b) 冷夏でないとき，アイスクリームの売上が例年より大きくなる確率はいくつか？

c) 冷夏でないとき，水着の売上が例年より大きくなる確率はいくつか？

d) 水着の売上が例年より大きいとき，アイスクリームの売上も大きい確率はいくつか？

e) ある年のアイスクリームの売上は例年より小さく，水着の売上は大きかったとする．この年が冷夏であった確率はいくつか？

f) 確率変数 X と Y の共分散はいくつか？

(4) 簡易検査によって疑わしい人を抽出し，その後精密検査を行う癌検診について考える．簡易検査としては A,B の 2 種があり，それぞれについて以下のことが判っているとする．

- 検査 A によって，癌でないにも関わらず精密検査が必要と判断される確率は 0.1(10%) である．
- 検査 A によって，癌である人は全て精密検査が必要と判断される．
- 検査 A によって精密検査が必要とされた人の中で，癌の人の割合は 0.04(4%) である．
- 検査 B によって，癌でないにも関わらず精密検査が必要と判断される確率は 0.02(2%) である．
- 検査 B によって，癌である人の 0.8(80%) が精密検査が必要と判断される．
- 検査 A,B での判断結果は独立な事象である．

以上の状況を踏まえて以下の問に答えなさい．

a) 癌の人の割合はどのくらいか？

b) 検査 B で精密検査は必要ないと判断されたにもかかわらず癌である人の割合はいくつか？

c) 検査 B で精密検査が必要と判断された人のどのくらいの割合が癌であるか？

d) 検査 A,B でともに精密検査が必要と判断された人のどのくらいの割合が癌であるか？

(5) ある病気の検査をする 2 種類の検査 A,B について以下のことが判っているとする．

- この病気に罹っている人の割合は全人口の 0.1(10%) である．

- 検査 A によって陽性 (病気に罹っている) と診断される人は 0.36(36%) である．
- 検査 A によって陽性と診断された人の中で本当に病気に罹っている人の割合は 0.25(25%) である．
- 検査 B によって病気の人が陽性と診断される確率は 0.8(80%) である．
- 検査 B によって病気でない人が陽性と診断される確率は 0.1(10%) である．
- 検査 A,B の診断結果は独立である．

以上の状況を踏まえて以下の問に答えなさい．

a) 検査 A によって病気の人が陽性と診断される確率はいくつか？

b) 検査 A によって病気でない人が陽性と診断される確率はいくつか？

c) 検査 B で陽性と診断された人の中で，病気の人の割合はいくつか？

d) 検査 A で陽性だが検査 B で陽性でないと診断された人の中で，病気の人の割合はいくつか？

e) 検査 A,B でともに陽性でないと診断された人の中で，病気の人の割合はいくつか？

(6) W 大の全学生に Web を用いて以下の調査を行った．

質問 1 あなたは奨学金を受け取っていますか？

質問 2 あなたはアルバイトをしていますか？

全学生のうち 20%の学生が回答し，回答者の中で

- 奨学金を受け取っている学生 (質問 1 に「はい」と回答) は 60%，
- アルバイトをしている学生 (質問 2 に「はい」と回答) は 50%，
- 奨学金を受けながらアルバイトもしている学生 (質問 1 と 2 に「はい」と回答) は 30%，

であった．回答率が低いので，別途全学生に対して聞き取り調査を行った結果

- 奨学金を受け取っている学生は 28%，
- 奨学金を受けながらアルバイトもしている学生は 14%，
- 奨学金も受けずアルバイトもしていない学生は 44%，

であることがわかった．

以下の問に答えなさい．

a) Web 調査に回答した学生のうち，奨学金は受けずにアルバイトをしている学生は何%か？

b) Web 調査に回答した学生のうち，奨学金も受けずアルバイトもしていない学生は何%か？

c) 全学生の中でアルバイトをしている学生は何%か？

d) Web 調査に回答しなかった学生のうち，アルバイトをしている学生は何%か？

e) Web 調査に回答しなかった学生のうち，奨学金を受け取っていない学生は何%か？

f) Web 調査に回答しなかった学生のうち，奨学金を受けながらアルバイトもしている学生は何%か？

g) Web 調査に回答しなかった学生のうち，奨学金は受けずにアルバイトをしている学生は何%か？

h) 奨学金を受け取っている学生のうちで Web 調査に回答した学生は何%か？

(7) ある大学の学生の猫好きと犬好きとカメラ好きの関係を調べることを考える．まず，全学生の 40%が利用している写真投稿サイトの猫と犬の写真を調べたところ以下のことがわかった．

- 猫を投稿する人 (猫好き) は犬を投稿する人 (犬好き) の丁度 2 倍である．
- 猫も犬も投稿しない人 (猫も犬も嫌い) は全体の 0.1 である．
- 猫と犬の両方を投稿する人 (猫も犬も好き) は全体の 0.3 である．

以降この投稿サイトを利用する学生を「カメラ好き」と考えることにする．

また，全学生に対して別途アンケート調査を行ったところ以下のことがわかった．

- 犬好きの人の割合は 0.46 である．
- 犬は好きだが猫は嫌いの人の割合は 0.16 である．
- 猫好きであるがカメラ好きでない (投稿したことのない) 人で犬好きと犬嫌いは同数である．

調査に誤差はなく，調査結果はそれぞれの条件での割合を正しく捉えているとして，以下の問に答えなさい．

a) 猫好きな人の割合はいくつか？

b) 真の猫好き (猫は好きだが犬は嫌い) な人の割合はいくつか？

c) カメラ嫌いの人の中で犬好きな人の割合はいくつか？

d) 猫好きな人の中でカメラ好きな人の割合はいくつか？

e) 猫も犬も好きな人の中でカメラ嫌いな人の割合はいくつか？

(8) A 先生は相撲好きで，お気に入りの二人の力士 B と C の勝敗が翌日の機嫌に大きく影響する．詳しく調べた結果

- A 先生の機嫌は「良い」か「悪い」の二通りしかない．
- B と C がともに勝つと機嫌が良い確率は 0.9 である．
- B か C のどちらか一方が勝つと機嫌が良い確率は 0.7 である．

- B と C がどちらも負けると機嫌が悪い確率は 0.9 である．
- B の勝率は 0.8，C の勝率は 0.75 である．
- B と C の勝敗は独立と考えてよい (B と C は対戦していない)．

であることが判った．このとき，以下の問に答えよ．

a) A 先生の機嫌が良いとき，前日に B と C がともに勝った確率はいくつか？

b) A 先生の機嫌が悪いとき，前日に B と C のどちらか一方しか勝てなかった確率はいくつか？

c) A 先生の機嫌が良いにもかかわらず前日に B と C のどちらも負けた確率はいくつか？

(9) アイスクリームの売上と電力消費量が前年より増加するかどうかと，夏の暑くなるか涼しくなるかの関係を調べたところ

- アイスクリームの売上は 0.6 の確率で前年より増加する．
- 前年より暑くなるとアイスクリームの売上は 0.8 の確率で増加する．
- 電力消費量は 0.65 の確率で前年より増加する．
- 前年より涼しいと電力消費量は 0.5 の確率で減少する．
- 前年より暑くなる確率は 0.6 である．
- 夏の暑さで条件付けたときアイスクリームの売上と電力消費量の増減は独立である．

ことが判った．このとき，以下の問に答えよ．

a) アイスクリームの売上は増加し電力消費量は減少したとき，前年より暑くなった確率はいくつか？

b) アイスクリームの売上は減少し電力消費量は増加したとき，前年より暑くなった確率はいくつか？

(10) 高田馬場のラーメン屋 A の売上は，前の週のレストラン口コミサイト B と C における評価点の上下に強く影響される．詳しく調べた結果

- 売上，および評価点は上がるか下るかの二通りしかない．
- B, C 共に評価点が上がったときに売上が上がる確率は 0.9 である．
- B, C の一方のみ評価点が上がったときに売上が上がる確率は 0.6 である．
- B, C 共に評価点が下がったときに売上が下がる確率は 0.8 である．
- B の評価点が上がる確率は 0.5 である．
- C の評価点が上がる確率は 0.6 である．
- B の評価点が上がるとき $X = 1$，下がるとき $X = 0$，C の評価点が上がるとき $Y = 1$，下がるとき $Y = 0$ となる確率変数 X, Y を考えると，X, Y の共分散は 0.1 である．

であることが判った．このとき，以下の問に答えよ．

a) B, C 共に評価点が上がる確率はいくつか？

b) A の売上が下がったとき，B, C 共に評価点が下がった確率はいくつか？

c) A の売上が上がったとき，B, C の一方のみ評価点が上がった確率はいくつか？

(11) 人間の声は波の性質を持っているので，一般には

$$s(t) = A(t)\sin(\omega(t)t + \theta(t)),$$

のように表される．図 5.3 に示したように $A(t)$ は点線で表された時間によって変化する振幅を表し，$\omega(t), \theta(t)$ は (図には陽に現れていないが) 時間によって変化する角速度・位相を表している．

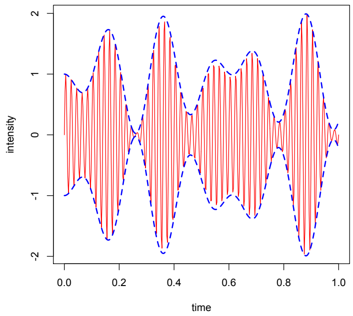

図 5.3: 音声信号の概念図．

実際の信号は図 5.4 のような波形をしており，ある程度長い時間で考えると振幅・角速度・位相が激しく変動しているが，短い時間だけ切り出して考えると振幅・角速度・位相は変化せず一定と看做すことができ，通常 A, ω, θ を時不変とした

$$s(t) = A\sin(\omega t + \theta)$$

で十分良く近似できる．

また，人間が感じる音の大きさはエネルギーによって決まるが，波のエネルギーは振幅の 2 乗に比例するので

$$(\text{音の大きさ}) \propto A^2$$

となる．

さて二人の人が同時に同じ音程で同じ大きさの声を出した場合を考えよう．上の考察から，ある短い時間においてはそれぞれの声は

$$s_1(t) = A\sin(\omega t + \theta_1)$$
$$s_2(t) = A\sin(\omega t + \theta_2)$$

で近似できる．この場合，一見振幅が $2A$ となるので，音の大きさは $(2A)^2/A^2 = 4$ 倍となりそうであるが，我々が通常経験するように音の大きさは 2 倍にしか感じられない．

このことを，位相 θ_1, θ_2 が $[0, 2\pi]$ 上で独立に，かつ一様にばらつくことを仮定して説明せよ．

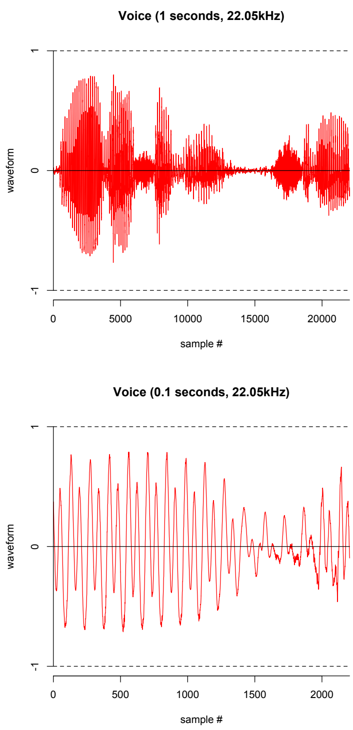

図 5.4: 実際の音声信号．長い時間で見た場合 (上) と短い時間を拡大した場合 (下)

ヒント

(1) 問題文に沿って K 先生の機嫌に関する事象と S 選手の成績に関する事象の関係を，条件付確率・周辺確率として記述し，Bayes の定理を用いる．

(2) 前問と同様に W 先生の機嫌に関する事象を，条件付確率・周辺確率として記述し，Bayes の定理を用いる．

(3) 前問と同様に事象を整理して条件付確率・周辺確率を求め，必要な量を計算すればよい．

(11) 位相が独立に一様に分布するとして，二人の声の和の 2 乗平均 (分散) を計算すればよい．

独立変数の和の性質 6

6.1 大数の法則

例えば再現性が保証されている物理実験では，同じ実験を繰り返し行うことによって多数の測定値を観測し，その算術平均を用いて測定精度を高めるといった操作が行われる．これは確率的なゆらぎに起因する確率変数の不確実な部分が，同一の確率法則に従う多数の確率変数を足し合わせることによって相殺し取り除かれるという著しい性質によっている．この性質は大数の法則と呼ばれ，統計的な推測において利用される確率論の重要な概念である．

6.1.1 大数の法則

以下では確率変数の列 (可算無限でも良い) を

$$\{X_n\} = \{X_1, X_2, X_3, \dots\}$$

と書く．多くの問題では X_n は同じ分布 P^X に従う場合を扱うが，一般に X_n はそれぞれ異なる分布 P^{X_n} に従っていて構わない．確率変数の従う分布に関する条件はその都度明記することにする．確率変数 X_n の平均や分散などをまとめて表記するときは $\{\mathbb{E}[X_n]\}$ や $\{\mathrm{Var}(X_n)\}$ のように $\{\cdot\}$ を付けて書く．また，確率変数列に関する性質を書く場合も同様とする．例えば，「$\{\mathrm{Var}(X_n)\}$ が有界である」とは，分散は非負の値なので上限が有界であることを意味し，

$$\sup_n \mathrm{Var}(X_n) < \infty$$

となることである．

以上の記法の準備のもと, 大数の法則は以下のように述べられる．

定理 6.1 (大数の法則)．$\{X_n\}$ を確率変数列として

$$S_n = \sum_{k=1}^{n} X_k$$

とする．また，$\{X_n\}$ は独立で，$\{\mathrm{Var}(X_n)\}$ は有界とする．任意の $\varepsilon > 0$ に対して，$n(\varepsilon)$ を十分大きくとれば，すべての $n > n(\varepsilon)$ に対して

$$\Pr\left(\frac{|S_n - \mathbb{E}[S_n]|}{n} > \varepsilon\right) < \varepsilon$$

とすることができる．

以下に大数の法則の簡略な証明を示す．

まず準備として次の不等式を証明する．

定理 6.2 (Čebyšev の不等式). 任意の確率変数 X において任意の $a > 0$ に対して

$$\Pr(|X(\omega) - \mathbb{E}[X]| > a) \leq \frac{1}{a^2}\mathrm{Var}(X)$$

が成り立つ.

証明.

$$A = \{\omega \big| |X(\omega) - \mathbb{E}[X]| > a\}$$

とする. 分散の性質から

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2, \Omega] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2, A + A^c] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2, A] + \mathbb{E}[(X - \mathbb{E}[X])^2, A^c] \\
&\geq \mathbb{E}[(X - \mathbb{E}[X])^2, A] \geq a^2 P(A)
\end{aligned}$$

であるので, 題意が証明される. □

証明. まず $a > 0$ として Čebyšev の不等式より

$$\Pr(|S_n - \mathbb{E}[S_n]| > a) \leq \frac{1}{a^2}\mathrm{Var}(S_n)$$

が成り立つ. $\{X_n\}$ の独立性より

$$\mathrm{Var}(S_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k) \leq n \cdot \max_k \mathrm{Var}(X_k)$$

がいえるので, $a = n\varepsilon$ と置き換えることによって

$$\Pr\left(\frac{|S_n - \mathbb{E}[S_n]|}{n} > \varepsilon\right) \leq \frac{1}{n\varepsilon^2} \max_k \mathrm{Var}(X_k)$$

が成り立つ. ここで $n \to \infty$ とすると右辺は $\{\mathrm{Var}(X_n)\}$ の有界性より 0 になる. したがって任意の $\varepsilon > 0$ に対して

$$\lim_{n\to\infty} \Pr\left(\frac{|S_n - \mathbb{E}[S_n]|}{n} > \varepsilon\right) \to 0$$

が成り立つ. 以上より題意が証明される. □

6.1.2 大数の強法則

大数の法則と同様な主張ではあるが, より強い意味での収束性を主張する大数の強法則は以下のように述べられる.

定理 6.3 (Kolmogorov の大数の強法則). $\{X_n\}$ を確率変数列として

$$S_n = \sum_{k=1}^{n} X_k$$

とする．$\{X_n\}$ が独立で，$\{\mathrm{Var}(X_n)\}$ が有界ならば

$$\frac{S_n - \mathbb{E}[S_n]}{n} \to 0 \ a.s.$$

が成り立つ．

「強法則」というのは上の S_n の収束の仕方について，概収束という強い意味での収束性が保証されることを意味している．一方，前節の大数の法則は確率収束という弱い意味で収束を扱っているため，大数の弱法則という場合もある．確率変数の収束については，概収束，平均収束，確率収束，法則収束という異なる収束概念があるが，本講義ではこれ以上詳しくは取り扱わない．なお，強法則の厳密な証明は繁雑なので，興味のある者は成書を参照されたい．

応用上良く使われる形は以下のように確率変数が同じ分布に従う場合である．この定理により，多数の観測を行いその算術平均を取ることによって対象となる確率変数の平均値を精度良く求めることができることが保証される．

定理 6.4 (同分布の場合の大数の強法則)**.** 確率変数列 $\{X_n\}$ が独立で同じ分布に従い，$\{\mathbb{E}[|X_n|]\}$ が有界であるとする．$\{X_n\}$ の平均値を μ とすると，

$$\frac{S_n}{n} \to \mu \quad a.s.$$

が成り立つ．

分散が有界であるという条件が絶対値の平均が有界という条件に緩められていることに注意する (絶対値の大きな値は 2 乗するとより大きくなるので，2 乗平均が有界という条件の方が絶対値の平均が有界より厳しい)．目的とする確率変数がこの条件を満すならば，十分多くの観測を行うことができるとき，すなわち n が十分大きくとれるとき，平均値 μ の近似値として観測値の算術平均

$$\frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

を使って良いことが定理により保証される．

さて十分多いとはどの程度の数なのかという問題が残るが，これに対しては次節で述べる中心極限定理が部分的な解答を与えてくれる．

6.2 中心極限定理

中心極限定理は大数の法則を精密化したものである．

6.2.1 中心極限定理

まず一般的な場合の定理の主張を述べておく．

定理 6.5 (Lindeberg の中心極限定理)**.** $\{X_n\}$ は独立で，その分散が有界，$S_n = \sum_{k=1}^{n} X_k$ とする．すべての $\varepsilon > 0$ に対し

$$\frac{1}{\mathrm{Var}(S_n)} \sum_{k=1}^{n} \mathbb{E}\left[(X_k - \mathbb{E}[X_k])^2, |X_k - \mathbb{E}[X_k]| \geq \varepsilon\sqrt{\mathrm{Var}(S_n)}\right] \to 0 \ (n \to \infty)$$

ならば，

$$T_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$$

の確率法則は，$n \to \infty$ のとき標準正規分布 $N(0, 1)$ に近づく．

練習問題 (1)

この証明は次節に纏めておく．

上の定理の条件は Lindeberg の条件と呼ばれる．問題ごとにこの条件が成り立つかどうか調べるのは大変であるが，応用上は確率変数が同分布に従う場合が多く，以下の形の定理が重要である．

定理 6.6 (同分布の場合の中心極限定理)**.** $\{X_n\}$ が独立で，正の分散を持つ同じ分布に従うときには中心極限定理が成り立つ．

正の分散を持つとは確率変数が定数ではないことを意味する．また，同分布の場合には Lindeberg の条件が 1 つの確率変数を用いて

$$\frac{1}{\mathrm{Var}(X)} \mathbb{E}\left[(X - \mathbb{E}[X])^2, |X - \mathbb{E}[X]| \geq \varepsilon\sqrt{n\mathrm{Var}(X)}\right] \to 0 \quad (n \to \infty)$$

と書かれる．左辺は，平均値から $\sqrt{n\mathrm{Var}(X)}$ 以上離れた見本点が得られるという事象 $|X - \mathbb{E}[X]| \geq \varepsilon\sqrt{n\mathrm{Var}(X)}$ の起こる確率を表すが，これは n が大きくなるとともに 0 に近づくので，同分布であれば自動的に Lindeberg の条件は満足される．

さて定理の主張するところを大雑把に言えば，たとえ X_n が正規分布とは掛け離れた分布に従っていたとしても，n 個の確率変数の和 S_n はその平均値 $\mathbb{E}[S_n]$ を中心に分散 $\mathrm{Var}(S_n)$ でほぼ正規分布に従うということである．この状況は平均 $\mathbb{E}[X]$，分散 $\mathrm{Var}(X)$ の同分布に従う確率変数の和を考えてみると理解しやすい．この場合

$$T_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{n\mathrm{Var}(X)}} = \frac{S_n - \mathbb{E}[S_n]}{n} \cdot \frac{1}{\sqrt{\mathrm{Var}(X)/n}}$$

が平均 0，分散 1 の標準正規分布に従うので，簡単な変換でわかるように n 個の確率変数の算術平均

$$\frac{S_n}{n} = \frac{\sum_{k=1}^{n} X_k}{n} = \bar{X}$$

はほぼ

$$\text{平均: } \mathbb{E}[X], \quad \text{分散: } \frac{\mathrm{Var}(X)}{n}$$

の正規分布に従うことになる．n が有限のとき S_n の分布は完全には正規分布ではないが，n が大きくなるにつれてだんだん正規分布に近づいていく (図 6.1 参照)．したがって n がある程度大きければ例えば S_n/n が標準偏差の α 倍 $(\pm\alpha\sigma)$ の区間

$$\left[\mathbb{E}[X]-\alpha\sqrt{\frac{\mathrm{Var}(X)}{n}},\ \mathbb{E}[X]+\alpha\sqrt{\frac{\mathrm{Var}(X)}{n}}\right]$$

に入る確率は正規分布を用いて近似され，ほぼ

$$\begin{aligned}\Pr\left(\mathbb{E}[X]-\alpha\sqrt{\frac{\mathrm{Var}(X)}{n}}<\frac{S_n}{n}<\mathbb{E}[X]+\alpha\sqrt{\frac{\mathrm{Var}(X)}{n}}\right)\\ \simeq\frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{\alpha}e^{-\frac{x^2}{2}}dx\end{aligned}$$

練習問題 (1)

となる．具体的な数値としては

$$\begin{aligned}\alpha&=1 &\quad \frac{1}{\sqrt{2\pi}}\int_{-1}^{1}e^{-\frac{x^2}{2}}dx&=0.6827\\ \alpha&=2 &\quad \frac{1}{\sqrt{2\pi}}\int_{-2}^{2}e^{-\frac{x^2}{2}}dx&=0.9545\\ \alpha&=3 &\quad \frac{1}{\sqrt{2\pi}}\int_{-3}^{3}e^{-\frac{x^2}{2}}dx&=0.9973\end{aligned}$$

のようになる．n が大きくなるにしたがって分散は小さくなり，$n\to\infty$ では 0 となる．これは確率変数の算術平均が平均値に一致することを言っており，中心極限定理は大数の法則を精密化したものとみることもできる．

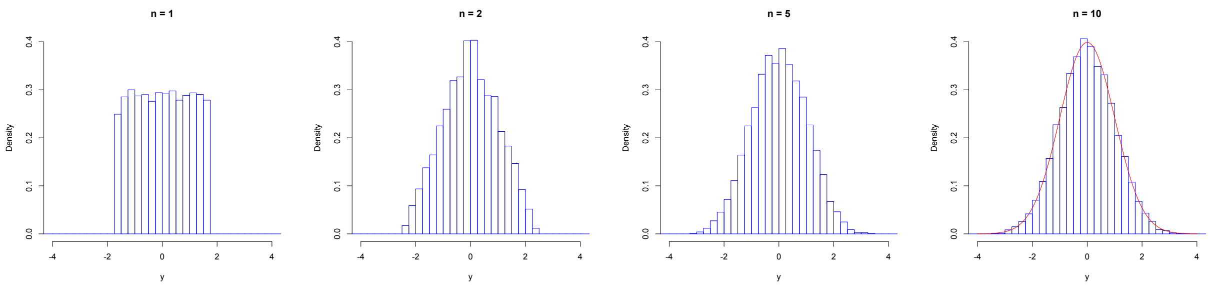

図 6.1: 中心極限定理の成り立つ様子：$[-\sqrt{3},\sqrt{3}]$ 上の独立な一様乱数 X_i (平均 0，分散 1) を用いて $Y=\sum_{i=1}^{n}X_i/\sqrt{n}$ を作り，histogram を描いたもの．縦軸は正規化して密度にしてある．n が大きくなるに従って正規分布の密度に近づいていく様子がわかる．$n=10$ では正規分布の密度を重ね合わせて描いた．

6.2.2 中心極限定理の証明

まず，証明のために以下の補題を用意する．

補題 6.7. 各 n に対して複素数列

$$\alpha_{n1},\alpha_{n2},\ldots,\alpha_{nN}\qquad (N=N(n)<\infty)$$

が与えられ，

$$\alpha_n = \sum_{k=1}^{N(n)} \alpha_{nk} \to \alpha \qquad (n \to \infty),$$
$$\beta_n = \max_k |\alpha_{nk}| \to 0 \qquad (n \to \infty),$$
$$\gamma_n = \sum_{k=1}^{N(n)} |\alpha_{nk}| < b \qquad (b \text{ は } n \text{ に無関係な定数})$$

が満たされているとする．このとき

$$\prod_{k=1}^{N(n)} (1+\alpha_{nk}) \to e^{\alpha} \quad (n \to \infty).$$

が成り立つ．

証明．題意を示すには

$$f_n = \log \prod_{k=1}^{N(n)} (1+\alpha_{nk}) = \sum_{k=1}^{N(n)} \log(1+\alpha_{nk})$$

とおいたとき，f_n が α に収束することを示せばよい．対数関数の Taylor 展開は

$$\log(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 + \cdots \quad (\text{収束半径は } |z|<1)$$

であるが，

$$\log(1+z) - z = \int_0^z \frac{1}{1+z'}dz' - \int_0^z 1dz' = \int_0^z \frac{-z'}{1+z'}dz'$$

であることに注意すると，$|z| < 1/2$ の場合には

$$\frac{1}{2} < |1+z| < \frac{3}{2} \quad \Leftrightarrow \quad \frac{2}{3} < \frac{1}{|1+z|} < 2$$

であるから

$$\begin{aligned} |\log(1+z) - z| &= \left| \int_0^z \frac{-z'}{1+z'}dz' \right| \\ &\le \left| \int_0^z \frac{z'}{|1+z'|}dz' \right| \\ &< \left| \int_0^z 2z'dz' \right| = |z^2| \end{aligned}$$

が言える．ここで $|\theta| < 1$ である適当な複素数 θ を用いると

$$\log(1+z) = z + \theta z^2 \quad |z| < \frac{1}{2},\ |\theta| < 1$$

と書けることがわかる (平均値の定理)．ただし θ は z に依存して決まることに注意する．これを用いると

$$\begin{aligned} f_n &= \sum_k \left\{ \alpha_{nk} + \theta_k \alpha_{nk}^2 \right\} \\ &= \alpha_n + \sum_k \theta_k \alpha_{nk}^2 \end{aligned}$$

となり，f_n と α の差は $|\theta_k| < 1$ に注意すると

$$\begin{aligned}
|f_n - \alpha| &\le |\alpha_n - \alpha| + \sum_k \left|\theta_k \alpha_{nk}^2\right| \\
&\le |\alpha_n - \alpha| + \max_k |\alpha_{nk}| \cdot \sum_k |\theta_k \alpha_{nk}| \\
&\le |\alpha_n - \alpha| + \beta_n \gamma_n \\
&\to 0 \quad (n \to \infty)
\end{aligned}$$

で上界が評価されるので，題意が証明された． □

この補題は

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

をより一般的な場合に拡張したものになっている．

また，証明の途中で用いた Taylor 展開の上限の評価は平均値の定理の応用である．一般に関数 $f(x)$ の点 a のまわりにおける Taylor 展開は

$$\begin{aligned}
f(x) = f(a) + (x-a)\frac{f'(a)}{1!} + (x-a)^2\frac{f''(a)}{2!} + \cdots \\
+ (x-a)^{n-1}\frac{f^{(n-1)}(a)}{(n-1)!} + (x-a)^n \frac{f^{(n)}(\xi)}{n!}
\end{aligned}$$

と書くことができるので，これを用いて示してもよい．ただし ξ は

$$\xi = a + \theta(x-a), \quad 0 < \theta < 1$$

であり，a と x の間の数である．

この補題を用いて中心極限定理の証明を行う．

まず

$$Y_{nk} = \frac{X_k - \mathbb{E}[X_k]}{\sqrt{\mathrm{Var}(S_n)}}$$

とおくと，T_n は

$$T_n = \sum_{k=1}^n Y_{nk}$$

と表される．このとき

$$\begin{aligned}
\mathbb{E}[Y_{nk}] &= 0, \\
\mathrm{Var}(Y_{nk}) &= \mathbb{E}[Y_{nk}^2], \\
\mathbb{E}[T_n] &= 0, \\
\mathrm{Var}(T_n) &= \sum_{k=1}^n \mathrm{Var}(Y_{nk}) = \sum_{k=1}^n \mathbb{E}[Y_{nk}^2] = \frac{\mathrm{Var}(S_n)}{\mathrm{Var}(S_n)} = 1
\end{aligned}$$

となるので，定理の仮定である Lindeberg の条件は

$$\sum_{k=1}^n \mathbb{E}[Y_{nk}^2, |Y_{nk}| \ge \varepsilon] \to 0 \quad (\forall \varepsilon > 0,\ n \to \infty)$$

と書くことができる．

さて $\{Y_{nk}, k=1,2,\dots,n\}$ は独立であるから，T_n の特性関数は独立変数の平均値の乗法性より

$$\mathbb{E}[e^{izT_n}] = \prod_{k=1}^{n} \mathbb{E}[e^{izY_{nk}}]$$

と分解できる．

補題を使うため，以下のように α_{nk} を定める．

$$\alpha_{nk} = \mathbb{E}[e^{izY_{nk}}] - 1$$

(平均の分解 $E(\cdot) = \mathbb{E}[\cdot, |Y_{nk}| < \varepsilon] + \mathbb{E}[\cdot, |Y_{nk}| \geq \varepsilon]$ を用いる)

$$= \mathbb{E}[e^{izY_{nk}}, |Y_{nk}| < \varepsilon] + \mathbb{E}[e^{izY_{nk}}, |Y_{nk}| \geq \varepsilon] - 1$$

(指数関数を適当な項まで Taylor 展開する)

$$= \mathbb{E}\left[1 + izY_{nk} - \frac{1}{2}z^2Y_{nk}^2 + \frac{1}{6}\theta_3|zY_{nk}|^3, |Y_{nk}| < \varepsilon\right]$$

$$+ \mathbb{E}\left[1 + izY_{nk} + \frac{1}{2}\theta_2 z^2 Y_{nk}^2, |Y_{nk}| \geq \varepsilon\right] - 1 \quad (|\theta_2|, |\theta_3| \leq 1)$$

(もう一度 $\mathbb{E}[\cdot] = \mathbb{E}[\cdot, |Y_{nk}| < \varepsilon] + \mathbb{E}[\cdot, |Y_{nk}| \geq \varepsilon]$ を用いる)

$$= iz\mathbb{E}[Y_{nk}] - \left(\frac{1}{2}z^2\mathbb{E}[Y_{nk}^2] - \frac{1}{2}z^2\mathbb{E}[Y_{nk}^2, |Y_{nk}| \geq \varepsilon]\right)$$

$$+ \frac{1}{6}|z|^3\mathbb{E}[\theta_3|Y_{nk}|^3, |Y_{nk}| < \varepsilon] + \frac{1}{2}\mathbb{E}[\theta_2 z^2 Y_{nk}^2, |Y_{nk}| \geq \varepsilon]$$

($\mathbb{E}[Y_{nk}] = 0$ および 3 乗のうち一つを ε で押さえる)

$$= -\frac{1}{2}z^2\mathbb{E}[Y_{nk}^2] + \frac{1}{2}z^2(1+\theta_2')\mathbb{E}[Y_{nk}^2, |Y_{nk}| \geq \varepsilon] + \frac{1}{6}|z|^3\theta_3'\varepsilon\mathbb{E}[|Y_{nk}|^2]$$

(ただし $|\theta_2'|, |\theta_3'| \leq 1$)

ただし指数関数の Taylor 展開を $n-1$ 次でとめたとき

$$e^{iz} = 1 + iz + \frac{(iz)^2}{2!} + \frac{(iz)^3}{3!} + \cdots + \frac{(iz)^{n-1}}{(n-1)!} + f(z)$$

の末尾の項 $f(z)$ はその n 階微分が

$$\frac{d^n}{dz^n}f(z) = \frac{d^n}{dz^n}e^{iz} = i^n e^{iz}$$

となることを用いると

$$f(z) = \int_0^z \int_0^{z_1} \int_0^{z_2} \cdots \int_0^{z_{n-1}} i^n e^{iz_n} dz_n dz_{n-1} \cdots dz_1$$

となることから

$$\begin{aligned}
|f(z)| &= \left| \int_0^z \int_0^{z_1} \int_0^{z_2} \cdots \int_0^{z_{n-1}} i^n e^{iz_n} dz_n dz_{n-1} \cdots dz_1 \right| \\
&\leq \left| \int_0^z \int_0^{z_1} \int_0^{z_2} \cdots \int_0^{z_{n-1}} |i^n e^{iz_n}| dz_n dz_{n-1} \cdots dz_1 \right| \\
&= \left| \int_0^z \int_0^{z_1} \int_0^{z_2} \cdots \int_0^{z_{n-1}} 1 dz_n dz_{n-1} \cdots dz_1 \right| \\
&= \frac{|z|^n}{n!}
\end{aligned}$$

と評価できるので，適当な複素数 θ を用いて

$$\begin{aligned}
e^{iz} &= 1 + iz + \frac{1}{2}\theta_2 |z|^2 & |\theta_2| \leq 1 \\
e^{iz} &= 1 + iz - \frac{1}{2}z^2 + \frac{1}{6}\theta_3 |z|^3 & |\theta_3| \leq 1
\end{aligned}$$

さらに一般には

$$e^{iz} = \sum_{k=0}^{n-1} \frac{(iz)^k}{k!} + \frac{\theta |z|^n}{n!} \qquad |\theta| \leq 1$$

と展開できることを使っている．またこの場合 θ_2, θ_3 は Y に依存するが，平均値の定理を用いると

$$\mathbb{E}[\theta(Y)Y^2] = \theta' \mathbb{E}[Y^2], \quad |\theta'| \leq 1$$

のように適当な θ_2', θ_3' で置き換えられることに注意する．

この α_{nk} が補題の条件を満たしていることを確認すれば証明は終わる．まず Lindeberg の条件より任意の ε に対して

$$\max_k \mathbb{E}[Y_{nk}^2, |Y_{nk}| \geq \varepsilon] \leq \sum_k \mathbb{E}[Y_{nk}^2, |Y_{nk}| \geq \varepsilon] \to 0 \quad (n \to \infty)$$

$$\begin{aligned}
\max_k \mathbb{E}[Y_{nk}^2] &= \max_k \mathbb{E}[Y_{nk}^2, |Y_{nk}| \leq \varepsilon] + \max_k \mathbb{E}[Y_{nk}^2, |Y_{nk}| \geq \varepsilon] \\
&\leq \varepsilon^2 + \sum_k \mathbb{E}[Y_{nk}^2, |Y_{nk}| \geq \varepsilon] \to \varepsilon^2 \quad (n \to \infty)
\end{aligned}$$

が成り立つ．また θ_2', θ_3' は絶対値が 1 以下なので，それぞれを 1 で置き換えたものが上界を与えることに注意すると

$$\max_k |\alpha_{nk}| \leq \frac{1}{2} z^2 \varepsilon^2 + \frac{1}{6} |z|^3 \varepsilon^3 \to 0 \quad (n \to \infty, \ \varepsilon \to 0)$$

が言える．なお，ε は任意に取ってよいのでいくらでも小さくできることに注意する．

また，$\sum \mathbb{E}[Y_{nk}^2] = 1$ に注意すると

$$\sum_k \alpha_{nk} = -\frac{1}{2}z^2 + \frac{1}{2}z^2(1+\theta_2'') \sum_k \mathbb{E}[Y_{nk}^2, |Y_{nk}| \geq \varepsilon] + \frac{1}{6}|z|^3\theta_3''\varepsilon \quad (|\theta_2''|, |\theta_3''| \leq 1)$$

であるが，これも上と同様にして

$$\sum_k \alpha_{nk} \to -\frac{1}{2}z^2 \quad (n \to \infty,\ \varepsilon \to 0)$$

であることが言える．

最後に $\sum_k \mathbb{E}[Y_{nk}^2] = \sum_k |\mathbb{E}[Y_{nk}^2]| = 1$ であることを用いると

$$\sum_k |\alpha_{nk}| \leq \frac{1}{2}z^2 + z^2 + \frac{1}{6}|z|^3\varepsilon$$

と評価できるので，n によらず有界であることが判る．

以上より，上に定めた α_{nk} という系列は補題の条件を満たしているので，T_n の特性関数は

$$\mathbb{E}[e^{izT_n}] = \prod_k \mathbb{E}[e^{izY_{nk}}] = \prod_k (1+\alpha_{nk}) \to e^{-z^2/2}$$

に収束することが判る．これは標準正規分布の特性関数

$$\frac{1}{\sqrt{2\pi}} \int e^{izx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{(x-iz)^2+z^2}{2}} dx = e^{-z^2/2}$$

と一致する． □

まとめ

練習問題 (2) (3) (4) (5) (6) (7) (8) (9) (10)

- 大数の法則とその条件
- 中心極限定理とその条件

練習問題

(1) 中心極限定理において定義された T_n の分布が $n \to \infty$ のとき標準正規分布に従うことから，S_n/n の従う分布の平均と分散を求めよ．

(2) n 個の独立な確率変数 $X_1,\ X_2, \ldots, X_n$ があり，いづれも

$$P(X_i = +1) = \frac{1}{2}$$
$$P(X_i = -1) = \frac{1}{2}$$

という確率法則に従うとき，以下の問いに答えよ．

a) 確率変数 Z を

$$Z = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

で定義するとき，その分散，および特性関数を求めよ．

b) 確率変数 Z において，$n \to \infty$ としたときの特性関数と Z の従う確率分布の密度関数を求めよ．

(3) n 個の独立な確率変数 X_1，$X_2, \ldots, X_n$ があり，いづれも確率密度関数

$$f(x) = \begin{cases} \frac{1}{2}, & -1 \leq x \leq 1 \\ 0, & \text{それ以外} \end{cases}$$

で表される分布に従うとき，以下の問いに答えよ．

a) 確率変数 Z を

$$Z = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

で定義するとき，その分散，および特性関数を求めよ．

b) 確率変数 Z において，$n \to \infty$ としたときの特性関数と Z の従う確率分布の密度関数を求めよ．

(4) $2n$ 個の独立な確率変数 U_1，$U_2, \ldots, U_{2n}$ があり，いずれも区間 $[0, 1]$ 上の一様分布に従っているとする．確率変数 Y, Z を

$$Y = \frac{U_1 - U_2 + \cdots + U_{2n-1} - U_{2n}}{2n} = \frac{1}{2n} \sum_{i=1}^{2n} (-1)^i U_i,$$

$$Z = \sqrt{n} Y$$

で定義するとき，以下の問いに答えよ．

a) Y, Z のそれぞれの特性関数を求めよ．

b) 前の問の確率変数 Y, Z において，$n \to \infty$ としたときの特性関数を求めよ．

c) 前の問の結果から確率変数 Y, Z はどのような分布に従うか答えよ．

(5) 以下の各問に答えよ．

a) 平均 0，分散 σ^2 の正規分布の特性関数を求めよ．

b) n 個の独立な確率変数 N_i，$i = 1, 2, \ldots, n$ が平均 0，分散 σ_i^2，$i = 1, 2, \ldots, n$ の正規分布に従っているとする．以下で定義する確率変数 S

$$S = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} N_i$$

の特性関数を求めよ．

c) 前問の結果から，異なる正規分布に従う独立な確率変数の和はどのような確率分布に従うか説明せよ．

d) 確率変数 X が次の式で表される確率密度関数に従っているとする．

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad (-\infty < x < \infty)$$

このとき X の特性関数を求めよ．

e) n 個の独立な確率変数 $X_i,\ i = 1, 2, \ldots, n$ が前問の確率密度関数で表される分布に従っているとする．以下で定義する確率変数 R

$$R = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

の特性関数を求めよ．

f) $n \to \infty$ とするとき「R は中心極限定理に従わない」ことを示せ．またそれは何故かを X の分散を計算することによって簡潔に説明せよ．

(6) 確率変数 U は区間 $(0, 1)$ 上の一様分布に従うとする．以下の問いに答えよ．

a) 確率変数 U の平均と分散を求めよ．

b) 確率変数 U と同じ分布に従う独立な 2 つの確率変数 U_1, U_2 を考える．

$$X = (U_1 - U_2)/2$$

とするとき，確率変数 X の平均と分散を求めよ．

c) 確率変数 N は正規分布に従い，その平均と分散は確率変数 X の平均と分散と等しいとする．確率変数 N の特性関数を求めよ．

d) 確率変数 X と同じ分布に従う独立な n 個の確率変数 $X_1, X_2, \ldots, X_n$ を考える．

$$T_n = \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}$$

とするとき，その平均と分散を求めよ．

e) 確率変数 T_n の特性関数 $\phi_n(s)$ を求めよ．

f) 特性関数 $\phi_n(s)$ の $n \to \infty$ における極限 $\lim_{n\to\infty} \phi_n(s)$ を求めよ．

(7) 確率変数 U は区間 $(-a, a)$ 上の一様分布 $(a > 0)$ に従うとする．以下の問いに答えよ．

a) 確率変数 U の分散を求めよ．

b) 確率変数 N は正規分布に従い，その平均と分散は確率変数 U の平均と分散と等しいとする．確率変数 N の特性関数 $\phi^N(s)$ を求めよ．

c) 確率変数 U と同じ分布に従う独立な n 個の確率変数 $U_1, U_2, \ldots, U_n$ を考える．

$$S_n = b \cdot \frac{U_1 + U_2 + \ldots + U_n}{n}, \quad (b > 0)$$

とするとき，S_n の分散が 1 となる b を求めよ．

d) 前問の確率変数 $U_1, U_2, \ldots, U_n$ を用いて

$$T_n = \frac{U_1 + U_2 + \ldots + U_n}{\sqrt{n}}$$

とするとき，確率変数 T_n の特性関数 $\phi^{T_n}(s)$ を求めよ．

e) 特性関数 $\phi^{T_n}(s)$ の $n \to \infty$ における極限 $\lim_{n\to\infty} \phi^{T_n}(s)$ を求めよ．

(8) 確率変数 X は，確率分布 (密度ではなく分布関数であることに注意)

$$f(x) = \begin{cases} \kappa e^{-x/\lambda}, & (x \geq 0) \\ 0, & (x < 0) \end{cases}$$

に従うとする．以下の問に答えなさい．

a) 上記の関数 f が確率密度となるように，κ を λ で表しなさい．

b) λ が与えられたときの X の平均 $\mathbb{E}[X]$ を求めなさい．

c) λ が与えられたときの X の分散 $\mathrm{Var}(X)$ を求めなさい．

d) X と同じ分布に従う n 個の独立な確率変数 $X_1, X_2, \ldots, X_n$ を得たとする．これらを用いて新たな確率変数

$$Z = \frac{\sqrt{n}(\bar{X} - \mathbb{E}[X])}{\sqrt{\mathrm{Var}(X)}}$$

を定義する．ただし

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

である．確率変数 Z の特性関数を求め，できるだけ式を見易くまとめなさい．

e) $n \to \infty$ となるとき，確率変数 Z はどのような分布に収束するか？ その確率密度関数を書き下しなさい．

(9) 確率変数 X は，確率分布 (密度ではなく分布関数であることに注意)

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

に従うとする．以下の問に答えなさい．

a) 確率変数 X の平均値

$$\mathbb{E}[X]=\sum_{k=0}^{\infty}kP(X=k)=\sum_{k=0}^{\infty}k\frac{\lambda^k e^{-\lambda}}{k!}$$

を求めなさい．
(ヒント: 指数関数のテーラー展開は以下のようになる．)

$$e^{\lambda}=1+\lambda+\frac{1}{2}\lambda^2+\frac{1}{3!}\lambda^3+\cdots=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$$

b) 確率変数 X の分散

$$\begin{aligned}\mathrm{Var}(X)&=\mathbb{E}[(X-E[X])^2]=\mathbb{E}[X^2]-\mathbb{E}[X]^2\\&=\sum_{k=0}^{\infty}k^2P(X=k)-\mathbb{E}[X]^2\end{aligned}$$

を求めなさい．
(ヒント: $\mathbb{E}[X(X-1)]$ を考えると計算が簡単になる．)

c) 確率変数 X の特性関数をを求めなさい．

d) X と同じ分布に従う n 個の独立な確率変数 $X_1, X_2, \ldots, X_n$ を得たとする．これらを用いて新たな確率変数

$$Z=\frac{\sqrt{n}(\bar{X}-\mathbb{E}[X])}{\sqrt{\mathrm{Var}(X)}}$$

を定義する．ただし

$$\bar{X}=\frac{X_1+X_2+\cdots+X_n}{n}$$

である．確率変数 Z の特性関数を求め，できるだけ式を見易くまとめなさい．

e) $n\to\infty$ となるとき，確率変数 Z はどのような分布に収束するか？ その確率密度関数を書き下しなさい．

(10) 確率変数 X は確率密度

$$f(x)=\alpha x e^{-x/\lambda}\quad(x\geq 0)$$

に従うとする．以下の問に答えなさい．

a) 以下の関係

$$\int_0^{\infty}x^n e^{-x}dx=n\int_0^{\infty}x^{n-1}e^{-x}dx$$

が成り立つことを示しなさい．

b) 関数 f が確率密度となるように，α を λ で表しなさい．
(ヒント: $n=0$ の場合を計算しておけば，$n>0$ は上の問題の結果を用いて簡単に計算できるので，あとは変数変換を行えばよい．)

c) λ が与えられたときの X の平均と分散を求めなさい．

d) λ が与えられたときの確率変数 X の特性関数をを求めなさい．
(ヒント: 以下の関係を利用して積分と微分の順序を入れ換えてよい．)

$$ixe^{isx} = \frac{d}{ds}e^{isx}$$

e) X と同じ分布に従う n 個の独立な確率変数 $X_1, X_2, \ldots, X_n$ を得たとする．これらを用いて新たな確率変数

$$Z = \sqrt{n}(\bar{X} - \mathbb{E}[X])$$

を定義する．ただし

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

で，$\mathbb{E}[X]$ は X の平均である．確率変数 Z の特性関数を求め，できるだけ式を見易くまとめなさい．

f) $n \to \infty$ となるとき，確率変数 Z はどのような分布に収束するか？ その確率密度関数を書き下しなさい．

ヒント

(2) X_i は 2 値しか取らないので，具体的に特性関数を書き下せば良い．

(3) $n \to \infty$ の極限を求めるには，Taylor 展開を適宜用いてよい．また極限操作において

$$\lim_{n\to\infty}\left(1+\frac{r(x)+s_n(x)}{n}\right)^n = e^{r(x)},$$

ただし $s_n(x)$ は $\lim_{n\to\infty} s_n(x) = 0$ となる関数

を使ってよい．

(4) $n \to \infty$ の極限を求めるには，Taylor 展開を用いて極限を取ってもよいし，中心極限定理を用いてもよい．

(5) X の特性関数を求めるには Fourier 変換・逆変換の関係，あるいは留数定理を用いる．

計算機による乱数の生成 A

最近は電子素子の熱雑音など物理的な現象を利用して乱数を生成する方法もある程度用いられるようになってきたが，最も手軽なものは線形合同法や M 系列を用いた擬似乱数の生成法である．例えば C 言語では `rand()` という関数が用意されており，一様分布に従う擬似乱数系列を取り出すことができる．また 1997 年に MT(Mersenne Twister) 法とよばれる性質の良い乱数列の生成法が提案され，モンテカルロ法などの数値実験において注目を集めている．これらの方法は規則的な方法で数字の系列を生成するため一見でたらめではあるが正しくは乱数ではないので**擬似乱数** (pseudorandom numbers) と呼ばれる．その反面繰り返し同じ系列を作り出せるため，再現性を必要とする計算機上での実験には都合がよい．

線形合同法や M 系列の詳しい説明は他書に譲るが，以下では一様擬似乱数から適当な分布に従う乱数を作る方法をいくつか挙げておく．

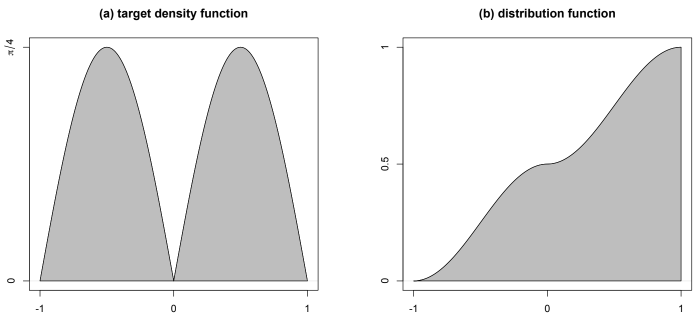

図 A.1: 以下の説明で用いる乱数の性質：目的とする確率変数は $[-1,1]$ 上で密度 $|\frac{\pi}{4}\sin(\pi x)|$ に従うものとする．(a) 確率密度関数，(b) 累積分布関数 (密度を -1 から x まで積分したもの)．

A.1 棄却法

定義 A.1 (棄却法)**.** 区間 $[a,b]$ で確率密度 $f(x)$ に従う乱数を n 個生成するためには次のような手続きを行えばよい．

step 1 計算機で生成される一様乱数を適当に線形変換して区間 $[a,b]$ 上の一様乱数 U を生成する．

step 2 同様にして区間 $[0, \max_x f(x)]$ 上の一様乱数 V を生成する．

step 3 $V \le f(U)$ なら U を受け入れ，そうでなければ捨てる．

step 4 受け入れられる U が n 個集められるまで，step 1-3 を繰り返す．

手続きが簡便なため良く用いられる．欠点としては，f の形によっては捨てられる確率が高くなり，多くの乱数を生成するときには効率が悪いことがある．

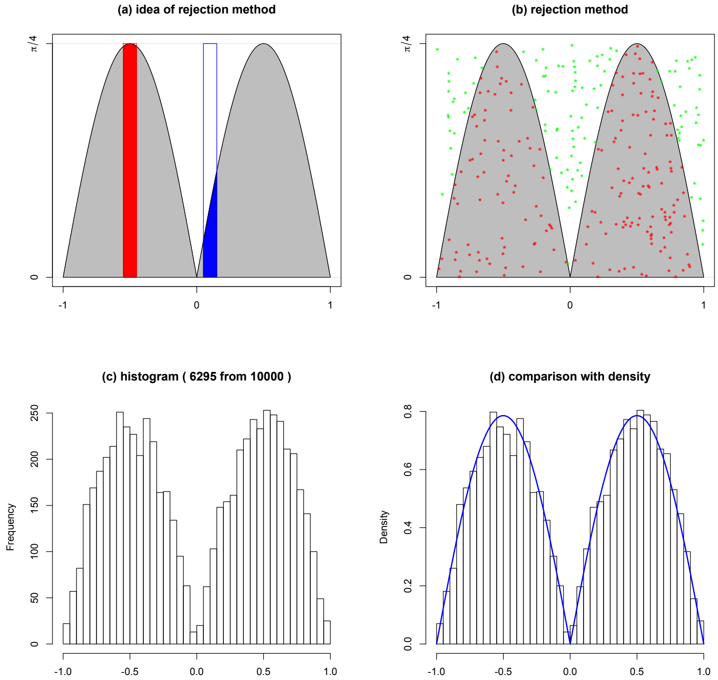

図 A.2: 棄却法による乱数の生成: (a) 棄却法では密度関数を含む 2 次元領域で一様乱数を生成し，密度関数で表される領域に入ったものだけ使われる．このため例えば密度の高い -0.5 付近 (赤い領域) ではほとんどのものが使われ，低い 0 付近 (青い領域) では棄てられるものが多い．(b) 実際の生成の様子．赤い点の x 座標が求める乱数として使われる．(c) 生成された乱数のヒストグラム (10000 点のうち使われた点のみ)．(d) 目的とする密度関数との比較．

A.2 逆関数法

定義 A.2 (逆関数法)．区間 $[a,b]$ で確率密度 $f(x)$ に従う乱数を n 個生成することを考える．

step 1 分布関数

$$F(x) = \int_a^x f(y)dy$$

を計算する．

step 2 計算機で生成される一様乱数を適当に線形変換して区間 $[0,1]$ 上の一様乱数 U を生成し，$F^{-1}(U)$ を計算する．

step 3 step 2 を繰り返し，$F^{-1}(U)$ を n 個集める．

このとき $F^{-1}(U)$ の従う確率分布は

$$\Pr(F^{-1}(U) \in E) = \Pr(U \in F(E))$$

であるが，具体的に区間 (c,d) に入る確率を計算してみると

$$\begin{aligned}
&\Pr(c < F^{-1}(U) < d) \\
&= \Pr(F(c) < U < F(d)) \\
&\quad (U \text{ は } [0,1] \text{ 上の一様分布に従っている}) \\
&= F(d) - F(c) \\
&\quad (U \text{ が区間 } [F(c), F(d)] \text{ に入る確率}) \\
&= \int_c^d f(x)dx
\end{aligned}$$

となることから，$F^{-1}(U)$ が求められる確率測度に従っていることがわかる．

この方法は区間 $[a,b]$ が有界でない，例えば $(-\infty,\infty)$ などであっても使うことができる．この点は棄却法より有利である．

一方，分布関数の逆関数が単純な関数で記述できない場合には，適当な方法により関数値を数値的に求めなくてはいけないが，その際精度と計算量が問題となることがある．

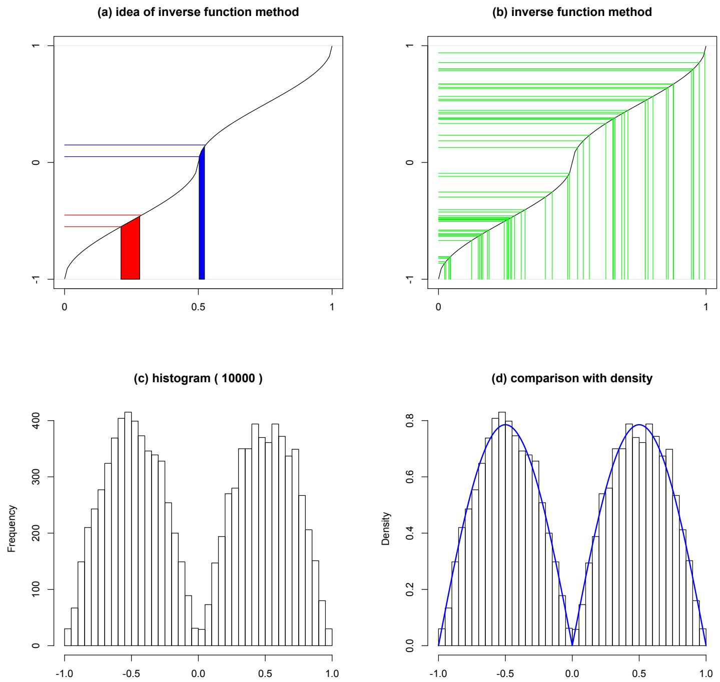

図 A.3: 逆関数法による乱数の生成: (a) 逆関数法では生成された一様乱数を，分布関数の逆関数によって変換する．逆関数の傾きに応じて変換後の密度が左右される．例えば傾きの小さい -0.5 付近に変換されるものは赤い領域で示される広い範囲のもので，傾きの大きい 0 付近に変換されるものは青い領域で示される狭い範囲のものとなる．(b) 実際の生成の様子．$[0,1]$ 上で生成された一様乱数が，どこに変換されたかを示している．(c) 生成された乱数のヒストグラム (10000 点)．(d) 目的とする密度関数との比較．

A.3 正規乱数の作り方

2 次元の標準正規分布を考える．

$$
\begin{aligned}
\Pr((X,Y)\in A) &= \int_A \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \times \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dxdy \\
&= \frac{1}{2\pi}\int_A e^{-\frac{x^2+y^2}{2}}dxdy
\end{aligned}
$$

このとき同時密度関数が二つの密度関数の積となっているので，X,Y は独立であることに注意する．X,Y を極座標 R,Θ $(0\le R<\infty, 0\le\Theta<2\pi)$ に変換すると

$$
\Pr((R,\Theta)\in A) = \frac{1}{2\pi}\int_A e^{-\frac{r^2}{2}}rdrd\theta
$$

となることから，確率変数 R,Θ は独立で，R は $[0,\infty)$ 上で確率密度が

$$
re^{-\frac{r^2}{2}}
$$

となる分布に従い，Θ は $[0,2\pi]$ 上で確率密度

$$
\frac{1}{2\pi}
$$

に従う一様分布になることがわかる．R の分布関数は

$$F(r) = \int_0^r se^{-\frac{s^2}{2}}ds = 1 - e^{-\frac{r^2}{2}}$$

であり，その逆関数は

$$F^{-1}(u) = \sqrt{-2\log(1-u)}$$

と計算できる．

定義 A.3 (Box-Muller 法 (Box-Muller transformation))．正規分布に従う乱数は次の手続きで生成することができる．

step 1 区間 $[0,1]$ 上の独立な一様乱数 U, V を生成する．

step 2 U, V を変換し

$$R = \sqrt{-2\log U}$$
$$\Theta = 2\pi V$$

R, Θ を得る．
(U が $[0,1]$ 上の一様乱数なら $1-U$ も $[0,1]$ 上の一様乱数)

step 3 さらに R, Θ を変換し

$$X = R\cos\Theta$$
$$Y = R\sin\Theta$$

二つの正規乱数 X, Y を得る．

step 4 n 個の乱数が集まるまで step 1-3 を繰り返す．

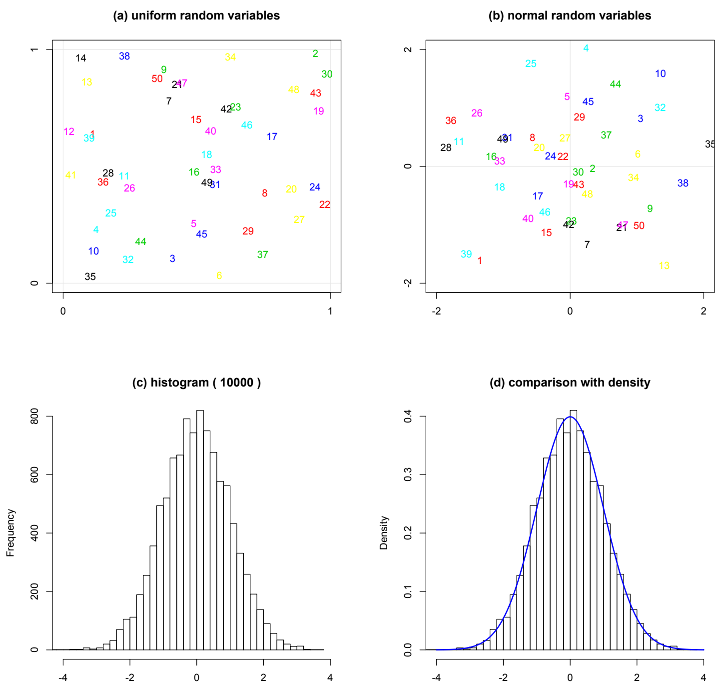

図 A.4: 一様乱数からの正規乱数の生成: (a) $[0,1]\times[0,1]$ 上の一様乱数 (50 点)，(b) 変換された一様乱数，(c) 変換後のヒストグラム (10000 点)，(d) 正規分布の密度関数との比較．

文献に関する補遺

本稿は主に

「岩波基礎数学選書 確率論」，伊藤清，岩波書店，1991.

および

「確率論」，伊藤清，岩波書店，1953.

を参考に，必要事項のみを絞り基本的な部分を概説した．時間の都合上数学的な厳密さには目を瞑っているので，確率論の理論的な側面をより詳しく，あるいはこれから確率過程をきちんと学びたい者はいずれ一読すると良いと思う．

確率の教科書としては

「確率入門」，鈴木 武，培風館，1997.

「確率 – 工学のための数学 5」，松葉 育雄，朝倉書店，2001.

「確率と統計」，藤澤 洋徳，朝倉書店，2006.

などは具体的な例が豊富で読み易く書かれていると思う．また確率論より確率過程に重きを置いてはいるが

「理工学者が書いた数学の本 確率と確率過程」，伏見正則，講談社，1987.

「物理・工学のための確率過程論」，小倉久直，コロナ社，1978.

なども参考になるかと思う．また乱数の生成法は計算機関連の教科書にまとめられているが，乱数の良さを検定する方法まで言及している

「乱数」，伏見正則，東京大学出版会，1989.

を参考に挙げておく．

確率論の基礎となる集合論や Lebesgue 積分については

「解析概論」，高木貞治，岩波書店，1961 (初版は 1938).

「ルベーグ積分入門」，伊藤清三，裳華房，1963.

「集合・位相入門」，松坂和夫，岩波書店，1968.

をはじめ多数の教科書が出ているので自分の読み易いものを探して欲しい．

最後に百科事典的ではあるが確率論の典型的な教科書として

「確率論とその応用」，W. Feller (河田龍夫監訳)，紀伊國屋書店，1960.

を挙げておく．これにはいろいろ面白い問題が載っている．

索引