

確率・統計入門

森 立平

2025-10-08

目次

前書き	2
第 I 部 確率論	3
第 1 章 はじめに	4
1.1 なぜ確率論と統計学を学ぶか？	4
1.2 本書の構成	4
1.3 その他の参考文献	4
第 2 章 集合論	5
2.1 集合	5
2.2 集合の関係	5
2.3 集合の演算	6
2.4 補集合	6
2.5 集合族	7
第 3 章 確率空間	9
3.1 確率論を数学的に定式化するには	9
3.2 有限集合上の確率空間	10
3.3 可算無限集合上の確率	10
3.4 すべての部分集合を可測にはできない	10
3.5 確率論の公理	11
3.6 確率の性質	12
第 4 章 確率変数と確率分布	15
4.1 確率変数	15
4.2 確率分布	16
4.3 離散確率分布の例	18
4.3.1 様々な離散確率分布	18
4.3.2 二項分布	19
4.3.3 幾何分布	19
4.3.4 超幾何分布	19

4.3.5	ポアソン分布	20
4.3.6	確率空間は何か?	21
4.4	連続確率分布の例	21
4.4.1	一様分布	21
4.4.2	正規分布 (ガウス分布)	21
4.4.3	指数分布	22
4.5	確率密度関数	22
第 5 章	期待値、分散、モーメント	24
5.1	期待値	24
5.2	分散	26
5.3	モーメントとモーメント母関数	27
第 6 章	複数の確率変数、多変量正規分布	29
6.1	複数の確率変数	29
6.2	条件付き確率	29
6.3	多変量正規分布	29
第 II 部	統計学	30
第 7 章	ベイズ推定	31
第 8 章	仮説検定	32
第 III 部	漸近論	33
第 9 章	大数の法則と集中不等式	34
第 10 章	正規分布と中心極限定理	35
第 11 章	サノフの定理、KL ダイバージェンス	36
参考文献		37

前書き

これは確率論と統計学の入門書である。確率論を数学的に取り扱うには通常は測度論とルベーグ積分を用いる。本書では測度論を学ぶ前の数学専攻の学生を対象に確率論と統計学の基礎を解説する。測度論とルベーグ積分を省略するため、しばしば積分と極限の交換などの等式を証明なしに用いる。後で測度論を学んだ後にぜひ振り返って欲しい。

第 I 部

確率論

第 1 章

はじめに

1.1 なぜ確率論と統計学を学ぶか？

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

1.2 本書の構成

1.3 その他の参考文献

第 2 章

集合論

2.1 集合

集合は対象の集まりである。例えば

$$A = \{1, 2, 3\}, \quad B = \{\text{赤}, \text{青}, \text{黄}, \text{緑}\}$$

のように表す。集合を構成するものを**要素**もしくは**元**という。

- $x \in A$ で要素 x は集合 A に含まれる
- $x \notin A$ で要素 x は集合 A に含まれない

を表す。

$|A|$ で集合 A の要素数を表すことにする。要素数が無限の集合を考えることもできる。例えば

$$A = \{n \in \mathbb{N} \mid n \text{ は偶数}\}, \quad B = \{x \in \mathbb{R} \mid x \text{ は無理数}\}$$

という集合は無限集合の例となる。何も要素を持たない集合 (要素数が零の集合) を**空集合**といい、 \emptyset で表す。

2.2 集合の関係

集合 A の要素がすべて集合 B に含まれるとき、 A を B の**部分集合** という。部分集合に関連する集合関係について記号を以下のように定義する。

$$\begin{aligned} A \subseteq B &\stackrel{\text{def}}{\iff} \forall x (x \in A \implies x \in B) && (A \text{ は } B \text{ の部分集合}) \\ A \supseteq B &\stackrel{\text{def}}{\iff} B \subseteq A && (A \text{ は } B \text{ の上位集合}) \\ A = B &\stackrel{\text{def}}{\iff} (A \subseteq B \wedge A \supseteq B) && (A \text{ と } B \text{ は等しい}) \\ A \neq B &\stackrel{\text{def}}{\iff} \neg(A = B) && (A \text{ と } B \text{ は等しくない}) \end{aligned}$$

また、

$$A \subsetneq B \stackrel{\text{def}}{\iff} (A \subseteq B \wedge A \neq B) \quad (A \text{ は } B \text{ の真部分集合})$$

$$A \supsetneq B \stackrel{\text{def}}{\iff} B \subsetneq A \quad (A \text{ は } B \text{ の真上位集合})$$

とする。部分集合や上位集合のような集合間の関係を**包含関係**という。集合 A, B について、 $A \subseteq B$ と $B \subseteq A$ のどちらかが成り立つとき、「 A と B の間に包含関係が成り立つ」という。

包含関係は以下の三条件を満たす。

- (反射律) $A \subseteq A$.
- (反対称律) $(A \subseteq B \wedge B \subseteq A) \iff A = B$.
- (推移律) $(A \subseteq B \wedge B \subseteq C) \implies A \subseteq C$.

2.3 集合の演算

複数の集合から新しい集合を作る演算がある。

$$A \cup B := \{x \mid x \in A \text{ または } x \in B\} \quad (\text{和集合})$$

$$A \cap B := \{x \mid x \in A \text{ かつ } x \in B\} \quad (\text{積集合})$$

$$A \setminus B := \{x \mid x \in A \text{ かつ } x \notin B\} \quad (\text{差集合})$$

これらの演算は以下の法則を満たす。

- (交換法則) $A \cup B = B \cup A, \quad A \cap B = B \cap A$.
- (結合法則) $(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C)$.
- (分配法則) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- (冪等法則) $A \cup A = A, \quad A \cap A = A$.
- (吸収法則) $A \cup (A \cap B) = A, \quad A \cap (A \cup B) = A$.
- $A \cup \emptyset = A, \quad A \cap \emptyset = \emptyset$.

和集合と積集合は結合法則を満たすことから、括弧を使わずに $A \cup B \cup C$ と表すことができる。集合 A と B が $A \cap B = \emptyset$ を満たすとき、**確率論文脈では、「 A と B は排反である」**(集合論文脈では「互いに素である」) という。

2.4 補集合

全体の集合 Ω というのが文脈上存在する場合は、

$$A^c := \Omega \setminus A \quad (\text{補集合})$$

と定義する。

補集合に関連して以下の法則が成り立つ。

- $\Omega^c = \emptyset, \quad \emptyset^c = \Omega.$
- $A \cup \Omega = \Omega, \quad A \cap \Omega = A.$
- $A \cup A^c = \Omega, \quad A \cap A^c = \emptyset.$
- (二重補集合の法則) $(A^c)^c = A.$
- (ド・モルガンの法則) $(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c$

2.5 集合族

集合 Ω について 2^Ω を Ω のすべての部分集合からなる集合を表す。

$$2^\Omega := \{A \subseteq \Omega\}.$$

これを Ω の**冪集合**という。例えば $\Omega = \{1, 2, 3\}$ のとき、

$$2^\Omega = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$$

である。 Ω が有限集合のとき、 $|2^\Omega| = 2^{|\Omega|}$ が成り立つ。また、 2^Ω の部分集合を Ω 上の**(部分) 集合族**と呼ぶ。集合 Λ の各 $\lambda \in \Lambda$ に対して集合 $A_\lambda \subseteq \Omega$ が存在するとき、集合族

$$\{A_\lambda \subseteq \Omega \mid \lambda \in \Lambda\} \subseteq 2^\Omega$$

を**添字集合 Λ で添字付けられた Ω 上の集合族**という。添字集合が有限集合の場合は集合族全体の和集合や積集合は二つの集合の和集合と積集合の定義を繰り返し用いることで定義できる。それらは以下のように表す。

$$\bigcup_{\lambda \in \Lambda} A_\lambda, \quad \bigcap_{\lambda \in \Lambda} A_\lambda.$$

添字集合 Λ が無限集合の場合は和集合と積集合を以下で定義する。

$$\begin{aligned} \bigcup_{\lambda \in \Lambda} A_\lambda &:= \{x \in \Omega \mid \exists \lambda \in \Lambda, \quad x \in A_\lambda\} \\ \bigcap_{\lambda \in \Lambda} A_\lambda &:= \{x \in \Omega \mid \forall \lambda \in \Lambda, \quad x \in A_\lambda\} \end{aligned}$$

この定義は Λ が有限集合の場合も正しいものである。この場合もド・モルガンの法則は成り立つ。つまり、

$$\begin{aligned} \left(\bigcup_{\lambda \in \Lambda} A_\lambda \right)^c &= \bigcap_{\lambda \in \Lambda} A_\lambda^c \\ \left(\bigcap_{\lambda \in \Lambda} A_\lambda \right)^c &= \bigcup_{\lambda \in \Lambda} A_\lambda^c \end{aligned}$$

が成り立つ。

証明: $x \in \Omega$ について、

$$\begin{aligned}x \in \left(\bigcup_{\lambda \in \Lambda} A_\lambda \right)^c &\iff x \notin \left(\bigcup_{\lambda \in \Lambda} A_\lambda \right) \\&\iff \neg (\exists \lambda \in \Lambda, \quad x \in A_\lambda) \\&\iff \forall \lambda \in \Lambda, \quad x \notin A_\lambda \\&\iff \forall \lambda \in \Lambda, \quad x \in A_\lambda^c \\&\iff x \in \bigcap_{\lambda \in \Lambda} A_\lambda^c.\end{aligned}$$

第 3 章

確率空間

3.1 確率論を数学的に定式化するには

確率は身近に現れる (感じられる) ものであるが、それを数学的に定式化することは自明な問題ではない。実際に確率論には複数の数学的定式化が存在する。その中で圧倒的に一般的なのが測度論的確率論と呼ばれる定式化である。測度というのは集合の「面積」のようなものであり、確率を測度と捉えるのが測度論的確率論である。これは多くの人間の直感にも自然なものであろう。

まず、確率を考える集合について考えよう。例えばコインを投げて表もしくは裏が出る確率を考えたいときは

$$\Omega = \{H, T\}$$

という集合になる。また、明日の天気確率を考えたいときは

$$\Omega = \{ \text{” 晴”}, \text{” 雨”}, \text{” 雪”} \}$$

という集合になるだろう。この Ω の部分集合に確率を与える関数 $P: 2^\Omega \rightarrow \mathbb{R}$ を定義しよう。偏りのないコインの場合は以下ようになる。

$$P(\emptyset) = 0, \quad P(\{H\}) = \frac{1}{2}, \quad P(\{T\}) = \frac{1}{2}, \quad P(\{H, T\}) = 1.$$

また、天気の場合は例えば以下ようになる。

$$\begin{aligned} P(\emptyset) &= 0, & P(\{\text{晴}\}) &= 0.7, & P(\{\text{雨}\}) &= 0.2, & P(\{\text{雪}\}) &= 0.1 \\ P(\{\text{晴}, \text{雨}\}) &= 0.9, & P(\{\text{雨}, \text{雪}\}) &= 0.3, & P(\{\text{雪}, \text{晴}\}) &= 0.8, & P(\{\text{晴}, \text{雨}, \text{雪}\}) &= 1. \end{aligned}$$

このように集合 Ω の部分集合に確率を与えることを考える。 Ω の要素一つずつに確率を与えれば十分であるようにも思えるが、 Ω が連続的な場合には Ω の一つの要素の確率は 0 になってしまうことが多い。例えば明日の正午の気温が $10\pi^\circ\text{C}$ になる確率は 0 であろう。そのため、 Ω の要素ではなく部分集合に確率を与えることにする。そのために測度論が適しているわけである。

3.2 有限集合上の確率空間

確率を考える集合を Ω とする。この Ω のことを**標本空間**という。また、 Ω の部分集合のことを**事象**という。そして、事象に確率を与える関数 $P: 2^\Omega \rightarrow \mathbb{R}_{\geq 0}$ を**確率測度**という。確率測度は以下の条件を満たす。

1. $P(\Omega) = 1$.
2. $\forall A, B \subseteq \Omega, \quad A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$.

一つ目の条件は全体の確率が1であるという意味の条件である。二つ目の条件は排反な二つの事象の和集合の確率はそれぞれの事象の確率の和であるという意味の条件である。この二つ目の条件を**有限加法性**という。例えば

$$P(\{\text{"晴"}, \text{"雨"}\}) = P(\{\text{"晴"}\}) + P(\{\text{"雨"}\})$$

という等式は「晴れもしくは雨になる確率 = 晴れになる確率 + 雨になる確率」という意味の等式になる。よって有限加法性が自然な条件であることが分かるだろう。また、これらのことから、 P は各要素 $\omega \in \Omega$ に対する確率 $P(\{\omega\})$ から一意に定まることが分かる。この標本空間と確率測度のペア (Ω, P) を**確率空間**という。

3.3 可算無限集合上の確率

標本空間 Ω が可算無限集合のときも、同様に確率測度を定義することもできるが、ここではより強い以下の条件を考える。

1. $P(\Omega) = 1$.
2. $\forall (A_n \subseteq \Omega)_{n \geq 0}, \quad \forall i \neq j, A_i \cap A_j = \emptyset \implies P\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n \geq 0} P(A_n)$.

ここで二つ目の条件における $n \geq 0$ という添字において n は非負整数全体をわたる。今後もこの記法を用いる。この二つ目の条件を**完全加法性**もしくは **σ -加法性**という。右辺の無限和において $P(A_n) \geq 0$ なので、無限和 $\sum_{n \geq 0} P(A_n)$ が存在するときこれは絶対収束する。よって、この無限和は事象列の並び順に依存しないことが分かる。完全加法性ではなく有限加法性だけを使って確率論を構築する試みもあるが、確率測度の連続性などの性質が失なわれるため標準的な確率論では完全加法性を課す。

3.4 すべての部分集合を可測にはできない

より一般に Ω が非可算無限集合の場合を考えよう。例えば $\Omega = [0, 1)$ の場合が考えられる。このとき、すべての $A \subseteq \Omega$ に確率を与えることができるのだろうか？ そうすると、ごく自然な性質を持つような確率測度が**存在しない**ことが、選択公理を認めると示される。

定理 3.1 (非可測集合の存在). $\Omega = [0, 1)$ とする。また、集合の平行移動を

$$A + x := \{a + x - \lfloor a + x \rfloor \mid a \in A\}$$

と定義する。このとき、

$$\forall x \in \Omega, A \subseteq \Omega, \quad P(A+x) = P(A) \quad (\text{平行移動不変性})$$

を満たす確率測度 $P: 2^{[0,1)} \rightarrow \mathbb{R}_{\geq 0}$ は存在しない。

証明. Ω 上の同値関係を $x \sim y \stackrel{\text{def}}{\iff} x-y \in \mathbb{Q}$ と定義する。選択公理より、この同値関係の同値類から一つずつ要素を含む集合 V が存在する (Vitali 集合)。任意の $x \in [0, 1)$ について、ある $z \in V$ が唯一存在して $x \sim z$ である。よって、任意の $x \in [0, 1)$ について、ある $z \in V$ と $q \in \mathbb{Q} \cap [0, 1)$ が唯一存在して $x = z + q - \lfloor z + q \rfloor$ であることから

$$[0, 1) = \bigcup_{q \in \mathbb{Q} \cap [0, 1)} (V + q)$$

であり、右辺は互いに排反である。よって条件を満たす P が存在すると仮定すると、

$$\begin{aligned} 1 = P([0, 1)) &= P\left(\bigcup_{q \in \mathbb{Q} \cap [0, 1)} (V + q)\right) \\ &= \sum_{q \in \mathbb{Q} \cap [0, 1)} P(V + q) \quad (\text{完全加法性}) \\ &= \sum_{q \in \mathbb{Q} \cap [0, 1)} P(V) \quad (\text{平行移動不変性}) \end{aligned}$$

ここで $P(V)$ をどのように定めても、それを無限回足して 1 にすることはできない。よって P は存在しない。 \square

3.5 確率論の公理

定理 3.1 より、

1. 選択公理
2. 確率測度の完全加法性
3. $[0, 1)$ 上の平行移動不変な確率測度
4. Ω のすべての部分集合に確率を与える

のどれかを諦めないといけない。標準的な確率論では 4 を諦める。以下に確率空間の厳密な定義を述べる。

定義 3.1 (完全加法族). Ω を集合とする。 Ω 上の集合族 $\mathcal{F} \subseteq 2^\Omega$ が以下を満たすとする。

1. $\Omega \in \mathcal{F}$.
2. $\forall A \in \mathcal{F}, \quad A^c = \Omega \setminus A \in \mathcal{F}$.
3. $\forall (A_n \in \mathcal{F})_{n \geq 0}, \quad \bigcup_{n \geq 0} A_n \in \mathcal{F}$.

このとき、 \mathcal{F} を Ω 上の完全加法族もしくは σ -加法族という。

定義 3.2 (確率空間). Ω を集合とし、 $\mathcal{F} \subseteq 2^\Omega$ を Ω 上の完全加法族とする。

また、 $P: \mathcal{F} \rightarrow \mathbb{R}_{\geq 0}$ が以下を満たすとする。

1. $P(\Omega) = 1$.

2. $\forall (A_n \in \mathcal{F})_{n \geq 0}, \quad \forall i \neq j, A_i \cap A_j = \emptyset \implies P\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n \geq 0} P(A_n)$.

このとき、 (Ω, \mathcal{F}, P) を確率空間という。また、 Ω を標本空間、 \mathcal{F} を事象集合、 P を確率測度という。

このように事象集合 \mathcal{F} の元についてのみ確率が与えられる。

ノート

ペア (Ω, \mathcal{F}) を可測空間という。確率空間 (Ω, \mathcal{F}, P) から $P(\Omega) = 1$ の条件を除いたものが一般の測度空間である。

定理 3.1 の証明では選択公理を用いたが、実際に選択公理を認めないと定理 3.1 が成立しないことが分かっている。以降、 Ω が非可算無限集合である場合も含めて、 $\mathcal{F} = 2^\Omega$ だと思ふことにする。これは正しくない場合もあるのだが、選択公理を使わない限り矛盾は導かれないので、問題になることはほとんどない。よって、以降は (Ω, P) を確率空間とする。

$\Omega = [0, 1)$ で任意の $0 \leq a < b \leq 1$ について $P([a, b)) = b - a$ であるような確率空間はとても基本的なものである。今後この性質を満たす確率空間が存在すると仮定して話を進める。

3.6 確率の性質

補題 3.1 (確率のいくつかの性質). 確率空間 (Ω, P) と任意の $A, B \subseteq \Omega$ について以下が成り立つ。

1. $P(A^c) = 1 - P(A)$.
2. $B \subseteq A \implies P(B) \leq P(A)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
4. $P(A \cup B) \leq P(A) + P(B)$ (ブールの不等式、union bound).

証明.

- 1.

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c).$$

- 2.

$$P(A) = P(B \cup (A \setminus B)) = P(B) + P(A \setminus B) \geq P(B).$$

- 3.

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

より、 $P(B \setminus A)$ を消去することで得られる。

4. 3 より自明

□

補題 3.2 (ユニオンバウンド). 確率空間 (Ω, P) と $(A_n \subseteq \Omega)_{n \geq 0}$ について、

$$P\left(\bigcup_{n \geq 0} A_n\right) \leq \sum_{n \geq 0} P(A_n).$$

証明. $B_0 := A_0$, $B_n := A_n \setminus \bigcup_{k=0}^{n-1} A_k$ とおくと、

$$\begin{aligned} P\left(\bigcup_{n \geq 0} A_n\right) &= P\left(\bigcup_{n \geq 0} B_n\right) \\ &= \sum_{n \geq 0} P(B_n) \\ &\leq \sum_{n \geq 0} P(A_n). \end{aligned}$$

□

定理 3.2 (確率測度の連続性). 確率空間 (Ω, P) と事象列 $A_0 \subseteq A_1 \subseteq \dots \subseteq \Omega$ について

$$P\left(\bigcup_{n \geq 0} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

また、事象列 $\Omega \supseteq A_0 \supseteq A_1 \supseteq \dots$ について

$$P\left(\bigcap_{n \geq 0} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

証明. 事象列 $A_0 \subseteq A_1 \subseteq \dots \subseteq \Omega$ について考える。 $B_0 := A_0$, $n \geq 1$ について $B_n := A_n \setminus A_{n-1}$ とおく。このとき、 $i \neq j$ について $B_i \cap B_j = \emptyset$ 。また $k \geq 0$ について、 $\bigcup_{n=0}^k B_n = \bigcup_{n=0}^k A_n = A_k$ が成り立つ。

$$\begin{aligned} P\left(\bigcup_{n \geq 0} A_n\right) &= P\left(\bigcup_{n \geq 0} B_n\right) = \sum_{n \geq 0} P(B_n) \\ &= \lim_{k \rightarrow \infty} \sum_{n=0}^k P(B_n) = \lim_{k \rightarrow \infty} P\left(\bigcup_{n=0}^k B_n\right) = \lim_{k \rightarrow \infty} P(A_k). \end{aligned}$$

事象列 $\Omega \supseteq A_0 \supseteq A_1 \supseteq \dots$ について考える。このとき、 $A'_n := A_n^c$ とおくと、 $A'_0 \subseteq A'_1 \subseteq \dots \subseteq \Omega$ を満たす。よって、

$$P\left(\bigcup_{n \geq 0} A'_n\right) = \lim_{n \rightarrow \infty} P(A'_n)$$

である。ド・モルガンの公式より、

$$\begin{aligned} P\left(\left(\bigcap_{n \geq 0} A_n\right)^c\right) &= \lim_{n \rightarrow \infty} P(A_n^c) \\ \Leftrightarrow 1 - P\left(\bigcap_{n \geq 0} A_n\right) &= \lim_{n \rightarrow \infty} (1 - P(A_n)) \\ \Leftrightarrow P\left(\bigcap_{n \geq 0} A_n\right) &= \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

□

第 4 章

確率変数と確率分布

4.1 確率変数

確率空間の上の様々な部分集合の確率を調べたい。そのためには確率変数を導入すると便利である。

定義 4.1 (確率変数). 確率空間 (Ω, P) について、関数 $X: \Omega \rightarrow \mathbb{R}$ を**確率変数**という。

また、任意の $A \subseteq \mathbb{R}$ について

$$\Pr(X \in A) := P(\{\omega \in \Omega \mid X(\omega) \in A\})$$

と定義する。さらに、任意の $a \in \mathbb{R}$ について

$$\Pr(X \geq a) := P(X \in \{x \in \mathbb{R} \mid x \geq a\}) = P(\{\omega \in \Omega \mid X(\omega) \geq a\})$$

と定義する。同様に $\Pr(X > a)$, $\Pr(X \leq a)$, $\Pr(X < a)$, $\Pr(X = a)$ なども定義される。

ノート

本当の確率論では確率変数は「確率空間から位相空間への写像で開集合の逆像が可測集合になるような関数」として定義される。また、文献によっては確率変数は「確率空間から可測空間への写像で可測集合の逆像が可測集合になるような関数」として定義されることもある。開集合族が生成する完全加法族をボレル集合族と呼ぶが、この方法で位相空間から可測空間を作ることができる。そうすると前者の定義は「確率空間から位相空間への写像でボレル集合の逆像が可測集合になるような関数」と等価なので、後者の定義に含まれる。後者の定義の中の可測空間が位相空間由来のものに限定したのが前者の定義であると言える。

任意の関数 $f: \mathbb{R} \rightarrow \mathbb{R}$ について $f \circ X$ は確率変数である。このとき、 $\Pr(f \circ X \in A)$ と書く代わりに $\Pr(f(X) \in A)$ と書く。

i ノート

本当の確率論の言葉で述べると任意のボレル関数 (開集合 (ボレル集合としても等価) の逆像がボレル集合となる関数) f について $f \circ X$ は確率変数である。

共通の確率空間 (Ω, P) 上の確率変数 X, Y についてその和 $X + Y$ や積 XY も確率変数である。

例 4.1 (コイン投げ). コインを 2 回独立に投げる場合の確率空間 (Ω, P) を以下で定義する。

- $\Omega = \{HH, HT, TH, TT\}$.
- $P(A) = \frac{|A|}{4}$.

確率変数 X_1 と X_2 を

$$\begin{array}{llll} X_1(HH) = 1 & X_1(HT) = 1 & X_1(TH) = 0 & X_1(TT) = 0 \\ X_2(HH) = 1 & X_2(HT) = 0 & X_2(TH) = 1 & X_2(TT) = 0 \end{array}$$

とすると、 $k \in \{1, 2\}$ について、 X_k は k 回目のコイン投げが表だったときに 1、裏だったときに 0 になる。また、 $X = X_1 + X_2$ は表が出た回数を表す確率変数となる。表が一回以上出る確率は以下のように表せる。

$$\Pr(X \geq 1) = P(\{\omega \in \Omega \mid X(\omega) \geq 1\}) = P(\{HH, HT, TH\}) = \frac{3}{4}.$$

4.2 確率分布

確率変数 X の像が高々可算集合のとき、 X を**離散型確率変数**という。離散型確率変数でない確率変数を**連続型確率変数**という。

X が離散型確率変数のとき、 A を X の像の部分集合とすると

$$\begin{aligned} \Pr(X \in A) &= P(\{\omega \in \Omega \mid X(\omega) \in A\}) \\ &= P\left(\bigcup_{x \in A} \{\omega \in \Omega \mid X(\omega) = x\}\right) \\ &= \sum_{x \in A} P(\{\omega \in \Omega \mid X(\omega) = x\}) \\ &= \sum_{x \in A} \Pr(X = x) \end{aligned}$$

であるので、

$$f_X(x) := \Pr(X = x) \quad \forall x \in \mathbb{R}$$

という関数を用いて

$$\Pr(X \in A) = \sum_{x \in A} f_X(x)$$

と表せる。この $f_X(x)$ を X の**確率質量関数**という。

X が連続型確率変数のとき、

$$\Pr(X \in A) = \int_{x \in A} f_X(x) dx \quad \forall A \subseteq \mathbb{R}$$

を満たす $f_X: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ が存在するとき、この $f_X(x)$ を X の**確率密度関数**という。しかし、確率密度関数は常に存在するとは限らない。

例 4.2. $\Omega = [0, 1)$ とし、 $P: 2^\Omega \rightarrow \mathbb{R}_{\geq 0}$ を $P([a, b)) = b - a$ を満たす確率測度とする。確率空間 (Ω, P) 上の確率変数 $X: [0, 1) \rightarrow \mathbb{R}$ を

$$X(\omega) = \begin{cases} 2\omega & \text{if } \omega < \frac{1}{2} \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

と定義する。

この時 X の像は $[0, 1)$ であるので X は連続型確率変数である。

また、 $\Pr(X = \frac{1}{2}) = P([\frac{1}{2}, 1) \cup \{\frac{1}{4}\}) = \frac{1}{2}$ である。よって X の確率密度関数は存在しない。

一方で、任意の確率変数 X について

$$F_X(x) := \Pr(X \leq x) \quad \forall x \in \mathbb{R}$$

を確率変数 X の**累積分布関数**という。累積分布関数は常に存在する。連続確率変数 X が確率密度関数を持つとき、

$$F_X(x) = \int_{-\infty}^x f_X(z) dz$$

という関係が成り立つ。よって累積分布関数が微分可能なときは

$$f_X(x) = \frac{dF_X}{dx}$$

とおいても構わない。確率密度関数はその積分値にのみ意味を持つ。そのため確率密度関数は一意には定まらない。

累積分布関数は定義より単調非減少関数であることが分かる。累積分布関数は連続とは限らないが右連続であることは以下のように確認できる。

$$\begin{aligned}
\lim_{n \rightarrow \infty} \Pr \left(X \leq x + \frac{1}{n} \right) &= \lim_{n \rightarrow \infty} P \left(\left\{ \omega \in \Omega \mid X(\omega) \leq x + \frac{1}{n} \right\} \right) \\
&= P \left(\bigcap_{n=0}^{\infty} \left\{ \omega \in \Omega \mid X(\omega) \leq x + \frac{1}{n} \right\} \right) \\
&= P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \\
&= \Pr(X \leq x)
\end{aligned}$$

例 4.2 の確率変数 X の累積分布関数は以下ようになる。

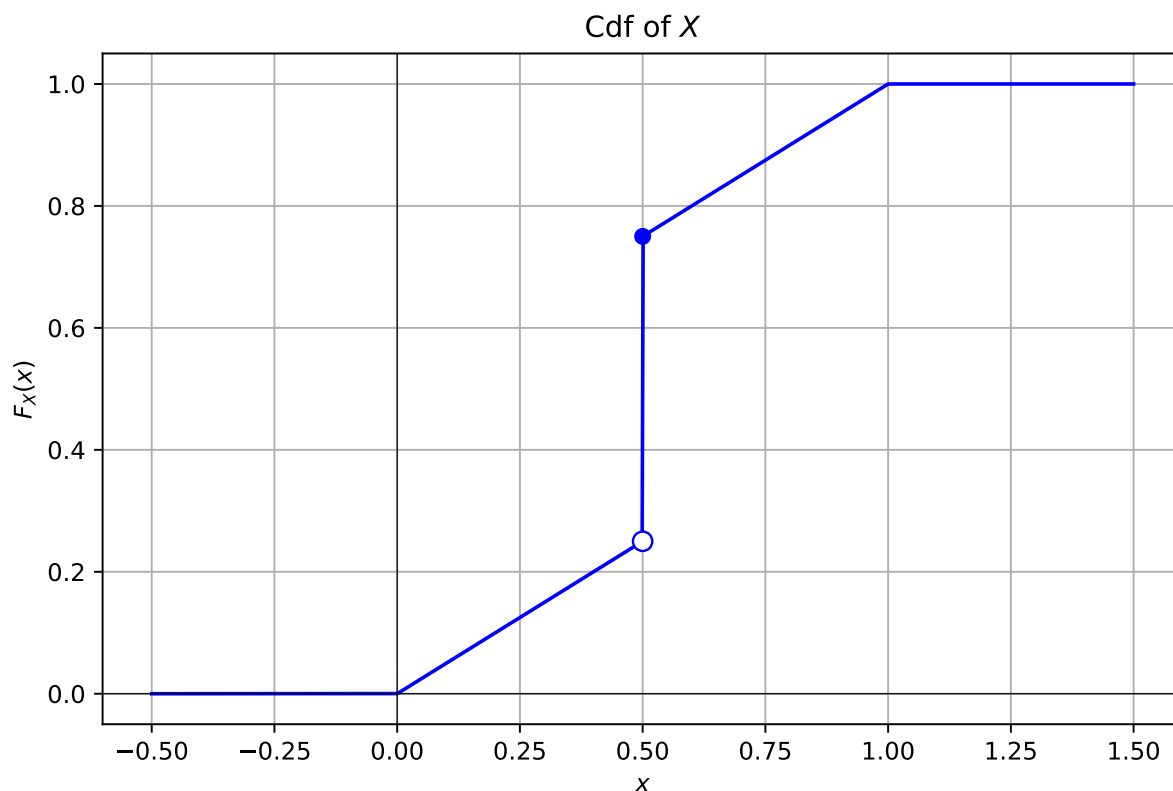


図 4.1: X の累積分布関数

4.3 離散確率分布の例

4.3.1 様々な離散確率分布

- 二項分布 $\text{Binom}(n, p)$: 表が出る確率が p のコインを n 回独立に投げたとき、表が出る回数の分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- 幾何分布: 表が出るまでに投げるコインの回数の分布

$$\Pr(X = k) = (1 - p)^{k-1}p.$$

- 超幾何分布: 袋の中に N 個のボールがあって、そのうち K 個が当たりとし、 n 個引いたときの当たりの個数の分布

$$\Pr(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}.$$

- ポアソン分布:

$$\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

4.3.2 二項分布

二項分布の確率質量関数は

$$f_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

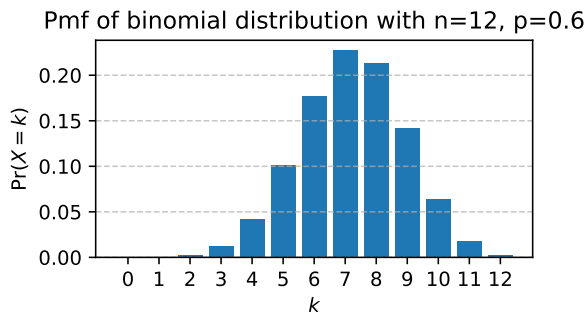


図 4.2: 二項分布の確率質量関数

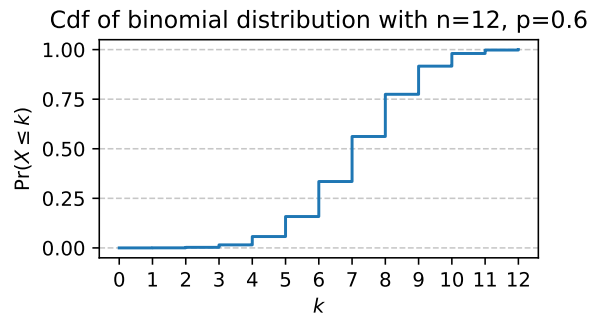


図 4.3: 二項分布の累積分布関数

4.3.3 幾何分布

$$f_X(k) = (1 - p)^{k-1}p, \quad \forall k \geq 1$$

4.3.4 超幾何分布

$$f_X(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad \forall k \in \{0, \dots, \min\{K, n\}\}$$

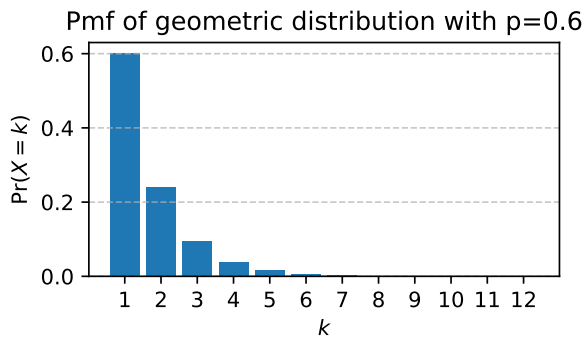


図 4.4: 幾何分布の確率質量関数

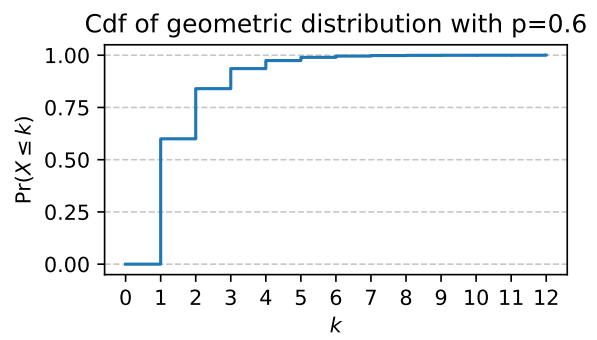


図 4.5: 幾何分布の累積分布関数

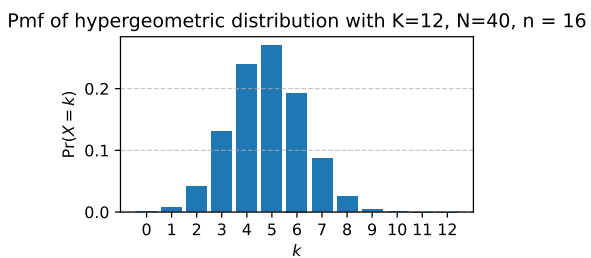


図 4.6: 超幾何分布の確率質量関数

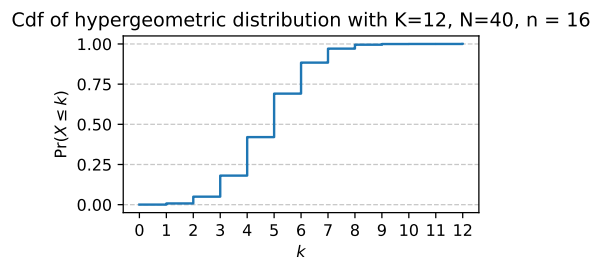


図 4.7: 超幾何分布の累積分布関数

4.3.5 ポアソン分布

$$f_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \forall k \geq 0$$

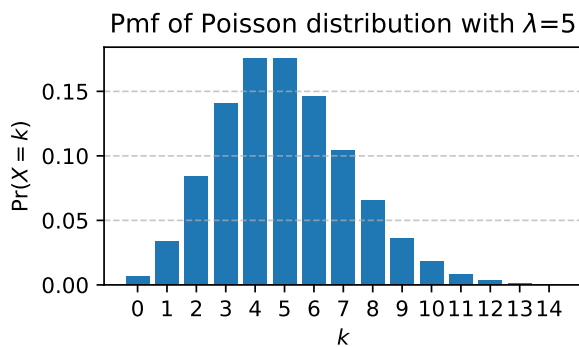


図 4.8: ポアソン分布の確率質量関数

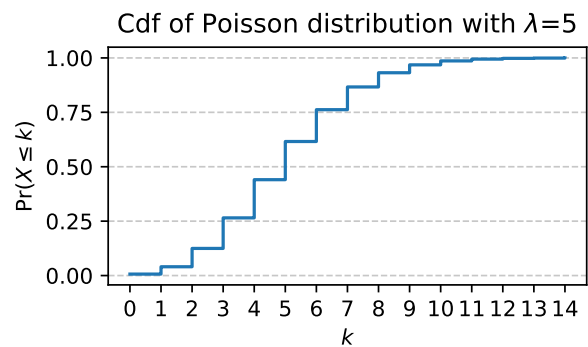


図 4.9: ポアソン分布の累積分布関数

4.3.6 確率空間は何か？

確率変数が従う確率分布のみに注目した議論をする場合、確率空間を陽に考えないことがしばしばある。

- 二項分布: n 回のコイン投げ、もしくは無限回のコイン投げ。
- 幾何分布: 無限回のコイン投げ。
- 超幾何分布: K 個の当たりと $N - K$ 個のはずれのボールの $\binom{N}{K}$ 通りの並び順。
- ポアソン分布: ？

4.4 連続確率分布の例

- 一様分布
- 正規分布
- 指数分布

4.4.1 一様分布

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases},$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } x \in [a, b] \\ 1 & \text{otherwise} \end{cases}.$$

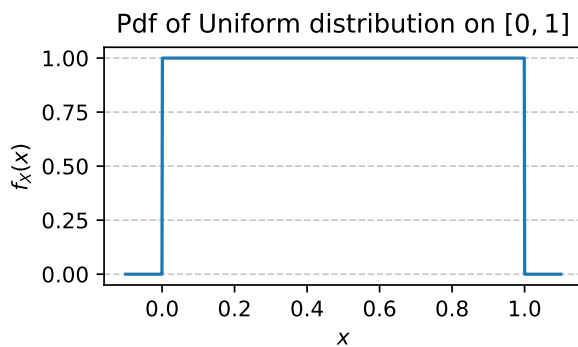


図 4.10: 一様分布の確率密度関数

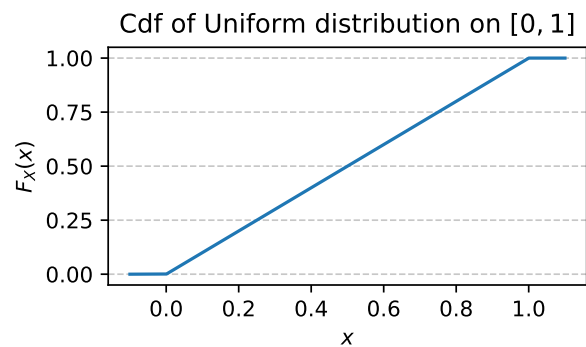


図 4.11: 一様分布の累積分布関数

4.4.2 正規分布 (ガウス分布)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

特に $\mu = 0, \sigma = 1$ のとき標準正規分布という。

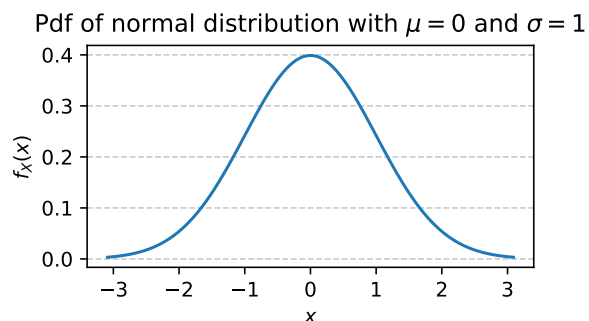


図 4.12: 標準正規分布の確率密度関数

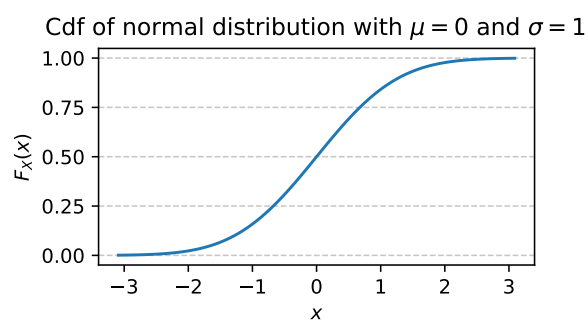


図 4.13: 標準正規分布の累積分布関数

4.4.3 指数分布

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{otherwise} \end{cases},$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{otherwise} \end{cases}$$

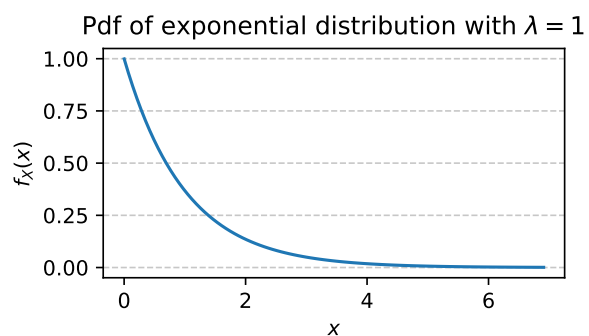


図 4.14: 指数分布の確率密度関数

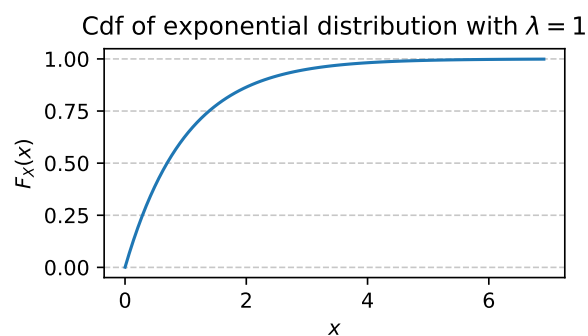


図 4.15: 指数分布の累積分布関数

4.5 確率密度関数

補題 4.1. 確率変数 X について、

$$\begin{aligned} f_{X+a}(x) &= f_X(x-a) & \forall a \in \mathbb{R} \\ f_{aX}(x) &= \frac{1}{|a|} f_X(x/a) & \forall a \in \mathbb{R}_{\neq 0} \end{aligned}$$

はそれぞれ $X+a$ と aX の確率密度関数になる。

証明. 関数 $f_Z: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ が

$$\Pr(Z \leq x) = \int_{-\infty}^x f_Z(z) dz$$

を満たすとき、 f_Z は確率変数 Z の確率密度関数となる。

任意の $a \in \mathbb{R}$ について、

$$\begin{aligned} \Pr(X + a \leq x) &= \Pr(X \leq x - a) \\ &= \int_{-\infty}^{x-a} f_X(z) dz \\ &= \int_{-\infty}^x f_X(z' - a) dz' \quad (z' = z + a). \end{aligned}$$

任意の $a > 0$ について、

$$\begin{aligned} \Pr(aX \leq x) &= \Pr(X \leq x/a) \\ &= \int_{-\infty}^{x/a} f_X(z) dz \\ &= \int_{-\infty}^x \frac{1}{a} f_X(z'/a) dz' \quad (z' = az). \end{aligned}$$

任意の $a < 0$ について、

$$\begin{aligned} \Pr(aX \leq x) &= \Pr(X \geq x/a) \\ &= \int_{x/a}^{\infty} f_X(z) dz \\ &= \int_x^{-\infty} \frac{1}{a} f_X(z'/a) dz' \quad (z' = az) \\ &= - \int_{-\infty}^x \frac{1}{a} f_X(z'/a) dz'. \end{aligned}$$

□

第 5 章

期待値、分散、モーメント

5.1 期待値

定義 5.1 (期待値). 離散確率変数 X が $A \subseteq \mathbb{R}$ に値を取るとき、その期待値は

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) \\ &= \sum_{x \in A} xP(\{\omega \mid X(\omega) = x\}) \\ &= \sum_{x \in A} x \Pr(X = x)\end{aligned}$$

と定義される。ここで、右辺の和が絶対収束しない場合は (適当な順番で和を取って収束したとしても) 期待値は定義されない。連続確率変数 X が確率密度関数 f_X を持つとき、その期待値は

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

と定義される。

補題 5.1 (期待値の性質).

1. 任意の $c \in \mathbb{R}$ について $\mathbb{E}[X + c] = \mathbb{E}[X] + c$
2. 任意の $c \in \mathbb{R}$ について $\mathbb{E}[cX] = c\mathbb{E}[X]$

証明. X が離散確率変数のとき、 X の値域を A とすると $X + c$ の値域は $A + c := \{a + c \mid a \in A\}$ なので、

$$\begin{aligned}\mathbb{E}[X + c] &= \sum_{x \in A+c} x f_{X+c}(x) \\ &= \sum_{x \in A+c} x \Pr(X + c = x) \\ &= \sum_{x' \in A} (x' + c) \Pr(X = x') \quad (x' = x - c) \\ &= \mathbb{E}[X] + c.\end{aligned}$$

X が連続確率変数のとき、

$$\begin{aligned}\mathbb{E}[X + c] &= \int_{-\infty}^{\infty} x f_{X+c}(x) dx \\ &= \int_{-\infty}^{\infty} x f_X(x - c) dx \\ &= \int_{-\infty}^{\infty} (x' + c) f_X(x') dx' \quad (x' = x - c) \\ &= \mathbb{E}[X] + c.\end{aligned}$$

$c = 0$ のとき $cX = 0$ なので、 $\mathbb{E}[cX] = 0 = c\mathbb{E}[X]$.

X が離散確率変数のとき、 X の値域を A とすると cX の値域は $cA := \{ca \mid a \in A\}$ なので、

$$\begin{aligned}\mathbb{E}[cX] &= \sum_{x \in cA} x f_{cX}(x) \\ &= \sum_{x \in cA} x \Pr(cX = x) \\ &= \sum_{x' \in A} cx' \Pr(X = x') \quad (x' = x/c) \\ &= c\mathbb{E}[X]\end{aligned}$$

X が連続確率変数のとき、

$$\begin{aligned}\mathbb{E}[cX] &= \int_{-\infty}^{\infty} x f_{cX}(x) dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{|c|} f_X(x/c) dx \\ &= \begin{cases} \int_{-\infty}^{\infty} cx' f_X(x') dx' & \text{if } c > 0 \\ \int_{\infty}^{-\infty} -cx' f_X(x') dx' & \text{if } c < 0 \end{cases} \quad (x' = x/c) \\ &= \int_{-\infty}^{\infty} cx' f_X(x') dx' \\ &= c\mathbb{E}[X]\end{aligned}$$

□

定理 5.1 (マルコフの不等式). 非負確率変数 X と $a > 0$ について

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

証明. 離散確率変数の場合の証明を与える。密度関数を持つ連続確率変数の場合も和を積分に変えた証明が成立する。

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{x \in A} f_X(x)x \\
 &= \sum_{x \geq a} f_X(x)x + \sum_{x < a} f_X(x)x \\
 &\geq \sum_{x \geq a} f_X(x)x \quad (X \geq 0) \\
 &\geq \sum_{x \geq a} f_X(x)a \\
 &= \Pr(X \geq a)a.
 \end{aligned}$$

□

5.2 分散

定義 5.2 (分散). 確率変数 X が期待値を持つとき、その分散を

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

と定義する。

確率変数 X が分散を持つとき、

$$\begin{aligned}
 \mathbb{V}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
 &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\
 &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\
 &= \mathbb{E}[X^2] - \mathbb{E}[X]^2
 \end{aligned}$$

である。分散は直感的には期待値からのはずれ具合を表す値である。

定理 5.2 (チェビシェフの不等式). 非負確率変数 X と $a > 0$ について

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

証明.

$$\begin{aligned}
 \Pr(|X - \mathbb{E}[X]| \geq a) &= \Pr((X - \mathbb{E}[X])^2 \geq a^2) \\
 &\leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{a^2} = \frac{\mathbb{V}[X]}{a^2}
 \end{aligned}$$

□

5.3 モーメントとモーメント母関数

定義 5.3 (モーメント). 確率変数 X と正の整数 $n \in \mathbb{Z}_{\geq 1}$ について、

$$\mathbb{E}[X^n]$$

を X の n 次モーメントという。

定義 5.4 (モーメント母関数 (積率母関数)). 確率変数 X について、

$$M_X(t) := \mathbb{E}[e^{tX}] \quad t \in \mathbb{R}$$

を X の **モーメント母関数**という。すべての $t \in \mathbb{R}$ で $M_X(t)$ が存在しない場合もある。また、

$$K_X(t) := \log M_X(t)$$

を X の **キュムラント母関数**という。

今後は以下の補題を認めることにする。証明にはルベーグ積分の知識が必要である。

補題 5.2. 確率変数 X について、ある $\epsilon > 0$ が存在し、モーメント母関数 $M_X(t)$ が $t \in (-\epsilon, \epsilon)$ で存在するとき、

$$\begin{aligned} M_X(t) &= \sum_{n \geq 0} \frac{\mathbb{E}[X^n]}{n!} t^n \\ \mathbb{E}[X^n] &= \left. \frac{d^n M_X(t)}{dt^n} \right|_{t=0}. \end{aligned}$$

証明. (概要)

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \mathbb{E}\left[\sum_{n \geq 0} \frac{(tX)^n}{n!}\right] \\ &= \sum_{n \geq 0} \mathbb{E}\left[\frac{(tX)^n}{n!}\right] \quad (\text{この無限和と期待値の交換が定理の条件より正当化される}). \end{aligned}$$

□

系 5.1. 確率変数 X について、ある $\epsilon > 0$ が存在し、モーメント母関数 $M_X(t)$ が $t \in (-\epsilon, \epsilon)$ で存在するとき、

$$\begin{aligned}\left.\frac{dK_X(t)}{dt}\right|_{t=0} &= \mathbb{E}[X] \\ \left.\frac{d^2K_X(t)}{dt^2}\right|_{t=0} &= \mathbb{V}[X].\end{aligned}$$

証明.

$$\begin{aligned}\left.\frac{dK_X(t)}{dt}\right|_{t=0} &= \left.\frac{M'_X(t)}{M_X(t)}\right|_{t=0} = \mathbb{E}[X] \\ \left.\frac{d^2K_X(t)}{dt^2}\right|_{t=0} &= \left.\frac{M''_X(t)M_X(t) - M'_X(t)^2}{M_X(t)^2}\right|_{t=0} = M''_X(0) - M'_X(0)^2 = \mathbb{V}[X].\end{aligned}$$

□

また、重要度は低くなるが、

$$\begin{aligned}\left.\frac{d^3K_X(t)}{dt^3}\right|_{t=0} &= \mathbb{E}[(X - \mathbb{E}[X])^3] \\ \left.\frac{d^4K_X(t)}{dt^4}\right|_{t=0} &= \mathbb{E}[(X - \mathbb{E}[X])^4] - 3\mathbb{V}[X]^2\end{aligned}$$

が成り立つ。一般に

$$\kappa_n(X) := \left.\frac{d^n K_X(t)}{dt^n}\right|_{t=0}$$

を X の n 次キュムラント と呼ぶ。

第 6 章

複数の確率変数、多変量正規分布

6.1 複数の確率変数

6.2 条件付き確率

6.3 多変量正規分布

第 II 部

統計学

第 7 章

ベイズ推定

第 8 章

仮説検定

第 III 部

漸近論

第 9 章

大数の法則と集中不等式

第 10 章

正規分布と中心極限定理

第 11 章

サノフの定理、KL ダイバージェンス

参考文献

Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.