

目次

第 1 章	モデルの枠組み	5
1.1	確率の数理的構造：Kolmogorov の公理	6
1.2	事象という概念：測度論からの結果	7
1.2.1	極限の定義	7
1.2.2	連続写像としての測度	8
1.2.3	初等的極限定理	9
1.2.4	確率不等式	10
1.3	確率空間の生息圏：可測空間の圏	10
1.3.1	Meas の特徴	10
1.3.2	Meas での極限構成	11
1.4	統計量：可測関数の空間上の作用素	11
1.5	確率不等式	12
1.6	σ -加法族が定める構造	12
第 2 章	確率変数の基本性質	14
2.1	確率測度の変換：条件付き確率と Bayes の定理	14
2.2	事象の独立性	15
2.3	確率変数と確率分布	16
2.4	確率変数の独立性	16
2.5	独立な確率変数	17
2.5.1	独立な確率変数に対する関手性	17
2.5.2	独立同分布	18
2.6	確率変数の和・積・商の分布	20
2.7	Markov 連鎖	20
2.8	大数の法則	20
2.8.1	独立同分布での大数の弱法則	21
2.8.2	独立同分布での大数の強法則	21
2.8.3	証明抽出と評価の精緻化	22
2.9	数学における大数の法則的現象	22
2.9.1	Weierstrass の多項式近似	22
2.10	中心極限定理	24
2.11	確率分布の収束	24
2.12	大偏差理論	25
第 3 章	離散確率分布	26
3.1	平均	26
3.2	分布の特性値	27
3.2.1	分布関数と確率関数	27
3.2.2	期待値と積率	28

3.3	特性関数と母関数	28
3.3.1	特性関数と確率母関数	28
3.3.2	分位点関数	29
3.3.3	母関数の概念の射程	29
3.4	一次元離散分布の例	30
3.4.1	デルタ分布	30
3.4.2	経験分布	30
3.4.3	Rademacher 分布	31
3.4.4	離散一様分布	31
3.4.5	二項分布	31
3.4.6	Poisson 分布	32
3.4.7	負の二項分布	34
3.4.8	Katz 族	35
3.4.9	超幾何分布	36
3.4.10	負の超幾何分布	36
3.4.11	対数分布	36
3.4.12	Ord 族	36
3.5	多次元離散分布の例	36
3.5.1	確率変数の積	36
3.5.2	多項分布	37
3.5.3	2 変量 Poisson 分布	37
3.5.4	負の多項分布	38
第 4 章	絶対連続確率分布	39
4.1	期待値	39
4.1.1	積分の定義	39
4.1.2	期待値の定義	39
4.1.3	期待値不等式	40
4.2	分布の特性値	40
4.2.1	確率密度関数	40
4.2.2	平均値と積率	40
4.2.3	共分散と相関	41
4.2.4	分散共分散行列	42
4.3	特性関数と母関数	43
4.3.1	特性関数と分布	43
4.3.2	分布の収束の特徴付け	43
4.3.3	特性関数と絶対積率	44
4.3.4	キュムラント母関数	44
4.3.5	分布関数	44
4.3.6	積率母関数	45
4.3.7	多次元確率変数の特性関数	45
4.4	一次元連続分布の例	45
4.4.1	一様分布	45
4.4.2	Gamma 分布	45
4.4.3	正規分布	46
4.4.4	Beta 分布	47
4.4.5	Cauchy 分布	47
4.4.6	Weibull 分布	48

4.4.7	対数正規分布	48
4.4.8	logistic 分布	48
4.4.9	Pareto 分布	49
4.5	確率変数の独立性と期待値	50
4.6	多次元連続分布	50
4.7	指数型分布族	50
4.8	指数分散モデル	50
4.9	コピュラ	50
4.10	確率分布と積分変換	50
4.11	不等式	50
第 5 章	多変量分布	51
5.1	多次元確率変数の分布	51
5.2	共分散	51
5.2.1	共分散の定義と例	51
5.2.2	期待値の拡張	52
5.2.3	積率の定義	53
5.3	特性関数と分布の収束	53
5.4	独立性	53
5.5	多変量連続分布	53
5.6	多変量正規分布	53
5.7	変数変換と確率密度関数	53
5.8	従属性	53
第 6 章	条件付き期待値	54
6.1	部分 σ -加法族に関する条件付き期待値	54
第 7 章	確率変数の収束	55
7.1	確率測度の収束	55
7.1.1	汎関数としての測度	55
7.1.2	収束の対応	55
7.1.3	位相構造	56
7.1.4	幾何構造	57
7.2	積分の拡張	57
7.3	概収束	57
第 8 章	確率測度の全て	58
8.1	一般の試行と確率測度	58
8.2	確率測度の拡張定理	58
8.3	確率測度の直積	58
8.4	標準確率空間	58
8.5	多変量分布	58
第 9 章	確率論の基礎概念	59
9.1	可分完全確率測度	59
9.2	事象と確率変数	59
9.3	条件付き期待値作用素	59
第 10 章	確率過程	60

10.1	関数空間 C と D	60
10.2	確率過程に関する一般事項	60
10.3	情報と情報増大系	60
10.4	停止時	60
10.5	離散時変数のマルチンゲール	60
第 11 章	線型推測論	61
第 12 章	統計的決定理論	62
12.1	枠組み	62
12.2	十分性と完備性	63
12.2.1	十分統計量	63
12.3	指数型分布族	63
12.4	統計的推定	63
第 13 章	統計的仮説検定	64
13.1	枠組みと例	64
13.1.1	枠組み	64
13.1.2	ランダム化検定	64
13.1.3	形式化	65
第 14 章	大標本理論	66
第 15 章	漸近展開とその応用	67
第 16 章	Martingale	68
16.1	Introduction	68
16.2	semimartingale	68
参考文献		69

第 1 章

モデルの枠組み

Probability theory is concerned with mathematical models of phenomena that exhibit randomness, or more generally phenomena about which one has incomplete information. 確率論の概念を数学的な公理に落とし込む段階に於ける形式科学的な議論をまとめる。^{†1} ランダム性・不確定性に関する概念のうち、例えば、独立性以前のエントロピーなどの概念は殆ど数学的に確定する。

試行 とは、類等式の双対のような^{†2}、標本空間の分割で、事象とはそのうちの 1 つである。この分割上の割合の分配なる観念を人類は確率と呼ぶが、これは正規化された測度に他ならない。^{†3}

確率変数 とは標本空間上の実線型空間に入る射で、標本空間上に同値類による分割を定める。 \mathbb{R}^n -値点の考え方を導入することで、数学理論を観測値の概念に持ち込むことができる。その背景には射があり、線型代数・微分積分学の関手がそこまで還元しているのである。こうして、 X は \mathbb{R}^n の元であるとも見るし、射 $X: \Omega \rightarrow \mathbb{R}^n$ とも見る。

確率 とは標本空間上の非負で正規化された加法的集合関数で、確率分布とは確率変数によるこの測度の押し出しである。この対応を分布から辿る時「確率変数が分布に従う」という（当該の分布を押し出すような確率変数を 1 つ取る、という条件と同値）。

エントロピー とは、系で考える全ての試行の期待驚愕度の上限である。
これらの類比は Fréchet の研究において初めて完全に示された。

「有限試行によって確率論の考え方を理解したならば、解析学の手法を用いて、現代確率論の対象である無限試行に進むのは容易である。」注意力の質の方が重要である。

統計モデルとしての多様な確率分布族と、それらに対する種々の統計推測法について解説する。多くの例を通じ、受講者が確率統計の基本事項に習熟することを目標とする。確率的な構造の表現からはじめ、確率の性質、確率変数と確率分布、独立性等の用語を準備し、離散確率分布とその例と計算法、連続分布とその例、確率変数の期待値、変数変換の公式、混合分布、指数型分布族、多次元分布の基礎について解説する。前半の確率の基礎概念の導入の後、確率モデルの推定について紹介する。不偏推定が統計推測の数理的構造を理解するための例となる。十分性、因子分解定理、完備性、ラオ・ブラックウェルの定理、レーマン・シェフェの定理、統計的決定理論の枠組み、ベイズ推定について説明する予定である。

^{†1} Notice that in this respect probability theory has a similar status as (other(?!)) theories of physics: there is a mathematical model (measure theory here as the model for probability theory, or for instance symplectic geometry as a model for classical mechanics) which can be studied all in itself, and then there is in addition a more or less concrete idea of how from that model one may deduce statements about the observable world (the average outcome of a dice role using probability theory, or the observability of the next solar eclipse using Hamiltonian mechanics).<https://ncatlab.org/nlab/show/probability+theory>

^{†2} 類等式は群作用による終域の分解だが、試行は可測関数による始域の分解

^{†3} もっと無限な概念が扱えたなら、測度の語もっと細かく分類すべきかもしれない。

1.1 確率の数理的構造：Kolmogorov の公理

測度論的確率論という枠組み

確率空間は測度空間の一種と見る．確率変数は関数空間の元と見て，平均は作用素と見て構成する．^a

^a こうして，量子論と確率論が関数解析を舞台として合流する．

記法 1.1.1 (mutually exclusive).

- (1) 集合 A, B が排反であるとは，互いに素であることをいう．この時の和事象を $A + B, A \sqcup B$ で表し直和という．[2] では直和も $\sum_{i=1}^n A_i$ と表す． $P(A + B) = P(A) + P(B), P(A \cup B) = P(A) + P(B) - P(A \cap B)$ と書き分けられる．
- (2) 族 $(A_i)_{i \in I}$ が排反であるとは， $\forall i, j \in I \ i \neq j \Rightarrow A_i \cap A_j = \emptyset$ であることをいう．
- (3) $A \subset B$ のときの差事象 $A \setminus B$ を固有差といい， $B - A$ と書く．

定義 1.1.2 (sample space, sample, event, total event).

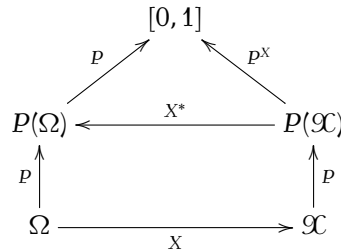
- (1) 標本空間 Ω の元 $\omega \in \Omega$ を標本 (点) という．^{†4}
- (2) σ -加法族 $\mathcal{F} \subset P(\Omega)$ の元を事象と呼ぶ．^{†5} 事象としての $\Omega \in \mathcal{F}$ を全事象という．

定義 1.1.3 (probability). σ -代数 \mathcal{F} に対して，次を満たす測度 $P: \mathcal{F} \rightarrow [0, 1]$ を確率測度または試行 T の確率法則という．

- (1) $P(A) \geq 0 \ (\forall A \in \mathcal{F})$.
- (2) $P(\Omega) = 1$.
- (3) (σ -加法性) $P(\cup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i)$, ただし族 $(A_i): \mathbb{N} \rightarrow \mathcal{F}$ は $\forall i \neq j \ A_i \cap A_j = \emptyset$ を満たす．

定義 1.1.4 (random variable, probability distribution / law). 確率空間を (Ω, \mathcal{F}) , $(\mathcal{X}, \mathcal{G})$ を可測空間とする．

- (1) \mathcal{X} -値確率変数とは，可測関数 $X: \Omega \rightarrow \mathcal{X}$ をいう．
- (2) これに沿って引き起こされる像測度 $X_*P = P^X: P(\mathcal{X}) \supset \mathcal{G} \rightarrow [0, 1]$ を $P^X(B) := P(X^{-1}(B))$ で定める．^{†6}



これを X の確率分布または確率法則という．

- (3) 特に，確率 P は，恒等写像 $\text{id}_\Omega: \Omega \rightarrow \Omega$ についての確率分布 P^{id_Ω} である．
- (4) 像 $\text{Im } X$ も同様の記法 Ω^X で表し，これを X の標本空間という．
- (5) Ω^X は混合試行 T_X の標本空間と考える． X の定める同一視による商空間を考えているために「混合」という．

要諦 1.1.5 (Kolmogorov 1933). すなわち，確率空間から出る射が確率変数で，それにより押し出される測度が確率分布である．

記法 1.1.6. $a \in \mathcal{X}$ として，確率 $P(X^{-1}(a)) \in [0, 1]$ を $P(X = a)$ と表す． $B \subset \mathcal{X}$ として，確率 $P(X^{-1}(B)) \in [0, 1]$ を $P(X \in B)$ と表す．これは $E[1_B(X)]$ にも等しい．

定義 1.1.7 (joint distribution, marginal distribution).

^{†4} 見本空間，見本点ともいう．

^{†5} 部分集合を条件や関係だけでなく事象と捉えるのは，集合論の初歩でもやるようだ．

^{†6} 逆像写像 X^* により，なんとか反変に見えるから許したが，nLab では $f_*\mu$ で表されている．

- (1) 確率変数の列 X_1, \dots, X_n に対して, 積写像 (X_1, \dots, X_n) を同時分布または結合分布という.
- (2) 同時分布から一部の確率変数を消去してえる分布を周辺分布という.^{†7}

定義 1.1.8 (可積分).

- (1) 実確率変数 $X: \Omega \rightarrow \mathbb{R}$ が可積分であるとは, $E[|X|] = \int_{\Omega} |X(\omega)| dP(\omega) < \infty$ すなわち, $X \in L^1$ であることをいう.
- (2) $L^1(\Omega, \mu) := \left\{ f \in \text{Map}(\Omega, \overline{\mathbb{R}}) \mid \int_{\Omega} |f(\omega)| \mu(d\omega) < \infty \right\}$ と表す.

定義 1.1.9 (stochastic process). 確率変数の族 $(X_t)_{t \in T}$ を, 特に T が全順序集合のとき, 確率過程と言う.

例 1.1.10. Brownian motion, Ornstein-Uhlenbeck process and Lévy process.

1.2 事象という概念：測度論からの結果

測度論の復習

測度とは, 部分集合 = 事象に対して「全体に占める割合」を定める. 類等式ではないが, 全事象をどのように分解するかが特徴量となる (エントロピー).

1.2.1 極限の定義

極限とは「十分遠くでは変わらない」ことである. 距離空間では点列の極限が距離で定義できたが, \mathcal{F} には擬距離しか定まらない. しかし, \mathcal{F} には特別に, 列の極限の定義が, 論理によって定められる. そしてこれは, (\mathcal{F}, Δ) が距離を定めるときの「点列」の定義に一致する. 「完備」は, ここに於て通じ合っているが, いずれにしろ距離概念が暗黙にあるということを捉える試みがいまだに続いている.

定義 1.2.1 (上極限事象, 下極限事象, 極限事象). 事象列 $(A_n): \mathbb{N} \rightarrow \mathcal{F}$ に対して, 次の事象が定義できる.

- (1) $\limsup_{n \rightarrow \infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{n \geq m} A_n$ を上極限事象という. 「事象列のうち無限個が起きる」という条件 i.o. を表す.
- (2) $\liminf_{n \rightarrow \infty} A_n := \bigcup_{m=1}^{\infty} \bigcap_{n \geq m} A_n$ を下極限事象という. 「事象列のうちある番号から先の事象が全て起きる」という条件 f.e. を表す.
- (3) 2つの集合が一致するとき, 集合列 (A_n) は収束するといい, $\lim_{n \rightarrow \infty} A_n$ を極限集合という.

補題 1.2.2 (事象としての意味論: infinitely often, with finite exceptions). $\{A_n\}_{n \in \mathbb{N}} \subset P(X)$ を集合列とする.

- (1) $\limsup_{n \rightarrow \infty} A_n = \{x \in X \mid \#\{n \in \mathbb{N} \mid x \in A_n\} = \infty\}$.
- (2) $\liminf_{n \rightarrow \infty} A_n = \{x \in X \mid \#\{n \in \mathbb{N} \mid x \notin A_n\} < \infty\}$.

[証明].

(1)

$$\begin{aligned} x \text{ が } \#\{n \in \mathbb{N} \mid x \in A_n\} = \infty \text{ を満たす} &\Leftrightarrow \{n \in \mathbb{N} \mid x \in A_n\} \text{ は非有界である} \\ &\Leftrightarrow \forall n \in \mathbb{N} \exists m \geq n \ x \in A_m \Leftrightarrow x \in \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m. \end{aligned}$$

(2)

$$\begin{aligned} x \text{ が } \#\{n \in \mathbb{N} \mid x \notin A_n\} < \infty \text{ を満たす} &\Leftrightarrow \{n \in \mathbb{N} \mid x \notin A_n\} \text{ は有界である} \\ &\Leftrightarrow \exists n \in \mathbb{N} \forall m \geq n \ x \in A_m \Leftrightarrow x \in \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_m. \end{aligned}$$

^{†7} 表の欄外 (margin) に行や列の和を記載することから周辺 (marginal) と呼ばれるようになった.

補題 1.2.3 (σ -代数の構造を求めて). 有限な測度を備えた空間 (X, \mathcal{F}, μ) について,

- (1) 対称差について, $d(A, B) := \mu(A \Delta B)$ ($A, B \in \mathcal{F}$) と定めると, これは擬距離関数となる.
- (2) \mathcal{F} が σ -代数として完備であるとき, (\mathcal{F}, d) の距離等化によって得る距離空間は完備である.

[証明].

- (1) μ は有限としたから, $\text{Im } d \subset [0, \infty)$ である. 対称性と $d(A, A) = 0$ は明らか. 三角不等式については, $A, B, C \in \mathcal{F}$ について,

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

と直和分解できるが, それぞれについて,

$$(A \setminus C) \cup B \subset A \setminus B, \quad (A \setminus C) \cap B \subset B \setminus C, \quad (C \setminus A) \cup B \subset C \setminus B, \quad (C \setminus A) \cap B \subset B \setminus A$$

より, $\mu(A \Delta C) \leq \mu(A \Delta B) + \mu(B \Delta C)$.

- (2) 略.

■

要諦 1.2.4. はじめは技術的で煩瑣な手続きに思えた測度空間の完備化であるが, これにより $\mu: \mathcal{F} \rightarrow [0, \infty)$ が完備距離空間の射に, これが定める積分が有界線型作用素 $L^p(X) \rightarrow \mathbb{R}$ となる.

1.2.2 連続写像としての測度

測度 $\mu: \mathcal{F} \rightarrow [0, 1]$ は, \mathcal{F} 上の擬距離 $d(A, B) := \mu(A \Delta B)$ について連続であることは示せる. しかし, 前節で定義した「集合列の極限」が定める位相は, この擬距離が誘導する位相とは異なる. 注 1.2.7 のように, ∞ と 0 は区別できないフィルターとなっている.

命題 1.2.5 (測度の連続性と劣加法性). 一般の測度空間 $(\Omega, \mathcal{F}, \mu)$ の可測集合の列 $(A_i): \mathbb{N} \rightarrow \mathcal{F}$ について, 次が成り立つ.

- (1) $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$ のとき, $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\bigcup_{n=1}^{\infty} A_n)$.
- (2) $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ かつ $\mu(A_1) < \infty$ のとき, $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\bigcap_{n=1}^{\infty} A_n)$.
- (3) (subadditivity) $\mu(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mu(A_i)$ ($n \in \mathbb{N}$).

[証明].

- (1) $B_n := A_n - A_{n-1}$ とおくと (ただし $A_0 = \emptyset$ とする), $(B_n)_{n \in \mathbb{N}}$ は排反であり, $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n$. よって,

$$\begin{aligned} \mu(\bigcup_{n=1}^{\infty} A_n) &= \sum_{n=1}^{\infty} \mu(B_n) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(B_k) = \lim_{n \rightarrow \infty} \mu(A_n) \end{aligned}$$

- (2) 族 $(A_n^c)_{n \in \mathbb{N}}$ が単調増加列であることに気をつけて,

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

- (3) $B_n := \bigcup_{k=1}^n A_k$ とおくと, これは単調列である. 極限は不等式を保存するから,

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mu(B_n) \leq \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(A_k).$$

要諦 1.2.6 (連続写像としての測度). 確率測度 $\mu : \mathcal{F} \rightarrow [0, 1]$ は, の射を定める. 一方で, 一般の測度 $\mu : \mathcal{F} \rightarrow [0, \infty)$ については, 条件 $\mu(A_1) < \infty$ が必要となる.

注 1.2.7. (2) で $\mu(A_1) = \infty$ を許すならば, $A_i := 2^i \mathbb{N}$ とすることで, $\lim_{n \rightarrow \infty} \mu(A_n) = \infty, \mu(\cap_{n \in \mathbb{N}} A_n) = 0$ を満たす列が作れてしまう.

測度の連続性が実際にどういう位相について連続なのかを考え途中. 特に, (1) と (2) で, $\mu(A_1) < \infty$ の条件が必要になる非対称性がどこから来るのか探していて, おそらく $[0, \infty]$ に入れる位相から. これは収束空間の知識が必要か? 2つの「完備」の概念は, フィルターの意味で一緒なのではないか?

1.2.3 初等的極限定理

$\limsup : \mathcal{F}^{\mathbb{N}} \rightarrow \mathcal{F}$ の性質を考える.

命題 1.2.8 (集合に関する Fatou の補題). $\limsup : \mathcal{F}^{\mathbb{N}} \rightarrow \mathcal{F}$ は次を満たす.

- (1) \mathcal{F} の列 (A_n) について, $\mu\left(\liminf_{n \rightarrow \infty} A_n\right) \leq \liminf_{n \rightarrow \infty} \mu(A_n)$.
- (2) \mathcal{F} の列 (A_n) について, $\mu(\cup_{n \in \mathbb{N}} A_n) < \infty$ ならば, $\limsup_{n \rightarrow \infty} \mu(A_n) \leq \mu\left(\limsup_{n \rightarrow \infty} A_n\right)$.

定理 1.2.9 (Borel-Cantelli lemma).

- (1) 列 $(A_i) : \mathbb{N} \rightarrow \mathcal{F}$ が独立であろうと無かろうと, 一般の測度 μ について,

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty \Rightarrow \mu(\limsup_{n \rightarrow \infty} A_n) = 0, \quad \mu(\liminf_{n \rightarrow \infty} A_n^c) = \mu(X)^{\dagger 8}.$$

- (2) 列 $(A_i) : \mathbb{N} \rightarrow \mathcal{F}$ が独立であるとき, $\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\limsup_{n \rightarrow \infty} A_n) = 1, \quad P(\liminf_{n \rightarrow \infty} A_n^c) = 0$.

[証明].

- (1) $B_m := \cup_{n=m}^{\infty} A_n$ と定めると, これは単調減少列であるから $\mu(\cap_{m=1}^{\infty} B_m) = \lim_{m \rightarrow \infty} \mu(B_m)$. 和が収束する列は 0 に収束する

($\lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mu(A_n) = 0$) ことに注意して,

$$\begin{aligned} \mu(\limsup_{n \rightarrow \infty} A_n) &= \mu(\cap_{m=1}^{\infty} B_m) = \lim_{m \rightarrow \infty} \mu(B_m) && \text{単調列の積 1.2.5(3)} \\ &\leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mu(A_n) = 0. && \text{劣加法性 1.2.5(4)} \end{aligned}$$

また, de Morgan の法則より,

$$\begin{aligned} P(\liminf_{n \rightarrow \infty} A_n^c) &= P(\cup_{n \rightarrow \infty} \cap_{m=n}^{\infty} A_m^c) \\ &= 1 - P(\cap_{n \rightarrow \infty} \cup_{m=n}^{\infty} A_m) = 1 - P(\limsup_{n \rightarrow \infty} A_n) = 1. \end{aligned}$$

- (2) (1) と同様にして, $P(\limsup_{n \rightarrow \infty} A_n) = P(\cap_{n=1}^{\infty} \cup_{k \geq n} A_k) = \lim_{n \rightarrow \infty} P(\cup_{k \geq n} A_k)$ である. 最右辺を評価すると,

$$\begin{aligned} 1 - P(\cup_{k \geq n} A_k) &= P(\cap_{k \geq n} A_k^c) \\ &\leq P(\cap_{k \geq n}^p A_k^c) && \cap_{k \geq n} A_k^c \subset \cap_{k \geq n}^p A_k^c \end{aligned}$$

^{†8} 確率測度ならば $\mu(X) = 1$ である

$$\begin{aligned}
&= \prod_{k=n}^p P(A_k^c) && \text{独立性} \\
&= \prod_{k=n}^p (1 - P(A_k)) \leq \exp \left(- \sum_{k=n}^p P(A_k) \right) \xrightarrow{p \rightarrow \infty} e^{-\infty} = 0. && 1 - x \leq e^{-x}
\end{aligned}$$

もう一つの結論も de Morgan の定理から従う。

■

要諦 1.2.10 (どうやら洗練された証明はこの一通りである)。(2) が極めて非自明であるが、余事象を自在に使いこなして解析関数を持ち出して不等式評価へ。複利の式だね。

例 1.2.11. 無限の猿定理はこの補題 (2) の特別な場合である。

系 1.2.12. 事象列 (A_n) が次の (1) と、(2)(a),(2)(b) のいずれかの 2 条件を満たすならば、 $P \left(\limsup_{n \rightarrow \infty} A_n \right) = 0$, $\lim_{n \rightarrow \infty} P(A_n) = 0$ が成り立つ。

- (1) $\liminf_{n \rightarrow \infty} P(A_n) = 0$.
- (2) (a) $\sum_{n \in \mathbb{N}} P(A_n^c \cap A_{n+1}) < \infty$.
- (b) $\sum_{n \in \mathbb{N}} P(A_n \cap A_{n+1}^c) < \infty$.

1.2.4 確率不等式

命題 1.2.13 (Bonferroni の不等式).

$$P \left(\bigcap_{i=1}^n A_i \right) \geq \sum_{i=1}^n P(A_i) - (n-1)$$

等号成立条件は、 A_1^c, \dots, A_n^c が背反のとき。

[証明]. A_1^c, \dots, A_n^c について、劣加法性より。

■

1.3 確率空間の生息圏：可測空間の圏

Meas は Set 上位相的である。また、完備かつ余完備である。最先端は monad と valuation の理論だと見受けられる。^a

^a There is at least some similarity of the concept of random variables to usage of the function monad ("reader monad") in the context of monads in computer science. <https://ncatlab.org/nlab/show/random+variable>

1.3.1 Meas の特徴

定義 1.3.1 (topological category). 代数的構造を集合演算などから抽出し、空間的概念を得る構成 (数 Ord, 位相空間 Top, 可測空間 Meas, 多様体 Diff) を形式化する。^{†9} 束 $U: C \rightarrow D$ を考える。C の対象を空間とよび、射を射と呼ぶ。D の対象を代数とよび、射を準同型と呼ぶ。

C が D 上の位相的な圏であるとは、任意の代数 $X \in D$ と任意の D-射の族 $f_i: X \rightarrow U(S_i) \in D$ に対して、initial lift $(T, m_i: T \rightarrow S_i) \in C$ が存在することをいう。lift とは、次の条件を満たす C の対象と射の組である：

$$\forall T' \in C \exists g': U(T') \rightarrow X \exists m'_i: T' \rightarrow S_i \quad g' \circ f_i = U(m'_i) \Rightarrow \exists ! n: T' \rightarrow T \quad U(n) = g' \wedge n \circ m_i = m'_i.$$

C の対象は D の対象と $U: C \rightarrow D$ が定める initial structure / weak structure の組として表せる。

^{†9} 全て $D = \text{Set}$ の例だが、 $D = \text{Grp}$ として、位相群などを考えても良い。

例 1.3.2 (Meas). $D = \text{Alg}_\sigma \subset \text{Set}$ とすると, $C = \text{Meas}$ である.

定義 1.3.3 (Borel 可測空間). 関手 $\mathcal{B} : \text{Top} \rightarrow \text{Meas}$ に対して, $\mathcal{B}(S)$ を Borel σ -加法族という.

1.3.2 Meas での極限構成

Meas は cartesian category ではないのは周知の事実である (直積測度の選択はエントロピー最大の公理を必要とする恣意的なものであり, Fubini の定理で議論される). そこでも出来る直積構成がある. Kolmogorov product は一般の symmetric semicartesian monoidal category^a で定義され, 無限次元のテンソル積を構成する方法である. cartesian である場合は, 直積の概念と一致する.

^a CMC とは直積によって圏のモノイド構造 = テンソル積が与えられる圏であるが, 一般にテンソル積の単位が終対象によって与えられる時, もうすでに十分「直積っぽい」ということで SMC という.

命題 1.3.4 (完備かつ余完備). Meas は完備かつ余完備である. すなわち, 全ての図式は極限と余極限をもつ.

定義 1.3.5 (filtered category). filtered category とは, 任意の有限な図式が余錐を持つような圏をいう. 有向集合の圏化された概念である.

定義 1.3.6 (lattice of projections, lattice of finite projections, Kolmogorov product). $(C, \otimes, 1)$ を symmetric semicartesian monoidal category とする.^{†10}

- (1) C の対象の有限列 $(X_n)_{n \in \mathbb{N}}$ について, 射影のなす有限 Boole 代数 (を細い圏とみなした圏) B から C への関手 $B \rightarrow C$ が存在する. これを射影の束という.
- (2) C の対象の族 $(X_i)_{i \in I}$ について, 任意の有限部分集合 $F, S \subset I$, $S \subset F$ に対して $\bigoplus_{i \in F} X_i \rightarrow \bigoplus_{j \in S} X_j$ の形をした射影のなす束からの関手 $B \rightarrow C$ が存在する. これを有限な射影の束という.
- (3) 有限な射影の束 $B \rightarrow C$ は cofiltered diagram である. この cofiltered limit が存在するとき, これをコルモゴロフ積という.

1.4 統計量: 可測関数の空間上の作用素

標本空間 \mathcal{X}^n 上の関数 $\mathcal{X}^n \rightarrow \mathbb{R}$ を統計量という. これは像測度 P^X に依らずに算出される.

確率測度は, 線型汎関数の空間 $P(\Omega) \subset C(\Omega)^*$ を, 積分によって定める. したがって, この値や, 他の積分核を採用した線型汎関数を用いて, 分布を関数解析の方法で特徴づけることを考える.^aこれが分布の特徴量の考え方であり, 統計量はその有限な / 経験的な場合である. 前者に母をつけて区別することもある.

^a まるで Ergode 理論のような理論展開である.

定義 1.4.1 ((population) mean / expectation, (population) variance). 作用素 $\text{Meas}(X, \mathbb{R}) \rightarrow \mathbb{R}$ を定める. 一般に $\text{Meas}(X, V) \rightarrow V$ で定まる.

- (1) $E[X] := \int_{\Omega} X(\omega) P(d\omega)$ が実数であるとき, 特に離散の場合は $\sum_{\omega \in \Omega} X(\omega) P\{\omega\}$ を X の (母) 平均という.^{†11}
- (2) $E(X, A) := E(X1_A)$ と定める.
- (3) $V[X] := E[(X - EX)^2]$ が実数であるとき, これを分散という.
- (4) $V(X, Y) := E[(X - EX)(Y - EY)] = E[XY] - E[X]E[Y]$ を共分散 $\text{Meas}(X, \mathbb{R}) \times \text{Meas}(X, \mathbb{R}) \rightarrow \mathbb{R}$ という.
- (5) $\sigma(X) = \sqrt{V(X)}$ を標準偏差という.

^{†10} すなわち, C は終対象 1 をもち, これを単位とするモノイド構造 $\otimes : C \times C \rightarrow C$ を備える.

^{†11} 標本平均と区別していう. 平均は位置 (location) 母数の例で, 分散は尺度 (scale) 母数の例である.

(6) $R(X, Y) := \frac{V(X, Y)}{\sigma(X)\sigma(Y)}$ を相関係数という.

定義 1.4.2 (entropy). Shannonn のエントロピーとは自己情報量 $I(p) := -\log_2 p$ ($p \in \mathcal{G}$) が定める積分作用素 $H : \text{Meas}(\mathcal{G}, \mathbb{R}) \rightarrow [0, \log \alpha]$ である.

1.5 確率不等式

定理 1.5.1 (Chebyshev の不等式).

- (1) $\forall a > 0 \ P\{|X(\omega) - EX| > a\sigma(X)\} \leq \frac{1}{a^2}$.
- (2) $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ を非負値単調増加関数とし, X を $E[|X|] < \infty, E[|f(X)|] < \infty$ を満たす確率変数とする. $\forall a \in \mathbb{R} \ f(a) > 0 \Rightarrow P(X \geq a) \leq \frac{E[f(X)]}{f(a)}$.

[証明].

- (1) $\sigma(X) = 0$ のとき $\sigma(X) = 0 \Leftrightarrow X(\omega) = EX$ a.s. であるから, 上の不等式は当然成り立つ.
- $\sigma(X) \neq 0$ のとき 求める事象を $A := \{\omega \in \Omega \mid |X(\omega) - EX| > a\sigma(X)\}$ とおくと,

$$\sigma(X)^2 = E(X - EX)^2 \geq E((X - EX)^2, A) \geq a^2 \sigma(X)^2 P(A)$$

と評価できる.

- (2) f が単調減少であることと, f が非負値であるから $\int_{\{X < a\}} f(X) dP \geq 0$ であることより,

$$\begin{aligned} E[f(X)] &= \int_{\{X \geq a\}} f(X) dP + \int_{\{X < a\}} f(X) dP \\ &\geq f(a) P(X \geq a). \end{aligned}$$

■

要諦 1.5.2 (平均値から標準偏差の a 倍以上離れる確率は $\frac{1}{a^2}$ 以下である.). まさかそんなに当然な評価の変形だったのか.

1.6 σ -加法族が定める構造

実は, 確率論で直積測度を考えるときは, 暗黙にエントロピー最大の原理を仮定している. 確率論の枠組みではエントロピーを厳密に定義でき, これは系の状態を決定するために必要な情報量 = 乱雑さの小ささを計っている. Entropy is a measure of disorder, given by the amount of information necessary to precisely specify the state of a system.^a

^a <https://ncatlab.org/nlab/show/entropy>

注 1.6.1 (情報としての σ -加法族). σ -代数の細かさによって, 観察の粒度を表現する. 「情報」が追加されるたびに確率測度 P を取り替えるのではなく, P は一つの固定されたものと考えて「情報」の変化は σ -加法族の細分化として表現する. 新たに追加されていくのである. 例えば, 確率変数 X の値がわかれば X^2 の値もわかるが, 逆は成り立たない. すなわち, σ -加法族の引き戻しについて, $(X^2)^*(\mathcal{G}) \subsetneq X^*(\mathcal{G}) = (X^3)^*(\mathcal{G})$.^{†12} したがって, 外から観測できる範囲は (Ω, \mathcal{F}, P) の σ -部分代数を自然になすので, これは情報を表すと考えると自然である.^{†13} こうして, 確率変数 X が \mathcal{F}_t 可測である, ということ, 時刻 t 時点の情報で値が (確率 1 で) 定まっていることになる.^{†14} そこで, 問題は平均 $E[X_n | \mathcal{F}_m]$ ($m < n$) を求めることとなる. したがって, このモデルでは条件付き確率だけである.

^{†12} この考え方を確率変数が定める filtration $\sigma[X_0, \dots, X_n]$ という.

^{†13} 例えば事象 $B \in \mathcal{F}$ の与える情報は σ -加法族 $\{\emptyset, \Omega, B, \Omega \setminus B\}$ で表されると考えられる.

^{†14} サイコロを 2 回振る事象は, 2 回降った時点で値が 0, 1 の 2 値に収束するはずである. 情報 \mathcal{F}_t がどこから来たかは不問とする. この間に σ -代数の構造があるため, 他の確率分布 $\mathcal{G} \rightarrow \mathbb{R}$ からの引き戻しや, 他の事象についての知識からも持って来れるはず.

定義 1.6.2 (関数が生成する σ -加法族). 関数 $X_1, \dots, X_n : \Omega \rightarrow \mathcal{X}$ に対して, これら全てを可測とするような Ω 上の最小の σ -加法族を $\sigma(X_1, \dots, X_n)$ で表す.

定義 1.6.3 (principle of maximum entropy).

- (1) $\Omega_1 \times \Omega_2$ が考えている試行 $T_1 \times T_2$ を試行の直結合という.
- (2) 次の図式を可換にする確率測度 \tilde{P} は一般に複数存在するが, $\tilde{P}\{(\omega_1, \omega_2)\} = P_1\{\omega_1\}P_2\{\omega_2\}$ を満たすものは Hahn-Kolmogorov の定理よりただ一つである.
- (3) これを直積測度の公理としても良いし, エントロピー最大の原理からも従う. 確率測度 P のエントロピーとは, $\epsilon(P) := \sum_{i=1}^m P\{a_i\} \log \frac{1}{P\{a_i\}}$ ($\Omega = \{a_i\}_{i \in [m]}$) と定まる.

定義 1.6.4 (surprisal / self-information / information content, expected surprisal, almost partition, entropy). (X, \mathcal{M}, μ) を確率空間とする.

- (1) 可測集合 $A \in \mathcal{M}$ の新鮮さとは, $\sigma_\mu(A) := -\log \mu(A) = \log \frac{1}{\mu(A)} (\geq 0)$ をいう. ただし, $\mu(A) = 0$ の時は $\sigma_\mu(A) = \infty$ とする. これは可測関数 $\sigma_\mu : \mathcal{M} \rightarrow [0, \infty]$ を定める. 「もしその事象 $A \in \mathcal{M}$ が起こったら観測者がどれほど驚くか」をモデリングしていると考えられる. $\mu(A) = 1$ のとき, $\sigma_\mu(A) = 0$ であり, $\mu(A) = 0$ のとき, $\sigma_\mu(A) = \infty$ である.
- (2) $h_\mu(A) := \sigma_\mu(A)\mu(A)$ を期待驚愕度という. このとき $\mu(A) = 0 \Rightarrow h_\mu(A) = 0$ である. これは $\mu(A) = e^{-1}$ のときに最大値 $e^{-1} \log e$ を取り, $h_\mu(\emptyset) = h_\mu(X) = 0$ を満たす上に凸な関数である.
- (3) 合併が full set, 任意の2つの共通部分が零集合となるような X の族を概分割という.
- (4) σ -加法族 \mathcal{M} のエントロピーとは,

$$H_\mu(\mathcal{M}) := \sup \left\{ \sum_{A \in \mathcal{F}} h_\mu(A) \in \mathbb{R}_{\geq 0} \mid \mathcal{F} \subset \mathcal{M}, |\mathcal{F}| < \omega, X = \bigcup \mathcal{F} \right\}$$

のことをいう. X の可測集合による有限な直和分割の期待驚愕度の和として定められる.^{†15} なお, 実際 \mathcal{F} は X の概分割で十分である.

要諦 1.6.5. \log は本質的ではない. $1 \mapsto 0, 0 \mapsto \infty$ を満たす関数で, $\frac{1}{x}$ より急激でないものならば良かったのではないかと. また, \log の底は分野によってさまざまである. Shanonn の entropy では 2,

^{†15} こういうものは本当に極限構成になっている. X の可算な分割は directed で h_μ は concave であるため.

第 2 章

確率変数の基本性質

前章で準備した公理と数学を用いて、どのように確率という概念に迫るかの数学手法を考える。

偏差値は、統計的な分布が正規分布で近似できることを暗黙裡に認めて算出している。これらの観念が定理として得られる。まず Bayes の定理や大数の弱法則が極めて簡潔に導出できる。これは一体何であるか？

これらの道具は全て数学が包含できる。確率という概念から本質的に数学的な構造を湛えた観念であり、さらには近似も数学的な行為である。

分布の近似のアルゴリズムが計算機自然であるとするなら、この世の中は関数解析である。量子論の頃から、人の使う数理モデルは本質的に漸近して行っている。

古典的な確率論を特徴付けてきたところの数学的な概念は、試行の独立性と確率変数の独立性の概念である。Laplace, Poisson, Chebyshev, Markov, von Mises, Bernstein の古典的研究は、独立な確率変数列についての基礎的な研究であり、Markov, Bernstein による現代的な研究は、完全な独立性よりも弱めた Markov 性の条件を課して考察している。このようにして、独立性の概念に、確率論固有の問題意識、少なくともその萌芽が見られる [3]。

確率論もどこか圏論的である。確率論で重要な概念である独立性とは、確率変数 = 射がある種の可換性を満たすことをいう。

2.1 確率測度の変換：条件付き確率と Bayes の定理

確率空間の万華鏡による拡大

離散の場合を考えると明らかであるが、確率空間とは「事象をどのように分割するか」が肝要になり、乗法を得ることとはそれについての知識の更新だと言える。その基本的な言葉は、「条件付き確率」にある。部分代数を取り出しているようなもので、再帰的な構造がある。ここでは可算な分割のみを考える（一般には条件付き期待値の議論になる）。

さらに進んだものだと、重点サンプリングや Girsanov 変換が、ファイナンスやフィルタリング問題で用いられる測度変換の例である。

定義 2.1.1 (conditional probability).

(1) $A, B \in \mathcal{F}, P(A) > 0$ とする。

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

を、事象 A の下での B の条件付き確率という。 $P_A(B)$ とも表す。^{†1} $P(A) = 0$ の時はこの値は任意に定めることで、2 変数測度 $P(\cdot | \cdot) : \mathcal{F} \times \mathcal{F} \rightarrow [0, 1]$ が定まる。

(2) $P(\cdot | A)$ は再び Ω 上の確率測度となる。

(3) $P_X(B) := P(B | X = \cdot)$ は X 上の可測関数となる。

^{†1} $i : A \hookrightarrow \Omega$ についての引き戻し測度である？

命題 2.1.2 (law of total probability : 全確率の分解). $(H_i)_{i \in I}$ を互いに背反で $\Omega = \bigsqcup_{i \in I} H_i$ を満たす事象の族とする .

$$\forall A \in \mathcal{F} \quad P(A) = \sum_{i \in I} P(A \mid H_i) \cdot P(H_i).$$

[証明]. $A = A \cap (\cup_{i \in I} H_i) = \cup_{i \in I} A \cap H_i$ より ,

$$\begin{aligned} P(A) &= P(\cup_{i \in I} A \cap H_i) \\ &= \sum_{i \in I} P(A \cap H_i) && \text{加法性} \\ &= \sum_{i \in I} P(A \mid H_i) P(H_i) && \text{定義 2.1.1} \end{aligned}$$

要諦 2.1.3. こうして 1.2 節で確認したとおり , 分割 $\sum H_i$ の取り方が測度の取り替えについて肝要になる .

定理 2.1.4 (Bayes). $(H_i)_{i \in I}$ を互いに背反で $\Omega = \bigsqcup_{i \in I} H_i$ を満たす事象の族とする . $P(A) > 0$ ならば ,

$$P(H_i \mid A) = \frac{P(A \mid H_i) P(H_i)}{\sum_{j \in I} P(A \mid H_j) P(H_j)}.$$

[証明].

$$\begin{aligned} P(H_i \mid A) &= \frac{P(H_i \cap A)}{P(A)} && \text{条件付き確率の定義 2.1.1} \\ &= \frac{P(A \mid H_i) P(H_i)}{\sum_{j \in I} P(A \mid H_j) P(H_j)} && \text{分母は定義 2.1.1 , 分子は全確率の分解則 2.1.2} \end{aligned}$$

要諦 2.1.5. 条件付き確率の第一引数・第二引数の入れ替え法則という意味で , 何かの変換則に見える . 情報の更新規則^{†2} $I = 1$ の場合は ,

$$P(H \mid E) = \frac{P(E \mid H) P(H)}{P(E)} \quad \Leftrightarrow \quad P(H \mid E) P(E) = P(E \mid H) P(H) = P(E \cap H)$$

となる .

2.2 事象の独立性

積への関手性

測度は純粹的に加法的な集合関数であって , 積との相互関係は一切考えなかった . ここで , 積への関手性は , 確率論的に重要な意味論を持つことをみる . 独立性は確率測度 \mathcal{P} の構造のみに依存する , 純粹に確率論的な概念である .

定義 2.2.1 (independent). 事象 $A, B \in \mathcal{F}$ について ,

- (1) 可測空間 (Ω, \mathcal{F}) で事象 $A, B \in \mathcal{F}$ が背反であるとは , $A \cap B = \emptyset$ であることをいう .
- (2) 測度空間 $(\Omega, \mathcal{F}, \mathcal{P})$ で事象 $A, B \in \mathcal{F}$ が独立であるとは , $\mathcal{P}(A \cap B) = \mathcal{P}(A) \mathcal{P}(B)$ であることをいう . $P(A) > 0$ のとき , $P(B \mid A) = P(B)$ に同値 .

注 2.2.2. $P(A), P(B) > 0$ のとき , A, B が背反であることと独立であることは両立し得ない (背反である) . これは , 「同時には起こらない」というのは従属関係であって , 無関係たり得ないとも捉えられる .

定理 2.2.3 (独立性の特徴付け). 2 つの事象 $X, Y \in \mathcal{F}$ について , 次の 3 条件は同値 .

^{†2} In quantum mechanics, the collapse of the wavefunction may be seen as a generalization of Bayes's Rule to quantum probability theory. This is key to the Bayesian interpretation of quantum mechanics.

- (1) X, Y は独立である .
 (2) $N := \{x \in \Omega^X \mid \exists y \in \Omega^Y P(Y = y | X = x) \neq P(Y = y)\}$ について , $P^X(N) = 0$.
 (3) 殆ど至る所 $P_X(Y = y) = P(Y = y)$.

[証明] .

(1) \Rightarrow (2) 任意の $x \in N$ について , X, Y の独立性より $P(X = x) > 0$ ならば $P_{X=x}\{Y = y\} = P\{Y = y\}$ が必要だから , $P\{X = x\} = 0$. よって , $P^X(N) = P\{X \in N\} = \sum_{x \in N} P\{X = x\} = 0$.

定義 2.2.4 (事象族の独立性).

- (1) 列 $(A_i)_{i \in \mathbb{N}} : \mathbb{N} \rightarrow \mathcal{F}$ が独立であるとは , $\forall 2 \leq k \leq n \quad \forall 1 \leq i_1 < \dots < i_k \leq n \quad P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$.
 (2) 族 $(A_i)_{i \in I}$ が独立であるとは , $\forall n \geq 2 \quad A_1, \dots, A_n$ が独立 であることとする .
 (3) 族 $(A_i)_{i \in I}$ が対独立であるとは , $\forall i \neq j \in I \quad A_i \cap A_j = \emptyset$.

例 2.2.5. 対独立であるが独立ではない例がある .

補題 2.2.6 (独立性の遺伝). 事象 A_1, \dots, A_n が独立であるとする .

- (1) 事象 A_1^c, \dots, A_n^c も独立である .
 (2) $A_1 \cup A_2, A_3 \cap A_4, A_5$ は独立である .

補題 2.2.7 (独立事象に対する確率の関手性). 列 $(A_i)_{i \in \mathbb{N}}$ が独立ならば ,

$$P(\cap_{n=1}^{\infty} A_n) = \prod_{n=1}^{\infty} P(A_n).$$

[証明] . $B_n := \cup_{k=1}^n A_k$ と定めると , これは単調列であるから ,

$$\begin{aligned} P(\cap_{n=1}^{\infty} A_n) &= \lim_{n \rightarrow \infty} P(B_n) \\ &= \lim_{n \rightarrow \infty} \prod_{k=1}^n P(A_k) = \prod_{n=1}^{\infty} P(A_n). \end{aligned}$$

2.3 確率変数と確率分布

前節で確率測度の変換を初等的な言葉で捉えた . ここで確率変数とは , 確率測度の変換を引き起こす射であると考えられる . 人間は , 主に実数値のものを考える . \mathbb{R} は数理モデルにおいて特別なのだ .

顕著な特徴として , 統計推測などにおいて , 確率変数 $X : \Omega \rightarrow \mathcal{X}$ が誘導する測度 P^X を観測することが問題となり , (Ω, \mathcal{F}, P) の特定の構造には執着しない . 特に , 経験過程論のように , (Ω, \mathcal{F}, P) を拡大して考えることがしばしばである .

2.4 確率変数の独立性

確率空間とは「分割」の定め方である . 確率変数が独立であるとは , これらが定める分割が , どの組み合わせも積確率について独立になることをいう . これは明らかに , 一般化された概念である .

定義 2.4.1 (independent). 確率変数 $X_1, \dots, X_n : \Omega \rightarrow \mathcal{X}_i$ が独立であるとは , 次を満たすことをいう :

$$\forall B_1 \in \mathcal{B}_1, \dots, B_n \in \mathcal{B}_n \quad \mathcal{P}^X(X_1 \in B_1, \dots, X_n \in B_n) = \mathcal{P}^X(X_1 \in B_1) \times \dots \times \mathcal{P}^X(X_n \in B_n).$$

ただし, $X := (X_1, \dots, X_n)$ を同時分布, P^X は直積測度 $P^{X_1} \times \dots \times P^{X_n}$ とした. とした. 事象 (A_i) の独立性は, $X_i = \chi_{A_i}$ の場合に当たる.

命題 2.4.2 (可測関数は独立性を保つ). 可測空間 $(\mathcal{Y}_i, \mathcal{G}_i)$ への射 $f_i: \mathcal{X}_i \rightarrow \mathcal{Y}_i$ について, 確率変数 X_1, \dots, X_n が独立ならば, $f_1(X_1), \dots, f_n(X_n)$ も独立である.

[証明]. 合成 $f_i \circ X_i$ も可測だから, 確かに $f_i(X_i)$ も確率変数である. 任意の $C_i \in \mathcal{G}_i$ について, $P^Y(f_i(X_i) \in C_i) = P^X(X_i \in f_i^{-1}(C_i))$ であるから,

$$\begin{aligned} P^Y(f_1(X_1) \in C_1, \dots, f_n(X_n) \in C_n) &= P^X(X_1 \in f_1^{-1}(C_1)) \times \dots \times P^X(X_n \in f_n^{-1}(C_n)) \\ &= P^Y(f_1(X_1) \in C_1) \times \dots \times P^Y(f_n(X_n) \in C_n). \end{aligned}$$

■

命題 2.4.3. 確率変数 $A_1, \dots, A_n: \Omega \rightarrow \mathcal{G}$ が独立であるとする. 任意の $I := [n] \supset J$ について, $\{A_j^c, A_k \mid j \in J, k \in I \setminus J\}$ は独立である.

[証明]. 写像 $f: \mathcal{G} \rightarrow \mathcal{G}$ を, A_j を A_j^c に写し, A_i を変えない $f(A_i) = A_i$ 写像として定義できたなら, この像変数も独立であることから従う.

■

2.5 独立な確率変数

確率変数が独立のとき, 作用素に種々の関手性が生じる

関数の積分と確率変数の期待値との間にある類似点が明らかになってきた. こうした類推はさらに拡張され, 独立な確率変数のさまざまな性質は, 対応する直交関数の性質と完全に類似しているものとみなされるようになった [3].

2.5.1 独立な確率変数に対する関手性

命題 2.5.1 (実空間の議論への持ち上げ). 実確率変数の族 $X := (X_1, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$ と任意の可測関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ に対して,

(1) $h := f \circ X: \Omega \rightarrow \mathbb{R}^n \rightarrow \mathbb{R}$ は実確率変数である.

(2) $E[h] = \int_{\mathbb{R}^n} f(x_1, \dots, x_n) P^X(dx_1, \dots, dx_n)$.

[証明]. 可測関数の合成は可測だから, h は当然実確率変数である.

f が特性関数の場合 $f = \chi_A$ ($A \in \mathcal{B}(\mathbb{R}^n)$) とする.

$$\begin{aligned} E[\chi_A \circ X] &= E[\chi_{X^{-1}(A)}] = 1 \cdot P(X^{-1}(A)) \\ &= P^X(A) = \int_{\mathbb{R}^n} \chi_A(x_1, \dots, x_n) P^X(dx). \end{aligned}$$

f が単関数の場合 積分の線形性より成り立つ.

f が一般の可測関数の場合 $f = f^+ - f^-$ について, それぞれの非負値単関数近似から, Lebesgue の優収束定理より.

■

系 2.5.2 (期待値が積を保つ条件). 実確率変数の族 $(X_i)_{i \in [n]}$ が独立である時, 任意の可積分な可測関数列 $(f_i)_{i \in [n]}$ に対して, $h := (f_1 \circ X_1) \cdot (f_2 \circ X_2) \cdots (f_n \circ X_n)$ とすると次が成り立つ: $E[h] = \prod_{i=1}^n E[f_i \circ X_i]$.

[証明]. 命題 2.5.1 より

$$E[h] = \int_{\mathbb{R}^n} f_1(x_1) \cdots f_n(x_n) P^X(dx_1, \dots, dx_n)$$

であるが, P^X は直積測度 $P^{X_1} \times \dots \times P^{X_n}$ であり 2.4.1, 測度空間 \mathbb{R}^n は σ -有限であるから, Fubini の定理より, $\prod_{i \in [n]} E[f_i(X_i)]$ に等しい.

■

系 2.5.3 (分散が和を保つ条件). $(X_i)_{i \in [n]}$ を二乗可積分で $(E[|X_i|^2] < \infty \ (i = 1, \dots, n))$, 互いに独立な確率変数の列とする. この時, 次が成り立つ: $\text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i]$.

[証明]. 2変数の場合,

$$\begin{aligned} V[X + Y] &= E[X + Y - E[X + Y]] = E[(X - EX) + (Y - EY)]^2 \\ &= V[X] + V[Y] + V(X, Y). \end{aligned}$$

であるから, 互いに独立な2変数の共分散は0であることを示せば良いが,

$$V(X, Y) = E[(X - EX)(Y - EY)] = E[XY] - E[X]E[Y] = 0.$$

■

要諦 2.5.4. ものすごく余弦定理っぽく, 内積の構造がある. 実際 Hilbert 空間の内積である. だから二乗可積分の条件があるのだ.

2.5.2 独立同分布

さらに理想的なクラスを定義する. 確率変数は分布を定めるが, 分布からそれを定める確率変数が存在するかはある種の逆問題で, 必ずしも自明でない.

定義 2.5.5 (independent and identically distributed). 実確率変数の列 $(X_i)_{i \in \mathbb{N}}$ が互いに独立であるだけでなく, 任意の分布 $P^{X_i} \ (i \in \mathbb{N})$ が等しい時, $(X_i)_{i \in \mathbb{N}}$ は独立同分布を持つという.

定理 2.5.6 (Kolmogorov extension theorem). 確率空間の列 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_n)$ が次の条件を満たすとする:

$$(\text{consistency}) \ \forall A \in \mathcal{B}(\mathbb{R}^n) \ P_n(A) = P_{n+k}(A \times \mathbb{R}^k).$$

この時, $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ 上の確率測度 P で, $\forall A \in \mathcal{B}(\mathbb{R}^n) \ P(\pi_n^{-1}(A)) = P_n(A)$ を満たすものが唯一つ存在する. ただし, $\pi_n := (\text{pr}_1, \dots, \text{pr}_n)$ と定めた.

[証明].

方針 任意の $A_n \in \mathcal{B}(\mathbb{R}^n)$ に対して, $\Lambda := \pi_n^{-1}(A_n)$ での値を $Q(\Lambda) := P_n(A_n)$ とする有限加法的な確率測度 $Q: \mathcal{B}(\mathbb{R}^\infty) \rightarrow [0, 1]$ を考える. 有限加法性の確認は, 族 $(A_n)_{n \in \mathbb{N}}$ に対して $m := \max_{n \in \mathbb{N}} \dim(A_n)$ として \mathbb{R}^m 上での P_n の有限加法性を考えれば良い. なお, この Q は well-defined である: $A_n \in \mathcal{B}(\mathbb{R}^n), A_m \in \mathcal{B}(\mathbb{R}^m)$ を用いて $\Lambda = \pi_n^{-1}(A_n) = \pi_m^{-1}(A_m)$ と2通りで表せる場合でも, 一貫性の条件より $Q(\Lambda) = P_n(A_n) = P_m(A_m)$ である. $\mathcal{A} := \{\pi_n^{-1}(A_n) \in \mathbb{R}^\infty \mid A_n \in \mathcal{B}(\mathbb{R}^n), n \in \mathbb{N}\}$ は集合体であることは同様に示す. すると, Q が \mathcal{A} 上で σ -加法的であることを示せば, \mathbb{R}^∞ は σ -有限であるから, $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R}^\infty)$ 上への一意的な延長が存在し, Q の定め方よりこれが条件を満たす (Hopf-Kolmogorov の拡張定理). すると, 補題より, 任意の単調減少列 $(A_n)_{n \in \mathbb{N}}$ について, $\alpha := \lim_{n \rightarrow \infty} Q(A_n) > 0 \Rightarrow \cap_{n \in \mathbb{N}} A_n \neq \emptyset$ を示せば良い.

証明 単調減少列 (Λ_n) を任意にとると, $\Lambda_n = \pi_{n_i}^{-1}(A_{n_i})$ と表せる. $\max_{n \in \mathbb{N}} n_i$ が存在するならば同様に示す. $\{n_i\}_{n \in \mathbb{N}} \subset \mathbb{N}$ は非有界とする. すると, 部分列をとって添字を打ち直すことより, $\Lambda_n = \pi_n^{-1}(A_n)$ とし得る. Borel 集合の位相的正則性より, $C_n \subset A_n, P_n(A_n \setminus C_n) < \frac{\alpha}{2^{n+1}}$ を満たすコンパクト集合の列 $(C_n \subset \mathbb{R}^n)_{n \in \mathbb{N}}$ が取れる. $D_n := \pi_n^{-1}(C_n)$ と定めると, $D_n \subset \Lambda_n, Q(\Lambda_n \setminus D_n) < \frac{\alpha}{2^{n+1}}$ を満たす. $\bar{D}_n := \cap_{k=1}^n D_k$ とおくと,

$$\begin{aligned} Q(\bar{D}_n) &= Q(\Lambda_n) - Q(\Lambda_n \setminus \bar{D}_n) \\ &\geq Q(\Lambda_n) - \sum_{k=1}^n Q(\Lambda_k \setminus D_k) \geq \frac{\alpha}{2} \end{aligned}$$

より, $\bar{D}_n \neq \emptyset$ である. よって, 空でない閉集合の単調減少列の極限は空ではなく, $\cap_{k=1}^\infty D_k \neq \emptyset$.

こうして, $D_k \subset \Lambda_k$ であって, $\emptyset \subsetneq \cap_{k=1}^\infty D_k \subset \cap_{k=1}^\infty \Lambda_k$ が従う.

注 2.5.7. 一般の完備可分距離空間上の確率空間の列について成り立つ。

補題 2.5.8 (有限加法的確率空間の完全加法性の単調族による特徴付け). Q を有限加法的な確率測度, \mathcal{A} を集合体とする. この時, Q についての次の2条件は同値.

- (1) \mathcal{A} 上 σ -加法的である. すなわち, $A \in \mathcal{A}$ に収束する互いに素な \mathcal{A} -列 $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ について, $Q(A) = \sum_{n \in \mathbb{N}} Q(A_n)$.
- (2) 任意の \mathcal{A} の単調減少列 (A_n) に対して, $Q\left(\lim_{n \rightarrow \infty} A_n\right) = Q\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} Q(A_n)$.
- (3) 任意の \mathcal{A} の単調減少列 (A_n) に対して, $\lim_{n \rightarrow \infty} Q(A_n) > 0 \Rightarrow \bigcap_{n \in \mathbb{N}} A_n \neq \emptyset$ である.

[証明].

(1) \Rightarrow (2) \mathcal{A} 上の有限加法的測度 μ が \mathcal{A} 上 σ -加法的であることは, 任意の単調増加列 (A_n) について $\mu(\lim_{n \rightarrow \infty} A_n) = \mu(\bigcup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mu(A_n)$ を満たすことと同値. 任意の単調減少列 (A_n) に対して, その補集合の定める単調増加列を考えると,

$$Q(\lim_{n \rightarrow \infty} A_n) = Q(\bigcap_{n \in \mathbb{N}} A_n) = 1 - Q(\bigcup_{n \in \mathbb{N}} \overline{A_n}) = 1 - \lim_{n \rightarrow \infty} Q(\overline{A_n}) = \lim_{n \rightarrow \infty} Q(A_n).$$

最後の等号は Q の有限加法性 $\forall_{n \in \mathbb{N}} Q(A_n + \overline{A_n}) = Q(A_n) + Q(\overline{A_n}) = 1$ による.

(2) \Rightarrow (3) 自明.

(3) \Rightarrow (1) \mathcal{A} 内に収束する互いに素な列 (A_n) を任意に取り, $A := \bigcup_{n \in \mathbb{N}} A_n$ と定める. $B_1 := A, B_n := A \setminus (\bigcup_{i=1}^{n-1} A_i)$ と帰納的に定めると, これは \emptyset に収束する単調減少列である. よって,

$$\begin{aligned} \lim_{n \rightarrow \infty} Q(B_n) = 0 &\Leftrightarrow \lim_{n \rightarrow \infty} Q(A \setminus (\bigcup_{i=1}^{n-1} A_i)) = 0 \\ &\Leftrightarrow \lim_{n \rightarrow \infty} (Q(A) - Q(\bigcup_{i=1}^{n-1} A_i)) = 0 && \because Q \text{ の有限加法性} \\ &\Leftrightarrow Q(A) = \lim_{n \rightarrow \infty} Q(\bigcup_{i=1}^{n-1} A_i) = \sum_{n=1}^{\infty} Q(A_n). && \because Q \text{ の有限加法性} \end{aligned}$$

定理 2.5.9 (独立同分布をもつ確率変数の族の存在). \mathbb{R} 上の確率測度 μ を分布にもつ \mathbb{R}^∞ 上の独立同分布 $(X_i)_{i \in \mathbb{N}}$ が存在する.

[証明].

構成 確率空間の列 $((\mathbb{R}^l, \mathcal{B}(\mathbb{R}^l), P_l))_{l \in \mathbb{N}}$ を $P_l := \otimes_{i=1}^l \mu$ と定めると, 一貫性条件を満たすから Kolmogorov の拡張定理 2.5.6 より, \mathbb{R}^∞ 上の確率測度 P で $\forall_{A \in \mathcal{B}(\mathbb{R}^n)} P(\text{pr}_n^{-1}(A)) = P_n(A)$ を満たすものが定まる. これに対して, $X_n := \text{pr}_n : \mathbb{R}^\infty \rightarrow \mathbb{R}$ と定めれば良い.

確認 実際,

(1) (同分布) 任意の $i \in \mathbb{N}$ と $E_i \in \mathcal{B}(\mathbb{R})$ に対して, P の定め方より,

$$\begin{aligned} P^{X_i}(E_i) &= P(\{X_i \in E_i\}) = P(\bigcap_{j < i} \{X_j \in \mathbb{R}\} \cap \{X_i \in E_i\}) \\ &= P(\pi^{-1}(\mathbb{R}^{i-1} \times E_i)) = P_i(\mathbb{R}^{i-1} \times E_i) && P \text{ の定め方 (一貫性条件)} \\ &= 1 \cdot \mu(E_i). && P_i \text{ の定義} \end{aligned}$$

よって $\forall_{i \in \mathbb{N}} P^{X_i} = \mu$ であるから, $(X_i)_{i \in \mathbb{N}}$ は同分布.

(2) 任意の $i \in \mathbb{N}$ に対して, $\tilde{X} := (X_1, \dots, X_n)$ を同時分布とすると, 任意の $E = E_1 \times \dots \times E_n \in \mathcal{B}(\mathbb{R}^n)$ に対して,

$$\begin{aligned} P^{\tilde{X}}(E) &= P(\bigcap_{i=1}^n \{X_i \in E_i\}) = P(\pi_n^{-1}(E)) \\ &= P_n(E) = \prod_{i=1}^n \mu(E_i) = \prod_{i=1}^n P^{X_i}(E_i) \end{aligned}$$

より, 独立性も従う.

例 2.5.10 (Bernoulli sequence / process). 特に確率空間 Ω を介する射は、ヤヌス対象や2進法やTVではないが、特殊なクラスの確率変数である。 $P(\{X_i = 0\}) = p, P(\{X_i = 1\}) = 1 - p$ ($0 \leq p \leq 1$) なる分布 $P^X: \{0, 1\} \rightarrow [0, 1]$ に従う2値確率変数の列 $(X_i)_{i \in \mathbb{N}}$ を **Bernoulli 過程**という。その存在は独立同分布を持つ確率変数の族の存在 2.5.9 により保証される。最初の n 回のうちの成功数は二項分布に従い、 r 回成功するのに必要な回数は負の二項分布に従う。 $r = 1$ を幾何分布という。

2.6 確率変数の和・積・商の分布

確率変数の積 $(X_1, \dots, X_n): \Omega_1 \times \dots \times \Omega_n \rightarrow \mathcal{X}_1 \times \dots \times \mathcal{X}_n$ が引き起こす分布を同時・結合分布という。これは単に標準的な構成であるが、では、確率変数の演算は、測度の演算にどのように対応するのであろうか？独立な確率変数の和が引き起こす分布は、測度の畳み込みである。

定義 2.6.1 (convolution of measures). $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の測度 μ, ν の畳み込みとは、確率測度

$$\mu * \nu(E) := \int_{\mathbb{R}} \nu(E - y) \mu(dy) = \iint_{\mathbb{R}} \chi_E(x + y) \nu(dx) \mu(dy)$$

を指す。ただし、 $E - y := \{x - y \in \mathbb{R} \mid x \in E\}$ を E を平行移動した集合とした。^{†3}

命題 2.6.2. 2つの確率変数 X_1, X_2 は像測度 μ, ν を定め、互いに独立であるとする。この時、確率変数 $X_1 + X_2$ の像測度は畳み込み $\mu * \nu$ である。

[証明]。同時分布を $X := (X_1, X_2)$ とおくと、 X_1, X_2 は互いに独立であるから、像測度は直積測度に一致する： $P^X = P^{X_1} \otimes P^{X_2}$ 。よって、任意の事象 $E \in \mathcal{B}(\mathbb{R})$ について、

$$\begin{aligned} \mu * \nu(E) &= \iint_{\mathbb{R}} \chi_E(x + y) P^{X_1}(dx) P^{X_2}(dy) \\ &= \iint_{\mathbb{R}} \chi_E(x + y) P^X(dx dy) \\ &= P(X_1 + X_2 \in E) = P^X(E). \end{aligned}$$

■

2.7 Markov 連鎖

互いに独立な試行の列（確率変数の列）の一般化を考える。

2.8 大数の法則

大数の法則は物理学実験でも扱ったある種の自然現象であるが、これが定理として導けるような公理系を、我々は用意できたのである。実際は、いずれの大数の法則も、仮定は可積分性 $E[|X_1|] < \infty$ で十分。

定義 2.8.1 (convergence in probability, almost sure convergence).

- (1) 確率変数の族 $(Y_i)_{i \in \mathbb{N}}$ と確率変数 X について、次が成り立つ時、 (Y_n) は X に確率収束するという： $\forall \epsilon > 0 \lim_{n \rightarrow \infty} P(|Y_n - X| \geq \epsilon) = 0$ 。
- (2) 確率変数の族 $(Y_i)_{i \in \mathbb{N}}$ と確率変数 X について、次が成り立つ時、 (Y_n) は X に概収束するという： $P\left(\lim_{n \rightarrow \infty} Y_n = X\right) = 1$ 。

それぞれを形式化すると、

^{†3} 集合 E を平行移動しながら、元の位置から移動させた時の測度の変化を足し上げていく。

$$(1) \forall \epsilon_1 > 0 \forall \epsilon_2 > 0 \exists N > 0 \forall n \geq N P(|Y_n - X| \geq \epsilon_1) < \epsilon_2.$$

$$(2) \forall \epsilon > 0 \exists N > 0 \forall n \geq N P(Y_n = X) = 1 - \epsilon.$$

すると、(2)⇒(1) であるが、(1)⇒(2) は反例が構成できる。^{†4}

独立同分布の場合は期待値について

$$E \left[\left| \sum_{i=1}^n (X_i - \mu) \right|^{2k} \right] = \sum_{i_1, \dots, i_{2k}} E[(X_{i_1} - \mu) \cdots (X_{i_{2k}} - \mu)] \leq C n^k \quad \exists C \in \mathbb{R}$$

という評価が使えるので議論が簡単になる。

2.8.1 独立同分布での大数の弱法則

まず収束とは何かが問題になる。収束とは基本的に距離空間で定義される概念であるが、測度を用いて一般化することが出来るのであった。確率収束は概収束より弱いので、

記法 2.8.2. 確率空間 (Ω, \mathcal{F}, P) 上の独立同分布 $(X_i)_{i \in \mathbb{N}}$ に対して、 $S_n := \sum_{i=1}^n$ と定める。

定理 2.8.3 (weak law of large numbers). 独立同分布を持つ確率変数の族 $(X_i)_{i \in \mathbb{N}}$ が二乗可積分であるとする： $E[(X_1)^2] < \infty$ 。この時、次が成り立つ：

$$\forall \epsilon > 0 \lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{n} - E[X_1] \right| \geq \epsilon \right) = 0.$$

[証明]。Schwarz の不等式より、 $E[|X_1|] \leq \sqrt{E[(X_1)^2]} < \infty$ であるから、特に可積分である。任意の $\epsilon > 0$ と $n \in \mathbb{N}_{>0}$ に対して、 $X := |S_n - nE[X_1]|$, $f(a) = a^2$ とすると、Chebyshev の不等式 1.5.1 より、

$$P(|S_n - nE[X_1]| \geq n\epsilon) \leq \frac{E[|S_n - nE[X_1]|^2]}{(n\epsilon)^2} = \frac{nE[|X_1 - E[X_1]|^2]}{(n\epsilon)^2}$$

より結論を得る。なお、最右辺の変形は、独立な確率変数に対する分散の線型性 2.5.3 による。 ■

2.8.2 独立同分布での大数の強法則

確率収束は位相を定め、また距離化可能でもあるが、

定理 2.8.4 (strong law of large numbers). 独立同分布を持つ確率変数の族 $(X_i)_{i \in \mathbb{N}}$ が4乗可積分であるとする： $E[(X_1)^4] < \infty$ 。この時、次が成り立つ：

$$P \left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = E[X_1] \right) = 1.$$

[証明]。以下のことに注意する。

(1) Schwarz の不等式より $E[|X_1|] \leq (E[(X_1)^2])^{1/2} \leq (E[(X_1)^4])^{1/4} < \infty$ であるから特に可積分。

(2) 2次と4次の中心積率は

$$E[(X_1 - \mu)^4] \leq 8(E[(X_1)^4] + \mu^4) < \infty, \quad E[(X_1 - \mu)^2] \leq 2(E[(X_1)^2] + \mu^2) < \infty,$$

と評価できる。

Chebyshev の不等式で期待値へ還元 $A_n^\epsilon := \left\{ \omega \in \Omega \mid \left| \frac{S_n(\omega)}{n} - m \right| > \epsilon \right\}$ と置くと、 $f(a) = a^4$ についての Chebyshev の不等式 1.5.1 より、

$$P(A_n^\epsilon) = P \left(\left| \sum_{i=1}^n (X_i - \mu) \right| > \epsilon n \right) \leq \frac{E \left[\left(\sum_{i=1}^n (X_i - \mu) \right)^4 \right]}{(\epsilon n)^4}$$

^{†4} Y_n がかさばりながら X に近づくとき、 ϵ 範囲には必ず入るが、 $\epsilon/2$ 範囲には半分しか入らない、というような収束の仕方もあるはずである。

と評価できる．

期待値を抑える 4 次の中心積率の和は

$$\begin{aligned} E \left[\left(\sum_{i=1}^n (X_i - \mu)^4 \right) \right] &= \sum_{i_1, i_2, i_3, i_4=1}^n E[(X_{i_1} - \mu)(X_{i_2} - \mu)(X_{i_3} - \mu)(X_{i_4} - \mu)] \\ &= nE[(X_1 - \mu)^4] + 3n(n-1)(E[(X_1 - \mu)^2])^2 \quad \because \text{独立な変数の共分散は 0} \\ &\leq Cn^2 \quad \exists C \in \mathbb{R} \end{aligned}$$

と評価できるから，

$$\sum_{n=1}^{\infty} P(A_n^{\epsilon}) \leq \sum_{n=1}^{\infty} \frac{C}{(\epsilon n)^4} n^2 = \frac{C}{\epsilon^4} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

である．

Borel-Cantelli の補題 $B_{\epsilon} := \limsup_{n \rightarrow \infty} A_n^{\epsilon} = \left\{ \omega \in \Omega \mid \limsup_{n \rightarrow \infty} \left| \frac{S_n(\omega)}{n} - \mu \right| > \epsilon \right\}$ とすると，Borel-Cantelli の補題 1.2.9(1) より， $P(B_{\epsilon}) = 0$ ．

結論 以上より，補集合が

$$\begin{aligned} P \left(\left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} \left| \frac{S_n(\omega)}{n} - \mu \right| = 0 \right\}^c \right) &= P \left(\lim_{\epsilon \rightarrow 0} \left\{ \omega \in \Omega \mid \limsup_{n \rightarrow \infty} \left| \frac{S_n(\omega)}{n} - \mu \right| > \epsilon \right\} \right) \\ &\leq P(\lim_{n \rightarrow \infty} B_{\epsilon}) = \lim_{\epsilon \rightarrow 0} P(B_{\epsilon}) \because (B_{\epsilon})_{\epsilon > 0} \text{ は単調減少族より.} \\ &= 0. \end{aligned}$$

■

2.8.3 証明抽出と評価の精緻化

次の定理は Hausdorff が Bernoulli 列の場合について最初に証明した．

定理 2.8.5. $\forall_{k \in \mathbb{N}} E[|X_1|^k] < \infty$ のとき，任意の $\epsilon' > 0$ に対して， $P \left(\lim_{n \rightarrow \infty} \frac{S_n - n\mu}{n^{1/2+\epsilon'}} = 0 \right) = 1$ ． $\epsilon' = 1/2$ のときを大数の強法則という．

定理 2.8.6 (収束のオーダー：law of iterated logarithm (Khinchin)). $E[(X_1)^2] < \infty$ のとき， X_1 の分散を σ^2 とすると，

$$P \left(\limsup_{n \rightarrow \infty} \frac{|S_n| - n\mu}{\sqrt{2n \log \log n}} = \sigma \right) = 1$$

2.9 数学における大数の法則的現象

2.9.1 Weierstrass の多項式近似

Bernstein の基底関数 $b_{k,b}$ を Bernoulli 試行 $B(n, x)$ の確率 $b(k; n, x)$ を表しているを見ると，Weierstrass の多項式近似の議論は大数の法則の議論と同じ構造をしている．すなわち，各サンプル $k/n \in [0, 1]$ で重み付きに近似していけば， $k/n \rightarrow x$ に概収束するから， $f(x)$ は $P_n(x)$ で近似できる．すると，全ての関数は確率変数の退化（特殊化）なのかもしれない．となると，物理学理論が確率論化したのは自然で，いずれ全ての理論がそうなるであろうという新たな自然法則に向き合いつつあるのかもしれない．2 項展開の各項を Bernoulli 過程の確率を表す項 $b(k; n, x)$ と見る，という見方は，確率論を形式化した恩恵なのかもしれない．

定義 2.9.1 (Bernstein polynomial). $b_{k,b}(x) := \binom{n}{k} x^k (1-x)^{n-k}$ ($k \in n+1$) の形で表される多項式を， n 次の Bernstein の

(基底) 関数という．これらは n 次以下の多項式がなす実線型空間の基底をなし，1 の分割をなす： $\sum_{k=0}^n b_{k,n} = 1$ ．

これに対して, $B_n : C([0, 1]) \rightarrow \mathbb{R}[x]$ を $B_n(f) := \sum_{k=0}^n f\left(\frac{k}{n}\right) b_{k,n}$ とおくと, 一様位相において $\lim_{n \rightarrow \infty} B_n(f) = f$.

定理 2.9.2 (Weierstrass の多項式近似). 任意の連続関数 $f \in C([0, 1])$ について, 多項式の列 $(P_n)_{n \in \mathbb{N}}$ ($\deg P_n = n$) が存在して, $\lim_{n \rightarrow \infty} \max_{x \in [0, 1]} |P_n(x) - f(x)| = 0$.

[証明].

構成 任意の $x \in [0, 1]$ について, これを成功確率とする Bernoulli 試行 $B(n, x)$ に従う Bernoulli 列 $(X_i^x)_{i \in \mathbb{N}}$ を取る (独立同分布に従う確率変数の列の存在定理 2.5.9). これが定める確率変数を $S_n^x := \sum_{i=1}^n X_i^x$ とすると, この Bernoulli 試行 $B(n, x)$ の期待値の f による押し出しの期待値を $P_n(x) := E\left[f\left(\frac{S_n^x}{n}\right)\right]$ とすると, $k \in n+1$ 回成功する確率はそれぞれ $P(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k}$ と表せるため,

$$P_n(x) = \sum_{k=1}^n f\left(\frac{k}{n}\right) \cdot \binom{n}{k} x^k (1-x)^{n-k}$$

とも表せる.

検証 いま, $\delta : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ を $\delta(\epsilon) := \sup \{|f(x) - f(y)| \in \mathbb{R}_{\geq 0} \mid x, y \in [0, 1], |x - y| \leq \epsilon\}$ と定めると, $[0, 1]$ 上の関数は連続ならば一様連続だから, $\epsilon \rightarrow 0$ のとき $\delta(\epsilon) \rightarrow 0$. また, $M := \max_{x \in [0, 1]} |f(x)|$ とおくと, 任意の $\epsilon > 0$ に対して,

$$\begin{aligned} \max_{x \in [0, 1]} |f(x) - P_n(x)| &= \max_{x \in [0, 1]} \left| E\left[f(x) - f\left(\frac{S_n}{n}\right)\right] \right| & P_n(x) &= E\left[f\left(\frac{S_n}{n}\right)\right] \\ &\leq \max_{x \in [0, 1]} E\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \\ &= \max_{x \in [0, 1]} \left\{ \int_{\{\omega \in \Omega \mid \left|\frac{S_n(\omega)}{n} - x\right| \geq \epsilon\}} \left|f(x) - f\left(\frac{S_n}{n}\right)\right| dP \right. \\ &\quad \left. + \int_{\{\omega \in \Omega \mid \left|\frac{S_n(\omega)}{n} - x\right| < \epsilon\}} \left|f(x) - f\left(\frac{S_n}{n}\right)\right| dP \right\} \\ &\leq \max_{x \in [0, 1]} 2MP(|S_n(\omega)n - x| \geq \epsilon) + \delta(\epsilon) & \text{第一項は } M \text{ の } 2 \text{ 倍で, 第二項は } \delta(\epsilon) \text{ で抑えられる} \\ &\leq \max_{x \in [0, 1]} \frac{2Me[|X_1^x - x|^2]}{n\epsilon^2} + \delta(\epsilon) & \text{大数の弱法則 2.8.3 と同様 Chebyshev の不等式} \\ &\leq \frac{2M}{n\epsilon^2} + \delta(\epsilon) & B(1, x) \text{ の分散 } \leq x(1-x) \leq 1 \end{aligned}$$

と評価できる. すると, $\forall \epsilon > 0 \exists N > 0 \forall n \geq N \max_{x \in [0, 1]} |f(x) - P_n(x)| < \epsilon$ を得た.

■

要諦 2.9.3. $f \in C([0, 1])$ を, 確率空間 $([0, 1], \mathcal{B}([0, 1]), P)$ 上の実確率変数だと思つと, 少し難しすぎる. そこで, 離散的な確率空間へと引き戻して考え, これらの離散空間からの実確率変数の列 $\lim_{n \rightarrow \infty} (n^{-1})^* f$ の極限だと考える:

$$\begin{array}{ccccc} n+1 & \xrightarrow{\times \frac{1}{n}} & [0, 1] & \xrightarrow{f} & \mathbb{R} \\ \psi & & \psi & & \psi \\ k & \longmapsto & \frac{k}{n} & \longmapsto & f\left(\frac{k}{n}\right). \end{array}$$

すると, $f(x)$ の値は, 大数の法則より, 確率 $x \in [0, 1]$ で成功する Bernoulli 試行 $B(n, x)$ の期待値という確率変数 S_n/n の f による押し出しで, 近似できる.

2.10 中心極限定理

偏差値は、統計的な分布が正規分布で近似できることを暗黙裡に認めて算出している．どう考えても「極限分布の標準分解」とか呼ぶべきだと思うが、Pólya が 1920 年の論文で「確率論において中心的な役割を果たすであろう」ということから命名した．

2.11 確率分布の収束

記法 2.11.1. 標準的な確率空間 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率測度全体のなす集合を $\mathcal{P}(\mathbb{R}^n)$ で表す．この空間のでの収束は、全変動収束、各集合収束、弱収束の 3 つなどが考えられる．弱収束は分布収束とも呼ばれ、分布関数で特徴づけることができる．また分布関数は càdlàg 関数としての特徴付けを持つので、 $D \subset B$ 空間という対象への注目が自然に生じる． D 値確率過程もまた標準的な対象の一つである．

定義 2.11.2 (weak convergence of measure / convergence in distribution / law). $\mathcal{P}(\mathbb{R}^n)$ の列 $(\mu_k)_{k \in \mathbb{N}}$ が $\mu \in \mathcal{P}(\mathbb{R}^n)$ に弱収束する $\mu_k \rightarrow \mu$ ($k \rightarrow \infty$) とは、次が成り立つことをいう： $\forall f \in B(\mathbb{R}^n) \quad \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f(x) \mu_k(dx) = \int_{\mathbb{R}^n} f(x) \mu(dx)$.

定義 2.11.3 ((cumulative) distribution function). 確率測度 $\mu \in \mathcal{P}(\mathbb{R}^n)$ が定める (累積) 分布関数 $F : \mathbb{R}^n \rightarrow [0, 1]$ とは、 $F(x) := \mu((-\infty, x_1] \times \cdots \times (-\infty, x_n])$ ($x = (x_1, \dots, x_n) \in \mathbb{R}^n$) と定められる．

要諦 2.11.4. Euclid 空間上の Borel σ -加法族は $(a, b]$ の全体から生成されるから、Borel 確率測度は全て累積分布関数によって定まり、1 対 1 対応する．これは特性関数のような標準的な対応である．

補題 2.11.5. F を分布関数とする．

- (1) 右連続である： $\lim_{0 < h \rightarrow 0} F(x + h) = F(x)$.
- (2) 左極限が存在する： $\lim_{0 < h \rightarrow 0} F(x - h) \in \mathbb{R}$.
- (3) 単調増加である： $x \leq y \Rightarrow F(x) \leq F(y)$.
- (4) $\lim_{x \rightarrow \infty} F(x) = 1, \lim_{x \rightarrow -\infty} F(x) = 0$.

命題 2.11.6 (分布関数の特徴付け). $F : \mathbb{R} \rightarrow \mathbb{R}$ が補題 2.11.5 の 3 条件を満たすとする． $\mu((a, b]) := F(b) - F(a)$ ($a \leq b \in \mathbb{R}$) は $\mathcal{B}(\mathbb{R})$ 上の確率測度に一意的に延長する．

定理 2.11.7 (弱収束の特徴付け). 任意の $\mathcal{P}(\mathbb{R}^n)$ の列 $(\mu_k)_{k \in \mathbb{N}}$ と $\mu \in \mathcal{P}(\mathbb{R}^n)$ について、次の 6 条件は同値．

- (1) $\mu_k \rightarrow \mu$ ($k \rightarrow \infty$) .
- (2) (連続な点の上で分布が収束する) μ_k, μ の分布関数 F_k, F について、 $\forall x \in \mathbb{R}^n \quad \lim_{y \rightarrow x} F(y) = F(x) \Rightarrow \lim_{k \rightarrow \infty} F_k(x) = F(x)$.
- (3) 任意の開集合 O に対して、 $\liminf_{k \rightarrow \infty} \mu_k(O) \geq \mu(O)$.
- (4) 任意の閉集合 F に対して、 $\limsup_{k \rightarrow \infty} \mu_k(F) \leq \mu(F)$.
- (5) $\forall A \in \mathcal{B}(\mathbb{R}^n) \quad \mu(\partial A) = 0 \Rightarrow \lim_{k \rightarrow \infty} \mu_k(A) = \mu(A)$.
- (6) $\forall f \in C_0(\mathbb{R}^n) \quad \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f(x) \mu_k(dx) = \int_{\mathbb{R}^n} f(x) \mu(dx)$.

[証明] . $n = 1$ の場合について証明する．

(1) \Rightarrow (2) (1) 実は $\forall x \in \mathbb{R} \quad F_\mu(x - 0) \leq \liminf_{n \rightarrow \infty} F_{\mu_n}(x) \leq \limsup_{n \rightarrow \infty} F_{\mu_n}(x) \leq F_\mu(x)$ が成り立つから、 F が x において連続ならば、 $\lim_{n \rightarrow \infty} F_{\mu_n}(x)$ は存在し、 $F_\mu(x)$ に一致する．

(2)

$$f_k(y) := \begin{cases} 1, & y \in (-\infty, x], \\ -ky + kx + 1, & y \in \left(x, x + \frac{1}{k}\right), \\ 0, & y \in \left[x + \frac{1}{k}, \infty\right). \end{cases}$$

とすると, $(f_k)_{k \in \mathbb{N}}$ は $B(\mathbb{R})$ の列であり, $f_k \xrightarrow{k \rightarrow \infty} \chi_{(-\infty, x)}$. よって, μ_k は μ に弱収束するから, 任意の $n \in \mathbb{N}, x \in \mathbb{R}$ について,

$$F_{\mu_n}(x) \leq \int_{\mathbb{R}} f_k(y) \mu_n(dy) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} f_k(y) \mu(dy) \xrightarrow{k \rightarrow \infty} \int_{\mathbb{R}} \chi_{(-\infty, x)} \mu(dy) = F_{\mu}(x)$$

と評価できるから, $\limsup_{n \rightarrow \infty} F_{\mu_n}(x) \leq F_{\mu}(x)$.

(3)

$$g_k(y) := \begin{cases} 1, & y \in \left(-\infty, x - \frac{1}{k}\right], \\ -ky + kx + 1, & y \in \left(x - \frac{1}{k}, x\right), \\ 0, & y \in [x, \infty). \end{cases}$$

について同様の議論を繰り返す.

(2) \Rightarrow (1) F_{μ} の不連続点を $D_{\mu} := \{x \in \mathbb{R} \mid F_{\mu}(x-0) > F_{\mu}(x)\}$ と置くと, これは高々可算である. 実際,

$D_n := \{x \in \mathbb{R} \mid F_{\mu}(x) - F_{\mu}(x-0) \geq 1/n\}$ とおくと $D_{\mu} = \bigcup_{n=1}^{\infty} D_n$ であるが, $\text{Im } F_{\mu} \subset [0, 1]$ より, $|D_n| \leq n$ であるため. $f \in \mathcal{B}(\mathbb{R})$ を任意に取り, 任意の $\epsilon > 0$ に対して,

$$\begin{aligned} & \left| \int_{\mathbb{R}} f(x) \mu_n(dx) - \int_{\mathbb{R}} f(x) \mu(dx) \right| \\ & \leq \left| \int_{(-\infty, a) \cup (b, \infty)} f(x) \mu_n(dx) \right| + \left| \int_{(-\infty, a) \cup (b, \infty)} f(x) \mu(dx) \right| + \sum_{k=0}^{m-1} \left| \int_{[a_k, a_{k+1}]} f(x) \mu_n(dx) - \int_{[a_k, a_{k+1}]} f(x) \mu(dx) \right| \end{aligned}$$

を ϵ で抑えれば良い. ここで, $a, b \in \mathbb{R} \setminus D_{\mu}$ を $F_{\mu}(a) < \epsilon, 1 - F_{\mu}(b) < \epsilon$ を満たすように取れる (分布関数の極限 2.11.5(4)). すると, (2) の仮定から, 十分大きい $n \geq N$ について, $F_{\mu_n}(a) < 2\epsilon, 1 - F_{\mu_n}(b) < 2\epsilon$ が成り立つから, 最初の2項は $M := \|f\|_{\infty}$ として, $4M\epsilon + 2M\epsilon$ で抑えられる.

(1) \Rightarrow (3)

■

要諦 2.11.8. (2) \Rightarrow (1) は, 累積分布関数の尾の部分と山の部分で分けて評価する. 任意の有界な連続関数について, その期待値が収束するように確率分布も収束することは, 累積分布関数の殆ど至る所 (たぶん) での各点収束に等しい.

2.12 大偏差理論

分布を近似するにあたって, 偏差が大きい部分の挙動を捉える.

第 3 章

離散確率分布

確率測度 $P(X)$ の具体例を、 X が可算集合 $|\mathcal{X}| \leq \infty$ の場合について考える．特に、 $\mathcal{X} = \mathbb{N}$ の場合について考えれば十分である．この場合の確率測度を特に、離散分布という．一般の議論は絶対連続な測度を用いて行うが、現象論的な本質は離散の場合にすでに宿る．

確率変数とは、標本空間の取り替えの一般化であり、つまりは観測行為の形式化である．射による映り込みしか我々は観測することができないのだとしたら、これはまるで近傍座標である．そして、特に認知容易性が高い標本空間 $(\mathbb{R}, \mathcal{B}^1(\mathbb{R}), m)$ を 1 つ定めて特別視する．これが科学という営みである．

となると、根元事象 Ω は、多様体のように、むしろ抽象的存在であって、束の底空間として使うべきものである．すると、確率変数の定義は、適切な束の使い方として認められるもの = ファイバーが可測なもののみとなる．これは可測関数の定義である．(確率変数について、始域と終域を対等に扱うのはあまり良い案ではないのではないか? という思索である)．

注 (statistics). 確率分布を特徴づける変数を母数という．従来の頻度主義に基づく統計学では、これらの母数是不確定ではあるが何らかの値をもった定数であると考ええる．一方ベイズ主義の統計学では、母数を固有の分布を持つ確率変数と考え、その不確定性を確率分布で記述する．標本から求める値である統計量は、標本のもとになる母集団の母数の推定量として用いる．たとえば「標本平均」 \bar{X} は母集団の「平均」母数 μ の推定量である．

記法 3.0.1. コンパクトハウスドルフ群 X 上の連続関数のなす Banach 代数 $C(X)$ の双対空間 $M(X)$ は、Radon 測度の空間であり、測度の量み込みについて再び Banach 代数をなす．この内で確率測度のなす部分空間を $P(X) := \{P \in M(X) \mid \|P\| \leq 1, P(1) = 1\} \subset (C(X))^*$ で表すと、これは Dirac 測度を極点とする w^* -コンパクトな凸集合をなす．

3.1 平均

まさか、確率空間とは構造の入った多様体か．「観測」と「座標変換」の構造を持つ、極めて幾何的な setup である．

まず、分布を記述するにあたって最も基本的な射を定義する．前章のように確率空間の概念を定義すれば、期待値とは積分に他ならない．積分作用素を特に平均という．すると、あらゆる有界線型作用素は、何かしらの核についての期待値であり、これが積率という考え方につながる．こうして、確率測度の違いを、線型作用素をうまく選ぶことによって判別するという問題が統計推測であるという枠組みが見えてくる．

\mathbb{N} 上の恒等写像 $\text{id}_{\mathbb{N}}$ の確率測度による積分を平均といい、一般の確率変数 $X: \mathbb{N} \rightarrow \mathcal{X}$ の確率測度による積分を期待値という．積率などの諸概念も、一般の確率変数について拡張できる．

補題 3.1.1 (確率分布の特徴付け). こうして $X: \Omega \rightarrow \mathcal{X}$ を用いて押し出した終確率空間 $(\mathcal{X}, \mathcal{B}, P^X)$ を得る．これについて、次が成り立つ．

$$(1) \text{ (non-negative) } \forall_{U \subset \mathcal{X}} \int_U dP^X \geq 0.$$

$$(2) \text{ (normalized) } \int_{\mathcal{X}} dP^X = 1.$$

要諦 3.1.2 (information geometry). 確率分布全体の空間は距離構造を持つ (a statistical manifold)．例えば正規化されたガウス分布は、近傍座標 (μ, σ) をもつ 2 次元多様体である．これを研究する分野が情報幾何学である．

「私が興味を持ったのは、情報の幾何学的な構造であった。情報にだって、相互の距離があり位相的なつながり方があるだろう。こうした理論が無用なはずがない。」情報幾何は、確率分布の族を幾何学的空間としてとらえ、その不変な性質を調べるものである。」

定義 3.1.3 (expectation value). 確率空間 (Ω, \mathcal{F}, P) , 終確率空間 $(\mathcal{X}, \mathcal{B}, P^X)$ について、確率変数 $f: \Omega \rightarrow \mathcal{X}$ の事象 $B \in \mathcal{B}$ 上での期待値とは、

$$\langle X \rangle := \int_B f \cdot P.$$

注 3.1.4. 期待値の言葉で、分散や積率を特徴づけることができる： $\alpha_r(X) = \alpha_r(P^X)$. $\varphi_X(u) = E[e^{iuX}]$.

定義 3.1.5 (density). 確率測度の anti-integral を確率密度関数と呼ぶ。すなわち、可測関数 $X: \Omega \rightarrow \mathcal{X}$ であって、 $(\mathcal{X}, \mathcal{B})$ の標準測度 μ について^{†1}、任意の可測集合 $A \in \mathcal{F}$ に対して、 $P(X \in A) = \int_A X d\mu = \int_{X^*A} dP$ を満たすものである。

$$\begin{array}{ccc} & [0, 1] & \\ P \uparrow & \nwarrow X_* P & \\ (\Omega, \mathcal{F}) & \xrightarrow{X} & (\mathcal{X}, \mathcal{B}) \end{array}$$

要諦 3.1.6. 確率密度関数も確率変数であるが、積分による使用が想定されている特殊な道具である。Radon-Nikodym の定理により、可測空間 $(\mathcal{X}, \mathcal{B})$ の標準測度 μ を用いて $f := \frac{dX_*P}{d\mu}$ と構成できる。

3.2 分布の特性値

3.2.1 分布関数と確率関数

離散分布では、確率空間上の確率測度は確率関数に自然に退化する。これを確率質量関数という。つまり、確率測度と確率密度関数を同一視する。実用上の基本は $\mathcal{X} \subset \mathbb{R}$ が想定されるが、まずは一般的に与えられた確率測度 $\mu: \mathcal{F} \rightarrow \mathbb{R}$ から標準的に定められる、代表的な確率分布をみる。特に母関数の方法は Taylor 展開や解析接続での圧倒的成功例があり、特性関数の方法とは、確率密度関数の Fourier 変換に他ならない。何かしらの特徴量と考えられていて、ありがたがられる意味論を持つ。

定義 3.2.1 (discrete probability function, probability mass function).

- (1) 可算集合 \mathcal{X} 上の確率分布 $P^X: \mathcal{X} \rightarrow \mathbb{R}$ を離散分布という。
- (2) 離散分布 $\nu: \mathcal{X} \rightarrow \mathbb{R}$ に対して、 $p_x = p(x) := \nu(\{x\})$ と置くと、これは正定値性と正規化条件を満たす。これを確率（質量）関数と呼ぶ。^{†2}
 - (a) $\forall_{x \in \mathcal{X}} p_x \geq 0$.
 - (b) $\sum_{x \in \mathcal{X}} p_x = 1$.

補題 3.2.2. \mathcal{X} を可算集合として、次の標準的な対応が存在する。

$$P(\mathcal{X}) \simeq_{\text{Set}} \left\{ p \in \text{Map}(\mathcal{X}, [0, 1]) \mid \forall_{x \in \mathcal{X}} p_x \geq 0, \sum_{x \in \mathcal{X}} p_x = 1 \right\}$$

[証明] . 一般の連続分布の特別な場合 . ■

^{†1} Borel class $(\mathbb{R}^n, \mathcal{B})$ ならば、Lebesgue 測度である。

^{†2} 日本語ではよく確率関数と略されるが、英語だと mass は落とさない方がいいと wikipedia に注釈がある。

3.2.2 期待値と積率

標本平均ではなく確率分布の平均であることを強調するときは母平均 (population mean) という。なお標本の特性値は経験分布関数などを用いて定める。期待値などの位置母数 (location parameter) と分散などの尺度母数 (scale parameter) がある。

定義 3.2.3 (mean (value), variance, absolute moment, moment, central moment, standard deviation, skewness, kurtosis).

$X \subset \mathbb{R}$ とし、その上の確率分布を $\mu = (p_x)_{x \in \mathcal{X}} : \mathcal{X} \rightarrow \mathbb{R}$ とする。ここから新たな確率分布 = 測度を構成する。

- (1) 和 $\sum_{x \in \mathcal{X}} |x| p_x < \infty$ が絶対収束するとき、この和 $\mu := \sum_{x \in \mathcal{X}} x p_x$ を、確率分布 $\mu = (p_x)_{x \in \mathcal{X}}$ の平均という。
- (2) 和 $\sum_{x \in \mathcal{X}} |x|^2 p_x < \infty$ が絶対収束するとき、この和 $\sigma^2 := \sum_{x \in \mathcal{X}} (x - \mu)^2 p_x$ を、確率分布 $\mu = (p_x)_{x \in \mathcal{X}}$ の分散という。
- (3) 一般の $r \in \mathbb{N}$ について、 $\beta_r := \sum_{x \in \mathcal{X}} |x|^r p_x \in [0, \infty]$ とおく。これを r 次の絶対積率という。
- (4) 絶対積率が $\beta_r < \infty$ を満たすとき、確率変数 x^r の平均 $\alpha_r = \mu'_r := \sum_{x \in \mathcal{X}} x^r p_x$ を r 次の積率という。1 次の積率が平均 $\alpha_1 = \mu$ である。
- (5) 絶対積率が $\beta_r < \infty$ を満たすとき、 $\mu_r := \sum_{x \in \mathcal{X}} (x - \mu)^r p_x$ を r 次の中心積率という。2 次の中心積率が分散である： $\mu_2 = \sigma^2$ 。その平方根を標準偏差という。
- (6) 確率分布 $\frac{\mu_3}{\mu_2^{3/2}}$ を歪度という。^{†3}
- (7) 確率分布 $\frac{\mu_4}{\mu_2^2}$ を尖度という。^{†4}

3.3 特性関数と母関数

$\mathbb{Z}_+ := \mathbb{N}$ 上の離散分布を扱う際には、確率母関数もよく用いられる。 $\varphi(u) = g(e^{iu})$ なので、特性関数の微分と確率母関数の微分は本質的に等価で、便利な方を使えば良い。

3.3.1 特性関数と確率母関数

定義 3.3.1 (characteristic function, probability generating function).

- (1) 実数上の確率変数 $\varphi(u) := \sum_{x \in \mathcal{X}} e^{iux} p_x : \mathbb{R} \rightarrow \mathbb{C}$ を、確率関数 $\mu = (p_x)_{x \in \mathcal{X}}$ の特性関数または Fourier 変換という。
 $|e^{iux}| = 1$ より、これは必ず収束する。^{†5}
- (2) 特に $\mathcal{X} = \mathbb{N} \subset \mathbb{R}$ のとき、実数上の確率変数 $g(z) := \sum_{x \in \mathbb{N}} p_x z^x$ を確率母関数という。特性関数は、これに $z = e^{iu}$ と変数変換を合成した場合である： $\varphi(u) = g(e^{iu})$ 。^{†6}

補題 3.3.2 (fractional moment : 分散公式・積率は特性関数の微分・平均と分散は確率母関数の微分)。

- (1) (分散公式) $\sigma^2 = \alpha_2 - \mu^2 = \sum_{x \in \mathcal{X}} x^2 p_x - \left(\sum_{x \in \mathcal{X}} x p_x \right)^2$ 。
- (2) $\beta_r < \infty$ のとき、 $\alpha_r = i^{-r} \varphi^{(r)}(0)$ 。

^{†3} 奇関数である 3 乗を使って歪みを測るアイデアである。

^{†4} 裾の長さ。

^{†5} 理論解析の極みのような存在である。well-defined であり、一般性を持ち、また性質が理想的である。また、確率変数が確率密度関数を持つ場合、特性関数と密度関数は互いにもう一方のフーリエ変換になっているという意味で双対である。

^{†6} これは特性関数の自然数への特長である。e なら持ち出さなくても良い場合、 $\varphi(u) = g(e^{iu})$ という関係がある。

- (3) g が $z = 1$ で項別微分可能であるとき, 階乗モーメントが $g^{(r)}(1) = \sum_{x=r}^{\infty} x(x-1)\cdots(x-r+1)p_x$ である.
- (4) 特に, 平均と分散について $\alpha_1 = g'(1), \mu_2 = g''(1) + g'(1) - (g'(1))^2$ が成り立つ.

[証明].

(1)

$$\begin{aligned}\sigma^2 &= \sum_{x \in \mathcal{G}} (x - \mu)^2 p_x \\ &= \sum_{x \in \mathcal{G}} x^2 p_x - 2\mu \underbrace{\sum_{x \in \mathcal{G}} x p_x}_{=\mu} + \mu^2 \underbrace{\sum_{x \in \mathcal{G}} p_x}_{=1} \\ &= \alpha_2 - \mu^2.\end{aligned}$$

- (2) $\beta_r < \infty$ ならば, $\varphi^{(r)}(u) = \sum_{x \in \mathcal{G}} (ix)^r e^{iux} p_x$ は収束し (すなわち項別微分可能で), 関数 $\varphi^{(r)}: \mathbb{R} \rightarrow \mathbb{R}$ を定める. $u = 0$ として, $\varphi^{(r)}(0) = i^r \sum_{x \in \mathcal{G}} x^r p_x = i^r \alpha_r$.

- (3) 一般に, 項別微分可能ならば $g^{(r)}(z) = \sum_{x=0}^{\infty} x(x-1)\cdots(x-r+1)p_x z^{x-r}$ であるが, いま $\mathcal{G} = \mathbb{Z}$ としているから, $z = 1$ のとき, $g^{(r)}(1) = \sum_{x=r}^{\infty} x(x-1)\cdots(x-r+1)p_x$ である. また, 特に $g'(1) = \mu, g''(1) = \sum_{x \in \mathcal{G}} x(x-1)p_x$ であるから, (1) より, $\sigma^2 = \underbrace{g''(1) + g'(1)}_{=\sum_{x \in \mathcal{G}} (x(x-1)+x)p_x} - (g'(1))^2$

■

3.3.2 分位点関数

定義 3.3.3 (quantile function).

- (1) 累積分布関数 F が連続かつ狭義単調増加であるとき, 逆関数 F^{-1} を分位点関数と定める.
- (2) 一般の場合, $F_L^{-1}(u) := \inf \{x \in \mathbb{R} \mid F(x) \geq u\}, F_R^{-1}(u) := \sup \{x \in \mathbb{R} \mid P(X \geq x) \geq 1 - u\}$ と定める.

補題 3.3.4.

- (1) F_L^{-1} は左連続である.
- (2) F_R^{-1} は右連続である.

定義 3.3.5. 分布関数 $F(x)$ において, $x \rightarrow \pm\infty$ としたときの収束の速さを裾の重さという. 期待値や分散は積分で定義されるために, 裾の重さに影響を受けやすい.

3.3.3 母関数の概念の射程

母関数 (generating function) とは「数列の定める関数」である.^a 「数列を各項ごとに調べるよりも一度に扱った方が物事が見えてくることが多いので, 母関数を導入して, その性質を調べることは数学の常套手段です。」保型関数もその大成功例である. そこで, 確率分野でも Fourier 展開を考えると, その係数が積率である. 離散数学, 物理学, 統計学へ.

^a 母関数の「母」は, 関数を母親と見て数列の各項を子供と見立てて, そのように呼んでいます. (ちなみに, 英語では generating function という味わいのない呼び方をします.)

注 3.3.6 (generating function). 一般に母関数とは, Knuth の第一章に乗っているくらいに組合せ数学でよく使われる手法で, 今回の離散変数のような列 $(f_n)_{n \in \mathbb{N}}$ という対象に対して定義される, その rig R 上の形式的冪級数の集合 $R[[z]]$ の元 $f(z) = \sum_{n=0}^{\infty} f_n z^n$

のことをいう．これにより，数列が関数として生まれ変わったこととなり，こちらを解析することができる．また，元の数列も微分の言葉で再現できる． $R = \mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ などが主に考えられ，後2者については Taylor 展開の理論が， \mathbb{C} では解析接続の理論が模範としてあり，そこでは母関数とは解析関数のことをいう．最初にこの方法を始めたのは一般線型回帰問題を解くために^{†7} A. de Moivre が創始し，James Stirling, Euler, Laplace が応用した．

例 3.3.7 (母関数の応用). 関手 $\text{Seq} \rightarrow \text{Fun}$ に他ならない． $G(z)$ が列 (a_n) の母関数で， $H(z)$ が列 (b_n) の母関数とする．

- (1) 加算： $\alpha G(z) + \beta H(z)$ ．
- (2) シフト： $z^m G(z)$ ．
- (3) 乗算： GH ．

(1),(2) の組み合わせが，漸化式を解くという行為である．

定義 3.3.8 (moment). 一般に，関数 $f: \mathbb{R} \rightarrow \mathbb{R}$ の， $x = c$ を中心とす n 次の積率を

$$\mu_n^{(c)} = \int_{-\infty}^{\infty} (x - c)^n f(x) dx$$

と定める． f を密度関数とする測度の重心は $\mu = \frac{\mu_1^{(0)}}{\mu_0^{(0)}}$ と表せる．ほとんどの場合，中心 c は重心 = 平均に取る．

3.4 一次元離散分布の例

3.4.1 デルタ分布

Shannon のエントロピーとは自己情報量 $I(p) := -\log_2 p$ ($p \in \mathcal{C}$) が定める積分作用素 $H: \text{Meas}(\mathcal{C}, \mathbb{R}) \rightarrow [0, \log \alpha]$ であるが，これが最小になるときの確率分布である．

定義 3.4.1 (degenerated / delta distribution). ある一点のみで 1 をとる確率質量関数が定める確率分布を退化分布と呼ぶ．デルタ関数 δ_a ($a \in X$) がこの測度 $\epsilon_a(A) := 1_{\{B \in \mathcal{A} \mid a \in B\}}$ ($A \in \mathcal{A}$) を定める．これは自然な埋め込み $X \hookrightarrow \mathcal{M}(X)$ とも考えられる．^{†8}

3.4.2 経験分布

経験分布関数と順序統計量は一対一対応する．

定義 3.4.2 (order statistic, empirical distribution function, resampling from sample, bootstrap method). $X_1, \dots, X_n \sim F$, i.i.d. とする．

- (1) これらの確率変数の値を小さい順に並べ替えて得る列 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ を順序統計量という．
- (2) 順序統計量の第 i 成分 $X_{(i)}$ を第 i 順序統計量といい，第一順序統計量を最小値，第 n 順序統計量を最大値という．
- (3) 特定の値 $x \in \Omega$ に対して， x 以下となる観測値の割合を返す関数 $F_n(x) := \frac{1}{n} |\{i \in [n] \mid X_i \leq x\}|$ を経験分布関数という．^{†9}
- (4) すでに得られた標本から再び標本抽出を行うことを標本からのリサンプリングとよぶが，これは経験分布 F_n に従う確率変数を観測することにあたる．
- (5) すでに得られた標本 $(x_i)_{i \in [n]}$ の経験分布 F_n からリサンプリングを繰り返し， F_n を代用して仮想的な標本の取り直しを行う方法をブートストラップ法という．

注 3.4.3. 大数の強法則によって， $P \left[\lim_{n \rightarrow \infty} |F_n(x) - F(x)| = 0 \right] = 1$ が従うが，より強い「一様大数の法則」ともいふべき結果がある．

^{†7} 数列とその母関数の対応は線型同型の代表例に他ならない．

^{†8} 例えば弱位相について，この埋め込みの像の凸包は稠密である．

^{†9} 点 $x_i \in \Omega$ に確率 $1/n$ を持つような離散確率分布の累積分布関数となっているので，記法も寄せている．

定理 3.4.4 (Glivenko-Cantelli (1933)).

$$P \left[\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0 \right] = 1.$$

3.4.3 Rademacher 分布

機械学習研究者の十八番．Rademacher 過程は，幅 1 のランダムウォークと考えられる．

定義 3.4.5 (Rademacher distribution).

(1) \mathbb{Z} 上の確率質量関数

$$f(k) = \frac{1}{2}(\delta(k-1) + \delta(k+1))$$

によって定まる離散分布を，**Rademacher 分布**という．

(2) Rademacher 確率分布に従う独立同分布な確率変数列 $\epsilon_1, \dots, \epsilon_n$ に対して， $X := \sum_{i=1}^n \epsilon_i$ と定めた $X : \mathbb{R}^n \rightarrow \mathbb{R}$ を

Rademacher 過程という．

注 3.4.6. X が Rademacher 分布に従う確率変数ならば， $\frac{X+1}{2} \sim B(1/2)$ ．

命題 3.4.7 (Van Zuijlen's bound).

$$P \left(\left| \frac{\sum_{i=1}^n X_i}{\sqrt{n}} \right| \leq 1 \right) \geq \frac{1}{2}.$$

3.4.4 離散一様分布

整数 $1, 2, \dots, N$ から k 個の標本が非復元抽出され、離散一様分布と同様に、標本の抽出のされ方に整数による差はないとする。ここで未知の最大値 N を推定する問題が生じる。このような問題を一般に German tank problem (ドイツ戦車問題) と呼び、第二次世界大戦中のドイツでの戦車生産数の最大値を推定するという問題に由来する。

定義 3.4.8 (discrete uniform distribution). $0 < |\mathcal{X}| < \infty$ を満たす有限集合上の，定値関数 $p = \frac{1}{|\mathcal{X}|}$ を確率関数として定まる測度 $U : \text{FinSet} \rightarrow \mathcal{M}(\mathcal{X})$ を離散一様分布という．

要諦 3.4.9. Shannonn のエントロピーとは自己情報量 $I(p) := -\log_2 p$ ($p \in \mathcal{X}$) が定める積分作用素 $H : \text{Meas}(\mathcal{X}, \mathbb{R}) \rightarrow [0, \log \alpha]$ であるが，これが最大になるときの確率分布である．

3.4.5 二項分布

600 人の中で 1 年を 365 日として，今日誕生日の人が x 人である確率は， $b \left(x; 600, \frac{1}{365} \right)$ となる．

定義 3.4.10 (binomial distribution). 可測空間 $(n+1, P(n+1))$ 上のパラメータ x, p の二項分布 $B(n, p) : \mathbb{N} \times [0, 1] \rightarrow \mathcal{M}(n+1)$ は，成功確率が p で一定な試行 (Bernoulli 試行という) を独立に n 回続けるという確率空間 2^n における確率分布であり，確率変数を成功回数 $x \in n+1$ とすると，確率質量関数 $b : n+1 \rightarrow [0, 1]$ は $b(x; n, p) = \binom{n}{x} p^x q^{n-x}$ と表される． $\text{Bernoulli}(1, p) := B(1, p)$ を Bernoulli 分布という．

命題 3.4.11 (二項分布の平均と分散).

- (1) 平均について， $\alpha_1(b(n, p)) = np$ ．
- (2) 分散について， $\mu_2(b(n, p)) = npq$ ．

[証明].

(1)

$$\begin{aligned}
 \mu &= \sum_{x=0}^n x \binom{n}{x} p^x q^{n-x} \\
 &= \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
 &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \\
 &= np \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!} p^y q^{n-1-y} \quad y := x-1 \\
 &= np \sum_{y=0}^{n-1} b(n-1, p) = np.
 \end{aligned}$$

(2) 確率関数 $(b(x; n, p))_{x \in n+1}$ が定める母関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ は

$$\begin{aligned}
 g(z) &= \sum_{x=0}^n b(x; n, p) z^x \\
 &= \sum_{x=0}^n \binom{n}{x} p^x q^{n-x} z^x = (pz + q)^n
 \end{aligned}$$

と表せる. $g'(z) = np(pz + q)^{n-1}$, $g''(z) = n(n-1)p^2(pz + q)^{n-2}$ より, 分散公式 3.3.2 から,

$$\sigma^2 = g''(1) + g'(1) - g'(1)^2 = n(n-1)p^2 + np - n^2p^2 = npq.$$

■

要諦 3.4.12. 確率母関数が, 二項展開の式に一致することから, ここに予想だにできなかった抜け道がある.

3.4.6 Poisson 分布

極限分布のお手本

所与の時間間隔で, 確率が線型に減衰する現象の観測回数を確率変数としたモデルである. 放射線物質から一定期間に放射される粒子の数の経時変化は Poisson 過程であり, 一定期間に起こる事故の数など. 逆に発生間隔は指数分布となる.

3.4.6.1 極限分布としての Poisson 分布

定義 3.4.13 (Poisson distribution (1838)). 可測空間 $(\mathbb{N}, P(\mathbb{N}))$ 上の **Poisson** 分布 $\text{Pois}(\lambda): (0, \infty) \rightarrow \mathcal{M}(\mathbb{N})$ とは, 確率質量関数 $p(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}: \mathbb{N} \rightarrow [0, 1]$ が定める確率分布である. 母数 λ は, 単位時間当たりの事象の平均発生回数などの割合と見なされる場合は, 到着率と呼ばれる. 平均も分散も λ に一致する.

命題 3.4.14 (Poisson's limit theorem). 減衰する確率 $p(n) := \frac{\lambda}{n}$ についての二項分布 $B(n, p(n))$ は, $\lim_{n \rightarrow \infty} b(x; n, p) = \frac{\lambda^x}{x!} e^{-\lambda} (\forall x \in \mathbb{N})$ を満たす.

[証明].

$$\begin{aligned}
 b\left(x; n, \frac{\lambda}{n}\right) &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
 &= \frac{n(n-1) \cdots (n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
 &= \frac{\lambda^x}{x!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \left(1 + \frac{-\lambda}{n}\right)^{n/(-\lambda) \cdot (-\lambda)} \left(1 - \frac{\lambda}{n}\right)^{-x}
 \end{aligned}$$

$$\xrightarrow{n \rightarrow \infty} \frac{\lambda^x}{x!} e^{-\lambda}$$

この収束は一様収束である。

命題 3.4.15 (Poisson 分布の確率母関数). Poisson 分布 $\text{Pois}(\lambda)$ の定める確率母関数は $g(z) = e^{\lambda(z-1)}$ である。

[証明]. 確率変数列 $(p(x; \lambda))_{x \in \mathbb{N}} = \left(\frac{\lambda^x}{x!} e^{-\lambda} \right)_{x \in \mathbb{N}}$ が定める母関数は,

$$\begin{aligned} g(z) &= \sum_{x=0}^{\infty} \left(\frac{\lambda^x}{x!} e^{-\lambda} \right) z^x \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda z)^x}{x!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}. \end{aligned}$$

歴史 3.4.16. 歴史的に有名な事例としては、ロシア生まれでドイツで活躍した経済学者、統計学者のポルトケヴィッチによる「プロイセン陸軍で馬に蹴られて死亡した兵士数」の例が知られている。ポルトケヴィッチは著書 "Das Gesetz der kleinen Zahlen" (The Law of Small Numbers)^[4] において、プロイセン陸軍の 14 の騎兵連隊の中で、1875 年から 1894 年にかけての 20 年間で馬に蹴られて死亡する兵士の数について調査しており、1 年間当たりに換算した当該事案の発生件数の分布が母数 0.61 のポアソン分布によく従うことを示している。

3.4.6.2 ポアソン過程

定義 3.4.17 (Poisson process). $\lambda > 0$ を定数として、次の 3 条件を満たす確率過程 $(P_k(t))_{k \in \mathbb{N}}$ を **Poisson 過程** という：時間 $(0, t]$ に点事象が起こる回数が k 回である確率を $P_k(t)$ とし、

- (1) 区間 $(t, t + \Delta t]$ に 1 回だけ事象が起こる確率は $\Delta t \searrow 0$ のとき $\lambda \Delta t + o(\Delta t)$ である。
- (2) 区間 $(t, t + \Delta t]$ に 2 回以上事象が起こる確率は $o(\Delta t)$ である。
- (3) 区間 $(t, t + \Delta t]$ に点事象が起こる回数は $(0, t]$ で起こる回数とその起こり方に「独立」である。^{†10}

補題 3.4.18. Poisson 過程は存在する。

命題 3.4.19. Poisson 過程 $(P_k(t))_{k \in \mathbb{N}}$ は \mathbb{N} 上の Poisson 分布 $\text{Pois}(\lambda t)$ である。

[証明].

関係式の導出 区間 $(t, t + \Delta t]$ で点事象が 1 回起こる確率が $\lambda \Delta t + o(\Delta t)$ 、2 回以上起こる確率が $o(\Delta t)$ より、0 回起こる確率が $1 - \lambda \Delta t - o(\Delta t)$ だから、

$$P_k(t + \Delta t) = P_k(t)(1 - \lambda \Delta t - o(\Delta t)) + P_{k-1}(t)(\lambda \Delta t - o(\Delta t)) + o(\Delta t)$$

となる。ただし、 $k \in \mathbb{N}, P_{-1}(t) = 0$ とした。 $\Delta t \searrow 0$ を考えることで、微分方程式

$$P'_k(t) = \lambda(P_{k-1}(t) - P_k(t)), \quad P_{-1}(t) = 0$$

を得る。

分布の導出 まず $k = 0$ とすると、

$$P'_0(t) = -\lambda P_0(t), \quad P_0(0) = 1$$

より、 $P_0(t) = e^{-\lambda t}$ 。次に $k = 1$ とすると、

$$P'_1(t) = \lambda(e^{-\lambda t} - P_1(t)), \quad P_1(0) = 0$$

より、 $P_1(t) = \lambda t e^{-\lambda t}$ 。 $k = 2$ とすると、

$$P'_2(t) = \lambda(\lambda t e^{-\lambda t} - P_2(t)), \quad P_2(0) = 0$$

より、 $P_2(t) = \frac{(\lambda t)^2}{2} e^{-\lambda t}$ 。以下帰納的に、 $P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$ を得る。

^{†10} 確率過程の独立はまだ定義していない。

[別証明]. 過程 $(P_k(t))_{k \in \mathbb{N}}$ の定める確率母関数 $g(z, t) = \sum_{k=0}^{\infty} P_k(t) z^k$ が, Poisson 分布 $\text{Pois}(\lambda t)$ の確率母関数 $e^{\lambda t(z-1)}$ に一致することを示しても良い.

3.4.7 負の二項分布

定義 3.4.20 (negative binomial distribution, geometric distribution).

- (1) 成功の確率が p の試行を独立に繰り返す Bernoulli 試行 2.5.10 に於て, k 回成功するまでに必要な失敗の回数 x が測度空間 $(\mathbb{N}, P(\mathbb{N}))$ 上に定める分布を負の二項分布 $\text{NB}(k, p) : \mathbb{Z}_+ \times [0, 1] \rightarrow \mathcal{M}(\mathbb{N})$ という. その確率関数は $p(x; k, p) = \binom{x+k-1}{x} p^k q^x$ ($x \in \mathbb{Z}_+$) と表せる.
- (2) 特に $k = 1$ の場合を幾何分布 $G(p) := \text{NB}(1, p)$ といい, 確率関数は $p(x; 1, p) = p q^x$ ($x \in \mathbb{Z}_+$) と表せる.
- (3) 負の二項分布 $\text{NB}(k, p)$ は $k \in (0, \infty)$ 上に延長できる.

命題 3.4.21 (負の二項分布の確率母関数・平均・分散).

- (1) 負の二項分布 $\text{NB}(b, p)$ の確率母関数は $g(z) = p^k (1 - qz)^{-k} = \left(\frac{p}{1 - qz} \right)^k$ で表せる.
- (2) 負の二項分布 $\text{NB}(b, p)$ の平均は $\alpha_1 = \frac{kq}{p}$, 分散は $\mu_2 = \frac{kq}{p^2}$ である.

[証明].

- (1) $|zq| < 1$ すなわち $|z| < \frac{1}{q}$ のとき, $(1 - qz)^{-k}$ の $z = 0$ における Taylor 展開を考えると

$$\begin{aligned} (1 - qz)^{-k} &= 1 + k(qz) + \frac{k(k+1)}{2!} (qz)^2 + \cdots = \sum_{x=0}^{\infty} \frac{(-k)(-k-1) \cdots (-k-x+1)}{x!} (-qz)^x \\ &= \sum_{x=0}^{\infty} \frac{(x+k-1)(x+k-2) \cdots (k+1)k}{x!} (qz)^x \\ &= \sum_{x=0}^{\infty} \binom{x+k-1}{x} (qz)^x. \end{aligned}$$

$$\left. \frac{(-k)(-k-1) \cdots (-k-x+1)}{x!} \right|_{x=0} = 1 \text{ としたことに注意. よって, } \text{NB}(k, p) \text{ の確率母関数は } g(z) = p^k (1 - qz)^{-k}.$$

- (2) 確率母関数と平均・分散の関係 3.3.2(3) より,

$$\begin{aligned} g'(z) &= kp^k q (1 - qz)^{-k-1}, & \mu &= g'(1) = \frac{kp^k q}{p^{k+1}} = \frac{kq}{p}, \\ g''(z) &= k(k+1)p^k q^2 (1 - qz)^{-k-2}, & \sigma^2 &= g''(1) + g'(1) - (g'(1))^2 \\ & & &= \frac{k(k+1)q^2}{p^2} + \frac{kq}{p} - \frac{k^2 q^2}{p^2} \\ & & &= \frac{kq(p+q)}{p^2} = \frac{kq}{p^2}. \end{aligned}$$

命題 3.4.22 (Poisson 近似). 平均 $\lambda := \frac{kq}{p} \in (0, \infty)$ を一定にして, 成功数を表す母数を $k \rightarrow \infty$ とすると (このとき失敗率 q は極めて小さくなる), Poisson 分布 $\text{Pois}(\lambda)$ に収束する: $\text{NB}(k, p) \rightarrow \text{Pois}(\lambda)$ ($k \rightarrow \infty$).

[証明]. 確率関数が $\tilde{p}(x) = \frac{\lambda^x}{x!} e^{-\lambda}$ に収束することを見れば良い.

$$p(x; k, p) = \binom{x+k-1}{x} p^k q^x$$

$$\begin{aligned}
&= \frac{(x+k-1) \cdots (k+1)k}{x!} p^k q^x \\
&= \frac{(x+k-1) \cdots (k+1)k}{k^x} x! \left(\frac{kq}{p}\right)^x p^{k+x} \\
&\xrightarrow[\lambda=\text{const.}]{k \rightarrow \infty} 1 \cdot \frac{\lambda^x}{x!} e^{-\lambda}.
\end{aligned}$$

実際,

$$p^{k+x} = \left(\frac{p}{p+q}\right)^{k+x} = \left(1 + \frac{\lambda}{k}\right)^{-(k+x)} \xrightarrow{k \rightarrow \infty} e^{-\lambda}.$$

■

注 3.4.23 (幾何分布の無記憶性). $X \sim G(\theta)$ とするとき,

$$P(X \geq m+n | X \geq m) = P(X \geq n) \quad (m, n \in \mathbb{N})$$

が成り立つ. これは, 時刻 m までに成功していないことはその後の成功までの待ち時間の分布に影響しないことを意味しているとみなせる.

3.4.8 Katz 族

定義 3.4.24 (Katz family). \mathbb{N} 上の離散分布の族であって, $a, b \in \mathbb{R}$ を用いた漸化式

$$p_0 > 0, \quad p_x = \left(a + \frac{b}{x}\right) p_{x-1}, \quad x \in \mathbb{N}$$

を満たす族を **Katz 族** という. この漸化式を満たすことを, $(a, b, 0)$ -クラスの分布であるという.

例 3.4.25.

- (1) $\delta_0 : a + b = 0, p_0 = 1$ のとき.
- (2) $B(n, \theta) : a = -\frac{\theta}{1-\theta}, \quad b = \frac{(n+1)\theta}{1-\theta}, \quad p_0 = (1-\theta)^n.$
- (3) $\text{Pois}(\lambda) : a = 0, b = \lambda, p_0 = e^{-\lambda}.$
- (4) $NB(k, \theta) : a = 1-\theta, b = (k-1)(1-\theta), p_0 = \theta^k.$

命題 3.4.26 (実際の母数はかなり狭い). \mathbb{N} 上の確率分布 $p = (p_x)_{x \in \mathbb{N}}$ が Katz 族であるとする. このとき, p は上の例の 4 つの場合に限られる.

- (1) $a + b = 0$ ならば, $p = \delta_0$.
- (2) $a = 0, b > 0$ ならば, $p = \text{Pois}(b)$.
- (3) $a + b > 0, a \in (0, 1)$ ならば, $p = NB((a+b)/a, 1-a)$.
- (4) $a + b > 0, a \in (-\infty, 0)$ ならば, $\exists n \in \mathbb{N} \ a(n+1) + b = 0, p = B(n, -a/(1-a))$.

特に, p がデルタ測度 δ_0 でなければ, $a + b > 0$ かつ $a < 1$ であり, 確率母関数は

$$g(z) = \begin{cases} \left(\frac{1-az}{1-a}\right)^{-\frac{a+b}{a}}, & a \neq 0, \\ \exp(b(z-1)), & a = 0. \end{cases}$$

と表せる.

3.4.9 超幾何分布

二項分布を2からの復元抽出だと思えば、これを非復元抽出にするとより複雑な関数形が得られる。これは主に個体群生態学で使用される標識再捕獲法などで使われる。

定義 3.4.27 (hypergeometric distribution). 成功 n 個, 失敗 $N - n$ 個が入った多重集合から r 個玉を取り出したときにそのう

ち成功が x 個である確率 $p(x; N, n, r) = \frac{\binom{n}{x} \binom{N-n}{r-x}}{\binom{N}{r}}$ が, $X := \{x \in \mathbb{N} \mid \max\{0, r - N + n\} \leq x \leq \min\{r, n\}\}$ として可測

空間 $(X, P(X))$ 上に定める確率分布 $H(N, n, r)$ を超幾何分布という。

命題 3.4.28.

- (1) 平均は $\alpha_1 = r \frac{n}{N}$.
- (2) 分散は $\mu_2 = r \left(\frac{N-r}{N-1} \right) \left(\frac{n}{N} \right) \left(1 - \frac{n}{N} \right)$.

例 3.4.29 (mark and recapture method における Lincoln-Peterson 推定量). 個体数 N を推定するために, n 匹を捕まえて, 標識をつけて再放流する. 十分な時間経過後に sampling し, 標識がついているものが r 匹だった場合, $\tilde{N} := \frac{nr}{x}$ によって N が推定できる. これは, 確率 $p(x; N, n, r)$ が最大になるような N の選び方で, 最尤法の考え方である。

3.4.10 負の超幾何分布

3.4.11 対数分布

定義 3.4.30. \mathbb{N} 上の分布・対数分布 $\text{Log}(\theta) := (p_x)_{x \in \mathbb{N}} (\theta \in (0, 1))$ とは, 確率関数

$$p_x = \begin{cases} 0 & x = 0, \\ \frac{-1}{\log(1-\theta)} \frac{\theta^x}{x} & x \in \mathbb{N}_{>0}, \end{cases}$$

で定まる分布をいう。

命題 3.4.31. $\text{Log}(\theta)$ の確率母関数は $g(z) = \frac{\log(1-\theta z)}{\log(1-\theta)}$ である。

3.4.12 Ord 族

定義 3.4.32 (Ord family). \mathbb{Z} 上の分布であって, 確率関数が差分方程式

$$p_x - p_{x-1} = \frac{(a-x)p_x}{(a+b_0) + (b_1-1)x + b_2x(x-1)}$$

を満たすものの全体を **Ord 族**という。

例 3.4.33. Katz 族と, 超幾何分布, 負の超幾何分布は Ord 族である。

3.5 多次元離散分布の例

3.5.1 確率変数の積

定義 3.5.1 (marginal distribution, joint / simultaneous distribution). $I := \{1, \dots, d\}$ について, 確率変数 $X_j : \Omega \rightarrow \mathcal{X}_j (j \in I)$

の積 $X := (X_j)_{j \in I} : \Omega \rightarrow \mathcal{X} := \prod_{j=1}^d \mathcal{X}_j$ は d 次元確率変数である。

- (1) $J \subsetneq I$ について, 分布 $(p_J(x_j))_{x_j \in \mathcal{X}_J}$ を X_I の周辺分布と呼ぶ .
 (2) これに対して X の分布自体のことを, 区別して結合分布または同時分布という .

定義 3.5.2 (covariance, correlation coefficient). 2 つの実確率変数 $X: \Omega \rightarrow \mathcal{X}, Y: \Omega \rightarrow \mathcal{Y}$ について,

- (1) $\text{Cov}[X, Y] := \mathcal{E}[(X - \mathcal{E}[X])(Y - \mathcal{E}[Y])]$ を共分散と呼ぶ .
 (2) $\rho(X, Y) := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}}$ を相関係数と呼ぶ .

補題 3.5.3.

- (1) $E[X^2] < \infty, E[Y^2] < \infty$ のとき, または, $E[|X|] < \infty, E[|Y|] < \infty$ かつ X, Y が独立のとき, 共分散は存在する .
 (2) $E[X^2] < \infty, E[Y^2] < \infty, \text{Var}(X), \text{Var}(Y) > 0$ のとき, 相関係数は存在し, $\text{Im } \rho \subset [-1, 1]$ を満たす .

3.5.2 多項分布

二項分布は多項分布の周辺分布であった .

定義 3.5.4 (multinomial distribution). $I = \{E_1, \dots, E_k\}$ とし,

$$T_I := \left\{ x_I = (x_j)_{j \in I} \in \mathbb{N}^k \mid \sum_{j=1}^k x_j = n \right\}$$

とする . 1 回の試行で E_1, \dots, E_k のいずれかが起こるとし, それぞれの生起確率を $p_1, \dots, p_k, \sum_{j=1}^k p_j = 1$ とする . このとき, E_1, \dots, E_k が起きた回数が x_1, \dots, x_k 回である確率は

$$f_I(x_I; p) = \frac{n!}{x_1! \dots x_k!} \prod_{j=1}^k p_j^{x_j} \quad (x_I = (x_j)_{j \in I} \in T_I)$$

である . これを確率関数とする T_I 上の離散分布を, パラメータ $(n, p) \in \mathbb{N} \times [0, 1]^k$ の多項分布 $\text{Mult}(n, p)$ という .

命題 3.5.5. $g_I(z_I; p) = (p_1 z_1 + \dots + p_k z_k)^n$ が確率母関数である .

[証明]. 多項定理より . ■

$$\text{命題 3.5.6. } \text{Cov}[X_i, X_j] = \begin{cases} np_i(1 - p_i) & i = j \\ -np_i p_j & i \neq j \end{cases}$$

3.5.3 2 変量 Poisson 分布

定義 3.5.7 (bivariate Poisson distribution). 確率変数 U_1, U_2, U_3 が独立で $U_i \sim \text{Pois}(\lambda_i)$ を満たすとする . このとき, $X_1 = U_1 + U_3, X_2 = U_2 + U_3$ で定まる確率変数 $(X_1, X_2): \Omega \rightarrow \mathbb{N}^2$ が定める分布を 2 変量 **Poisson** 分布といい, $\text{BPois}(\lambda_1, \lambda_2, \lambda_3)$ で表す . 周辺分布は $X_1 \sim \text{Pois}(\lambda_1 + \lambda_3), X_2 \sim \text{Pois}(\lambda_2 + \lambda_3)$ である .

命題 3.5.8 (分散や混合積率は確率母関数の項別微分で求める).

- (1) 確率母関数は $g(z_1, z_2) = \exp(\lambda_1(z_1 - 1) + \lambda_2(z_2 - 1) + \lambda_3(z_1 z_2 - 1))$.
 (2) $\text{Cov}[X_1, X_2] = \lambda_3$.

3.5.4 負の多項分布

負の二項分布 $NB(k, p)$ の確率母関数は, $\hat{q} := 1/p, \hat{p} := q/p = (1-p)/p$ とすると,

$$g(z) = p^k(1 - qz)^{-k} = (\hat{q} - \hat{p}z)^{-k}$$

と表わせ, この形で一般化を考える.

定義 3.5.9. $k > 0, P_i > 0$ をパラメータとする負の多項分布 $NM(k, (P_1, \dots, P_d))$ とは, $Q := 1 + \sum_{i=1}^d P_i$ とするとき, 確率母関数

$$g(z_1, \dots, z_d) = \left(Q - \sum_{i=1}^d P_i z_i \right)^{-k}$$

が定める \mathbb{N}^d 上の分布を言う.

命題 3.5.10 (確率母関数の微分からわかること).

$$(1) \text{Cov}[X_i, X_j] = \begin{cases} kP_i(1 + P_i) & i = j \\ kP_iP_j & i \neq j \end{cases}$$

第 4 章

絶対連続確率分布

離散分布というクラスは、極限構成について閉じていないという意味で、理論的な実用性がない。極限について議論するには、絶対連続分布と実数というところに行き着く。このクラスは「絶対連続」なる名前だが、確率が積分によって表される確率分布という意味である。分布の特性値は、確率分布が定める積分に、適切な積分核を挟んで得られるものであることが明確になる。

4.1 期待値

一般の集合上の関数の期待値作用素の定義には、Lebesgue 積分を用いる。経験分布論では、さらに拡張された線型作用素を用いることも考える。

4.1.1 積分の定義

命題 4.1.1. X を確率変数とする。 $0 < p < q$ について、 $|X|^q$ が可積分ならば、 $|X|^p$ も可積分である。

[証明]。 $\forall x \in \mathbb{R} \quad |x|^p \leq 1 + |x|^q$ である。これと、Lebesgue の優収束定理より。 ■

命題 4.1.2 (Markov の不等式). 非負可測関数 X と任意の $p > 0, q \geq 0, \epsilon > 0$ について、

$$\int X^q 1_{\{X \geq \epsilon\}} d\mu \leq \epsilon^{-p} \int 1_{\{X \geq \epsilon\}} X^{p+q} d\mu \leq \epsilon^{-p} \int X^{p+q} d\mu.$$

[証明]。 $\epsilon^p 1_{\{X \geq \epsilon\}} X^q \leq 1_{\{X \geq \epsilon\}} X^{p+q} \leq X^{p+q}$ と、積分の単調性より。 ■

4.1.2 期待値の定義

定義 4.1.3 (expectation / expected value / mean (value)). 確率変数 $X : \Omega \rightarrow \mathbb{R}$ が確率測度 P について可積分のとき、 $E[X] := \int_{\Omega} X(\omega) P(d\omega)$ を X の期待値または平均 (値) という。 $E_P[X]$ とも表す。

定義 4.1.4 (r -th moment, r -th central moment, r -th absolute moment). 確率変数 $X : \Omega \rightarrow \mathbb{R}$ について、

- (1) $\alpha_r := \mu'_r = E[X^r]$ を X の r 次の積率と呼ぶ。
- (2) $\mu_r := E[(X - E[X])^r]$ を X の r 次の中心積率と呼ぶ。
- (3) $\beta_r := E[|X|^r]$ を X の r 次の絶対積率と呼ぶ。
- (4) μ_2 を X の分散とよび、 $\text{Var}[X]$ で表す。
- (5) $\sqrt{\text{Var}[X]}$ を標準偏差と呼ぶ。

命題 4.1.5 (分散の性質). 2 乗可積分な実確率変数 $X : \Omega \rightarrow \mathbb{R}$ について、

- (1) (分散公式) $\text{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$ 。
- (2) (2 次斉次性) $\forall a, b \in \mathbb{R} \quad \text{Var}[aX + b] = a^2 \text{Var}[X]$ 。

4.1.3 期待値不等式

記法 4.1.6. 積分論の記号を，期待値に対しても流用する．

- (1) $p \in (0, \infty)$ について， $\|X\|_p = (E[|X|^p])^{1/p}$ と定める．
- (2) $\|X\|_\infty := \text{ess. sup}|X|$ とする．

定理 4.1.7 (測度論における結果). X, Y を確率変数とする．

- (1) (Hölder) $\forall p, q \in (1, \infty) \quad \frac{1}{p} + \frac{1}{q} = 1 \Rightarrow \|XY\|_1 \leq \|X\|_p \|Y\|_q$. 特に， $p = q = 2$ のとき Schwarz の不等式．
- (2) (Minkowski) $\forall p \in [1, \infty) \quad \|X + Y\|_p \leq \|X\|_p + \|Y\|_p$.
- (3) $\forall p, q \in (0, \infty) \quad p < q \Rightarrow \|X\|_p \leq \|X\|_q$.
- (4) (Markov) 任意の非減少可測関数 $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ について， $\varphi(A) > 0 \Rightarrow E[|X|1_{\{|Y| \geq A\}}] \leq \frac{E[|X|\varphi(|Y|)1_{\{|Y| \geq A\}}]}{\varphi(A)} \leq \frac{E[|X|\varphi(|Y|)]}{\varphi(A)}$.
特に， $P[|X - E[X]| \geq A] \leq \frac{\text{Var}[X]}{A^2}$ (Chebyshev) .
- (5) (Jensen) 任意の开区間上の凸関数 $\psi : I \rightarrow \mathbb{R}$ について， $X, \psi(X)$ が可積分かつ $P[X \in I] = 1 \Rightarrow \psi(E[X]) \leq E[\psi(X)]$.

4.2 分布の特性値

測度論による確率の特微量の理解

$(\mathbb{R}, \mathcal{B}_1)$ 上の分布について改めて特性値を定義すると，これは「確率変数の特性値」の一般化となっている．確率変数の期待値は，それが誘導する分布の平均のことである．

積分は可測関数と測度についての2変数関数とするならば，これは後者を引数とするとより一般的になるという不思議な状況を物語ってはいないか？これが期待値作用素の限界ということか？

4.2.1 確率密度関数

定義 4.2.1 (probability density function). \mathbb{R} 上の分布 ν が Lebesgue 測度 dx に関して絶対連続であるとき^{†1}，その Radon-Nikodym 微分 f を確率密度関数とよび， ν を (絶対) 連続分布という．

注 4.2.2. 確率密度関数は Lebesgue 零集合の差を除いて一意に定まる．

補題 4.2.3. 任意の可測関数 $g : \mathbb{R} \rightarrow \mathbb{R}$ と絶対連続分布 ν について，

$$E[g(X)] = \int_{\mathbb{R}} g(x) \nu(dx) = \int_{\mathbb{R}} g(x) f(x) dx.$$

すなわち，左辺または右辺が存在すればもう一方も存在し，値が一致する．

4.2.2 平均値と積率

定義 4.2.4 (expectation). $(\mathbb{R}, \mathcal{B}_1)$ 上の確率測度 ν について， $\mu := \int_{\mathbb{R}} x \nu(dx)$ を， ν の期待値または平均 (値) という．

定義 4.2.5. $(\mathbb{R}, \mathcal{B}_1)$ 上の確率測度 ν について，

- (1) $\alpha_r := \mu'_r = \int_{\mathbb{R}} x^r \nu(dx)$ を ν の r 次の積率と呼ぶ．
- (2) $\mu_r := \int_{\mathbb{R}} (x - \mu)^r \nu(dx)$ を ν の r 次の中心積率と呼ぶ．

^{†1} Lebesgue 零集合上の確率が零であることが定義．

- (3) $\beta_r := \int_{\mathbb{R}} |x|^r \nu(dx)$ を ν の r 次の絶対積率と呼ぶ .
 (4) μ_2 を ν の分散と呼ぶ .
 (5) $\sqrt{\mu_2}$ を標準偏差と呼ぶ .

命題 4.2.6 (変数変換公式:well-definedness). $(\mathcal{G}, \mathcal{A})$ を可測空間とし, X を \mathcal{G} -値確率変数とする . このとき ,

$$\forall g \in \text{Meas}(\mathcal{G}, \mathbb{R}) \quad \int_{\Omega} g(X(\omega)) P(d\omega) = \int_{\mathcal{G}} g(x) P^X(dx).$$

すなわち, 左辺または右辺のいずれかの積分が存在すればもう一方も存在し, 値が等しくなる .

系 4.2.7 (積率の well-definedness). $g(x) = x^r$ とすれば ,

$$E[X^r] = \int_{\Omega} X(\omega)^r P(d\omega) = \int_{\mathbb{R}} x^r P^X(dx).$$

すなわち, $\alpha_r(X) = \alpha_r(P^X)$.

定義 4.2.8 (skewness, kurtosis). 位置母数でも尺度母数でもないものとして, 分布の形状を表すと考えられる次の母数がある .

- (1) $\mathfrak{s} = \gamma_1 := \frac{\mu_3}{\mu_2^{3/2}}$ を歪度と呼ぶ .
 (2) $\mathfrak{k} = \gamma_2 := \frac{\mu_4}{\mu_2^2}$ を尖度と呼ぶ .

4.2.3 共分散と相関

定義 4.2.9 (covariance, correlation coefficient).

- (1) X, Y, XY が P -可積分のとき ,

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$$

を X, Y の共分散という .

- (2) $X, Y \in L^2$ で $\text{Var}[X]\text{Var}[Y] \neq 0$ のとき, $\rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}}$ を相関係数という .

- (3) $X = (X_i)_{i \in [r]}, Y = (Y_j)_{j \in [c]}$ が r, c 次元の 2 乗可積分確率変数とするととき, 共分散行列は

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])^{\top}] = (E[(X_i - E[X_i])(Y_j - E[Y_j])])_{(i,j) \in [r] \times [c]} = (\text{Cov}[X_i, Y_j])_{ij}$$

で定まる $r \times c$ 行列をいう . $\text{Cov}[X, Y] = \text{Cov}[Y, X]^{\top}$ が成り立つ .

- (4) $\text{Var}[X] := \text{Cov}[X, X]$ で定まる $r \times r$ 行列を, 分散共分散行列という .
 (5) $\text{Corr}[X] = (\rho(X_i, X_j))_{i,j \in [n]}$ で定まる $r \times r$ 行列を相関行列という . これは, 対角要素が 1 に基準化された無次元量だと考えられる .

要諦 4.2.10. $X, Y \in L^2$ は十分条件である .

命題 4.2.11. $X, Y, Z \in L^2$ について ,

- (1) $\text{Cov}[X, Y] = \text{Cov}[Y, X]$.
 (2) $\text{Cov}[aX + bY, Z] = a\text{Cov}[X, Z] + b\text{Cov}[Y, Z]$.
 (3) $\text{Cov}[X, 1] = 0$. 特に, $\text{Cov}[aX + b, Y] = a\text{Cov}[X, Y]$.
 (4) (共分散公式) $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$.
 (5) $\text{Cov}[X, X] = \text{Var}[X] \geq 0$. 等号成立条件は $X = E[X]$ a.s.
 (6) (Schwarz の不等式) $|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}$.

命題 4.2.12 (Pearson の不等式). $X \in L^1$ について, $\text{Var}[X] > 0$ とする . このとき, $\mathfrak{s}^2(X) + 1 \leq \mathfrak{k}(X)$. 特に $\mathfrak{s}(X) = 0$ ならば, $\mathfrak{k}(X) \geq 1$.

命題 4.2.13 (共分散行列の双線型性).

(1) 可積分確率変数 $X_1, \dots, X_m, Y_1, \dots, Y_n$ の積 $X_i Y_j$ も可積分とする。このとき,

$$\forall_{a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{R}} \quad \text{Cov} \left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j \right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j)$$

(2) 各 X_i ($i \in [m]$) を r_i 次元確率変数, 各 Y_j ($j \in [n]$) を c_j 次元確率変数とする。すべての $X_i, Y_j, X_i Y_j \in L^1$ のとき, 任意の $r \times r_i$ 定数行列 A_i と $c \times c_j$ 定数行列 B_j について

$$\text{Cov} \left(\sum_{i=1}^m A_i X_i, \sum_{j=1}^n B_j Y_j \right) = \sum_{i=1}^m \sum_{j=1}^n A_i \text{Cov}(X_i, Y_j) B_j^\perp$$

注 4.2.14. 行列 M に関して, $|M| = (\text{Tr}(MM^\perp))^{1/2}$ とすると, $|X_i \otimes Y_j| = |X_i| |Y_j|$ だから, $|X_i| |Y_j|$ が可積分であることと, $|X_i \otimes Y_j|$ が可積分であることと, X_i, Y_j の要素のペアの積がすべて可積分であることは同値になる。

命題 4.2.15 (和の分散). $X_1, \dots, X_n \in L^2$ について,

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{(i,j): 1 \leq i < j \leq n} \text{Cov}(X_i, Y_j).$$

命題 4.2.16 (相関係数の値域). $\mu_X = E[X], \mu_Y = E[Y], \sigma_X = \sqrt{\text{Var}[X]}, \sigma_Y = \sqrt{\text{Var}[Y]}$ とする。

$$|\rho(X, Y)| \leq 1$$

等号成立条件は, $Y = \pm \frac{\sigma_Y}{\sigma_X} (X - \mu_X) + \mu_Y$ a.s. .

[証明]. $\sigma_X \sigma_Y \neq 0$ のとき,

$$0 \leq E[(\sigma_X^{-1}(X - \mu_X) \pm \sigma_Y^{-1}(Y - \mu_Y))^2] = 2(1 \pm \rho(X, Y))$$

より. ■

命題 4.2.17 (ランダムな双線型形式の期待値). m 次元確率変数 X と n 次元確率変数 Y とについて, $|[X], |Y|, |X||Y|$ を可積分とする。このとき, $m \times n$ -定数行列 A について,

$$E(X^T A Y) = \text{Tr}(\text{Cov}(X, Y)^\perp A) + E(X)^\perp A E(Y)$$

4.2.4 分散共分散行列

分散共分散行列と半正定値行列は同一視出来る。^a

^a <https://ja.wikipedia.org/wiki/分散共分散行列>

補題 4.2.18. 分散共分散行列 $\text{Var}[X]$ は半正定値である。

[証明]. $\forall_{u \in \mathbb{R}^{d'}} u^\perp \text{Var}[X] u = E[(u \cdot (X - E(X)))^2]$ より. ■

注 4.2.19 (退化した多次元確率変数). $\text{Var}[X]$ が正定値でないとして, $\exists_{u \in \mathbb{R}^{d'} \setminus \{0\}} u^\perp \text{Var}[X] u = E[(u^\perp (X - E(X)))^2] = 0$ である。すなわち, X は確率 1 で超平面 $u^\perp (X - E(X)) = 0$ 上に値を取る。

4.3 特性関数と母関数

積分変換

期待値作用素が積分によって定義されるため、積分変換による分析手法が肝要になる。今回は初めから確率測度について定義する。

逆転公式とは、母関数から元の確率分布を復元する算譜である。

4.3.1 特性関数と分布

定義 4.3.1 (characteristic function). $(\mathbb{R}, \mathcal{B}_1)$ 上の確率測度 ν に対して,

$$\varphi(u) := \int e^{iux} \nu(dx) = \int \cos(ux) \nu(dx) + i \int \sin(ux) \nu(dx)$$

により定まる関数 $\varphi: \mathbb{R} \rightarrow \mathbb{C}$ を特性関数という。 $|e^{iux}| = 1$ より、特性関数は常に存在する。

補題 4.3.2 (inversion formula). 確率測度 ν とその特性関数 φ について、区間 (a, b) で $\nu[\{a\}] = \nu[\{b\}] = 0$ ならば、

$$\nu[(a, b)] = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u) du.$$

定理 4.3.3 (特性関数は分布を特徴付ける). 特性関数の全体と確率測度の全体とに標準的な全単射が存在する。

[証明]. 反転公式 4.3.2 により、任意の開区間の測度は φ が一意に定める。 ■

4.3.2 分布の収束の特徴付け

定義 4.3.4 (convergence of probability measures). $(\mathbb{R}, \mathcal{B}_1)$ 上の確率測度 (ν_n) が ν に収束する: $\nu_n \rightarrow \nu$ ($n \rightarrow \infty$) とは、次が成り立つことをいう: $\forall g \in B(\mathbb{R}) \quad \int g(x) \nu_n(dx) \rightarrow \int g(x) \nu(dx)$ ($n \rightarrow \infty$) .

例 4.3.5.

$$\nu_n(A) := \sum_{j=1}^n \frac{1}{n} \delta_{\frac{j}{n}}(A) \quad (A \in \mathcal{B}(\mathbb{R}))$$

は一様分布 $U(0, 1)$ に収束する。実際、積分の定義より、

$$\int_{\mathbb{R}} g(x) \nu_n(dx) = \sum_{j=1}^n \frac{1}{n} g\left(\frac{j}{n}\right) \xrightarrow{n \rightarrow \infty} \int_0^1 g(x) dx.$$

定理 4.3.6. 確率測度 ν_n の特性関数を φ_n とする。 φ_n が関数 φ に各点収束し、かつ φ が原点で連続であるとき、 φ はある確率測度 ν の特性関数であり、 $\nu_n \rightarrow \nu$ である。

系 4.3.7 (Glivenko). 確率測度 ν_n, ν の特性関数を φ_n, φ とする。このとき、次の2条件は同値。

- (1) $\nu_n \rightarrow \nu$.
- (2) $\forall u \in \mathbb{R} \quad \varphi_n(u) \rightarrow \varphi(u)$.

[証明].

- (1) \Rightarrow (2) 定義より自明。
- (2) \Rightarrow (1) 定理による。

4.3.3 特性関数と絶対積率

命題 4.3.8. $|\beta_r| < \infty$ のとき, 特性関数 φ は C^r 級で,

$$\frac{d^r}{du^r} \varphi(u) = i^r \int e^{iux} x^r v(dx).$$

特に, $\varphi^{(r)}(0) = i^r \alpha_r$.

系 4.3.9 (平均と分散の特性関数による特徴付け).

- (1) $\alpha_1 = \frac{1}{i} \varphi'(0)$.
- (2) $\mu_2 = -\varphi''(0) + (\varphi'(0))^2$.

命題 4.3.10. 分布 v の特性関数 φ が $u = 0$ において $2r$ 次までの微分を持つとする. このとき, $|\beta_{2r}| < \infty$.

命題 4.3.11. $\beta_r < \infty$ とする. このとき, φ は $u = 0$ の周りで展開

$$\varphi(u) = \sum_{j=0}^r \alpha_j \frac{(iu)^j}{j!} + o(u^r) \quad (u \rightarrow 0)$$

を持つ.

4.3.4 キュムラント母関数

その性質を研究した Thorvald N. Thiele に因み、ティエレの半不変数 (semi-invariant) とも呼ぶ。^a積率との繋がりが深い.

^a 統計学の分野で尤度に関する初期の考察を行い、保険数学の分野で Hafnia 保険会社を設立し、数学部長を務め、デンマーク保険統計協会を設立した。

命題 4.3.12. $\beta_r < \infty$ とする. このとき, 関数 $\psi(u) := \log \varphi(u)$ は次の形の展開を持つ:

$$\psi(u) = \sum_{j=1}^r \kappa_j \frac{(iu)^j}{j!} + o(u^r) \quad (u \rightarrow 0).$$

定義 4.3.13 (cumulant). 関数 $\psi(u) := \log \varphi(u)$ をキュムラント母関数といい, この展開係数 κ_j を, 分布 v の j 次のキュムラントと呼ぶ. 分布の高次の形態を表す特性値である.

命題 4.3.14 (キュムラントの積率による表現). 可積分性を仮定する.

- (1) $\kappa_1 = \alpha_1$.
- (2) $\kappa_2 = \alpha_2 - \alpha_1^2$.
- (3) $\kappa_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = \mu_3$.
- (4) $\kappa_4 = \alpha_4 - 3\alpha_2^2 - 4\alpha_1\alpha_3 + 12\alpha_1^2\alpha_2 - 6\alpha_1^4 = \mu_4 - 3\mu_2^2$.

4.3.5 分布関数

定義 4.3.15 (distribution function).

- (1) 実確率変数 X に対して, $F^X(x) := P[X \leq x]$ により定まる関数 $F^X: \mathbb{R} \rightarrow [0, 1]$ を (累積) 分布関数という.
- (2) $(\mathbb{R}, \mathcal{G}_1)$ 上の確率測度 v に対して, $F_v(x) := v((-\infty, x]) = \int_{-\infty}^x f(y)dy$ により定まる関数 $F^X: \mathbb{R} \rightarrow [0, 1]$ を (累積) 分布関数という.

命題 4.3.16. $v_1 = v_2$ と $F_{v_1} = F_{v_2}$ は同値.

命題 4.3.17. \mathbb{R} 上の分布 $(\nu_n), \nu$ の分布関数を F_n, F とする. 次の2条件は同値.

- (1) $\nu_n \rightarrow \nu$.
- (2) F の任意の連続点 $x \in \mathbb{R}$ について, $F_n(x) \rightarrow F(x)$.

4.3.6 積率母関数

定義 4.3.18 (moment generating function). $(\mathbb{R}, \mathcal{B}_1)$ 上の確率測度 ν に対して, $M_\nu(t) := \int_{\mathbb{R}} e^{tx} \nu(dx)$ により定まる関数 $\mathbb{R} \rightarrow \mathbb{R}$ を積率母関数という.

4.3.7 多次元確率変数の特性関数

複素数値関数が特性関数になるためには Bochner の定理が十分条件を与える.

4.4 一次元連続分布の例

4.4.1 一様分布

定義 4.4.1 (uniform distribution). $(\mathcal{C}, \mathcal{B}_1)$ 上の一様分布 $U(a, b)$ ($a < b \in \mathbb{R}$) とは, 確率密度関数 f

$$f(x) := \frac{1}{b-a} 1_{(a,b)} = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{otherwise.} \end{cases} \quad \nu[A] := \int_A f(x) dx \quad (A \in \mathcal{B}_1)$$

が定める確率分布をいう.

命題 4.4.2.

- (1) 特性関数は $\varphi(u) = \frac{e^{ibu} - e^{iau}}{(b-a)iu}$. ただし $\varphi(0) = 1$ とする.
- (2) $\alpha_1 = \frac{a+b}{2}$.
- (3) $\mu_2 = \frac{(b-a)^2}{12}$.
- (4) $\gamma_1 = 0, \gamma_2 = \frac{9}{5}$.

[証明].

$$(1) E[e^{iux}] = \int_a^b e^{iux} \frac{1}{b-a} dx \text{ なので.}$$

■

4.4.2 Gamma 分布

4.4.2.1 Gamma 分布の性質

定義 4.4.3 (gamma distribution). $(\mathcal{C}, \mathcal{B}_1)$ 上の **Gamma** 分布 $G(\alpha, \nu)$ ($\alpha, \nu \in \mathbb{R}_{>0}$) とは, 確率密度関数

$$f(x) = g(x; \alpha, \nu) := \frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x} 1_{x>0}$$

が定める分布をいう. ただし, Gamma 関数とは, $\Gamma(\nu) = \int_0^\infty t^{\nu-1} e^{-t} dt$.

命題 4.4.4.

- (1) $\varphi(u) = \frac{1}{(1 - \frac{i u}{\alpha})^\nu}$.

- (2) $\alpha_1 = \frac{v}{\alpha}$.
 (3) $\mu_2 = \frac{\alpha^2}{2}$.
 (4) $\gamma_1 = \frac{2}{\sqrt{v}}, \gamma_2 = 3 + \frac{6}{v}$.

[証明] .

(1)

$$\begin{aligned}\varphi(u) &= \int_0^\infty e^{iux} \frac{1}{\Gamma(v)} \alpha^v x^{v-1} e^{-\alpha x} dx \\ &= \int_0^\infty \frac{1}{\Gamma(v)} \alpha^v x^{v-1} e^{-\alpha(1-\frac{iu}{\alpha})x} dx\end{aligned}$$

でこの後 path を考えるらしい .

4.4.2.2 指数分布

定義 4.4.5 (exponential distribution, chi-square distribution).

- (1) $\text{Exp}(\gamma) := G(\gamma, 1)$ を母数 γ の指数分布という .
 (2) $\chi^2(k) := G\left(\frac{1}{2}, \frac{k}{2}\right)$ を自由度 k のカイ 2 乗分布という .^{†2}

4.4.3 正規分布

定義 4.4.6 (normal distribution). $(\mathcal{X}, \mathcal{B}_1)$ 上の正規分布 $N(\mu, \sigma^2)$ ($\mu \in \mathbb{R}, \sigma > 0$) とは, 確率密度関数

$$\phi(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

が定める分布をいう .

命題 4.4.7.

- (1) $\int_{\mathbb{R}} \phi(x; \mu, \sigma^2) dx = 1$.
 (2) $\varphi(u) = \exp\left(i\mu u - \frac{1}{2}\sigma^2 u^2\right)$.
 (3) $\alpha_1 = \mu, \mu_2 = \sigma^2$.
 (4) $\gamma_1 = 0, \gamma_2 = 3$.
 (5) 中心積率は $\mu_{2r+1} = 0, \mu_{2r} = (2r)!(2^r r!)^{-1} \sigma^{2r}$.
 (6) キュムラントは $\kappa_1 = \mu, \kappa_2 = \sigma^2, \kappa_r = 0$ ($r \geq 3$) .^{†3}

[証明] .

- (1) gamma 積分に変換するとわかる .
 (2)
 (3) 特性関数より 4.3.9 ,

$$\begin{aligned}\alpha_1 &= \frac{1}{i} \varphi'(0) = \frac{1}{i} (i\mu) = \mu, \\ \mu_2 &= -\varphi''(0) + (\varphi'(0))^2 = \sigma^2 - (i\mu)^2 + (i\mu)^2 = \sigma^2.\end{aligned}$$

^{†2} 統計推測で重要なクラスである .

^{†3} この 3 次以上のキュムラントが消えることが正規分布の特徴で, 積率による中心極限定理の証明に利用される .

定義 4.4.8 (standard normal distribution). $N(0, 1)$ を標準正規分布という .

補題 4.4.9.

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1)$$

[証明].

$$\begin{aligned} P\left(\frac{X - \mu}{\sigma} \leq Z\right) &= P(X \leq \mu + \sigma Z) \\ &= \int_{-\infty}^{\mu + \sigma Z} \phi(x; \mu, \sigma^2) dx. \end{aligned}$$

注 4.4.10. この操作を基準化という . ある種の affine 変換だという . ある意味での等質性を表す .

4.4.4 Beta 分布

定義 4.4.11 (beta distribution). $(\mathcal{X}, \mathcal{B}_1)$ 上の (第1種) ベータ分布 $B_E(\alpha, \beta)$ ($\alpha, \beta \in \mathbb{R}_{>0}$) とは , 確率密度関数

$$f(x) := \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{(0,1)}(x)$$

が定める確率分布をいう . ただし , $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$.

命題 4.4.12.

$$\begin{aligned} (1) \quad \alpha_1 &= \frac{\alpha}{\alpha + \beta} . \\ (2) \quad \mu_2 &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} . \\ (3) \quad \gamma_1 &= \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}} . \\ (4) \quad \gamma_2 &= \frac{3(\alpha + \beta + 1)(\alpha^2\beta + \alpha\beta^2 + 2\alpha^2 - 2\alpha\beta + 2\beta^2)}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)} . \end{aligned}$$

[証明].

(1)

$$\begin{aligned} \alpha_1 &= \int_0^1 x \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} \underbrace{\int_0^1 \frac{1}{B(\alpha + 1, \beta)} x^{\alpha} (1-x)^{\beta-1} dx}_{=1} \\ &= \frac{\Gamma(\alpha + 1)\Gamma(\beta)}{\Gamma(\alpha + 1 + \beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha + \beta} . \end{aligned}$$

注 4.4.13. beta 分布は種々の分布が表現できるので便利 , R で遊ぶと良い .

4.4.5 Cauchy 分布

定義 4.4.14 (Cauchy distribution). $(\mathcal{X}, \mathcal{B}_1)$ 上の **Cauchy** 分布 $C(\mu, \sigma)$ ($\mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$) とは , 確率密度関数

$$f(x) = \frac{1}{\pi\sigma \left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)}$$

が定める分布をいう . μ は位置母数 , σ は尺度母数と解釈される .

命題 4.4.15.

- (1) 裾が重く ($1/x^2$ のオーダー), 平均と分散は存在しない.
 (2) $\varphi(u) = \exp(i\mu u - \sigma|u|)$. 明らかに $u = 0$ で微分可能でない.

歴史 4.4.16. Cauchy(1853) によると考えられていたが, Poisson が 1824 年にすでに注目していた.

4.4.6 Weibull 分布

Weibull 分布は寿命分布として用いられ, また極値分布としても現れる.

定義 4.4.17 (Weibull distribution). $(\mathcal{X}, \mathcal{B}_1)$ 上の **Weibull** 分布 $W(\nu, \alpha)$ ($\nu, \alpha \in \mathbb{R}_{>0}$) とは, 確率密度関数

$$f(x) = \frac{\nu}{\alpha} \left(\frac{x}{\alpha}\right)^{\nu-1} \exp\left(-\left(\frac{x}{\alpha}\right)^\nu\right) 1_{x>0} = \int_{-\infty}^x f(y)dy$$

が定める分布である. α が尺度母数, ν が形状母数と解釈される.

命題 4.4.18.

- (1) 分布関数は $F(x) = \left\{1 - \exp\left(-\left(\frac{x}{\alpha}\right)^\nu\right)\right\} 1_{x>0}$.
 (2) $\alpha_1 = \alpha \Gamma\left(\frac{\nu+1}{\nu}\right)$.
 (3) $\mu_2 = \alpha^2 \left\{\Gamma\left(\frac{\nu+2}{\nu}\right) - \Gamma\left(\frac{\nu+1}{\nu}\right)^2\right\}$.

命題 4.4.19. $X \sim \text{Exp}\left(\frac{1}{\alpha^\nu}\right)$ ならば, $X^{\frac{1}{\nu}} \sim W(\nu, \alpha)$.

4.4.7 対数正規分布

定義 4.4.20 (log-normal distribution). $(\mathcal{X}, \mathcal{B}_1)$ 上の対数正規分布 $L_N(\nu, \sigma)$ ($\nu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$) とは, 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} 1_{x>0}$$

が定める分布である.

命題 4.4.21.

$$\log X \sim N(\mu, \sigma^2) \Leftrightarrow X \sim L_N(\mu, \sigma)$$

命題 4.4.22.

- (1) $\alpha_1 = \exp\left(\mu + \frac{\sigma^2}{2}\right)$.
 (2) $\mu_2 = (e^{\sigma^2} - 1) \exp(2\mu + \sigma^2)$.
 (3) $\gamma_1 = (\omega + 2)\sqrt{\omega - 1}$. ただし, $\omega := \exp(\sigma^2)$ とした.
 (4) $\gamma_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3$.

4.4.8 logistic 分布

定義 4.4.23 (logistic distribution). $(\mathcal{X}, \mathcal{B}_1)$ 上のロジスティック分布 $\text{Log}(\mu, \sigma)$ ($\mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$) とは, 分布関数

$$F(x) = \frac{1}{1 + \exp\left(-\frac{x-\mu}{\sigma}\right)}$$

で, 確率密度関数

$$f(x) = \frac{1}{\sigma} \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{\left\{1 + \exp\left(\frac{x-\mu}{\sigma}\right)\right\}^2}$$

が与える分布をいう．よって，分布は $x = \mu$ に関して対称である．

補題 4.4.24.

$$I(r) := \int_0^\infty x^r \frac{e^{-x}}{(1+e^{-x})^2} dx = (1-2^{-(r+1)})\Gamma(r+1)\zeta(r) \quad (r > 1)$$

[証明] .

$$\begin{aligned} I(r) &= \int_0^\infty x^r \frac{e^{-x}}{(1+e^{-x})^2} dx = \int_0^\infty \frac{rx^{r-1}}{1+e^x} dx \\ &= \int_0^\infty rx^{r-1}e^{-x} \sum_{j=0}^\infty (-1)^j e^{-jx} dx \\ &= \Gamma(r+1) \sum_{n=1}^\infty (-1)^{n-1} \frac{1}{n^r} \quad (r > 0) \end{aligned}$$

ここで， $\zeta(r) = \sum_{n=1}^\infty \frac{1}{n^r}$ ($r > 1$) について

$$(1-2^{-(r+1)})\zeta(r) = \sum_{n=1}^\infty (-1)^{n-1} n^{-r}$$

が成り立つから，最後の変形を得る．

命題 4.4.25.

(1) 中心積率は，奇数について $\mu_r = 0$ ，2以上の偶数について

$$\mu_r = 2\sigma^r \Gamma(r+1) \sum_{n=1}^\infty (-1)^{n-1} \frac{1}{n^r} = 2\sigma^r (1-2^{-(r+1)})\Gamma(r+1)\zeta(r).$$

$$(2) \alpha_1 = \mu, m_2 = \frac{\pi^2 \sigma^2}{3}, \mu_4 = \frac{7\pi^4 \sigma^4}{15}.$$

$$(3) \gamma_1 = 0, \gamma_2 = 4.2.$$

$$(4) \varphi(u) = e^{i\mu u} \frac{\pi \sigma u}{\sinh(\pi \sigma u)}.$$

[証明] .

(1) 奇数の時は対称性より．

$$(2) \zeta(2) = \frac{\pi^2}{6}, \zeta(4) = \frac{\pi^4}{90} \text{ より }.$$

(3) $-R, R, -R+2\pi i, R+2\pi i$ を頂点にもつ長方形についての留数計算より．

4.4.9 Pareto 分布

パレート分布は所得の分布に当てはまるという．

定義 4.4.26. $(\mathcal{X}, \mathcal{B}_1)$ 上のパレート分布 $P_A(b, a)$ ($a, b \in \mathbb{R}_{>0}$) とは，分布関数を

$$F(x) = 1 - \left(\frac{b}{x}\right)^a 1_{x \geq b}$$

確率密度関数を

$$f(x) := ab^a x^{-(a+1)} 1_{x \geq b}$$

とする分布をいう．

命題 4.4.27.

$$(1) \alpha_1 = ab(a-1)^{-1} \ (a > 1) .$$

$$(2) \mu_2 = ab^2(a-1)^{-2}(a-2)^{-1} \ (a > 2) .$$

$$(3) \gamma_1 = 2\frac{a+1}{a-3}\sqrt{\frac{a-2}{a}} \ (a > 3) .$$

$$(4) \gamma_2 = \frac{3(a-2)(3a^2+a+2)}{a(a-3)(a-4)} \ (a > 4) .$$

4.5 確率変数の独立性と期待値

4.6 多次元連続分布

4.7 指数型分布族

4.8 指数分散モデル

4.9 コピュラ

4.10 確率分布と積分変換

4.11 不等式

第 5 章

多変量分布

値域を \mathbb{R}^d に一般化し，線形代数的な本性を暴き出す．

5.1 多次元確率変数の分布

離散確率分布を table にまとめた時，分布表の中の部分が結合分布となり，縁の部分（合計欄）が周辺分布となる．

定義 5.1.1 (joint / simultaneous distribution, marginal distribution). $X = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ を確率変数とする．

- (1) $(\mathbb{R}^d, \mathcal{B}_d)$ 上の確率分布 P^X を X_1, \dots, X_d の結合分布または同時分布という．
- (2) 各 X_i の分布 P^{X_i} を周辺分布という． $P^{(X_{i_1}, \dots, X_{i_k})}$ も， $(X_{i_1}, \dots, X_{i_k})$ の周辺分布という．

例 5.1.2 (multinomial distribution). k 個の背反な事象 $\sum_{i=1}^k A_i = \Omega$ について， $P(A_i) = p_i$ とする．各 A_i の起こる回数を X_i とすると，確率ベクトル (X_1, \dots, X_k) の分布を k 項分布 $M(n; p_1, \dots, p_k)$ という． $k = 2$ の場合は二項分布に等しくなる．また，各 X_i の周辺分布は 2 項分布 $B(n, p_i)$ である．多項分布は必ず $\sum_{i=1}^k X_i = n$ という線型関係を満たすので，この意味で退化した分布である．

5.2 共分散

5.2.1 共分散の定義と例

定義 5.2.1 (covariance, correlation coefficient). 2 乗可積分実確率変数 X, Y について，

- (1) $\text{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])]$ を共分散という．
- (2) $\rho(X, Y) := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}}$ を相関係数という．

命題 5.2.2. 2 乗可積分実確率変数 X, Y, Z について，

- (1) $\text{Cov}[X, Y] = \text{Cov}[Y, X]$.
- (2) $\forall a, b \in \mathbb{R} \quad \text{Cov}[aX + bY, Z] = a\text{Cov}[X, Z] + b\text{Cov}[Y, Z]$.
- (3) $\text{Cov}[X, X] = \text{Var}[X] \geq 0$. 等号が成立するならば $X = E[X]$ a.s. .
- (4) $\text{Cov}[X, 1] = 0$. $\forall a, b \in \mathbb{R} \quad \text{Cov}[aX + b, Y] = a\text{Cov}[X, Y]$.
- (5) $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$.

例 5.2.3 (多項分布の共分散). $X = (X_1, \dots, X_k) \sim M(n; p_1, \dots, p_k)$ とする．

- (1) $E[X_i] = np_i$.
- (2) $\text{Var}[X_i] = np_i(1 - p_i)$.

$$(3) \operatorname{Cov}[X_{i_1}, X_{i_2}] = -np_{i_1}p_{i_2} \ (i_1 \neq i_2 \in [k]) .$$

[証明]. 多項定理より,

$$(e^{u_1}p_1 + \cdots + e^{u_k}p_k)^n = \sum_{x_1, \dots, x_k} {}^*P^X[\{(x_1, \dots, x_k)\}] e^{u_1x_1 + \cdots + u_kx_k} .$$

ただし, * は線形関係 $x_1 + \cdots + x_k = n$ を満たす x_1, \dots, x_k についての和とする.

$$(1) \text{ 両辺の } (u_1, \dots, u_k) = (0, \dots, 0) \text{ における } u_i \text{ 偏微分係数より, } np_i = \sum_{x_1, \dots, x_k} {}^*x_i P^X[\{(x_1, \dots, x_k)\}] = E[X_i] .$$

$$(2) \text{ 同様に } u_i \text{ の 2 階微分を考えて, } E[X_i^2] = n(n-1)p_i^2 + np_i \text{ を得る. よって, } \operatorname{Var}[X_i] = E[X_i^2] - (E[X_i])^2 = np_i(1-p_i) .$$

$$(3) i_1 \neq i_2 \text{ に関して, 順に } u_{i_1}, u_{i_2} \text{ での偏微分を考えることにより, } E[X_{i_1}X_{i_2}] = n(n-1)p_{i_1}p_{i_2} . \text{ よって, 共分散公式 5.2.2 より,}$$

$$\operatorname{Cov}[X_{i_1}, X_{i_2}] = E[X_{i_1}X_{i_2}] - E[X_{i_1}]E[X_{i_2}] = -np_{i_1}p_{i_2} .$$

■

要諦 5.2.4. 証明が技巧的すぎる, 多項定理に指数関数を代入する.

5.2.2 期待値の拡張

定義 5.2.5 (covariance matrix, variance-covariance matrix). $X : \Omega \rightarrow \mathbb{R}^d$ を確率変数とする.

(1) X が可積分のとき, 項別積分

$$E[X] = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_d] \end{bmatrix}$$

を X の平均ベクトルという.

(2) X, Y が 2 乗可積分のとき,

$$\operatorname{Cov}[X, Y] = \begin{bmatrix} \operatorname{Cov}[X_1, Y_1] & \cdots & \operatorname{Cov}[X_1, Y_s] \\ \vdots & \ddots & \vdots \\ \operatorname{Cov}[X_d, Y_1] & \cdots & \operatorname{Cov}[X_d, Y_s] \end{bmatrix}$$

を X, Y の共分散行列という.

(3) $\operatorname{Cov}[X, X]$ を X の分散共分散行列または分散行列と呼ぶ.

注 5.2.6. この共分散行列は、シンプルではあるが、非常に多岐にわたる分野でとても有用なツールである。分散共分散行列からは、データの相関を完全に失わせるような写像を作る変換行列を作ることができる。これは、違った見方をすれば、データを簡便に記述するのに最適な基底を取っていることになる。(分散共分散行列のその他の性質やその証明については、en:Rayleigh quotient を参照) これは、統計学では主成分分析 (PCA) と呼ばれており、画像処理の分野では、カルーネン・レーベ変換 (KL-transform) と呼ばれている。^{†1}

記法 5.2.7 (期待値作用素の拡張). 期待値作用素 E を行列値確率変数 $M = [M_{ij}] \in M_{ij}(\mathbb{R})$ 上に対しても $E : M_{ij}(\mathbb{R}) \rightarrow M_{ij}(\mathbb{R}); E[M] = [E[M_{ij}]]$ と拡張すると, 平均ベクトルは $E[M]$, 共分散行列は $\operatorname{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])^T]$ と表せる.

命題 5.2.8.

$$(1) \text{ 可積分 } d \text{ 次元確率変数 } X, m \times d \text{ 行列 } A, a \in \mathbb{R}^m \text{ に対して, } E[AX + a] = AE[X] + a .$$

$$(2) \text{ 2 乗可積分 } d \text{ 次元確率変数 } X, \text{ 2 乗可積分 } s \text{ 次元確率変数 } Y \text{ に対して, } \operatorname{Cov}[X, Y] = \operatorname{Cov}[Y, X]^T . \text{ また, } \operatorname{Cov}[X, Y] = E[XY^T] - E[X]E[Y]^T .$$

$$(3) \text{ 2 乗可積分 } d \text{ 次元確率変数 } X, Y, \text{ 2 乗可積分 } s \text{ 次元確率変数 } Z \text{ に対して, } \operatorname{Cov}[X + Y, Z] = \operatorname{Cov}[X, Z] + \operatorname{Cov}[Y, Z] .$$

$$(4) \text{ 2 乗可積分 } d \text{ 次元確率変数 } X, m \times d \text{ 行列 } A, a \in \mathbb{R}^m, \text{ 2 乗可積分 } s \text{ 次元確率変数 } Y, b \in \mathbb{R}^n \text{ に対して, } \operatorname{Cov}[AX + a, BY + b] = A\operatorname{Cov}[X, Y]B^T .$$

^{†1} <https://ja.wikipedia.org/wiki/分散共分散行列>

命題 5.2.9 (確率ベクトルの2次形式の平均). X を d 次元2乗可積分確率変数, G を $d \times d$ 定数行列とする. このとき,

$$E[X^T G X] = \text{Tr}(G \text{Var}[X]) + E[X]^T G E[X].$$

5.2.3 積率の定義

積率の次元もベクトル値 $n \in \mathbb{N}^d$ で指定できる.

記法 5.2.10. $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ に対して, $\partial_i := \frac{\partial}{\partial x_i}$ とし, $n := (n_1, \dots, n_d) \in \mathbb{Z}_+^d$ に対して,

$$\begin{aligned} |n| &:= n_1 + \dots + n_d, & n! &:= n_1! \dots n_d!, \\ x^n &:= x_1^{n_1} \dots x_d^{n_d}, & \partial^n &:= \partial_1^{n_1} \dots \partial_d^{n_d}. \end{aligned}$$

ただし, $x_j^0 = 1, \partial^0 = 1$ とよむ.

定義 5.2.11 (moment, central moment). d 次元確率変数 $X = (X_1, \dots, X_d)$ に対して, 可積分性の仮定の下で,

- (1) $\alpha_n := E[X^n]$ を n 次積率という.
- (2) $\mu_n := E[(X - E[X])^n]$ を n 次中心積率という.

定義 5.2.12. $(\mathbb{R}^d, \mathcal{B}_d)$ 上の確率測度 ν に対して, 可積分性の仮定の下で,

- (1) $\mu := \left(\int_{\mathbb{R}^d} x_i \nu(dx) \right) \in \mathbb{R}^d$ を平均 (ベクトル) という.
- (2) $\alpha_n := \int_{\mathbb{R}^d} x^n \nu(dx)$ を n 次の積率という.
- (3) $\nu_n := \int_{\mathbb{R}^d} (x - \mu)^n \nu(dx)$ を n 次の中心積率という.
- (4) 特に $(\mu_n)_{|n|=2}$ を ν の分散共分散行列という.

5.3 特性関数と分布の収束

5.4 独立性

5.5 多変量連続分布

5.6 多変量正規分布

5.7 変数変換と確率密度関数

5.8 従属性

第 6 章

条件付き期待値

6.1 部分 σ -加法族に関する条件付き期待値

条件付き期待値の存在は Radon-Nykodym の定理により，一意性は零集合を除いて一致する．

定義 6.1.1. 次の条件を満足する確率変数 Y を， \mathcal{G} に関する X の条件付き期待値と呼ぶ．

- (1) Y は \mathcal{G} -可測．
- (2) Y は可積分で，任意の $A \in \mathcal{G}$ に対して $\int_A X(\omega)P(d\omega) = \int_A Y(\omega)P(d\omega)$ ．

この Y を $E[X|\mathcal{G}]$ で表す．

補題 6.1.2 (well-definedness).

- (1) $E[X|\mathcal{G}]$ は存在する．
- (2) $E[X|\mathcal{G}]$ は P -零集合を除いて一意である．

例 6.1.3 (条件付き確率). $(\Omega_j)_{j \in \mathbb{N}}$ を Ω の分割で， $P[\Omega_j] > 0$ とする． $\mathcal{G} := \sigma[\Omega_j | j \in \mathbb{N}]$ と定めると，可積分確率変数 X に関して，

$$E[X|\mathcal{G}] = \sum_{j \in \mathbb{N}} \frac{E[X1_{\Omega_j}]}{P(\Omega_j)} 1_{\Omega_j} \quad \text{a.s.}$$

第 7 章

確率変数の収束

確率変数に収束の概念が多様に定義できる．その方法は確率変数の可測性に依らないので，なるべく一般的な形で述べる．

7.1 確率測度の収束

確率変数は，確率測度を押し出す．この構造を用いて確率変数の収束「法則収束」を定義するから，まずは測度の収束を論じる．

測度を，関数の双対空間の元と考える．すると，測度のなす空間 $M(X)$ とは作用素の空間 $C(X)^*$ であるから，そこで収束の議論をするのは函数解析学に他ならない．測度の弱収束とは w^* -収束に他ならず，確率測度のなす空間 $P(X)$ はこの位相についてコンパクトである．経験過程は極点の凸結合であると捉えられる．極点の凸結合の極限で真の分布を探そうとする営みが経験過程論であるか！？

7.1.1 汎関数としての測度

記法 7.1.1. X を Hausdorff 空間， X の Borel 集合体を \mathcal{A} ，Borel σ -集合体を \mathcal{B} で表し， \mathcal{A} 上の有限加法的な正則有限符号付測度の空間を $M_f(X)$ ， \mathcal{B} 上の σ -加法的な正則有限符号付測度の空間を $M_\sigma(X)$ で表す．これらに全変動ノルムを入れると，加法的集合関数は有界変動だから，ノルム空間となる．明らかに， $M_f(X) \subset C_b(X)^*$ である．

命題 7.1.2 (Alexandrov). X を任意の正規空間とする．任意の有界汎関数 $\Lambda \in C_b(X)^*$ に対して，ただ一つの $\mu \in M_f(X)$ が存在して， $\forall f \in C_b(X) \quad \Lambda(f) = \int_X f d\mu$ を満たし， $\|\Lambda\| = \|\mu\|$ を満たす．すなわち，Banach 空間として $C_b(X)^* \simeq_{\text{Ban}} M_f(X)$ ．

命題 7.1.3 (Riesz-Markov-Kakutani). コンパクトハウスドルフ空間 X について， $C_b(X) = C(X)$ であることに注意する．

- (1) 任意の有界汎関数 $\Lambda \in C_b(X)^*$ に対して，ただ一つの $\mu \in M_\sigma(X)$ が存在して， $\forall f \in C_b(X) \quad \Lambda(f) = \int_X f d\mu$ を満たし， $\|\Lambda\| = \|\mu\|$ を満たす．
- (2) さらに Λ が正である場合，対応する μ も正值である．

7.1.2 収束の対応

古典的な弱収束の概念は，Borel 確率空間に値を取る Borel 確率変数について定義された．しかし，可分でない空間に値を取る場合は，Borel 完全加法族は非常に大きいため，Borel 可測性は条件として強すぎる．たとえそのような関数の極限が Borel 可測であっても，である（この重要な例が経験過程である）．これは非常に不自然で，関数解析的な視点からは地を這っているように無明である．Skorokhod は可分となるような距離を見つけたが，この方向の対処よりも抜本的な解決がほしい．Dudley が，一様ノルムについての Banach 空間と見たまま，弱収束の定義を拡張する方向を開拓し，Hoffmann-Jorgensen が外積分の基盤を作った．

定義 7.1.4. Borel 可測空間 $(X, \mathcal{B}(X))$ 上の確率測度の空間 $P(X) \subset C_b(X)^*$ の列 (μ_n) について，

- (1) 各点収束 $\forall_{E \in \mathcal{G}(X)} \mu_n(E) \rightarrow \mu(E)$ は弱収束に対応する .
- (2) $\forall_{f \in C_b(X)} \int_X f d\mu_n \rightarrow \int_X f d\mu$, 平均で表すと $\forall_{f \in C_b(X)} E[f(X_n)] \rightarrow E[f(X)]$ は $*$ -弱収束 ($\sigma(C_b(X)^*, C_b(X))$ -位相) に対応する .
しかしこれは「測度の弱収束」と呼ばれる . これを $\mu_n \Rightarrow \mu$ と表す . これは確率変数列の分布収束・法則収束に対応する .
- (3) さらに条件を弱めて $\forall_{f \in C_c(X)} \int_X f d\mu_n \rightarrow \int_X f d\mu$ は $\sigma(C_b(X)^*, C_c(X))$ -位相に対応し , 漠収束と呼ばれる .

注 7.1.5. $*$ -弱収束の特徴付けを portmanteau 定理と呼ぶのであった .

7.1.3 位相構造

記法 7.1.6. Borel 確率測度のなす空間を $P(X)$ と表す . 距離空間 X 上の Dirac 測度全体の集合を $\Delta \subset P(X) \subset C_b(X)^*$ で表す .

補題 7.1.7. X は , $*$ -弱位相における Δ と位相同型である .

補題 7.1.8. Δ は $P(X)$ 内の点列閉集合である . すなわち , Δ 内の点列は Δ 内に w^* -収束する .^{†1}

補題 7.1.9. X を全有界な距離空間とする . X 上の有界な一様連続関数の集合 $U(X) \subset C_b(X)$ は , 一様ノルムの下で可分な Banach 空間となる .

定理 7.1.10. 次の 2 条件は同値 .

- (1) X は可分な距離空間である .
- (2) $P(X)$ は距離付け可能で可分である .

例 7.1.11 (Prokhorov 距離). $B_\epsilon(E) := \{x \in X \mid \rho(x, E) < \epsilon\}$ と表す .

$$\eta(\mu, \nu) := \inf \{ \epsilon > 0 \mid \forall_{E \in \mathcal{G}(X)} \mu(E) \leq \nu(B_\epsilon(E)) + \epsilon \wedge \nu(E) \leq \mu(B_\epsilon(E)) + \epsilon \}$$

は距離を定め , これが定める位相は $*$ -弱位相に一致する .

定理 7.1.12 (可算稠密部分集合の構成). X を可分距離空間 , D をその可算な稠密部分集合とする . このとき , D の有限部分集合を台とするような確率測度のなす集合 $F(D)$ は , $P(X)$ において稠密である .

系 7.1.13. D が可分である時 , Dirac 測度としての標準的な埋め込み $D \hookrightarrow \mathcal{M}(D)$ の像の凸包は稠密である .

定理 7.1.14 (コンパクト性). 距離空間 X について , 次の 2 条件は同値 .

- (1) X はコンパクトである .
- (2) $P(X)$ は w^* -コンパクトである .

定理 7.1.15 (完備性). 可分な距離空間 X について , 次の 2 条件は同値 .

- (1) X は完備である .
- (2) $P(X)$ は完備である .

定理 7.1.16. D がポーランドである時 , $\mathcal{M}(D)$ もポーランドである .

定義 7.1.17 (uniformly tight). 確率測度の族 $\Gamma \subset P(X)$ が一様に緊密であるとは , $\forall_{\epsilon > 0} \exists_{K \subset X}^{\text{cpt}} \forall_{\mu \in \Gamma} \mu(K) \geq 1 - \epsilon$.

定理 7.1.18 (緊密性とは , 相対コンパクト性の特徴付けである). X を完備可分距離空間とする . 確率測度の族 $\Gamma \subset P(X)$ について , 次の 2 条件は同値 .

- (1) Γ は $*$ -弱位相について相対コンパクトである .
- (2) Γ は一様に緊密である .

^{†1} 一般に部分集合 $A \subset X$ の閉包 \bar{A} は点列閉包 B を含むが , 点列閉包と一致するのは一般に X が距離空間の場合のみ .

7.1.4 幾何構造

経験過程は重要な意味を持つ。これを、極点の凸結合として幾何学的に説明できないか？経験過程の有限性は実用性に通じるが、これは組み合わせ論的な本質も備えているのではないか？

命題 7.1.19. コンパクトハウスドルフ空間 X 上の Banach 代数 $C(X) = C_b(X)$ を考える、但しノルムは一様ノルムとした。 $C(X)$ の双対空間を $M(X)$, $P(X) := \{\mu \in M(X) \mid \|\mu\| \leq 1, \mu(1) = 1\}$ を確率測度のなす部分空間とする。

- (1) $P(X)$ は $M(X)$ の凸集合である。
- (2) $P(X)$ は w^* -コンパクトである。
- (3) $P(X)$ の極点は Dirac 測度 δ_x ($x \in X$) , $\forall f \in C(X) \delta_x(f) = f(x)$ である。

[証明].

- (1) $\mu_1, \mu_2 \in P(X)$ と $\lambda \in (0, 1)$ を任意にとると, $\lambda\mu_1 + (1 - \lambda)\mu_2 \in P(X)$ がわかる。
- (2) 線型汎函数 $ev_1 : M(X) \rightarrow \mathbb{F}; \mu \mapsto \mu(1)$ は w^* -位相について連続である $M(X) = (C(X))^*$ の閉単位球 B^* は w^* -コンパクト?? である。

■

要諦 7.1.20. コンパクトハウスドルフ空間 X 上の確率測度は、 $C(X)$ 上において、有限な台を持つ測度 (= Radon charge が定める積分の空間) によって各点近似 (各点収束の位相で近似) が出来る。

7.2 積分の拡張

まずは、関数解析の力を借りて、積分を非可測な関数に対しても延長する、Hoffmann-Jorgensen の理論を展開する。この議論は一般の局所コンパクトハウスドルフ空間 X 上の Radon 積分 $\int : C_c(X) \rightarrow \mathbb{R}$ に関して展開できる。

定義 7.2.1 (outer integral). (Ω, \mathcal{A}, P) を確率空間, $T : \Omega \rightarrow \overline{\mathbb{R}}$ を写像とする。

- (1) $E^*[T] := \inf \{E[U] \in \mathbb{R} \mid U \geq T, U \in \text{Meas}(\Omega, \overline{\mathbb{R}}), E[U] \in \mathbb{R}\}$ を、押し出された測度 T_*P に関する $\overline{\mathbb{R}}$ 上の Radon 上積分とする。
- (2) 写像列 $(X_n : \Omega_n \rightarrow \Omega)$ が Borel 可測関数 $X : \Omega_\infty \rightarrow \Omega$ に弱収束するとは、 $\forall f \in C_b(\Omega) E^*[f(X_n)] \rightarrow E[f(X)]$ を満たすことをいう。

7.3 概収束

第 8 章

確率測度の全て

8.1 一般の試行と確率測度

定義 8.1.1 (null set, complete).

- (1) $P(N) = 0$ を満たす P -可測集合 N を, P -零集合という.
- (2) 任意の P -零集合の, 任意の部分集合も全て P -可測 (したがって零) である時, P を完備確率測度という.

命題 8.1.2 (Lebesgue expansion). 任意の確率測度 P に対し, 完備拡張は必ず存在し, 最小のものが一意に定まる. これを P の Lebesgue 拡大という.

8.2 確率測度の拡張定理

定義 8.2.1 (regular probability measure). 位相空間上の確率測度を考える.

- (1) 定義域が Borel class と一致するものを, **Borel 確率測度**という.
- (2) Borel 確率測度の Lebesgue 拡大を正則確率測度という.

8.3 確率測度の直積

8.4 標準確率空間

多様体に対する Euclid 空間のように, 標準的な空間を用意するために, 同型の理論を整備し, (\mathbb{R}, μ) の同型類を特徴づける.

定義 8.4.1 (canonical measure). P を Ω 上の完備な確率測度とする. (Ω, P) が, 正則確率測度を備えた実数空間 (\mathbb{R}, μ) に同型である時, (Ω, P) を標準確率空間と呼ぶ.

定義 8.4.2 (pushforward measure). S 上の測度 P の, 可測関数 $f: S \rightarrow T$ による像測度を, Pf^{-1} または f_*P で表す.

補題 8.4.3. P が完備ならば Q も完備である.

8.5 多変量分布

第 9 章

確率論の基礎概念

9.1 可分完全確率測度

Kolmogorov は晩年完全性を追加した．伊藤清の教科書では可分性も追加した．完全性は像測度に遺伝する．可分性はどこで効いてくるかはわからない．

定義 9.1.1 (perfect (Kolmogorov)). 完備な確率空間 (S, μ) 上の任意の μ -可測関数 $f: S \rightarrow \mathbb{R}$ に対し，像測度 μ_*f が正則になるとき，確率測度 μ を完全という．

定義 9.1.2 (separating family, separable).

- (1) S 上の集合族 \mathcal{A} が分離族であるとは， $\forall_{s_1 \neq s_2 \in S} \exists_{A \in \mathcal{A}} 1_A(s_1) \neq 1_A(s_2)$ を満たすことをいう．
- (2) 完備確率測度 μ について， $\text{dom}(\mu)$ が可算分離族を含むとき， μ を可分という．

定理 9.1.3. μ を S 上の確率測度， $f: S \rightarrow T$ を可測関数とする．像測度 $\nu := \mu f^{-1}$ は完全である．

9.2 事象と確率変数

記法 9.2.1 (extension). 条件 $\alpha \subset \Omega$ について， α を成立させるような元からなる集合 $\{\alpha\} := \{\omega \in \Omega \mid \alpha(\omega)\}$ を α の外延という．すると， $\{X(\omega) \leq \alpha\} = X^{-1}((-\infty, \alpha])$ などと表せる．

定義 9.2.2 (almost surely). $P\{\alpha\} = 1$ のとき，事象 α はほとんど確実に起こるといい， $\alpha(\omega)$ a.s. と書く．

定義 9.2.3 (random variable). (Ω, P) 上の P -可測実関数を実確率変数と呼ぶ．

9.3 条件付き期待値作用素

第 10 章

確率過程

10.1 関数空間 C と D

10.2 確率過程に関する一般事項

10.3 情報と情報増大系

定義 10.3.1 (adapted). $F = (F_t)_{t \in T}$ を情報系, $X = (X_t)_{t \in T}$ を確率過程とする.

- (1) $\forall t \in T$ X_t が \mathcal{F}_t -可測のとき, X は F に適合するという.
- (2) $\forall t \in T$ X_t が \mathcal{F}_{t-1} -可測のとき, X は F で可予測であるという.

10.4 停止時

10.5 離散時変数のマルチンゲール

定理 10.5.1 (Doob – Meyer decomposition theorem). 任意の (\mathcal{F}_n) -劣マルチンゲール (X_n) は, \mathcal{F} -マルチンゲールな $M = (M_n)$ と可予測な増大列 $A = (A_n)$ で $A_0 = 0, A_n \in L^1(\mathcal{F}_{n-1})$ ($n = 1, 2, \dots$) とに一意的に分解される: $X_n = M_n + A_n$.

第 11 章

線型推測論

線型推測論が統計モデルの線形代数である .

第 12 章

統計的決定理論

統計的決定問題が設定されたとき，決定問題の下で起こる確率現象を解明するのが数理統計学の課題であって，特定の決定関数を無条件に是とするものではない．

12.1 枠組み

決定すべきは関数 = 射である．その文脈は 8-組で表現できる．

定義 12.1.1 (statistical decision problem, sample space, decision / action space, loss function, nonrandomized, randomized decision function (Abraham Wald)). 次の 8-組 $(\mathcal{X}, \mathcal{A}, \mathcal{P}, \Theta, \mathcal{D}, \mathcal{B}, W, \Delta)$ を統計的決定問題という．

- (1) $(\mathcal{X}, \mathcal{A})$ は可測空間で，標本空間という．
- (2) Θ はパラメータの空間で， $\mathcal{P} := (P_\theta)_{\theta \in \Theta}$ は $(\mathcal{X}, \mathcal{A})$ の確率分布の族．
- (3) $(\mathcal{D}, \mathcal{B})$ は可測空間で，決定空間または行動空間という．
- (4) 第二変数について可測な関数 $W : \Theta \rightarrow \text{Meas}(\mathcal{D}, \mathbb{R}_+)$ を損失関数という．
- (5) Δ は決定関数の族とする．
 - (a) 可測写像 $\delta : \mathcal{X} \rightarrow \mathcal{D}$ を非確率的決定関数という．
 - (b) 第一変数については $(\mathcal{D}, \mathcal{B})$ 上の確率測度，第二変数については \mathcal{X} 上の \mathcal{A} -可測関数となる関数 $\delta : \mathcal{B} \times \mathcal{X} \rightarrow [0, 1]$ を確率的決定関数という．

注 12.1.2. 非確率決定関数 $\delta : \mathcal{X} \rightarrow \mathcal{D}$ が与えられたとき，任意の $B \in \mathcal{B}$ に対して $\tilde{\delta}(B|x) := 1_{\{\delta(x) \in B\}}$ で $\tilde{\delta} : \mathcal{B} \times \mathcal{X} \rightarrow [0, 1]$ が定まるから，確率的決定関数の方がより一般的な設定となる．

例 12.1.3 (平均値の推定量の決定). $(\mathcal{X}, \mathcal{A}) := (\mathbb{R}^n, \mathcal{B}_n)$ $\mathcal{P}_1 := \left\{ P = P_*^n \in \mathcal{M}(\mathbb{R}^n) \mid P_* \text{ は } \mathbb{R} \text{ 上の確率分布で } \int_{\mathbb{R}} x^2 P_*(dx) < \infty \right\}$ ， $\Theta := \mathcal{P}_1$ とする．決定空間は平均値の全体としたいから $(\mathcal{D}, \mathcal{B}) := (\mathbb{R}, \mathcal{B}_1)$ ．推定量 $\Delta := \{T_1, T_2, T_3\}$ はそれぞれ $T_1 := X_1$, $T_2 := \frac{1}{n} \sum_{j=1}^n X_j$, $T_3 := X_n$ という非確率的決定関数の族で，損失関数は $W(P, a) := (a - \mu_1(P))^2$ とする．

すると，損失関数 W から定まる危険関数は

$$\begin{aligned} R(P, T_i) &= \int W(P, T_i(x_1, \dots, x_n)) P(dx_1, \dots, dx_n) \quad (P \in \mathcal{P}_1) \\ &= \text{Var}_P[T_i] \end{aligned}$$

となる．こうして， $R(P, T_1) = R(P, T_3) = \text{Var}_P[X_1]$, $R(P, T_2) = \frac{\text{Var}_P[X_1]}{n}$ となり， T_2 が危険関数 R を最小にするとわかる．

定義 12.1.4 (risk function, admissible). 危険関数 $R : \Theta \times \Delta \rightarrow \mathbb{R}$ を次のように定める．

- (1) 確率的な決定関数 $\delta : \mathcal{B} \times \mathcal{X} \rightarrow [0, 1]$ については， $R(\theta, \delta) := \int_{\mathcal{X}} \int_{\mathcal{D}} W(\theta, a) \delta(da|x) P_\theta(dx)$ と定める．
- (2) 非確率的な決定関数 $\delta : \mathcal{X} \rightarrow \mathcal{D}$ については， $R(\theta, \delta) := \int_{\mathcal{X}} W(\theta, \delta(x)) P_\theta(dx)$ と定める．

2 つの決定関数 $\delta_1, \delta_2 \in \Delta$ について，

- (1) $\forall \theta \in \Theta \ R(\theta, \delta_1) \leq R(\theta, \delta_2)$ が成り立つとき, δ_1 と δ_2 は同程度に良い決定関数であるという.
- (2) 同程度に良い決定関数がさらに $\exists \theta_0 \in \Theta \ R(\theta_0, \delta_1) < R(\theta_0, \delta_2)$ を満たすとき, δ_1 は δ_2 より一様に良い決定関数であるという.
- (3) 「一様に良い」という関係は順序を定める. Δ に最大元が存在せず, 極大元 δ_* のみが存在するとき, 許容的であるという.

12.2 十分性と完備性

12.2.1 十分統計量

12.3 指数型分布族

12.4 統計的推定

第 13 章

統計的仮説検定

13.1 枠組みと例

13.1.1 枠組み

定義 13.1.1 (hypothesis testing, null hypothesis, alternative hypothesis, significance level, critical region). 確率空間 (\mathcal{X}, P) に対して, 事前に与えられた設定 $\alpha \in (0, 1)$, H_0, H_1 から定められる部分集合 $\mathcal{X}_1 \subset \mathcal{X}$ を検定という.

- (1) 仮説検定において, 仮説とは, 母数または確率分布に関する (メタ) 条件をいう.
- (2) 検定される仮説 H_0 を帰無仮説といい, それに対立する仮説 H_1 を対立仮説という.
- (3) 文脈としては, H_0 を棄却し H_1 の成立を示すことで社会的な意味を見出そうという文脈に当てはまるように H_0, H_1 を選ぶ.
- (4) $P(H_0)$ の値に意味を持たせるための指標を有意水準という. $\alpha = 0.05, 0.01$ などが用いられる.
- (5) $\mathcal{X}_1 \subset \mathcal{X}$ を α と H_0, H_1 の意義から事前に定めて棄却域といい, 観測値 x が $x \in \mathcal{X}_1$ のとき棄却し, $x \notin \mathcal{X}_1$ のとき採択する.

例 13.1.2. $\alpha = 0.05$ とする. 20 回中事象 E が 13 回起こったとする. 帰無仮説 H_0 を $P(E) = 1/2$ とし, 対立仮説 H_1 を $P(E) > 1/2$ とする. すると, 棄却域は $x_0 \in \mathbb{N}$ を用いて $\mathcal{X}_1 = \{x \in [20] \mid x \geq x_0\}$ と定めるのが良いと考えられる. いま,

$$\sum_{k=14}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20} \approx 0.0577$$

より, $13 < x_0$ であって, 観測値 13 は \mathcal{X}_1 には入らないから, 棄却できない. 「確率 0.05 程度の事象が偶々起こった」と解釈する方が選択される.

定義 13.1.3 (type I error, type II error). H_0 が正しいのに H_0 を棄却してしまう誤りを第一種の誤りまたは偽陽性といい, H_1 が正しいのに H_0 を採択してしまう誤りを第二種の誤りまたは偽陰性という.

13.1.2 ランダム化検定

検定の構成法の例をあげる. 検定 = 棄却域の良さは, 第一種の誤りの確率を有意水準 α で制限した後に, その範囲内で第二種の誤りの確率を最小化することで得られる.

例 13.1.4. 例 13.1.2 について, $x_0 = 15$ とすると,

$$\sum_{k=15}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20} \approx 0.0207$$

より, 棄却域のサイズは 5% を大きく切って 2% 程度となってしまう. そこで, $14 < x_0 < 15$ にあたる棄却域を構成したいが, $x_0 \in \mathbb{Z}$ の値は変えられない. そこで, 棄却域をランダム化する.

$$\binom{20}{k} \left(\frac{1}{2}\right)^{20} \varphi + \sum_{k=15}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20} = \alpha = 0.05$$

を満たす $\varphi \in (0, 1)$ を用いて, 観測値 $x = 14$ に対しては確率 φ で H_0 を棄却し, 確率 $1 - \varphi$ で H_0 を採択することとする.

13.1.3 形式化

定義 13.1.5 (test / critical (function), size, level- α -test, power function, uniformly most powerful (UMP) test). 確率分布族 $\mathcal{P} = (P_\theta)_{\theta \in \Theta} \subset \mathcal{M}(\mathcal{X}, A)$ を考える.

- (1) 棄却域 $\mathcal{X}_1 \subset \mathcal{X}$ の定めるパラメータ空間の分割 $\Theta = \Theta_0 + \Theta_1$ を考え, 帰無仮説を $H_0 : \theta \in \Theta_0$ で表し, 対立仮説を $H_1 : \theta \in \Theta_1$ と表す.
- (2) $|\Theta_i| = 1$ の場合を単純仮説といい, そうでない場合を複合仮説という.^{†1}
- (3) 可測関数 $\varphi : \mathcal{X} \rightarrow [0, 1]$ を検定 (関数) という. $\exists A \in \mathcal{A} \varphi = 1_A$ であるとき検定 φ は非確率的である.
- (4) $\sup_{\theta \in \Theta_0} E_\theta[\varphi]$ を φ の大きさという. 大きさが α 以下の検定を水準 α 検定という. 水準 α 検定の全体を $\Phi_\alpha \subset \text{Meas}(\mathcal{X}, [0, 1])$ で表す.
- (5) $\beta_\varphi : \Theta_1 \rightarrow [0, 1]; \theta \mapsto E_\theta[\varphi]$ を検定出力関数という. $1 - \beta_\varphi(\theta)$ は第二種の誤りの確率を表す.
- (6) $\forall \varphi \in \Phi_\alpha \beta_{\varphi_0}(\theta) \geq \beta_\varphi(\theta)$ を満たす検定 $\varphi_0 \in \Phi_\alpha$ を一様最強力検定という.

要諦 13.1.6 (統計的決定理論の枠組みでの解釈). 決定空間は, 採択する仮説の添字からなる空間 $\mathcal{D} := 2 = \{0, 1\}$ である (0 が受容, 1 が棄却). 損失関数は

$$W(\theta, a) = \begin{cases} 1_{\{1\}}(a), & \theta \in \Theta_0, \\ 1_{\{0\}}(a), & \theta \in \Theta_1. \end{cases}$$

で, 検定関数 φ は確率的決定関数

$$\delta_\varphi(dz|x) = \varphi(x)\epsilon_1(dz) + (1 - \varphi(x))\epsilon_0(dz)$$

に対応する. ただし, ϵ_a はデルタ測度とした. したがって危険関数は

$$\begin{aligned} R(\theta, \delta_\varphi) &= \int_{\mathcal{X}} \int_{\mathcal{D}} W(\theta, z) \delta_\varphi(dz|x) \epsilon_0(dz) \\ &= \begin{cases} E_\theta[\varphi], & \theta \in \Theta_0, \\ 1 - E_\theta[\varphi], & \theta \in \Theta_1. \end{cases} \end{aligned}$$

となり, $\Phi_\alpha = \left\{ \varphi \in \text{Meas}(\mathcal{X}, [0, 1]) \mid \sup_{\theta \in \Theta_0} R(\theta, \delta_\varphi) \leq \alpha \right\}$ と表せる. 決定関数の全体は $\Delta \subset \Phi_\alpha$ となる.

Δ の中に入る一番自然な順序は一様順序で, これによる最大元が存在するときこれを UMP と呼ぶ.

定義 13.1.7 (nuisance parameter, goodness of fit test).

- (1) Θ が 2 次元以上で, 特定のパラメータにしか興味がないとき, 分割 $\Theta_1 + \Theta_2$ は商空間となる. この時の興味のない母数を局外母数または攪乱母数という.
- (2) 帰無仮説を「確率変数が正規分布に従う」という命題にする場合, 特に分布の適合度検定と呼ぶ. 適合度検定や「確率分布が独立である」などの数学的な仮定が現実のデータと矛盾がないかチェックするための検定のクラスを統計的モデルの診断という.

^{†1} Θ が 1 次元で, $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$ の形のときを両側検定, $H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_0$ の形のときを片側検定という.

第 14 章

大標本理論

非線形モデル・非ガウスモデルに対しては，統計量の分布が複雑なため，標準的な統計的決定理論を適用するのは困難であるが，これに替わって漸近的方法に基づく大標本理論が有力な解析手段となる．ここで，正規近似に基づく 1 次の漸近理論を扱う．

第 15 章

漸近展開とその応用

中心極限定理は分布の正規近似であるが，Edgeworth による展開はその近似をより精密にしたものである．標本数がそれほど大きくない実際の状況では，分布のより精密な近似を基礎として統計量を構成することが必要になる．

今日，漸近展開法も確率過程にまでその領域を広げ，新しい確率統計学が発展しつつあるが，このような新領域を理解する上でも，独立確率変数列における現象を理解することが重要である．

第 16 章

Martingale

16.1 Introduction

回帰モデル $X_i = f(X_{i-1}, \dots, X_{i-p}) + \epsilon_i$ において，サンプリングが均等でないときなど， ϵ_i は何か連続的な確率過程を積分して定まる，と考えると数理モデルとして非常に自然である．連続関数 f について，

$$Y_t = Y_0 + \int_0^t f(Y_s) ds + W_t$$

とし， $W_{t_i} - W_{t_{i-1}} \sim \mathbb{N}(0, t_i - t_{i-1})$ を標準 Weiner 過程とする．

このようなモデルのうち，特に株価の対数を Y_t とおいたときに使われるパラメトリックモデルに，Vasicek 過程

$$Y_t = Y_0 - \int_0^t \alpha_1(Y_s - \alpha_2) ds + \beta W_t$$

などがあり，離散的観測 $\{Y_{t_0}, \dots, Y_{t_n}\}$ に基づいて未知パラメータ $\alpha_1, \alpha_2, \beta$ の推定を考える．このときにマルチンゲール理論が使える．

その理由は，martingale というクラスの形式的定義が，自然に統計モデルの「ノイズの直交性」の拡張となっていると考えられるためである．これは独立性の仮定による代数規則 $E[\epsilon_i \epsilon_j] = 0$ の抽出となっているのである．

大きな応用分野として生存解析における censored data^{†1}の解析がある．このとき， N_t を死亡数， Y_t を censor されずに残っている観測対象数，癌の再発時刻の分布関数を F ，密度関数を f とすると，

$$N_t - \int_0^t \alpha(s) Y_s ds \quad \alpha(t) = \frac{f(t)}{1 - F(t)}$$

はマルチンゲールになる． α はハザード関数といい，患者が時刻 t で生存しているという条件の下，その時間に死亡する条件付き確率となる．このマルチンゲールの期待値は常に 0 だから， N_t の不偏推定量が見つかったことになる．なお，

$$\int_0^t \frac{1}{Y_s} (dM_s - \alpha(s) Y_s ds)$$

もマルチンゲールとなることがわかる．

16.2 semimartingale

^{†1} 消息不明になる瞬間があること．癌の再発データにおいて，他の原因による死亡など．

参考文献

- [1] Vladimir Voevodsky "Notes on categorical probability"
- [2] 伊藤清 『確率論』
- [3] Kolmogorov 『確率論の基礎概念』
- [4] 吉田朋広 『数理統計学』(朝倉書店, 2006)
- [5] 竹村彰道 『現代数理統計学』(学術図書, 2020)
- [6] 久保川達也 『現代数理統計学の基礎』(共立出版, 2017)
- [7] 西山陽一 『マルチンゲール理論による統計解析』(近代科学社, 2011)
- [8] 丸山徹 - 確率測度の*弱収束 <https://core.ac.uk/download/pdf/145720102.pdf>