目次

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・不確定性に関係する概念のうち，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 $X:\Omega\to\mathbb{R}^n$ 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，系で考えう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．

1.1.1　Kolmogorov の公理

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 $A + B, A \coprod B$ で表し直和という．[4] 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 $P : \mathcal{F} \to [0, 1]$ 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 $(A_i) : \mathbb{N} \to \mathcal{F}$ 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 $(\Omega, \mathcal{F})$，$(\mathcal{X}, \mathcal{B})$ を可測空間とする．

(1) $\mathcal{X}$-**値確率変数**とは，可測関数 $X : \Omega \to \mathcal{X}$ をいう．

(2) これに沿って引き起こされる像測度 $X_* P = P^X : P(\mathcal{X}) \supset \mathcal{B} \to [0, 1]$ を $P^X(B) := P(X^{-1}(B))$ で定める：[†6]

$$\begin{tikzcd}
 & {[0,1]} & \\
P(\Omega) \arrow[ur, "P"] & & P(\mathcal{X}) \arrow[ul, "P^X"'] \arrow[ll, "X^*"'] \\
\Omega \arrow[u, "P"] \arrow[rr, "X"'] & & \mathcal{X} \arrow[u, "P"']
\end{tikzcd}$$

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 $\mathrm{id}_\Omega : \Omega \to \Omega$ についての確率分布 P^{id_Ω} である．

(4) 像 $\mathrm{Im}\, X$ 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．

[†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．

定義 1.1.7 (joint distribution, marginal distribution).

(1) 確率変数の列 $X_1, \cdots, X_n$ に対して，積写像 $(X_1, \cdots, X_n)$ 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 $X : \Omega \to \mathbb{R}$ 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) $L^1(\Omega, \mu) := \left\{ f \in \mathrm{Map}(\Omega, \overline{\mathbb{R}}) \mid \int_\Omega |f(\omega)|\mu(d\omega) < \infty. \right\}$ 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．

補題 1.1.10. 関数の族 $\{X_n\} \subset \mathrm{Map}(\Omega, \mathbb{R})$ について，次の 2 条件は同値.

(1) (X_n) は確率過程である．
(2) $X : \mathbb{N} \times \Omega \to \mathbb{R}$ は $\mathcal{B}(\mathbb{N}) \times \mathcal{F}/\mathcal{B}(\mathbb{R})$-可測である．

1.1.2　σ-加法性

事象とその確率なる概念が数学では σ-加法性に当たることを考察するのにもっとも直感的な例として，コイントスを繰り返す行為を一つの空間内で表現したい場合，部分集合の族は自然 σ-代数をなす．

例 1.1.11 (確率空間を取り直すのではなく, σ-加法族を発展させる). 素朴に, n 回試行を行ったときの標本空間は $\Omega_n := \mathrm{Map}([n], 2)$ であり，事象のなす σ-加法族は $P(\Omega_n)$ である．

一方で，$\Omega := \mathrm{Map}(\mathbb{N}, 2)$ を全体集合とすると，n 回試行を行ったときの事象のなす部分集合族は，$\omega^n \in \Omega_n$ に対して，$A(\omega^n) := \{\omega \in \Omega \mid \forall_{i \in [n]}\ \omega_i = \omega_i^n\}$ とすると，対応 $A_n : P(\Omega_n) \to P(\Omega); B \mapsto A(B) := \cup_{\omega^n \in B} A(\omega^n)$ を用いて，$\mathcal{F}_n := \mathrm{Im}\, A_n \subset P(\Omega)$ は，σ-加法族の構造を持つ．逆に，それ以外の構造は見出し難い．

定義 1.1.12 (cylinder set). この枠組みを一般化する．可測空間 $(S, \mathcal{S})$ に値を取る試行を繰り返す試行の標本空間を $\Omega := S^{\mathbb{N}}$ とする．その部分空間 $C \subset \Omega$ であって，ある $n \in \mathbb{N}$ を用いて

$$C = C(t_1, \cdots, t_n; A_1, \cdots, A_n) := \{\omega \in \Omega \mid \forall_{i \in [n]}\ \omega_{t_i} \in A_i\} \qquad \text{ただし}, t_i \in \mathbb{N}, A_i \in \mathcal{S}$$

と表せるとき，これを**柱状集合**という．これは第 t_i 回の試行の結果を指定して得られる事象である．Ω 上の σ-代数としては，柱状集合をすべて含むものを取る．

1.2　有界測度論の確率論的解釈

有界測度論

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）．そこで，有界測度論の枠組みを確率論的に解釈する方法を以下に示す．

[†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．

1.2.1　極限の定義

事象列には極限なる演算を定義したい

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，$(\mathcal{F}, \triangle)$ 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 $(A_n) : \mathbb{N} \to \mathcal{F}$ に対して，次の事象が定義できる．

(1) $\limsup_{n\to\infty} A_n := \cap_{m=1}^{\infty} \cup_{m\ge n} A_m$ 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) $\liminf_{n\to\infty} A_n := \cup_{m=1}^{\infty} \cap_{m\ge n} A_m$ 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，$\lim_{n\to\infty} A_n$ 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) $\limsup_{n\to\infty} A_n = \{x \in X \mid \#\{n \in \mathbb{N} \mid x \in A_n\} = \infty\}$.
(2) $\liminf_{n\to\infty} A_n = \{x \in X \mid \#\{n \in \mathbb{N} \mid x \notin A_n\} < \infty\}$.

[証明] **.**

(1)

$$\begin{aligned} x \text{ が} \#\{n \in \mathbb{N} \mid x \in A_n\} = \infty \text{ を満たす} &\Leftrightarrow \{n \in \mathbb{N} \mid x \in A_n\} \text{ は非有界である} \\ &\Leftrightarrow \forall_{n\in\mathbb{N}} \exists_{m\ge n}\ x \in A_m \Leftrightarrow x \in \cap_{n\in\mathbb{N}} \cup_{m\ge n} A_m. \end{aligned}$$

(2)

$$\begin{aligned} x \text{ が} \#\{n \in \mathbb{N} \mid x \notin A_n\} < \infty \text{ を満たす} &\Leftrightarrow \{n \in \mathbb{N} \mid x \notin A_n\} \text{ は有界である} \\ &\Leftrightarrow \exists_{n\in\mathbb{N}} \forall_{m\ge n}\ x \in A_m \Leftrightarrow x \in \cup_{n\in\mathbb{N}} \cap_{m\ge n} A_m. \end{aligned}$$

■

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 $(X, \mathcal{F}, \mu)$ について，

(1) 対称差について，$d(A, B) := A \triangle B$ $(A, B \in \mathcal{F})$ 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，$(\mathcal{F}, \triangle)$ 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，$\mathrm{Im}\, d \subset [0, \infty)$ 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，

$$A \triangle C = (A \setminus C) \setminus B \sqcup (A \setminus C) \cap B \sqcup (C \setminus A) \setminus B \sqcup (C \setminus A) \cap B$$

と直和分解できるが，それぞれについて，

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

より，$\mu(A \triangle C) \le \mu(A \triangle B) + \mu(B \triangle C)$.
(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 $\mu : \mathcal{F} \to [0, \infty)$ 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 $L^p(X) \to \mathbb{R}$ となる．

1.2.2　連続写像としての測度

確率の素朴な加法性に対して，その σ-加法性も考えたいとは，数学的には連続性を考えたいということである．$\mathcal{F}$ 内に定義した極限の構造に対して，これを保存することをいう．
測度 $\mu:\mathcal{F}\to[0,1]$ は，$\mathcal{F}$ 上の擬距離 $d(A,B):=\mu(A\triangle B)$ 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)**.** 可測空間 $(\Omega,\mathcal{F})$ の可測集合の列 $(A_i):\mathbb{N}\to\mathcal{F}$ について，次が成り立つ．

(1) 次の 4 条件は同値.
   (a) $\mu:\mathcal{F}\to[0,1]$ は測度である．
   (b) 単調増加列 (A_n) に対して，$\lim_{n\to\infty}\mu(A_n)=\mu(\cup_{n=1}^\infty A_n)$.
   (c) 単調減少列 (A_n) に対して，$\lim_{n\to\infty}\mu(A_n)=\mu(\cap_{n=1}^\infty A_n)$.
   (d) $\emptyset$ に収束する単調減少列 (A_n) に対して，$\lim_{n\to\infty}P(A_n)=0$.
(2) (subadditivity) 測度 μ に対して，$\mu(\cup_{i=1}^n A_i)\le\sum_{i=1}^n\mu(A_i)\ (n\in\mathbb{N})$.

［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，$\cup_{n\in\mathbb{N}}A_n=\cup_{n\in\mathbb{N}}B_n$．よって，

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

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

$$\begin{aligned}P\left(\cap_{n=1}^\infty A_n\right)&=1-P\left(\cup_{n=1}^\infty A_n^{\complement}\right)\\&=1-\lim_{n\to\infty}P(A_n^{\complement})=1-\left(1-\lim_{n\to\infty}P(A_n)\right)=\lim_{n\to\infty}P(A_n).\end{aligned}$$

(3) $B_n:=\cup_{k=1}^n A_k$ 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，

$$P(\cup_{n=1}^\infty A_n)=\lim_{n\to\infty}P(B_n)\le\lim_{n\to\infty}\sum_{k=1}^n P(A_k).$$

■

要諦 1.2.6 (連続写像としての測度)**.** 確率測度 $\mu:\mathcal{F}\to[0,1]$ は，の射を定める．一方で，一般の測度 $\mu:\mathcal{F}\to[0,\infty)$ 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，$A_i:=2^n\mathbb{N}$ とすることで，$\lim_{n\to\infty}\mu(A_n)=\infty,\mu(\cap_{n\in\mathbb{N}}A_n)=0$ 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　Borel-Cantelli

有界測度論の初等的な結論に，確率論的な解釈を与えることが出来る．

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

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

(1) $\mathcal{F}$ の列 (A_n) について，$\mu\left(\liminf_{n\to\infty}A_n\right)\le\liminf_{n\to\infty}\mu(A_n)$.

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

定理 1.2.9 (Borel-Cantelli lemma)**.**

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

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

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

［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 $\mu(\cap_{m=1}^{\infty}B_m)=\lim_{m\to\infty}\mu(B_m)$．和が収束する列は 0 に収束する $(\displaystyle\lim_{m\to\infty}\sum_{n=m}^{\infty}\mu(A_n)=0)$ ことに注意して，

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

また，de Morgan の法則より，

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

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

$$\begin{aligned}1-P(\cup_{k\ge n}A_k)&=P(\cap_{k\ge n}A_k^{\complement})\\&\le P(\cap_{k\ge n}^{p}A_k^{\complement}) &&\cap_{k\ge n}A_k^{\complement}\subset\cap_{k\ge n}^{p}A_k^{\complement}\\&=\prod_{k=n}^{p}P(A_k^{\complement}) &&\text{独立性}\\&=\prod_{k=n}^{p}(1-P(A_k))\le\exp\left(-\sum_{k=n}^{p}P(A_n)\right)\xrightarrow{p\to\infty}e^{-\infty}=0. &&1-x\le e^{-x}\end{aligned}$$

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．

---

†8 確率測度ならば $\mu(X)=1$ である

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，$P\left(\limsup_{n\to\infty} A_n\right) = 0, \lim_{n\to\infty} P(A_n) = 0$ が成り立つ．

(1) $\liminf_{n\to\infty} P(A_n) = 0$.

(2) (a) $\sum_{n\in\mathbb{N}} P(A_n^\complement \cap A_{n+1}) < \infty$.

(b) $\sum_{n\in\mathbb{N}} P(A_n \cap A_{n+1}^\complement) < \infty$.

1.2.4　Hewitt-Savage

定義 1.2.13. $X_1, X_2, \cdots \in \mathrm{Meas}(\Omega, \mathbb{R})$ を確率変数列とする．事象 $\{\omega \in \Omega \mid \forall_{i\in\mathbb{N}}\ X_i(\omega) \in A\} \in \mathcal{F}$ が**交換可能**であるとは，有限個の入れ替えについて

$$\forall_{n\in\mathbb{N}}\ \forall_{\sigma\in S_n}\ \ \{X_1, X_2, \cdots \in A\} \subset \{X_{\sigma_1}, X_{\sigma_2} \cdots \in A\}$$

が成り立つことをいう．

補題 1.2.14. $E \in \mathcal{F}$ が i.i.d. 列 (X_n) に関する交換可能な事象であるとき，$P[E] \in \{0, 1\}$ である．

1.2.5　Bonferroni

有界測度論としては初等的でも，確率論的には非自明で名前がほしい結果は多い

命題 1.2.15 (Bonferroni の不等式)**.**

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

等号成立条件は，$A_1^\complement, \cdots, A_n^\complement$ が背反のとき．

[証明] ．$A_1^\complement, \cdots, A_n^\complement$ について，劣加法性より． ■

1.3　確率変数論

有界測度空間への米田埋め込みが，確率論の世界への跳躍である．確率変数とは，$(\mathbb{R}, \mathcal{B}_1(\mathbb{R}))$ への表現に他ならない．この観点から，合成 $g \circ X_i$ を $g(X_i)$ などと書く上に，始域 Ω は自由に取って考察することとなる．
また，測度を無理やり可測関数の押し出しと見る場面は少ないかもしれないが，確率論においては，たしかに確率変数は確率分布を一般化する．

命題 1.3.1 (確率変数の構成)**.** 確率変数 $X_1, \cdots, X_n$ と可測関数 $g : \mathbb{R}^n \to \mathbb{R}$ に対して，$Y := g(X_1, \cdots, X_n)$ は再び可測関数である．

1.4　分布関数論

$\mathbb{R}$ 上の Radon 積分には，Lebesgue-Stieltjes 積分という名のついた古典的構成法があるのであった．これにより，分布関数とこれが定める積分（すなわち測度）とは一対一対応する．

定義 1.4.1 (distribution function)**.** 実確率変数 X の定める $(\mathbb{R}, \mathcal{B}_1(\mathbb{R}))$ 上の確率測度 μ について，関数 $F : \mathbb{R} \to [0, 1]; x \mapsto \mu((-\infty, x])$ を，X または μ の**分布関数**という．

注 1.4.2. $F(x) := \mu((-\infty, x))$ と定めると，左半連続なバージョンで同様の議論が展開可能である．

補題 1.4.3 (分布関数の特徴付け)**.** $(\mathbb{R}, \mathcal{B}_1(\mathbb{R}))$ 上の確率測度 μ の分布関数を F とする．

(1) F は広義単調増加である．
(2) $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to\infty} F(x) = 0$ をみたす有界関数である．
(3) 右半連続である．
(4) $F(x-) := \lim_{y\nearrow x} F(y)$ と表すと，$\mu(\{x\}) = F(x) - F(x-)$ を満たす．

要諦 1.4.4. 右半連続関数 $F : \mathbb{R} \to [0,1]$ に対して，Stieltjes 積分 $\int (-)dF : C_c(\mathbb{R}) \to \mathbb{R}$，またその延長として Lebesgue-Stieltjes 積分 $\int (-)dF : \mathcal{L}^1(\mathbb{R}) \to \mathbb{R}$ が定まる．すなわち，確率分布 μ が一意に定まる．逆に，$\mathbb{R}$ 上の任意の Radon 積分に対して，これに等しい分布関数とそれが定める Stieltjes 積分が存在するから，確率分布と分布関数は一対一対応する．

定義 1.4.5 (discrete, absolute continuous, density, singular)**.** 分布関数 $F : \mathbb{R} \to \mathbb{R}$ について，

(1) F が**不連続**であるとは，Lebesgue-Stieltjes 測度 dF がデルタ測度の可算和も許した凸結合で表せるときをいう．
(2) F が**絶対連続**であるとは，Lebesgue-Stieltjes 測度 dF が，Lebesgue 測度 dx に対して絶対連続であることをいう：$dF \ll dx$．このとき，F は密度関数 p を定める：$F(x) = \int_{-\infty}^{x} p(y)dy$．
(3) F が**特異**であるとは，F は関数として連続であるが，Lebesgue-Stieltjes 測度 dF が Lebesgue 測度 dx と互いに特異であるときをいう：$\exists_{A\in\mathcal{B}_1(\mathbb{R})}\ dF(A) = 1 \land m(A) = 0$．

注 1.4.6. F が連続であることと，これが定める Radon 積分が連続であること $\forall_{x\in X} \int [\{x\}] = 0$ は一般に同値になる．

例 1.4.7. Cantor 関数などは，特異型分布関数の例である．

定理 1.4.8 (Lebesgue decomposition)**.** 任意の分布関数 $F : \mathbb{R} \to \mathbb{R}$ は，不連続分布関数 F_1，絶対連続分布関数 F_2，特異分布関数 F_3 が一意的に存在して，これらの線型結合で表せる．

1.5　確率測度の Fourier 逆変換

特性関数の収束は確率分布の弱収束に対応する．これを用いて中心極限定理が証明される．

定義 1.5.1. 確率測度 $\mu \in P(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ の**特性関数** $\varphi : \mathbb{R} \to \mathbb{R}$ とは，

$$\varphi(\xi) = \varphi_\mu(\xi) := \int_{\mathbb{R}} e^{i\xi x}\mu(dx)$$

をいう．確率変数の特性関数は，これが定める分布の特性関数とする：$\varphi(\xi) = \varphi_X(\xi) := E[e^{i\xi X}]$.

定理 1.5.2 (一意性定理)**.** 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．すなわち，任意の $\mu_1, \mu_2 \in \mathcal{P}^d$ について，次は同値．

(1) $\mu_1 = \mu_2$.
(2) $\varphi_{\mu_1} = \varphi_{\mu_2}$.

[証明] ．反転公式 9.3.2 により，任意の開区間の測度は φ が一意に定める．■

1.6　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

1.6.1　Meas の特徴

定義 1.6.1 (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 $U : C \to D$ 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 $f_i : X \to U(S_i) \in D$ に対して，initial lift $(T, m_i : T \to S_i) \in C$ 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：

$$\forall_{T' \in C}\ \exists_{g' : U(T') \to X}\ \exists_{m_i' : T' \to S_i}\ g' \circ f_i = U(m_i') \Rightarrow \exists!_{n : T' \to T}\ U(n) = g' \wedge n \circ m_i = m_i'.$$

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

例 1.6.2 (Meas)**.** $D = \mathrm{Alg}_\sigma \subset \mathrm{Set}$ とすると，$C = \mathrm{Meas}$ である．

定義 1.6.3 (Borel 可測空間)**.** 関手 $\mathcal{B} : \mathrm{Top} \to \mathrm{Meas}$ に対して，$\mathcal{B}(S)$ を Borel σ-加法族という．

1.6.2　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 1.6.4 (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 1.6.5 (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 1.6.6 (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 $(X_n)_{n \in N}$ 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 $B \to C$ 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 $\oplus_{i \in F} X_i \to \oplus_{j \in S} X_j$ の形をした射影のなす束からの関手 $B \to C$ 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 $B \to C$ 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．

[†9] 全て $D = \mathrm{Set}$ の例だが，$D = \mathrm{Grp}$ として，位相群などを考えても良い．
[†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 $\otimes : C \times C \to C$ を備える．

1.7　可測関数の空間上の線型汎関数としての測度

統計的問題では線型汎関数の推定が主眼となる所以である

標本空間 $\mathcal{X}^n$ 上の関数 $\mathcal{X}^n \to \mathbb{R}$ 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．
したがって，平均を推定する問題は，線型汎関数を推定する問題と同値である．

[a] まるで Ergode 理論のような理論展開である．

1.7.1　確率変数に対する線型作用素

確率変数なる枠組みを採用すると，確率測度は $\mathrm{Meas}(X, \mathbb{R})$ 上の線型作用素を定め，これを E で表す．ここから派生した作用素が重要な特徴量を定める．

定義 1.7.1 ((population) mean / expectation, (population) variance)**.** 線型作用素 $\mathrm{Meas}(X, \mathbb{R}) \to \mathbb{R}$ を定める．一般に $\mathrm{Meas}(X, V) \to V$ 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) $\mathrm{Var}[X] = V[X] := E[(X - EX)^2]$ が実数であるとき，これを**分散**という．
(4) $\mathrm{Cov}(X, Y) = V(X, Y) := E[(X - EX)(Y - EY)] = E[XY] - E[X]E[Y]$ を**共分散** $\mathrm{Meas}(X, \mathbb{R}) \times \mathrm{Meas}(X, \mathbb{R}) \to \mathbb{R}$ という．
(5) $\sigma(X) = \sqrt{V(X)}$ を**標準偏差**という．
(6) $R(X, Y) := \dfrac{V(X, Y)}{\sigma(X)\sigma(Y)}$ を**相関係数**という．

1.7.2　他の測度に関する線型作用素

定義 1.7.2 (entropy)**.** Shanonn 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 $H : \mathrm{Meas}(\mathcal{X}, \mathbb{R}) \to [0, \log \alpha]$ である．

1.7.3　統計的推論

統計的推論において，確率論の枠組みをどのように用いるかを比較対象する．

定義 1.7.3 (frequentist probability)**.** 現実の事象の確率とは，仮想的に試行を繰り返したときの，相対頻度の極限として得られる極限分布として想定し，我々が眼前にする現象はこれの実現値であるという仮定をおいて行われる統計的推論である．
　想定した真の確率分布の特性量の計算や，真の確率分布に対する標本からの推定などが含まれる．

例 1.7.4 (頻度主義的推論)**.** 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．この仮定が数理的に妥当であることは，確率論の結果による．

[†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．

1.7.4　積分作用素に関する確率不等式

積分作用素に関する測度論的結果に，確率論的な解釈を与える．

定理 1.7.5 (Chebyshev の不等式)．$\psi:\mathbb{R}\to\mathbb{R}_{\ge 0}$ を非負な Borel 可測な関数とする $\psi\ge 0$.

$$\forall_{A\in\mathcal{B}(\mathbb{R})}\quad P[X\in A]\le\frac{E[\psi(X)]}{\inf_{x\in A}\psi(x)}.$$

[証明]．

$$\forall_{x\in\mathbb{R}}\quad\left(\inf_{x\in A}\psi(x)\right)1_{\{x\in A\}}\le\psi(x)$$

が成り立つ．両辺の期待値を取れば良い．■

系 1.7.6.

(1) 特に $\psi(x)=|x|^p\ (p>0), A:=\{|x|\ge\epsilon\}$ とすると，$P[|X|\ge\epsilon]\le\dfrac{1}{\epsilon^p}E[|X|^p]$.

(2) $\forall_{a>0}\ P\{|X(\omega)-EX|>a\sigma(X)\}\le\dfrac{1}{a^2}$.

(3) $f:\mathbb{R}\to\mathbb{R}_{\ge 0}$ 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．$\forall_{a\in\mathbb{R}}\ f(a)>0\Rightarrow P(X\ge a)\le\dfrac{E[f(X)]}{f(a)}$.

[証明]．

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

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

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 $\displaystyle\int_{\{X<a\}}f(X)dP\ge 0$ であることより，

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

■

要諦 1.7.7 (平均値から標準偏差の a 倍以上離れる確率は $\dfrac{1}{a^2}$ 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．

補題 1.7.8. 凸関数 $\psi:\mathbb{R}\to\mathbb{R}$ について，$X,\psi(X)\in\mathcal{L}^1(X)$ のとき，

$$\psi(E[X])\le E[\psi(X)].$$

補題 1.7.9. 共役指数 $p,q\in[1,\infty]$ について，$\|fg\|_1\le\|f\|_p\|g\|_q$．特に $p=q=2$ のとき $E[XY]\le\sqrt{E[X^2]E[Y^2]}$ で，さらに $Y=1$ のとき $(E[X])^2\le E[X^2]$.

1.7.5　Lebesgue 空間の包含関係

定理 1.7.10. 確率空間において，$p<q\in[1,\infty]$ のとき，$\mathcal{L}^q(X)\subset\mathcal{L}^p(X)$ かつ $\|-\|_p\le\|-\|_q$ である．

1.8　σ-加法族が定める構造（要再考）

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 1.8.1 (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，$(X^2)^*(\mathcal{B}) \subsetneq X^*(\mathcal{B}) = (X^3)^*(\mathcal{B})$.[†12] したがって，外から観測できる範囲は $(\Omega, \mathcal{F}, P)$ 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らけである．

定義 1.8.2 (関数が生成する σ-加法族)**.** 関数 $X_1, \cdots, X_n : \Omega \to \mathcal{X}$ 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 $\sigma(X_1, \cdots, X_n)$ で表す．

定義 1.8.3 (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 $\widetilde{P}$ は一般に複数存在するが，$\widetilde{P}\{(\omega_1, \omega_2)\} = P_1\{\omega_1\}P_2\{\omega_2\}$ 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 1.8.4 (surprisal / self-information / information content, expected surprisal, almost partiton, entropy)**.** $(X, \mathcal{M}, \mu)$ 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 $\sigma_\mu : \mathcal{M} \to [0, \infty]$ 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，

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

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 1.8.5. 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．また，

[†12] この考え方を確率変数が定める filtration $\sigma[X_0, \cdots, X_n]$ 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 $\mathcal{X} \to \mathbb{R}$ 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．
[†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．

log の底は分野によってさまざまである．Shanonn の entropy では 2,

第 2 章

有界測度論

確率分布を測度として定式化したが，まずは，有界な測度（を正規化したもの）としての純粋な性質を調べる．有界な測度は直ちに局所有限でかつ内部正則であるので，Radon 測度である．そして Radon 測度が有限であるとき，（外部）正則な Borel 測度である．

(1) 測度の収束には種々の位相が考えられ，修羅の相を呈している．ここで重要になる有界測度特有の概念に緊密性があり，有界測度の列が緊密であるとき，３つの関数のクラス $C_c(X), C_0(X), C_b(X)$ が定める収束概念は一致する．その上，緊密な確率測度列の収束先は再び確率測度となる．

(2) 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．

(3) 次に，確率変数の標準的な構成法を測度論の形で与える．実際に確率測度なる対象が存在することの保証も数学的には肝要である．

2.1 確率測度の収束

Banach 空間論の発展を後押しした消息

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．

測度の弱収束の理論は，A. D. Alexandorff と Prohorov による．As pointed out by A.D. Alexandro ff himself, a source of his abstract work in general measure theory was his research on the geometry of convex bodies.

凸解析と関数解析が肝要となる．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！？

2.1.1 汎関数としての測度

測度の双対空間と前双対空間

測度の空間の位相は，双対空間から入れるから，測度の空間の双対空間と predual とを考察しておく必要がある．

この消息ははじめ Radon 1913[a]によって研究され，その位相を「弱収束」と呼んだ．1907 年に Riesz の表現定理が独立に証明され，Hilbert 空間の研究が進んでいた頃からの用語で，Banach が「弱収束」と「$*$-弱収束」を一般のノルム空間上に定義した 1929 年よりも前の用語である．

[a] Radon, J. Theorie und Anwendungen der absolut additiven Mengenfunktionen. Sitz. Akad. Wiss. Wien, Math.-naturwiss. Kl. IIa. 1913. B. 122. S. 1295 – 1438. [245, 246]

記法 2.1.1. 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で有限な符号付測度の空間を $M_f(X)$，$\mathcal{B}$ 上の σ-加法的な正則で有限な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，有限な加法的集合関数は有界変動だから，そのままノルム空間となる．明らかに，$M_f(X) \subset C_b(X)^*$ である．

命題 2.1.2 (Alexandrov (1940)). 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，$\forall_{f\in C_b(X)}\ \Lambda(f) = \int_X f d\mu$ を満たし，$\|\Lambda\| = \|\mu\|$ を満たす．すなわち，Banach 空間として等長同型である $C_b(X)^* \simeq_{\mathrm{Ban}} M_f(X)$.

命題 2.1.3 (Riesz-Markov-Kakutani representation theorem). 局所コンパクトハウスドルフ空間 X 上の有限 Radon 電荷の空間 $M(X)$ にノルム $\|\Phi\| = |\Phi|(1)$ を考えたものは，Banach 空間 $(C_0(X))^*$ に等長同型である．

命題 2.1.4 (definition of Radon charge). 局所コンパクト空間 X 上の空間 $C_c(X)$ の双対空間 $(C_c(X))^*$ の正部分は，Radon 電荷の空間 $M(X)$ に同型である：$(C_c(X))^*_+ \simeq M(X)$.

注 2.1.5. あるいは，$C_c(X)^*$ の元は局所有限 $\forall_{K\mathrm{cpt}X}\ \mu(K) < \infty$ を満たす正則測度になるという定理とも見れる．

2.1.2　測度の収束の定義と特徴付け

測度の空間は Banach 空間であるが，それ以上に具体的対象である．

一般に測度の空間は Banach 空間となり，これの Banach 空間としての弱収束と，「測度の弱収束」は**相違する**．そこで，Banach 空間としての弱収束と同じ文脈で議論したい場合，有界測度論では Bourbaki の言葉を借りて「**狭収束**」と呼ばれる．しかし X が局所コンパクト空間であるとき，測度の空間に対して前節で議論したような表現定理が成り立つ．すると，測度という概念の性質上，双対空間の部分空間のクラス $C_c(X) \subset C_0(X) \subset C_b(X) \subset M(X)^*$ に応じて，特別な位相が入るのである．$C_b(X)$ が定める強い位相を**弱収束**といい，$C_c(X), C_0(X)$ が定める弱い位相を混同して**漠収束**という．３つとも言うなれば Banach 空間の w^*-位相であり，この表現では区別が出来ない．それぞれ $\sigma(M(X), C_b(X)), \sigma(M(X), C_c(X)), \sigma(M(X), C_0(X))$-位相と言うよりほかはない．ここでは，$\sigma(M(X), C_c(X))$-位相を漠位相ということとする．

定義 2.1.6 (strong, weak, narrow, wide / vague / weak star). 可測空間 $(X, \mathcal{B})$ 上の有界変動を持つ符号付き Radon 測度の空間 $M_b(X)$[†1] は，全変動 $|\mu|$ が定めるノルムについて Banach 空間をなす．他の位相で代表的なものは以下の通りである．[†2]

(1) まず各可測集合上での収束 (setwise convergence) が考えられる．
(2) ノルム収束は，変動収束 (convergence in variation) または強収束とも言う．
(3) 任意の有界線型汎関数 $F \in M_b(X)^*$ について $F(\mu_n) \to F(\mu)$ を満たすとき，**弱収束**という．
(4) X が位相空間で，$\mathcal{B}$ が Borel σ-代数であるとき，任意の有界連続関数 $f \in C_b(X)$ について $\mu_n f \to \mu f$ を満たすとき，**狭収束** (narrow topology) という．[†3] X が局所コンパクトハウスドルフ空間であるとき，μ_n が μ に狭収束するならば，$\forall_{A\in\mathcal{B}(X)}\ |\mu|(\partial A) = 0$ が必要であり，これが狭収束を特徴付ける 2.1.14
(5) X が局所コンパクトハウスドルフ空間で，$\mathcal{B}$ が Borel σ-代数であるとき，更に別の位相が定まる．任意のコンパクト台を持つ連続関数 $f \in C_c(X)$ について $\mu_n f \to \mu f$ を満たすとき，**広収束** (wide topology) または w^*-**収束**という．一般に狭位相より弱いが，X が Hausdorff のとき，確率測度の空間 $P(X)$ には同じ位相を定める．なお，$P(X)$ は広位相についてコンパクトとは限らない，すなわち収束先は $\mu(X) = 1$ を満たさないことがある．[†4]

注 2.1.7.

例 2.1.8 (退化する例：コンパクトハウスドルフ空間). X がコンパクトハウスドルフであるとき，Riesz の表現定理は同型 $M_b(X) = (C(X))^*$ を引き起こす．このとき，$C(X) = C_c(X) = C_0(X) = C_b(X)$ であるから，狭位相，広位相はいずれも（関数解析の意味で）w^*-位相と一致する．[†5]

例 2.1.9 (局所コンパクトハウスドルフ空間での例). X が局所コンパクトハウスドルフ空間であるとき，Riesz-Markov-Kakutani

[†1] 符号付き測度は，有限ならば有界変動を持つから，これは測度が有限だと言っている：$\mathrm{Im}\,\mu \subset \mathbb{R}$.
[†2] https://encyclopediaofmath.org/wiki/Convergence_of_measures
[†3] Bourbaki による造語で，いまでは稀．元々は"convergence étroite".
[†4] 例は極点の列 (δ_n) が与える．X がコンパクトであるとき，$\mu(X) = 1$ を必ず満たす．
[†5] なお，一般に $C(X)$ は回帰的でなく，双対空間 $(M_b(X))^*$ は $C(X)$ より広いから，w^*-位相は弱位相より弱いことは関数解析の結果である．

の定理より $M_b(X)=(C_0(X))^*$ を得て 2.1.3，同時に $M(X)=(C_c(X))^*_+$ を満たす 2.1.4．なお，$M_b(X)$ は有限 Radon 測度の空間で，$M(X)$ は Radon 測度全体の空間とした．よって，いずれも（特定の predual に対する）w^*-位相を備えるが，位相としては別物である．δ_n は 0 に C_c-収束するが，C_0 では収束しない：$f(x)=\sin(x)/x$ を取ると収束しない．

一方で，$C_b(X)$ の双対空間の中には，測度として表せないものも存在する 2.1.2．これは $M(X)\subsetneq M_f(X)=(C_b(X))^*$2.1.2 という意味では，$w^*$-位相の相対位相である．

例 2.1.10 (実数上の測度の収束)**.** $X=\mathbb{R}$ のとき，$1_{[a,b)}$ という形の特整関数が生成する部分空間は，$C_0(\mathbb{R})$ 上稠密であるから，漠位相は「ある稠密部分集合 $D\subset\mathbb{R}$ が存在して，$\forall_{a<b\in D}\ \mu_n((a,b])\to\mu((a,b])$ を満たすことと同値になる．

要諦 2.1.11 (緊密性とは，この消息によって要求される確率論的概念である)**.** こうして，(δ_n) の例などを省くためには，確率測度の列がある種のコンパクト性を満たす必要がある．緊密性を $\forall_{\epsilon>0}\ \exists_{K_\epsilon\mathrm{cpt}X}\ \forall_{n\in\mathbb{N}}\ \mu_n(K_\epsilon^\complement)<\epsilon$ と定めると，意味は the sequence of measures are all almost supported in a compact set, so there is no possibility of mass "escaping at infinity" as it was the case with (δ_n).[†6] 特に，Radon 測度 (μ_n) が緊密ならば μ_n は有限である．

こうして，標準的な設定が出来上がる．有界測度の列が緊密であるとき，３つの関数のクラス $C_c(X),C_0(X),C_b(X)$ が定める収束概念は一致する．

定義 2.1.12 (weak convergence of measure, vague convergence)**.** 一般の局所コンパクトハウスドルフ空間 X について，Riesz の表現定理より $P(X)\subset M(X)\simeq C_0(X)^*$ と同一視出来る．この Borel 可測空間 $(X,\mathcal{B}(X))$ 上の確率測度の空間 $P(X)\subset C_0(X)^*$ の列 (μ_n) について，

(1) $\forall_{f\in C_b(X)}\ \int_X fd\mu_n\to\int_X fd\mu$，平均で表すと $\forall_{f\in C_b(X)}\ E[f(X_n)]\to E[f(X)]$ を，**測度の弱収束**と呼ぶ．これを $\mu_n\Rightarrow\mu$ と表す．
(2) さらに条件を弱めて $\forall_{f\in C_0(X)}\ \int_X fd\mu_n\to\int_X fd\mu$ は w^*-位相に対応し，**漠収束**と呼ばれる．[†7] これは上述の狭位相と広位相が退化して一つになったものである．

要諦 2.1.13. $C_b(X)$ を試験関数とした場合は，確率変数が無限遠へ飛ぶ場合のすべてを検知するが，$C_c(X)\subset C_0(X)$ を用いた場合は見逃す場合がある．

2.1.3　弱収束の特徴付け

定理 2.1.14 (Portmanteau 定理：弱収束の特徴付け (Alexandroff)[†8])**.** $\{\mu_n\}\subset P(\mathbb{R}),\mu\in P(\mathbb{R})$ について，次の５条件は同値．これは一般の距離空間 X について成り立つ．

(1) $\mu_n\Rightarrow\mu$.
(2) 任意の開集合 $G\overset{\text{open}}{\subset}\mathbb{R}$ について，$\liminf_{n\to\infty}\mu_n(G)\ge\mu(G)$.
(3) 任意の閉集合 $C\subset\mathbb{R}$ について，$\limsup_{n\to\infty}\mu_n(C)\le\mu(C)$.
(4) $A\in\mathcal{B}(\mathbb{R})$ について，$\mu(\partial A)=0\Rightarrow\lim_{n\to\infty}\mu_n(A)=\mu(A)$．この条件を満たす集合 A を P-**連続集合**という．[†9]
(5) μ_n,μ が定める分布関数 F_n,F について，$\forall_{x\in\mathbb{R}}\ \lim_{y\nearrow x}F(y)=x\Rightarrow\lim_{n\to\infty}F_n(x)=F(x)$.

また，次とも同値．

(6) 任意の一様連続な有界関数 f について，$P_nf\to Pf$．なお，有界性の仮定を外しても同値である．
(7) 任意の非負有界連続関数 f に関して，$\liminf_{n\to\infty}P_nf\ge Pf$.

定理 2.1.15 (その他の集合族による特徴付け)**.** 距離空間上の測度空間 $(X,\mathcal{B}(X))$ について，$\mathcal{A}\subset\mathcal{B}(X)$ を乗法族とする：$X\in\mathcal{A},\forall_{A,B\in\mathcal{A}}\ A\cap B\in\mathcal{A}$．任意の開集合が $\mathcal{A}$ の元の可算和で表せるとき，$\forall_{A\in\mathcal{A}}\ P_n(A)\to P(A)$ は $P_n\Rightarrow P$ を含意する．

定理 2.1.16 (３つ目の特徴付け)**.** 列 (P_n) が P に弱収束するための必要十分条件は，これが相対コンパクトであること，すなわち

[†6] https://math.stackexchange.com/questions/313986/are-vague-convergence-and-weak-convergence-of-measures-both-weak-convergence
[†7] これを弱位相と呼ぶこともあるが，$C_0(X)$ はほとんど回帰的である例がないので，不適切である．
[†8] I don' tknow who invented such a nonsensical name for Alexandro ff ' s theorem.[3]
[†9] このような集合は体／代数をなす．

任意の部分列 (P_{n_i}) が，P に弱収束する部分列を持つことである．

2.1.4　漠収束の特徴付け

命題 2.1.17 (漠収束の特徴付け)**.** $\{\mu_n\} \subset P(\mathbb{R}), \mu \in M(\mathbb{R})$ について，次の 2 条件は同値．

(1) (μ_n) は μ に漠収束し，$\mu \in P(\mathbb{R})$ を満たす．
(2) (μ_n) は μ に弱収束する．

2.1.5　一様収束のための十分条件

必要になったらで．

定理 2.1.18 (Ranga Rao (62))**.** X を可分な距離空間，(μ_n) をその上の確率測度の列とする．このとき，次の 2 条件は同値．

(1) $\mu_n \Rightarrow \mu$.
(2) 任意の関数族 $\mathcal{F} \subset C_b(X)$ であって，一様に有界で，同程度連続であるものについて，$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} |\mu_n f - \mu f| = 0$ が成り立つ．

2.1.6　連続写像定理

記法 2.1.19. 距離空間 (S_i, d_i) $(i = 1, 2)$ と，確率測度 $\nu, \nu_n \in P(S_1)$ と可測写像 $T : S_1 \to S_2$ について，

命題 2.1.20. $\nu_n \to \nu$ かつ T が連続ならば，$\nu_n^T \to \nu^T$.

補題 2.1.21. 可測写像の列 $\{T_n\} \subset \mathrm{Meas}(S_1, S_2)$ について，次の条件は同値．

(C)　x に収束する S_1 の任意の点列 (x_n) について，$T_n(x_n) \to T(x)$ が成り立つ．
(C')　$\forall_{\epsilon>0}\ \exists_{(n_0,\delta)\in\mathbb{N}\times\mathbb{R}_+}\ \forall_{(n,x')\in\mathbb{N}\times S_1}\ n \geq n_0 \wedge d_1(x, x') < \delta \Rightarrow d_2(T(x), T_n(x')) < \epsilon$.

定理 2.1.22. $\nu_n \to \nu$ とする．ある full set$C \in \mathcal{B}(S_1), \nu(C) = 1$ が存在して，この上で $x \in C$ は条件 (C) を満たすとする．このとき，$\nu_n^{T_n} \to \nu^T$.

系 2.1.23. 可測写像 $T : S_1 \to S_2$ は，$\nu(C) = 1$ なる可測集合 $C \in \mathcal{B}(S_1)$ 上各点連続であるとする．このとき，$\nu_n \to \nu$ ならば $\nu^T \to \nu^T$.

2.2　確率測度の空間

$P(X)$ の幾何学的性質が基礎的になってくるので，ここにまとめることとする．

記法 2.2.1. X を距離空間とし，その上の 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．

2.2.1　可分性と距離付け可能性

補題 2.2.2. X は，弱位相における Δ と位相同型である．

補題 2.2.3. Δ は $P(X)$ 内で点列コンパクトである．[†10]

[†10] 一般に部分集合 $A \subset X$ の閉包 $\overline{A}$ 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．

補題 2.2.4. 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 2.2.5 (確率測度の空間の弱位相の距離化可能性)**.** 次の 2 条件は同値．

(1) X は可分な距離空間である．
(2) 弱位相を備えた $P(X)$ は距離付け可能で可分である．

例 2.2.6 (Levy-Prokhorov 距離)**.** $B_\epsilon(E) := \{x \in X \mid \rho(x, E) < \epsilon\}$ と表す．

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

は距離を定め，これが定める位相は $*$-弱位相に一致する．対応する分布関数の空間の距離を Levy の距離といい，これが先に提出され，Prokhorov がこの形に一般化した．

2.2.2　稠密部分集合の遺伝

定理 2.2.7. X を可分距離空間，$E \subset X$ は稠密とする．E に含まれる可測集合上のみに台をもつ確率測度のなす部分空間

$$\{\mu \in P(X) \mid \mathrm{supp}\,\mu \subset P(E)\}$$

は $P(X)$ 内で稠密である．

2.2.3　コンパクト性

定理 2.2.8 (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 2.2.9. 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 2.2.10 (コンパクト性 (Bogoliubov and Krylov))**.** 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) $P(X)$ は弱コンパクトで距離化可能．

2.2.4　完備性

定理 2.2.11 (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 2.2.12. D がポーランドである時，$\mathcal{M}(D)$ もポーランドである．

2.2.5　相対コンパクト集合の特徴付け

ここで緊密性の概念が出現する．例えば mass が無限遠点に逃げていく列などが省かれる．

定義 2.2.13 (totally bounded / weakly compact)**.** 確率測度の族 $\Gamma \subset P(X)$ が**相対コンパクト**であるとは，任意の列について，ある部分列が存在してこれが弱収束することをいう．

要諦 2.2.14. 可分距離空間 X 上の確率測度の空間 $P(X)$ の弱位相は距離化可能で可分である 2.2.5．距離空間について，可分性と全有界性とは同値である．また，距離空間について，コンパクト性と点列コンパクト性と「完備かつ全有界」であることとは同値

だから，S が完備可分距離空間である場合，相対コンパクト性を全有界ともいう（完備距離空間において，コンパクト集合と閉集合とは同値）．

定義 2.2.15 ((uniformly) tight / tendue)**.** 確率測度の族 $\Gamma\subset P(X)$ が **(一様に) 緊密**[†11]であるとは，$\forall_{\epsilon>0}\ \exists_{K\overset{\text{cpt}}{\subset}X}\ \forall_{\mu\in\Gamma}\ \mu(K)\geq 1-\epsilon$.

定理 2.2.16 (Prokhorov：緊密性とは，相対コンパクト性の特徴付けである)**.** X を距離空間とする．確率測度の族 $\Gamma\subset P(X)$ について，(1)⇒(2) が成り立つ．X が完備かつ可分であるとき，(2)⇒(1) も成り立つ．

(1) Γ は（一様に）緊密である．
(2) Γ は弱 (狭) 位相について相対コンパクトである．

系 2.2.17 (漠収束の特徴付け)**.** 完備可分距離空間 X 上の確率測度列 (μ_n) について，次の 2 条件は同値. [†12]

(1) μ に弱収束する．
(2) μ に漠収束し，かつ (μ_n) は一様に緊密である．

［証明］． (1)⇒(2) は明らかだから，(2)⇒(1) を示す．漠収束の特徴付け 2.1.17 より，極限測度 μ が $P(X)$ に属することを示せば十分である．略． ■

2.2.6　コンパクト空間上の確率測度

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？
X がコンパクトのとき，その上の弱位相と w^*-位相は一致する．

命題 2.2.18. コンパクトハウスドルフ空間 X 上の Banach 代数 $C(X)=C_b(X)=C_0(X)=C_c(X)$ 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 $\delta_x\ (x\in X),\forall_{f\in C(X)}\ \delta_x(f)=f(x)$ である．

［証明］．

(1) $\mu_1,\mu_2\in P(X)$ と $\lambda\in(0,1)$ を任意に取ると，$\lambda\mu_1+(1-\lambda)\mu_2\in P(X)$ がわかる．
(2) 線型汎函数 $\mathrm{ev}_1:M(X)\to\mathbb{F};\mu\mapsto\mu(1)$ 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 2.2.19. 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．

2.3　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 $\int:C_c(X)\to\mathbb{R}$ に関して展開できる．

[†11] 測度が内部正則であることも「緊密」というので，それから見れば「一様に緊密」と言いたくなる．
[†12] https://math.stackexchange.com/questions/313986/are-vague-convergence-and-weak-convergence-of-measures-both-weak-convergence

定義 2.3.1 (outer integral)**.** $(\Omega, \mathcal{A}, P)$ を確率空間，$T:\Omega\to\overline{\mathbb{R}}$ を写像とする．

(1) $E^*[T] := \inf\left\{E[U]\in\mathbb{R} \mid U\geq T, U\in\mathrm{Meas}(\Omega,\overline{\mathbb{R}}), E[U]\in\mathbb{R}\right\}$ 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 $(X_n:\Omega_n\to\Omega)$ が Borel 可測関数 $X:\Omega_\infty\to\Omega$ に弱収束するとは，$\forall_{f\in C_b(\Omega)}\ E^*[f(X_n)]\to E[f(X)]$ を満たすことをいう．

2.4　確率変数列の収束

次に，測度の収束の一般化として，確率変数列に対して収束を定義する．測度の弱収束は，作用素の w^*-収束と同値で，確率変数列の法則収束と同値．

2.4.1　定義

定義 2.4.1 (almost sure convergence, converge in probability, converge in the mean of order p, converge in law / distribution)**.** (X_n) を確率空間 $(\Omega,\mathcal{F},P)$ 上の実確率変数の列とする．

(1) (X_n) が X に**概収束**する $X_n\to X$ a.s. とは，$P\left(\lim_{n\to\infty}X_n=X\right)=1$ が成り立つことをいう．

(2) (X_n) が X に**確率収束**する $X_n\xrightarrow{p}X$ とは，$\forall_{\epsilon>0}\ \lim_{n\to\infty}P(|X_n-X|>\epsilon)=0$ が成り立つことをいう．

(3) (X_n) が X に p **次平均収束**するとは，$\lim_{n\to\infty}E[|X_n-X|^p]=0\ (p\geq 1)$ を満たすことをいう．

(4) (X_n) が X に**法則収束**または**分布収束**する $X_n\xrightarrow{d}X$ とは，$\forall_{f\in C_b(\mathbb{R})}\ \lim_{n\to\infty}E[f(X_n)]=E[f(X)]$ を満たすことをいう．

注 2.4.2 (法則収束の特異性)**.** 概収束，確率収束，p 次平均収束は X_n と X の間に関係 $=$ や演算 $-$ が定義されていることが必要であるから，列 (X_n) は同一の確率空間上で定義されている必要があるが，法則収束は実はその必要はなく，純粋に実数上に引き起こす測度のみによって決まる概念であるから，始域は抽象化されている．

注 2.4.3. $(\mathcal{X}_n,\mathcal{A}_n,P_n)$ 上の S-値確率変数列 (X_n) とある定点 $c\in S$ に対して，

$$X_n\xrightarrow{p}c :\Leftrightarrow \forall_{\epsilon>0}\ \lim_{n\to\infty}P_n[d(X_n,c)>\epsilon]=0$$

も確率収束と呼ばれるが，この場合は法則収束 $P_n^{X_n}\Rightarrow\delta_c$ と同値になる．

2.4.2　収束の間の関係

$$\begin{tikzcd}[row sep=small]
\text{概収束} \arrow[dr] & & \\
& \text{確率収束} \arrow[r] & \text{法則収束} \\
p\ \text{次確率収束} \arrow[ur] & &
\end{tikzcd}$$

補題 2.4.4 (相互関係 1)**.** (X_n) が X に確率収束するならば，法則収束する．

定理 2.4.5 (相互関係 2)**.** $(X_n), X$ について，

(1) 概収束するならば，確率収束する．

(2) p 次平均収束するならば，確率収束する．

命題 2.4.6 (確率収束の特徴付け)**.** 可分距離空間 S-値確率変数 X_n, X について，次の 3 条件は同値．

(1) (X_n) は X に確率収束する．

(2) $E[d(X_n,X)\wedge 1]\to 0\ (n\to\infty)$.

(3) $\{X_n\}$ は概収束に関して相対コンパクト．

系 2.4.7 (弱めた逆が成り立つ)**.** X_n は X に確率収束，または p 次平均収束するとする．このときある部分列が存在して，概収束する．

2.4.3　確率収束の性質

命題 2.4.8. (S_i,d_i) $(i=1,2)$ を可分距離空間とする．X,X_n を S_1-値確率変数，$h:S_1\to S_2$ を可測写像とする．$P[X\in C]=1$ を満たす可測集合 $C\in\mathcal{B}(S_1)$ 上で h は連続であるとする．このとき，$X_n\xrightarrow{p}X\Rightarrow h(X_n)\xrightarrow{p}h(X)$.

2.4.4　連続写像定理とデルタ法

次は分布収束を考える．分布収束は，有界測度論の世界が流入した風景が広がっている．

定理 2.4.9 (連続写像定理)**.** $T,T_n:S_1\to S_2$ を可測写像とし，$P[X\in C]=1$ を満たす可測集合 $C\in\mathcal{B}(S_1)$ の上で条件 (C) が満たされるとする．このとき，$X_n\xrightarrow{d}X$ ならば，$T(X_n)\xrightarrow{d}T(X)$.

定理 2.4.10 (デルタ法)**.** 関数 $f:\mathbb{R}^{d_1}\supset A\to\mathbb{R}^{d_2}$ は $c\in A$ で微分可能であるとする．$\forall_{n\in\mathbb{N}}\;P[X_n\in A]=1$ を満たす d_1 次元確率変数列 (X_n) に対して，∞ に発散する係数列 $\{b_n\}\subset\mathbb{R}$ が定める線型変換は $b_n(X_n-c)\xrightarrow{d}Z$ を満たすとする．このとき，

$$b_n[f(X_n)-f(c)]-\partial_x f(x)b_n(X_n-c)\xrightarrow{p}0\quad(n\to\infty)$$

特に，$b_n[f(X_n)-f(c)]\xrightarrow{d}\partial_x f(c)Z$ $(n\to\infty)$ である．

要諦 2.4.11. 列 $(b_n[f(X_n)-f(c)])$ は $(b_n\partial_x f(c)(X_n-c))$ に漸近同等である．連続写像定理に対応する，ある種の連鎖律である．

2.4.5　分布収束と確率収束とタイト性

$O_p(1),o_p(1)$ の概念が先に定まる．$O_p(1)$ の概念はタイトともいうが，$\mathbb{R}$ に限れば，確率収束の意味で有界であるための条件に読める．

定理 2.4.12. X,X_n,Y_n を確率ベクトルまたは確率行列とする．c を定数ベクトルまたは定数行列とする．$X_n\xrightarrow{d}X,Y_n\xrightarrow{p}c$ のとき，次が成り立つ．

(1) $(X_n,Y_n)\xrightarrow{d}(X,c)$.
(2) $X_n+Y_n\xrightarrow{d}X+c$.
(3) $Y_nX_n\xrightarrow{d}cX$.
(4) c が正則行列のとき，$Y_n^{-1}X_n\xrightarrow{d}c^{-1}X$.

注 2.4.13. 定値関数ではない Y に対して $Y_n\xrightarrow{p}Y$ であっても，結合ベクトルの収束 $(X_n,Y_n)\xrightarrow{d}(X,Y)$ は一般には成り立たない．

系 2.4.14. X,X_n,Y_n が d 次元確率変数で，$X_n\xrightarrow{d}X$ かつ $X_n-Y_n\xrightarrow{p}0$ を満たすならば，$Y_n\xrightarrow{d}X$ が成り立つ．

定義 2.4.15 (asymptotically equivalent, bounded in probability / tight)**.**

(1) 条件 $X_n-Y_n\xrightarrow{p}0$ なる条件を，$(X_n),(Y_n)$ は**漸近同等**であるといい，$X_n\equiv^a Y_n$ と表すこととする．
(2) $X_n=o_p(1):\Leftrightarrow X_n\xrightarrow{p}0$ とし，(一様に) 緊密または**確率有界**であることを $X_n=O_p(1):\Leftrightarrow\forall_{\epsilon>0}\,\exists_{M>0}\,\sup_{n\in\mathbb{N}}P[|X_n|\geq M]<\epsilon$ と表す．

補題 2.4.16. $X_n\xrightarrow{d}X$ ならば，$X_n=O_p(1)$ である．特に，$X_n=o_p(1)$ ならば $X_n=O_p(1)$ である．

[証明] . 漠収束の特徴付け 2.1.17 より． ■

補題 2.4.17. $(X_n),(Y_n)$ を確率変数列とする．

(1) $X_n = O_p(1), Y_n = O_p(1)$ ならば，$X_n Y_n = O_p(1), X_n + Y_n = O_p(1)$.
(2) $X_n = O_p(1), Y_n = o_p(1)$ ならば，$X_n Y_n = o_p(1), X_n + Y_n = O_p(1)$.

定義 2.4.18. 正数列 $\{r_n\} \subset \mathbb{R}_{>0}$ に対して，$X_n = o_p(r_n) :\Leftrightarrow \frac{X_n}{r_n} = o_p(1)$ とし，$X_n = O_p(r_n) :\Leftrightarrow \frac{X_n}{r_n} = O_p(1)$.

2.4.6 安定収束

定義 2.4.19. 任意の確率収束する確率変数列 $Y_n \xrightarrow{p} Y$ に対して $(X_n, Y_n) \xrightarrow{d} (X, Y)$ が成り立つとき，$X_n \xrightarrow{d} X$ $(stably)$ と表し，**安定収束**するという．

例 2.4.20. 独立確率変数列 (X_n)，マルチンゲール (X_n) など，中心極限定理が従う多くの場合，これが法則収束するとき，安定収束する．

注 2.4.21. 安定収束の概念は非エルゴード的統計において最も基本的な役割を演じる．非エルゴード統計とは，条件付き Fisher 情報量の極限にランダム性が残る場合をいう．

2.4.7 弱収束が定める概収束列

定理 2.4.22 (Skorokhod representation theorem)**.** 距離空間 $(S, \mathcal{S})$ 上の確率測度の列 (P_n) は，可分な台をもつ P に弱収束するとする．このとき，確率空間 $(\Omega, \mathcal{F}, P)$ とその上の S-値確率要素の列 (X_n) が存在して，確率測度 (P_n) をそれぞれ誘導し，P に概収束する．また，$S = \mathbb{R}$ のとき，$\Omega = (0, 1)$，$X_n(y) = \inf \{z \in \mathbb{R} \mid P_n(-\infty, z] \geq y\}$ と取れる．

要諦 2.4.23. あくまで測度の $*$-弱収束が主軸である，とも捉えられる．

2.4.8 距離との関係

概収束を定める位相は距離付け不可能で，確率収束を定める位相は距離付け可能である．p 次平均収束とは，L^p-ノルムが定める位相についての収束と同値．
法則収束は，分布の空間 $P(\Omega)$ の上で距離付け可能である．

命題 2.4.24. 概収束は距離付け不可能である．

命題 2.4.25 (確率収束を定める距離)**.** $X, Y \in \mathrm{Meas}(\Omega, \mathbb{R})$ について，$d(X, Y) := E\left[\frac{|X - Y|}{1 + |X - Y|}\right]$ と定める．

(1) d は距離を定める．
(2) d について収束することと，確率収束することは同値．ただし，$X = Y$ a.s. を同一視する．

2.4.9 一様可積分性の定義と特徴付け

平均作用素はノルム連続であるが，各収束概念を保存することとの関係は込み入っている．そこで，E が収束を保つための十分条件を調べると同時に，概収束と p-次平均収束の関係を精査する．

定義 2.4.26 (uniformly integrable)**.**

(1) 列 (X_n) が**一様可積分**であるとは，

$$\lim_{\lambda \to \infty} \sup_{n \in \mathbb{N}} E[|X_n||\, |X_n| \geq \lambda] = 0$$

が成り立つことをいう．

(2) 族 $(X_\lambda)_{\lambda \in \Lambda}$ が**一様可積分**であるとは，

$$\lim_{A \to \infty} \sup_{\lambda \in \Lambda} E_\lambda[|X_\lambda| 1_{|X_\lambda|} \geq A] = 0$$

が成り立つことをいう．ただし，確率変数 X_λ の定義域を確率空間 $(\Omega_\lambda, \mathcal{F}_\lambda, P_\lambda)$ とし，これが定める積分を E_λ とした．

要諦 2.4.27. 列 $(|X_n|)$ の末尾の積分が，一様に 0 に近づくことをいう．また定義から，$\{X_n\}$ が一様可積分であることと $\{|X_n|\}$ が一様可積分であることとは同値で，このとき $\sup_{\lambda\in\Lambda} E_\lambda[|X_\lambda|] < \infty$ が成り立つ（これは一部特徴付ける）．つまり，分布族 $\{P_\lambda^{X_\lambda}\}$ の性質であって，例えば列 $X_n(x) := x1_{[0,n]}(x)$ は，P が正規分布のとき一様可積分であるが，Cauchy 分布であるときはそうではない．

補題 2.4.28 (一様可積分性の十分条件)**.** Y を可積分な確率変数とする．

(1) $\lim_{\lambda\to\infty} E[|Y||Y| \geq \lambda] = 0$.
(2) 列 (X_n) が $\forall_{n\in\mathbb{N}}\ |X_n| \leq Y$ を満たすならば，(X_n) は一様可積分である．

補題 2.4.29 (一様可積分性の特徴付け)**.** 列 (X_n) について，次の 2 条件は同値．

(1) (X_n) は一様可積分である．
(2) ある Borel 可測関数 $\psi : \mathbb{R} \to \mathbb{R}_+$ が存在して，$\lim_{|x|\to\infty} \dfrac{|x|}{\psi(x)} = 0, \sup_{n\in\mathbb{N}} E[\psi(X_n)] < \infty$ を満たす．

特に，$p > 1$ が存在して $\sup_{n\in\mathbb{N}} E[|X_n|^p] < \infty$ を満たすならば，(X_n) は一様可積分である．

系 2.4.30. $0 < p < q$ とするとき，$\sup_{\lambda\in\Lambda} E_\lambda[|X_\lambda|^q] < \infty$ ならば，$\{|X_\lambda|^p\}_{\lambda\in\Lambda}$ は一様可積分である．

補題 2.4.31 (一様可積分性の特徴付け)**.** 列 (X_n) が一様可積分であることと，次の 2 条件が成り立つこととは同値：

(1) $\sup_{n\in\mathbb{N}} E[|X_n|] < \infty$.
(2) $\forall_{\epsilon>0}\ \exists_{\delta>0}\ A \in \mathcal{F} \wedge P(A) < \delta \Rightarrow \sup_{n\in\mathbb{N}} E[|X_n||A] < \epsilon$.

2.4.10　収束概念が退化するための十分条件

一様可積分性が，次の系のような形で，p-次平均収束が他の収束概念と一致するための必要十分条件になる．

定理 2.4.32 (一般化された Lebesgue の優収束定理)**.** (X_n) は一様可積分とする．

(1) X_n が X に概収束するならば，1 次平均収束する：$\lim_{n\to\infty} E[|X_n - X|] = 0$．特に，極限 X は可積分で，$\lim_{n\to\infty} E[X_n] = E[X]$.
(2) X_n が X に確率収束することと，1 次平均収束することとは同値．

系 2.4.33 (一様可積分性の特徴付け)**.** 可積分な確率変数の列 (X_n) は X に概収束するとする．このとき次の 3 条件は同値：

(1) (X_n) は一様可積分である．
(2) (X_n) は X に 1 次平均収束する：$\lim_{n\to\infty} E[|X_n - X|] = 0$.
(3) $E[|X_n|] \to E[|X|]$ かつ $E[|X|] < \infty$.

系 2.4.34 (さらに一般化)**.** 任意の $q \geq 1$ について，$E[|X_n|^q] < \infty$ かつ $X_n \xrightarrow{p} X$ のとき，次の 3 条件は同値．

(1) $\{|X_n|^q\}_{n\in\mathbb{N}}$ は一様可積分である．
(2) $E[|X_n - X|^q] \to 0$.
(3) $E[|X_n|^q] \to E[|X|^q] < \infty$.

定理 2.4.35 (Dunford-Pettis)**.** 列 (X_n) について，次の 2 条件は同値．

(1) (X_n) は一様可積分である．
(2) $\{X_n\} \subset L^1(\Omega)$ は弱相対コンパクトである：$\forall_{Y\in L^\infty(\Omega)}\ \lim_{n\to\infty} E[X_nY] = E[XY]$.

2.5 不等式

2.5.1 確率不等式

記法 2.5.1. $(\Omega, \mathcal{F}, P)$ を確率空間とする．

命題 2.5.2 (Markov の不等式)**.** 非負確率変数 $X \geq 0$ と，単調増加関数 $g : \mathbb{R}_+ \to \mathbb{R}_+$ に対して，

$$\forall_{x\geq 0} \quad P(X \geq x) \leq \frac{E[g(X)]}{g(x)}$$

命題 2.5.3 (Jensen の不等式)**.** 確率変数 X と凸関数 $g : \mathbb{R} \to \mathbb{R}$ に対して，$X, g(X) \in \mathcal{L}^1(\Omega)$ ならば，

$$g(E[X]) \leq E[g(X)].$$

2.5.2 Lebesgue 空間のノルム不等式

$L^p(\mathbb{R}^d)$ $(p \in [1, \infty])$ はノルムに関して完備であるという事実は，Riesz-Fisher の定理と呼ばれている．

命題 2.5.4 (Holder)**.** $p, q, r \in [1, \infty]$ は $p^{-1} + q^{-1} = r^{-1}$ を満たすとする．このとき，$X \in L^p \wedge Y \in L^q$ ならば $XY \in L^r$ であって，

$$\|XY\|_r \leq \|p\|\|q\|.$$

系 2.5.5.

(1) Cauchy-Schwarz $\|XY\|_1 \leq \|X\|_2\|Y\|_2$.
(2) Lyapunov $0 < a < b$ かつ $\mu(\Omega) = 1$ のとき，$X \in L^b$ ならば $\|X\|_a \leq \|X\|_b$.

命題 2.5.6 (Minkowski)**.** $p \in [1, \infty), X, Y \in L^p$ について，

$$\|X + Y\|_p \leq \|X\|_p + \|X\|_p$$

2.5.3 畳み込みと不等式

記法 2.5.7. 一般に，局所コンパクトで単模である群 G の両側不変 Haar 測度 μ に対して，G 上の関数の畳み込みが考えられ，同じ結果が成り立つ．

定義 2.5.8.

$$f * g(x) := \int_{\mathbb{R}^d} f(x-y)g(y)dy$$

によって，$\mathrm{Meas}(\mathbb{R}^d, \mathbb{R})^2 \nrightarrow \mathrm{Meas}(\mathbb{R}^d, \mathbb{R})$ を定める．Young の不等式より，$L^p(\mathbb{R}^d) \times L^q(\mathbb{R}^d)$ 上に制限すると，積を定める．

命題 2.5.9 (Young)**.** $p, q, r \in [1, \infty], p^{-1} + q^{-1} = r^{-1} + 1$ とする．$f \in L^p(\mathbb{R}^d), g \in L^q(\mathbb{R}^d)$ のとき，f^*g は殆ど至るところ存在し，$f * g \in L^r(\mathbb{R}^d)$ を満たし，

$$\|f * g\|_r \leq \|f\|_p\|g\|_q.$$

2.6 一般の試行と確率測度

定義 2.6.1 (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 2.6.2 (Lebesgue expantion)**.** 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．

2.7　確率測度の拡張定理

定義 2.7.1 (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．

2.8　確率測度の直積

> 独立性は P の集合積に対する関手性だから，数学的には直積と関係が深い．

定義 2.8.1. $(\Omega_n, \mathcal{F}_n, P_n)$ を確率空間の列とする．$\Omega := \prod_{n\in\mathbb{N}} \Omega_n$ とし，$\mathrm{pr}_n : \Omega \to \Omega_n$ を射影とする．射影が Ω 上に定める σ-加法族を，積 σ-加法族という．これを Kolmogorov の σ-加法族ともいう．この上の確率測度を，直積確率測度という．

注 2.8.2. 射影 $q_n := (\mathrm{pr}_1, \cdots, \mathrm{pr}_n)$ による σ-加法族 $\mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ の引き戻し $q_i^*(A_i)$ $(A_i \in \mathcal{S}_i)$

$$C_A^{(n)} = \{\omega \in \Omega \mid (\omega_1, \cdots, \omega_n) \in A\} \quad A \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$$

を**柱状集合**という．柱状集合の全体 $\mathcal{C}$ は有限加法族で，これが生成する σ-加法族が Kolmogorov σ-加法族である．

補題 2.8.3.

(1) 直積確率測度は一意的である．
(2) 直積確率測度は存在する．

定理 2.8.4. Ω_n はいずれも Poland 空間で，$\mathcal{F}_n$ が Borel 集合族であるとき，コルモゴロフの σ-加法族 $\mathcal{F}$ は，Ω の直積位相が生成する Borel 集合族に一致する．

2.9　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，$(\mathbb{R}, \mu)$ の同型類を特徴づける．

定義 2.9.1 (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 $(\mathbb{R}, \mu)$ に同型である時，(Ω, P) を**標準確率空間**と呼ぶ．

定義 2.9.2 (pushforward measure)**.** S 上の測度 P の，可測関数 $f : S \to T$ による像測度を，Pf^{-1} または f_*P で表す．

補題 2.9.3. P が完備ならば Q も完備である．

第 3 章

確率論の基礎概念

単なる可測関数と測度としての範囲を超える確率変数・確率分布特有の概念と，その極限の取り扱いを議論する．大事な概念は独立性と収束である．

(1) 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．確率変数列の独立性は，確率測度が定める関係である．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 [5]．なお事象の独立性とは，測度 $P:\mathcal{F}\to[0,1]$ が，積 $\cap$ なる演算に対して可換（準同型）であること：$P(A\cap B)=P(A)\cdot P(B)$ と理解できる．

(2) 独立性は，事象に対する定義から，σ-加法族・確率変数族へと一般化され，独立な確率変数列なる対象が構成される．

(3)

3.1 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，情報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．分母の $P(A)$ は規格化条件で，再び確率測度を与えるため，これは確率測度の変換の一種である．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 3.1.1 (conditional probability).

(1) $A,B\in\mathcal{F},P(A)>0$ とする．

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

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 $P(\cdot\mid\cdot):\mathcal{F}\times\mathcal{F}\to[0,1]$ 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．

注 3.1.2 (conditional expectation). 新たな確率空間 $(A,\mathcal{F}\cap P(A),P_A)$ における積分作用素を，条件付き期待値という．これらの定義は，事象 $A\in\mathcal{F}$ を指定している裏で暗黙に指定されている σ-代数 $\mathcal{F}\cap P(A)$ について一般的に定義できる．

---

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

命題 3.1.3 (law of total probability：全確率の分解)**.** $(H_i)_{i\in I}$ を互いに背反で $\Omega = \coprod_{i\in I} H_i$ を満たす事象の族とする．

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

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

$$\begin{aligned} P(A) &= P\left(\cup_{i\in I}A\cap H_i\right) \\ &= \sum_{i\in I} P(A\cap H_i) && \text{加法性} \\ &= \sum_{i\in I} P(A\mid H_i)P(H_i) && \text{定義 3.1.1} \end{aligned}$$

■

要諦 3.1.4. 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 3.1.5 (Bayes)**.** $(H_i)_{i\in I}$ を互いに背反で $\Omega = \coprod_{i\in I} H_i$ 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］．

$$\begin{aligned} P(H_i\mid A) &= \frac{P(H_i\cap A)}{P(A)} && \text{条件付き確率の定義 3.1.1} \\ &= \frac{P(A\mid H_i)P(H_i)}{\sum_{j\in I}P(A\mid H_j)P(H_j)} && \text{分母は定義 3.1.1，分子は全確率の分解則 3.1.3} \end{aligned}$$

■

要諦 3.1.6. 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．

3.2 事象の独立性

積への関手性

測度は純粋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，純粋に確率論的な概念である．

3.2.1 互いに独立な事象

定義 3.2.1 (independent)**.** 事象 $A, B\in\mathcal{F}$ について，

(1) 可測空間 $(\Omega,\mathcal{F})$ 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 3.2.2. $P(A), P(B) > 0$ 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．

[†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.

補題 3.2.3 (補集合演算に関する関手性). $A, B \in \mathcal{F}$ について，次の 4 条件は同値.

(1) A, B は独立.
(2) $A^{\complement}, B$ は独立.
(3) $A, B^{\complement}$ は独立.
(4) $A^{\complement}, B^{\complement}$ は独立.

定理 3.2.4 (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.

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

［証明］.

(1)⇒(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，$P^X(N) = P\{X \in N\} = \sum_{x \in N} P\{X = x\} = 0$.

■

3.2.2　事象族の独立性

定義 3.2.5 (事象族の独立性, pairwisely independent).

(1) 列 $(A_i)_{i \in n} : n \to \mathcal{F}$ が独立であるとは，$\forall_{2 \le k \le n}\ \forall_{1 \le i_1 < \cdots < i_k \le n}\ P(A_{i_1} \cap \cdots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_k})$.
(2) 族 $(A_i)_{i \in I}$ が独立であるとは，任意の有限部分集合について $\forall_{n \ge 2}\ A_1, \cdots, A_n$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，$\forall_{i \neq j \in I}\ A_i \cap A_j = \emptyset$.

例 3.2.6. 対独立であるが独立ではない例がある.

補題 3.2.7 (独立性の遺伝). 事象 $A_1, \cdots, A_n$ が独立であるとする.

(1) 事象 $A_1^{\complement}, \cdots, A_n^{\complement}$ 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 3.2.8 (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，

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

［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，

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

■

3.2.3　σ-代数の独立性

定義 3.2.9 (σ-代数の独立性).

(1) $\mathcal{F}$ の部分 σ-代数 $\mathcal{F}_1, \mathcal{F}_2$ が独立であるとは，$\forall_{C_1 \in \mathcal{F}_1, C_2 \in \mathcal{F}_2}\ P(C_1 \cap C_2) = P(C_1) \cap P(C_2)$ を満たすことをいう.
(2) 部分代数の有限列 $(\mathcal{F}_k)_{k \in [n]}$ が独立であるとは，$\forall_{k \in [n]}\ \forall_{C_k \in \mathcal{F}_k}\ P(\cap_{k \in [n]} C_k) = \prod_{k \in [n]} P(C_k)$ を満たすことをいう.
(3) 部分代数族が独立であるとは，任意の有限部分集合が独立であることをいう.
(4) 部分代数族が対独立であるとは，任意の 2 つの部分代数 $\mathcal{F}_i, \mathcal{F}_j$ が互いに独立であることをいう.

3.3　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 $X:\Omega\to\mathcal{X}$ が誘導する測度 P^X を観測することが問題となり，$(\Omega,\mathcal{F},P)$ の特定の構造には執着しない．特に，経験過程論のように，$(\Omega,\mathcal{F},P)$ を拡大して考えることがしばしばである．

3.4　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（σ-加法族）が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．
独立性は P の集合積に対する関手性だから，数学的には直積によって構成する．

定義 3.4.1 (independent)**.** 確率変数 $X_1,\cdots,X_n:\Omega\to\mathcal{X}_i$ が独立であるとは，次を満たすことをいう：

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

ただし，$X:=(X_1,\cdots,X_n)$ を同時分布，P^X は直積測度 $P^{X_1}\times\cdots\times P^{X_n}$ 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 3.4.2 (可測関数は独立性を保つ)**.** 可測空間 $(\mathcal{Y}_i,\mathcal{C}_i)$ への射 $f_i:\mathcal{X}_i\to\mathcal{Y}_i$ について，確率変数 $X_1,\cdots,X_n$ が独立ならば，$f_1(X_1),\cdots,f_n(X_n)$ 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 $C_i\in\mathcal{C}_i$ 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，

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

■

命題 3.4.3. 確率変数 $A_1,\cdots,A_n:\Omega\to\mathcal{X}$ が独立であるとする．任意の $I:=[n]\supset J$ について，$\{A_j^\complement,A_k\mid j\in J,k\in I\setminus J\}$ は独立である．

[証明]． 写像 $f:\mathcal{X}\to\mathcal{X}$ を，A_j を $A_j^\complement$ 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． ■

3.5　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 [5].

3.5.1　確率変数の独立性

定義 3.5.1. S_k 値確率変数列 $(X_k)_{k\in[n]}$ が独立とは，

$$\forall_{A_1\in\mathcal{S}_1,\cdots,A_n\in\mathcal{S}_n}\ P[X_k\in A_k,k\in[n]]=\prod_{k=1}^nP[X_k\in A_k]$$

補題 3.5.2 (well-definedness)**.** 次の 2 条件は同値.

(1) (X_n) は独立.
(2) $(\sigma(X_n))$ は独立.

3.5.2　独立な確率変数に対する関手性

命題 3.5.3 (実空間の議論への持ち上げ)**.** 実確率変数の族 $X := (X_1, \cdots, X_n) : \Omega \to \mathbb{R}^n$ と任意の可測関数 $f : \mathbb{R}^n \to \mathbb{R}$ に対して,

(1) $h := f \circ X : \Omega \to \mathbb{R}^n \to \mathbb{R}$ は実確率変数である.
(2) $E[h] = \displaystyle\int_{\mathbb{R}^n} f(x_1, \cdots, x_n) P^X(dx_1, \cdots, dx_n)$.

[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.

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

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 3.5.4 (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，h も可積分で次が成り立つ：$E[h] = \displaystyle\prod_{i=1}^n E[f_i \circ X_i]$.

[証明]． 命題 3.5.3 より

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

であるが，P^X は直積測度 $P^{X_1} \times \cdots \times P^{X_n}$ であり 3.4.1，測度空間 $\mathbb{R}^n$ は σ-有限であるから，Fubini の定理より，$\displaystyle\prod_{i \in [n]} E[f_i(X_i)]$ に等しい. ■

系 3.5.5 (分散が和を保つ条件)**.** $(X_i)_{i \in [n]}$ を二乗可積分で：$E[|X_i|^2] < \infty$ $(i = 1, \cdots, n)$，対独立な確率変数の列とする．この時，次が成り立つ：$\mathrm{Var}\left[\displaystyle\sum_{i=1}^n X_i\right] = \displaystyle\sum_{i=1}^n \mathrm{Var}[X_i]$.

[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 3.5.6. 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.

3.5.3　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 3.5.7 (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 3.5.8 (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：

(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，$\pi_n:=(\mathrm{pr}_1,\cdots,\mathrm{pr}_n)$ 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 $Q:\mathcal{B}(\mathbb{R}^\infty)\to[0,1]$ を考える．有限加法性の確認は，族 $(A_n)_{n\in N}$ に対して $m:=\max_{n\in N}\dim(A_n)$ 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．$\mathcal{A}:=\left\{\pi_n^{-1}(A_n)\in\mathbb{R}^\infty \mid A_n\in\mathcal{B}(\mathbb{R}^n),n\in\mathbb{N}\right\}$ は集合体であることは同様にして明らか．すると，Q が $\mathcal{A}$ 上で σ-加法的であることを示せば，$\mathbb{R}^\infty$ 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，$\alpha:=\lim_{n\to\infty}Q(A_n)>0\Rightarrow\cap_{n\in\mathbb{N}}A_n\neq\emptyset$ を示せば良い．

証明 単調減少列 (Λ_n) を任意に取ると，$\Lambda_n=\pi_{n_i}^{-1}(A_{n_i})$ と表せる．$\max_{n\in\mathbb{N}}n_i$ が存在するならば同様にして自明だから，$\{n_i\}_{n\in\mathbb{N}}\subset\mathbb{N}$ 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，$\Lambda_n=\pi_n^{-1}(A_n)$ として良い．

Borel 集合の位相的正則性より，$C_n\subset A_n,P_n(A_n\setminus C_n)<\dfrac{\alpha}{2^{n+1}}$ 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，$D_n\subset\Lambda_n,Q(\Lambda_n\setminus D_n)<\dfrac{\alpha}{2^{n+1}}$ を満たす．$\overline{D}_n:=\cap_{k=1}^nD_k$ とおくと，

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

より，$\overline{D}_n\neq\emptyset$ 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 3.5.9. これは Hopf の拡張定理の一般化に当たる．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．さらに一般的な空間上については，σ-加法族の表現が変わる 2.8.4.

補題 3.5.10 (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，$Q\left(\lim_{n\to\infty}A_n\right)=Q\left(\bigcap_{n\in\mathbb{N}}A_n\right)=\lim_{n\to\infty}Q(A_n)$.

(3) 任意の $\mathcal{A}$ の単調減少列 (A_n) に対して，$\lim_{n\to\infty}Q(A_n)>0\Rightarrow\cap_{n\in\mathbb{N}}A_n\neq\emptyset$ である．

［証明］．

(1)⇒(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 $\mu(\lim_{n\to\infty}A_n)=\mu(\cup_{n\in\mathbb{N}}A_n)=$

$\lim_{n\to\infty} A_n$ 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，

$$Q(\lim_{n\to\infty} A_n) = Q(\cap_{n\in\mathbb{N}} A_n) = 1 - Q(\cup_{n\in\mathbb{N}} \overline{A_n}) = 1 - \lim_{n\to\infty} Q(\overline{A_n}) = \lim_{n\to\infty} Q(A_n).$$

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)⇒(3)　自明．

(3)⇒(1)　$\mathcal{A}$ 内に収束する互いに素な列 (A_n) を任意に取り，$A := \cup_{n\in\mathbb{N}} A_n$ と定める．$B_1 := A, B_n := A \setminus \left(\cup_{i=1}^{n-1} A_i\right)$ 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，

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

■

定理 3.5.11 (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 $\mathbb{R}^\infty$ 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 6.1.1 より，$\mathbb{R}^\infty$ 上の確率測度 P で $\forall_{A\in\mathcal{B}(\mathbb{R}^n)}\ P(\mathrm{pr}_n^{-1}(A)) = P_n(A)$ を満たすものが定まる．これに対して，$X_n := \mathrm{pr}_n : \mathbb{R}^\infty \to \mathbb{R}$ 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，

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

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

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

より，独立性も従う．

■

例 3.5.12 (Bernoulli sequence / process)**.** 特に確率空間 2 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．$P(\{X_i = 0\}) = p, P(\{X_i = 1\}) = 1 - p\ (0 \le p \le 1)$ なる分布 $P^X : \{0, 1\} \to [0, 1]$ 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 3.5.11 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．

3.5.4　独立確率変数列

定理 3.5.13. (μ_n) を $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度の列とする．このとき，ある確率空間 $(\Omega, \mathcal{F}, P)$ とその上の独立な確率変数列 (X_n) が存在して，$\forall_{n\in\mathbb{N}}\ X_n \sim \mu_n$ を満たす．

3.6 確率変数の和・積・商の分布

確率変数の積 $(X_1,\cdots,X_n):\Omega_1\times\cdots\times\Omega_n\to\mathcal{X}_1\times\cdots\times\mathcal{X}_n$ 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 3.6.1 (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 3.6.2. 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，

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

■

3.7 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 3.7.1 (perfect (Kolmogorov)). 完備な確率空間 (S,μ) 上の任意の μ-可測関数 $f:S\to\mathbb{R}$ 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 3.7.2 (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，$\mathrm{Dom}\,(\mu)$ が可算分離族を含むとき，μ を**可分**という．

定理 3.7.3. μ を S 上の確率測度，$f:S\to T$ を可測関数とする．像測度 $\nu:=\mu f^{-1}$ は完全である．

3.8 事象と確率変数

記法 3.8.1 (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，$\{X(\omega)\le a\}=X^{-1}((-\infty,a])$ などと表せる．

[†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．

3.9　条件付き期待値

条件付き期待値は素朴には，部分代数 $\mathcal{G}\subset\mathcal{F}$ について，各 $B\in\mathcal{G}$ 上で，$\mathcal{F}$ が定める測度を積分する演算である．構成論は Lebesgue 積分論で終わらせて居るため，定義は性質のみによって行い，零関数の差に目を瞑る．

3.9.1　動機

議論 3.9.1 (事象の条件付き確率が定める条件付き期待値)**.** ある事象 $B\in\mathcal{F}$ が定める条件付き確率 $P(-|B)$ は，$(\Omega,\mathcal{F})$ 上の確率測度である．これが定める積分を，条件付き期待値と呼べる．

$$E[X|B]:=\frac{E[X,B]}{P(B)}.$$

ただし，$E[X,B]=E[1_BX]$ とした．これは，積分範囲の制限と規格化に他ならない．これだけでは，概念の射程が限られる．

議論 3.9.2 (条件付き確率の一般化)**.** そこで，条件付き確率の概念を一般の σ-代数に一般化することで，条件付き期待値を一般化することを考える．まずは，有限な直和分割が生成する σ-代数を考える．

(B_i) を事象による Ω の有限な直和分割でいずれも零集合でないとする．$P[A|(B_i)]:=\sum_{i=1}^{n}P(A|B_i)1_{B_i}$ と定め，条件付き期待値は

$$E[X|(B_i)]:=\sum_{i=1}^{n}E[X|B_i]1_{B_i}$$

と定めると，これは先程の定義の凸結合が与える単関数となっている．(B_i) の生成する σ-代数を $\mathcal{G}$ で表すと，それぞれを $P[A|\mathcal{G}],E[A|\mathcal{G}]$ と表す．

議論 3.9.3 (測度論的議論)**.** 有限生成とは限らない一般の σ-代数 $\mathcal{G}$ に対する条件付き期待値を定義したい．単関数の無限和とは，積分に他ならない．実は裏技が存在して，満たすべき性質を指定するのみで，Radon-Nykodym の定理により，平均は一意的に定まる．

3.9.2　定義

$E[-|\mathcal{G}]:\mathrm{Meas}_{\mathcal{F}}(\Omega,\mathbb{R})\to L^1(\Omega,\mathcal{G})$ は，$\mathcal{F}$-可測な確率変数に対して，ある $\mathcal{G}$-可測な確率変数の同値類を与える．しかし，$E[X,B]=E[Y,B]$ を満たすため，確率変数としての本質は変わらない．すなわち，$\mathcal{G}$ に応じて，解像度を粗くするのである．

定義 3.9.4 (conditional expectation)**.** 次の条件を満足する確率変数 $Y:\Omega\to\mathbb{R}$ を，$\mathcal{G}$ に関する X の**条件付き期待値**と呼ぶ．

(1) Y は $\mathcal{G}$-可測で P-可積分．
(2) 任意の $B\in\mathcal{G}$ に対して $E[X,B]=E[Y,B]$ すなわち $\int_B X(\omega)P(d\omega)=\int_B Y(\omega)P(d\omega)$ を満たす．

この Y を $E[X|\mathcal{G}]$ で表す．X が可測関数の特性関数である場合，$P(A|\mathcal{G}):=E[1_A|\mathcal{G}]$ $(A\in\mathcal{F})$ を**条件付き確率**という．

例 3.9.5 (自明な例)**.**

(1) $\mathcal{G}=\mathcal{F}$ のとき，$E[X|\mathcal{F}]=X$ a.s. である．
(2) $\mathcal{G}=\{\emptyset,\Omega\}$ のとき，$E[X|\mathcal{G}]=E[X]$ で，定数関数である．

補題 3.9.6 (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．

［証明］． $Q(B) := E[X, B] = E1_B X\ (B \in \mathcal{G})$ をおくことで，Q は $(\Omega, \mathcal{G})$ 上の確率測度を定める．いま，$P|_{\mathcal{G}}$ に関して Q は絶対連続：$\forall_{B\in\mathcal{G}}\ P(B) = 0 \Rightarrow Q(B) = 0$ が成り立つから，Radon-Nikodym の定理より，ある $\mathcal{G}$-可測で P-可積分な関数 $Y : \Omega \to \mathbb{R}$ が存在して，$\forall_{B\in\mathcal{G}}\ Q(B) = \int_B Y(d\omega)P(d\omega)$ が成り立つ．よって，(1),(2) が成り立つ．■

例 3.9.7 (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，

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

が成り立つ．

3.9.3 性質

$E[-|\mathcal{G}] : \mathcal{L}^1(\Omega, \mathcal{F}) \to L^1(\Omega, \mathcal{G})$ は関数空間上の正な線型作用素である．１次平均収束を保つという意味で，$\mathcal{L}^1(\Omega, \mathcal{F})$ 上ノルム連続，すなわち有界である．
また，Hilbert 空間 $L^2(\Omega, \mathcal{F})$ 上で見ると，$E[-|\mathcal{G}]$ は，$\mathcal{G}$-可測関数がなす閉部分空間への直交射影となる．

補題 3.9.8. $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F})$ とする．

(1) 線形性：$\forall_{a,b\in\mathbb{R}}\ E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$ a.s.. [†4]
(2) 正：$X \geq 0$ a.s. $\Rightarrow E[X|\mathcal{G}] \geq 0$ a.s.. 特に，$X \leq Y$ a.s. $\Rightarrow E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$ a.s..
(3) $X \in \mathrm{Meas}_{\mathcal{G}}(\Omega, \mathbb{R})$，$XY \in \mathcal{L}^1(\Omega, \mathcal{F})$ のとき，$E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$ a.s.. 特に，$E[X|\mathcal{G}] = X$ a.s..
(4) $\mathcal{H} \subset \mathcal{G}$ を部分 σ-代数とする．$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}]$ a.s.. 特に，$E[E[X|\mathcal{G}]] = E[X]$.
(5) $\sigma(X)$ と $\mathcal{G}$ とが独立ならば，$E[X|\mathcal{G}] = E[X]$ a.s.. したがって，$f : \mathbb{R} \to \mathbb{R}$ を Borel 可測関数とすると，$f(X) \in \mathcal{L}^1(\Omega, \mathcal{F}) \Rightarrow E[f(X)|\mathcal{G}] = E[f(X)]$ a.s..

［証明］．

(1) 右辺 $Z := aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$ は定義より，$\mathcal{G}$-可測関数で P-可積分である．任意の $B \in \mathcal{G}$ について，

$$\begin{aligned} E[aX + bY, B] &= aE[X, B] + bE[Y, B] \\ &= aE[E[X|\mathcal{G}], B] + bE[E[Y|\mathcal{G}], B] \\ &= E[aE[X|\mathcal{G}] + bE[Y|\mathcal{G}], B] = E[Z, B] \end{aligned}$$

が成り立つ．よって，条件付き期待値の一意性より，$E[aX + bY|\mathcal{G}] = Z$ a.s..
(2) $Z := E[X|\mathcal{G}]$ は $\forall_{B\in\mathcal{G}}\ E[Z, B] = E[X, B] \geq 0$ a.s. を満たす．これは $Z \geq$ a.s. を含意する．
(3) 右辺は $\mathcal{G}$-可測で P-可積分だから，任意の $B \in \mathcal{G}$ について $E[XY, B] = E[XE[Y|\mathcal{G}], B]$ を示せば，条件付き期待値の一意性から従う．単関数の場合から示す．
(4) 両辺とも $\mathcal{H}$-可測で P-可積分だから，任意の $B \in \mathcal{H}$ について

$$E[E[E[X|\mathcal{G}], \mathcal{H}], B] = E[X, B]$$

を示せば，条件付き期待値の一意性から従う．この左辺はまず $E[E[X|\mathcal{G}], B]$ に一致する必要があるが，$B \in \mathcal{G}$ でもあるから，これらはさらに $E[X, B]$ なる右辺に一致する必要がある．
(5) 独立性は，$\forall_{B\in\mathcal{G}}\ E[X, B] = E[1_B X] = E[X]P(B) = E[E[X], B]$ を含意する 3.5.4．これは $E[X|\mathcal{G}] = E[X]$ a.s. を意味する．また，$f(X)$ と $\mathcal{G}$ も独立である 3.4.2.

■

命題 3.9.9 (Jensen の不等式)**.** $\psi : \mathbb{R} \to \mathbb{R}$ を凸関数とする．$X, \psi(X) \in \mathcal{L}^1(\Omega)$ のとき，

$$\psi(E[X|\mathcal{G}]) \leq E[\psi(X)|\mathcal{G}] \text{ a.s.}$$

[†4] 除外集合 $N_{a,b}$ は任意の $a, b \in \mathbb{R}$ について一様に取ることは一般には出来ない．

特に，$X \in \mathcal{L}^p(\Omega)$ のとき，$|E[X|\mathcal{G}]|^p \le E[|X|^p|\mathcal{G}]$ a.s..

命題 3.9.10 (条件付き期待値の連続性)**.** $\mathcal{L}^1(\Omega)$ の列 (X_n) が X に 1 次平均収束するとき，$E[X_n|\mathcal{G}]$ も $E[X|\mathcal{G}]$ に 1 次平均収束する．

定理 3.9.11 (直交射影としての条件付き期待値)**.** $L^2_{\mathcal{G}}$ を，$L^2(\Omega,\mathcal{F})$ 内の $\mathcal{G}$-可測関数がなす閉部分空間とする．$Y \in L^2(\Omega,\mathcal{F})$ について，

$$E\left[(Y - E[Y|\mathcal{G}])^2\right] = \min_{Z \in L^2_{\mathcal{G}}} E[(Y-Z)^2].$$

3.9.4　可測写像を与えたもとでの条件付き期待値

部分 σ-代数 $i : (\Omega,\mathcal{G}) \hookrightarrow (\Omega,\mathcal{F})$ から，一般の可測写像へさらに一般化する．

記法 3.9.12. $(\Omega,\mathcal{F},P)$ 上の可積分確率変数 $X \in \mathcal{L}^1(\Omega)$ と，可測空間 $(\mathcal{T},\mathcal{B})$ への可測写像 $T : \Omega \to \mathcal{T}$ を考える．

定義 3.9.13. 次の 2 条件を満たす関数 $g : \mathcal{T} \to \mathbb{R}$ を，**$T = t$ の下での X の条件付き期待値**という：

(1) g は $\mathcal{B}$-可測かつ P^T 可積分．
(2) $\forall_{B\in\mathcal{B}}\ \displaystyle\int_{T^{-1}(B)} X(\omega)P(d\omega) = \int_B g(t)P^T(dt)$.

g は P^T-零集合を除いて一意であり，これを $E[X|T=t]$ で表す．すなわち，

$$\forall_{B\in\mathcal{B}} \quad \int_{T^{-1}(B)} X(\omega)P(d\omega) = \int_B E[X|T=t]P^T(dt).$$

補題 3.9.14 (well-definedness)**.**

$$E[X|\sigma(T)] = E[X|T]\ P\text{- a.s.}$$

ただし，$\sigma[T] := \left\{T^{-1}(B) \in P(\Omega) \mid B \in \mathcal{B}\right\}$ で，$E[X|T](\omega) := E[X|T=t]|_{t=T(\omega)}$ とした．

定義 3.9.15 (条件付き確率)**.**

(1) 部分 σ-加法族 $\mathcal{G} \subset \mathcal{F}$ に関する事象 $A \in \mathcal{F}$ の条件付き確率を，$P[A|\mathcal{G}] := E[1_A|\mathcal{G}]$ で定める．
(2) $T = t$ の下での事象 $A \in \mathcal{F}$ の条件付き確率を，$P[A|T=t] := E[1_A|T=t]$ で定める．

注 3.9.16. これは規格化しておらず，実際に A 上に確率測度を定めるかどうかは問うていない．

3.9.5　正則条件付き確率

定義 3.9.17 (regular conditional probability)**.** 族 $(p(\omega,A))_{w\in\Omega,A\in\mathcal{F}}$ が次の 3 条件を満たすとき，**部分 σ-代数 $\mathcal{G}$ が与えられたときの正則条件付き確率**であるという：

(1) $\forall_{\omega\in\Omega}\ p(\omega,-) : \mathcal{F} \to [0,1]$ は確率測度を定める．
(2) $\forall_{A\in\mathcal{F}}\ p(-,A) : \Omega \to [0,1]$ は $\mathcal{G}$-可測．
(3) $\forall_{A\in\mathcal{F}}\ \forall_{B\in\mathcal{G}}\ P(A\cap B) = \displaystyle\int_B p(\omega,A)P(d\omega)$.

定義 3.9.18. 可測空間 $(\Omega,\mathcal{F})$ が条件 (S) を満たすとは，

(1) [6] では，ある距離 d によって完備可分距離空間となり，$\mathcal{F}$ はその Borel σ-加法族となること．
(2) [4] では，可分完全確率空間とする．
(3) Ikeda-Watanabe では，標準確率空間とする．

応用上 (1) で十分らしいので，これでいこう．

定理 3.9.19 (存在と一意性)**.** 条件 (S) を満たす可測空間 $(\Omega,\mathcal{F})$ 上の，任意の確率測度 P と任意の部分 σ-代数 $\mathcal{G}$ に対して，$\mathcal{G}$ が与えられたときの正則条件付き確率 $(p(\omega,A))_{\omega\in\Omega,A\in\mathcal{F}}$ は存在し，零集合の差を除いて一意である．

定義 3.9.20. $T:(\Omega,\mathcal{F},P)\to(\mathcal{T},\mathcal{B}), X:(\Omega,\mathcal{F},P)\to(\mathcal{X},\mathcal{A})$ を可測とする．$(p(t,A))_{t\in\mathcal{T},A\in\mathcal{A}}$ が，**$T=t$ の下での X の条件付き確率分布**であるとは，次の 3 条件を満たすことをいう：

(1) $\forall_{t\in\mathcal{T}}\ p(t,-):\mathcal{A}\to[0,1]$ は $(\mathcal{X},\mathcal{A})$ 上の確率測度である．
(2) $\forall_{A\in\mathcal{A}}\ p(-,A):\mathcal{T}\to[0,1]$ は $\mathcal{B}$-可測関数である．
(3) $\forall_{A\in\mathcal{A}}\ \forall_{B\in\mathcal{B}}\ P[X\in A,T\in B]=\int_B p(t,A)P^T(dt)$.

定理 3.9.21 (存在と一意性)**.** 条件 (S) を満たす可測空間 $(\mathcal{X},\mathcal{A})$ 上の，$T=t$ の下での X の条件付き確率分布は存在し，零集合の差を除いて一意である．

命題 3.9.22 (Fubini の類似)**.** $f:\mathcal{T}\times\mathcal{X}\to\mathbb{R}$ は $\mathcal{B}\times\mathcal{A}$-可測で，$P^{(T,X)}$-可積分であるとする．このとき，

$$\int_{\mathcal{T}\times\mathcal{X}} f(t,x)dP^{(T,X)}(t,x)=\int_{\mathcal{T}}\left[\int_{\mathcal{X}} f(t,x)p(t,dx)\right]dP^T(t).$$

特に，可積分実確率変数 X に対して，

$$E[X|T=t]=\int_{\mathbb{R}} xp(t,dx)\ P^T\text{- a.s.}$$

命題 3.9.23 (多変量正規分布の条件付き分布は再び多変量正規分布となる)**.** d_1 次元確率変数 X_1 と d_2 次元確率変数 X_2 の結合分布は d_1+d_2 変量正規分布で，$E[X_i]=\mu_i, \mathrm{Cov}[X_i,X_j]=\Sigma_{ij}$ とおく．$\Sigma_{11}\in M_{d_1}(\mathbb{R})$ が正則ならば，$X_1=x_1$ の下での X_2 の正則条件付き分布は，多変量正規分布 $N_{d_2}(\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1),\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma 12)$ である．

注 3.9.24. Σ_{11} が退化しているときも，Σ_{11}^{-1} を一般化逆行列とすれば，同様の結果が成り立つ．

3.10　0-1 法則

末尾事象は，$\mathcal{F}$ の部分 σ-代数の列 $(\mathcal{B}_k)$ のうち，無限個によって指定される事象（各有限部分列と独立な事象）をいう．例えば，級数 $\sum_{n=0}^\infty X_n$ が収束するという事象は末尾事象である．
これにより，「情報」的な概念を完全に σ-加法族に翻訳しつつある．Borel-Cantelli の補題の一般化であることは明らかである．

定義 3.10.1 (tail field)**.** $(\mathcal{B}_k)$ を，$\mathcal{F}$ の部分 σ-代数の列とする．このとき，

$$\mathcal{G}_k:=\sigma\left(\bigcup_{j=k}^\infty\right)\mathcal{B}_j,\qquad \mathcal{T}:=\bigcap_{k=1\infty}\mathcal{G}_k$$

と定まる $\mathcal{T}$ を，**末尾加法族**という．

定理 3.10.2 (Kolmogorov 0-1)**.** $\forall_{A\in\mathcal{T}}\ P(A)\in\{0,1\}$.

系 3.10.3. (X_n) を独立な確率変数列とする．見本平均の概収束極限 $Y:=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n X_k$ が存在するとする．このおき，Y は殆ど確実に定数である．

第 4 章

独立確率変数列の和

確率過程論への入門として，独立確率変数列の和に関して成り立つ極限定理を調べる．実際，大数の法則の一般化・精緻化が喫緊の問題であった 1930 年代の確率論では最重要分野であった．この理論の一般化が martingale である．

(1) 見本平均は，和を規格化したものである（新たな測度を考えているとも捉えられる）．この極限が収束することは末尾事象で，概収束か概発散かが起こる．実は確率 1 で収束する．
(2) 弱法則は，対独立性と分散の一様有界性を必要とする．強法則は独立性に関しては強い条件を要求するが，分散の有界性については弱められる．
(3) 実は応用上もっとも中心的な興味は，大数の法則の収束の速度に関する情報である．実は，速度は $O(1/\sqrt{n})$ であり，1 次の項の形も標準的に得られるが，収束の強さは法則収束までである．

大偏差原理を見ると，確率論は関数解析と結びついて，現代の物理学を生んだような，とてつもない表現力を持ち得る可能性を感じる．

4.1 確率不等式

定理 4.1.1 (Kolmogorov). 実確率変数列 $\{X_n\} \subset \mathcal{L}^2(\Omega, \mathcal{F}, P)$ は独立で，$E[X_n] = 0, \mathrm{Var}[X_n] < \infty$ を満たすとする．このとき，

$$\forall_{a>0} \quad P\left(\max_{1\le k\le n}\left|\sum_{i=1}^k X_i\right| \ge a\right) \le \frac{1}{a^2}\sum_{i=1}^n \mathrm{Var}[X_i]$$

が成り立つ．

要諦 4.1.2 (martingale の萌芽).

4.2 独立同分布に関する大数の法則

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 4.2.1 (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：$\forall_{\epsilon>0}\ \lim_{n\to\infty} P(|Y_n - X| \ge \epsilon) = 0$.
(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：$P\left(\lim_{n\to\infty} Y_n = X\right) = 1$.

それぞれを形式化すると，

(1) $\forall_{\epsilon_1>0}\ \forall_{\epsilon_2>0}\ \exists_{N>0}\ \forall_{n\ge N}\ P(|Y_n - X| \ge \epsilon_1) < \epsilon_2$.
(2) $\forall_{\epsilon>0}\ \exists_{N>0}\ \forall_{n\ge N}\ P(Y_n = X) = 1 - \epsilon$.

すると，(2)⇒(1) であるが，(1)⇒(2) は反例が構成できる．[†1]

独立同分布の場合は期待値について

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

という評価が使えるので議論が簡単になる．

4.2.1　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 4.2.2. 確率空間 $(\Omega,\mathcal{F},P)$ 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 4.2.3 (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：

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

［証明］． Schwarz の不等式より，$E[|X_1|]\le\sqrt{E[(x_1)^2]}<\infty$ 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 1.7.5 より，

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

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 3.5.5 による．■

4.2.2　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 4.2.4 (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：

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

［証明］． 以下のことに注意する．

(1) Schwarz の不等式より $E[|X_1|]\le(E[(X_1)^2])^{1/2}\le(E[(X_1)^4])^{1/4}<\infty$ であるから特に可積分．
(2) 2 次と 4 次の中心積率は

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

と評価できる．

Chebyshev の不等式で期待値へ還元　$A_n^\epsilon := \left\{\omega\in\Omega \,\middle|\, \left|\frac{S_n(\omega)}{n}-m\right|>\epsilon\right\}$ と置くと，$f(a)=a^4$ についての Chebyshev の不等式 1.7.5 より，

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

と評価できる．

[†1] 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．

期待値を抑える　4 次の中心積率の和は

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

と評価できるから，

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

である．

Borel-Cantelli の補題　$B_\epsilon := \limsup_{n\to\infty} A_n^\epsilon = \left\{\omega\in\Omega \;\middle|\; \limsup_{n\to\infty}\left|\frac{S_n(\omega)}{n}-\mu\right| > \epsilon\right\}$ 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

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

■

4.2.3　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 4.2.5. $\forall_{k\in\mathbb{N}}\ E[|X_1|^k]<\infty$ のとき，任意の $\epsilon'>0$ に対して，$P\left(\lim_{n\to\infty}\frac{S_n-n\mu}{n^{1/2+\epsilon'}}=0\right)=1$．$\epsilon'=1/2$ のときを大数の強法則という．

定理 4.2.6 (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，

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

4.3　一般の大数の弱法則

定義 4.3.1. 一般の $\mathcal{L}^1(\Omega,\mathcal{F},P)$ の列 (X_n) に対して，$Y_n := \frac{1}{n}\sum_{k=1}^{n}X_n, \overline{m}_n := \frac{1}{n}\sum_{k=1}^{n}m_k$ とおく．

(1) $\lim_{n\to\infty}P(|Y_n-\overline{m}_n|>\epsilon)=0$ が成り立つとき，大数の弱法則を満たすという．

(2) $P\left(\lim_{n\to\infty}|Y_n-\overline{m}_n|=0\right)=0$ が成り立つとき，大数の強法則を満たすという．

定理 4.3.2. $\{X_n\}\subset\mathcal{L}^1(\Omega,\mathcal{F},P)$ が次の 2 条件を満たすとき，大数の弱法則を満たす：

(1) (X_n) は対独立である．

(2) $\sup_{n\in\mathbb{N}}\mathrm{Var}[X_n]<\infty$．特に，$\{X_n\}\subset\mathcal{L}^2(\Omega,\mathcal{F},P)$ である．

なお，確率収束するだけでなく，特に 2 次平均収束する．

[証明]．任意の $\epsilon>0$ を取る．

$$
P[|\widetilde{Y}_n|>\epsilon] \le \frac{E[\widetilde{Y}_n^2]}{\inf_{|x|>\epsilon}x^2} \qquad \because \text{Chebyshev の不等式 } 1.7.5
$$

$$
\begin{aligned}
&= \frac{E[\widehat{Y}_n^2]}{\epsilon^2} = \frac{1}{\epsilon^2}E[(Y_n - \overline{m_n})^2] \\
&= \frac{1}{\epsilon^2}E\left[\frac{1}{n^2}\left(\sum_{k=1}^n X_k - \sum_{k=1}^n m_k\right)^2\right] \\
&= \frac{1}{\epsilon^2 n^2}\sum_{j,k=1}^n E[(X_j - m_j)(X_k - m_k)] \\
&= \frac{1}{\epsilon^2 n^2}\sum_{k=1}^n E[(X_k - m_k)^2] && \because 対独立 \\
&= \frac{1}{\epsilon^2 n^2}\sum_{k=1}^n \mathrm{Var}[X_k] \\
&\le \frac{1}{\epsilon^2 n}\sup_{n\in\mathbb{N}} \mathrm{Var}[X_n] \xrightarrow{n\to\infty} 0.
\end{aligned}
$$

結局証明では，$E[|Y_n - \overline{m_n}|^2] \xrightarrow{n\to\infty} 0$ を示しているので，2 次平均収束する． ■

4.4　一般の大数の強法則

定理 4.4.1 (Kolmogorov 1). $\{X_n\} \subset \mathcal{L}^1(\Omega, \mathcal{F}, P)$ が次の 2 条件を満たすとき，大数の強法則を満たす：

(1) (X_n) は独立である．

(2) $\displaystyle\sum_{n=1}^\infty \frac{1}{n^2}\mathrm{Var}[X_n] < \infty$.

定理 4.4.2 (Kolmogorov 2). 独立同分布に従う確率変数の族 $\{X_n\} \subset \mathcal{L}^1(\Omega, \mathcal{F}, P)$ は，大数の強法則を満たす．

4.5　数学における大数の法則的現象

4.5.1　Weierstrass の多項式近似

Bernstein の基底関数 $b_{k,b}$ を Bernolli 試行 $B(n, x)$ の確率 $b(k; n, x)$ を表していると見ると，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，$k/n \to x$ 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 Bernolli 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 4.5.1 (Bernstein polynomial). $b_{k,b}(x) := \begin{pmatrix} n \\ k \end{pmatrix} x^k(1-x)^{n-k}$ $(k \in n+1)$ 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：$\displaystyle\sum_{k=0}^n b_{k,n} = 1$.

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

定理 4.5.2 (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，$\displaystyle\lim_{n\to\infty} \max_{x\in[0,1]} |P_n(x) - f(x)| = 0$.

［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 3.5.11)．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 $P(S_n = k) = \begin{pmatrix} n \\ k \end{pmatrix} x^k(1-x)^{n-k}$ と表せるため，

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

とも表せる．

検証　いま，$\delta : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ 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，$\epsilon \to 0$ のとき $\delta(\epsilon) \to 0$．また，$M := \max_{x\in[0,1]} |f(x)|$ とおくと，任意の $\epsilon > 0$ に対して，

$$\begin{aligned}
\max_{x\in[0,1]} |f(x) - P_n(x)| &= \max_{x\in[0,1]} \left| E\left[f(x) - f\left(\frac{S_n}{n}\right)\right]\right| && P_n(x) = E\left[f\left(\frac{S_n}{n}\right)\right] \\
&\leq \max_{x\in[0,1]} E\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \\
&= \max_{x\in[0,1]} \left\{ \int_{\left\{\omega\in\Omega \mid \left|\frac{S_n(\omega)}{n} - \mu\right| \geq \epsilon\right\}} \left|f(x) - f\left(\frac{S_n}{n}\right)\right| dP \right. \\
&\qquad \left. + \int_{\left\{\omega\in\Omega \mid \left|\frac{S_n(\omega)}{n} - \mu\right| < \epsilon\right\}} \left|f(x) - f\left(\frac{S_n}{n}\right)\right| dP \right\} \\
&\leq \max_{x\in[0,1]} 2MP\left(|S_n(\omega)n - x| \geq \epsilon\right) + \delta(\epsilon) && \text{第一項は } M \text{ の 2 倍で，第二項は} \delta(\epsilon) \text{ で抑えられる} \\
&\leq \max_{x\in[0,1]} \frac{2Me[|X_1^x - x|^2]}{n\epsilon^2} + \delta(\epsilon) && \text{大数の弱法則 4.2.3 と同様 } Chebyshev \text{ の不等式} \\
&\leq \frac{2M}{n\epsilon^2} + \delta(\epsilon) && B(1,x) \text{ の分散} \leq x(1-x) \leq 1
\end{aligned}$$

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 4.5.3. $f \in C([0,1])$ を，確率空間 $([0,1], \mathcal{B}([0,1]), P)$ 上の実確率変数だと思うと，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 $\lim_{n\to\infty} (n^{-1})^* f$ の極限だと考える：

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

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．

4.6 物理学における大数の法則的現象

4.6.1 Maxwell 分布

n-粒子系の速度の分布を考えたい．熱力学的平衡状態についていくつかの仮定をおくと，正規分布のクラスとして，Maxwell-Boltzmann 分布を得る．

記法 4.6.1. 半径 $\sqrt{n}$ の球面 $\sqrt{n}S^{n-1} \subset \mathbb{R}^n$ 上の一様確率測度を $\sigma_n(dx)$ とする．$k \leq n$ について，$(x_1, \cdots, x_k)$ 上の周辺分布を k 次周辺分布といい，$\sigma_n^k \in \mathcal{P}(\mathbb{R}^k)$ で表す．$\mu := \mu_{0,1} \in \mathcal{P}(\mathbb{R})$ を平均 0，分散 1 の標準正規分布とする．$\mu^k \in \mathcal{P}(\mathbb{R}^k)$ を k 重直積とすると，平均 $0 \in \mathbb{R}^k$，共分散行列 $I_k \in M_k(\mathbb{R})$ を持つ $\mathbb{R}^k$ 上の正規分布となる．

要諦 4.6.2. この条件は，n-粒子系の速度を x_i とし，条件 $\sum_{i=1}^n x_i^2 = n$ を規格化されたエネルギー保存則とする．そして，一様確率測度は，等重率の原理なる作業仮説によって置かれる仮定であり，これらの条件の下で起こり得るすべての事象は等しい確率を持つとする．なお，このような仮定によって条件付けられた分布を微視的正準 Gibbs 分布 (microcanonical Gibbs distribution) という．

定理 4.6.3. $\forall_{k\in\mathbb{N}}\ \sigma_n^k \Rightarrow \mu^k\ (n\to\infty)$.

4.6.2　熱力学的極限

記法 4.6.4. $\Lambda_L := [-L, L]^d \subset \mathbb{R}^d$ に閉じ込められた N-粒子系を考える．単位体積あたりの粒子数 $\frac{N}{(2L)^d}$ は一定であるとして，$L, N\to\infty$ の極限を考えたい．これを熱力学的極限という．このときの $\mathbb{R}^d$ 上の N-粒子の分布を，**強さ λ の Poisson 点過程**という．

$S := \Lambda_L, p(A) := \frac{m(A)}{(2L)^d}\ (A\in\mathcal{B}(\mathbb{R}^d)\cap P(\Lambda_L))$ が定める確率空間 $(\Omega := S^N, P := \prod_{i=1}^N p)$ を考えると，P は正準 Gibbs 分布である．

定理 4.6.5. 部分空間 $D\subset\Lambda_L$ に対して，その範囲で発見される粒子数を表す確率変数 $n(D,-):\Omega\to\mathbb{Z}$ を

$$n(D,\omega) := |\{k\in[N] \mid \omega_k\in D\}|$$

と定めると，

$$\forall_{l\in\mathbb{N}}\quad \lim_{L,N\to\infty,\frac{N}{(2L)^d}\to\lambda} P(n(D)=l) = e^{-\lambda m(D)}\frac{(\lambda m(D))^l}{l!}$$

4.7　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．

4.7.1　標準化された部分和についての結果

定理 4.7.1. 独立同分布を持つ確率変数列 $\{X_n\}\subset\mathcal{L}^1(\Omega,\mathcal{F},P)$ は $E[X_n]=m\in\mathbb{R}, \mathrm{Var}[X_n]=v\in\mathbb{R}_+$ を満たすとする．このとき，確率変数 $Z_n := \frac{1}{\sqrt{n}}\sum_{k=1}^n (X_k - m)$ は，正規分布 $N(0,v)$ に法則収束する．特に，

$$\forall_{a<b\in\mathbb{R}}\quad \lim_{n\to\infty} P(a\le Z_n\le b) = \frac{1}{\sqrt{2\pi v}}\int_a^b e^{-\frac{x^2}{2v}}dx.$$

4.7.2　一般化

独立同分布に従う確率変数の分散が有限な場合，それらの和の確率分布は正規分布に収束する．これを種々の作業仮説をおいて示すことが出来る．一方で，確率変数が従う分布の裾が重く（$|x|^{-\alpha-1}$ $(0<\alpha<2)$ の冪乗)，分散が発散する場合，正規分布には収束せず，特性指数 α の安定分布に収束する．

定理 4.7.2. $(k_n)_{n\in\mathbb{N}}:\mathbb{N}\to\mathbb{N}$ に対して，$(\xi_{n,j})_{j\in[k_n]}$ を各 $n\in\mathbb{N}$ が定める d 次元独立確率変数列とする．次の 2 条件を仮定する．

(A1) $\xi_{n,j}\in\mathcal{L}^2(\mathbb{R}^d)$ は 2 乗可積分で，期待値ベクトルは零とする：$E[\xi_{n,j}]=0$.

(A2) (Lindeberg) $\Sigma_n := \sum_{j=1}^{k_n} \mathrm{Var}[\xi_{n,j}] \in M_d(\mathbb{R})$，$\Sigma := \lim_{n\to\infty} \Sigma_n$ としたとき，次が成り立つ：

$$\forall_{\epsilon>0}\ \lim_{n\to\infty} \sum_{j=1}^{k_n} E[|\xi_{n,j}|^2 1_{\{|\xi_{n,j}|\ge\epsilon\}}] = 0.$$

このとき，

$$\sum_{j=1}^{k_n} \xi_{n,j} \xrightarrow{d} N_d(0,\Sigma) \quad (n\to\infty)$$

要諦 4.7.3. 通常は $(k_n)_{n\in\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ を恒等写像として，$\xi_{n,j} := \dfrac{X_j}{\sqrt{n}}$ として Lindeberg 条件を満たすものを構成している．

例 4.7.4. $X_j = (Y_j, Z_j)$ は独立に同一の 2 変量正規分布 $N_2(\mu,\Sigma)$ に従うとする．$\rho := \rho(Y_j, Z_j)$ とする．標本相関係数 $\widehat{\rho}_n$ の漸近分布を求めたい．中心極限定理とデルタ法 2.4.10 を組み合わせることで，$\sqrt{n}(\widehat{\rho}_n - \rho) \xrightarrow{d} N(0,(1-\rho^2)^2)$ とわかる．

さらに良い結果を引き出すために，関数

$$g(\rho) := \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$$

による変換 $g(\widehat{\rho}_n)$ を考える．これを **Z 変換**という．これについてもう一度デルタ法を用いると，$\sqrt{n}(g(\widehat{\rho}_n) - g(\rho)) \xrightarrow{d} N(0,1)$ となり，漸近分散はパラメータ ρ に依存しなくなる．このような変換を**分散安定化変換**という．

4.8　経験分布に対する拡張

定理 4.8.1 (Glivenko-Cantelli). F_n を経験分布関数，F を真の分布関数とする．経験分布関数は，分布関数の一致推定量である：$P[\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0] = 1$.

議論 4.8.2. F_n の推定量としての誤差分布を考えたい．各点 $x \in \mathbb{R}$ 毎に見ると，$\sqrt{n}(F_n(x) - F(x))$ は，$\mathcal{X}^n$ 上に定まり，$\mathbb{R}$ 上に値を取る確率変数で，$N(0, F(x)(1-F(x)))$ に分布収束する．結合分布は多変量正規分布に収束する．では，$\sqrt{n}(F_n - F)$ 自体はどうか？これは，標本が与えられる毎に分布関数を確定させる $\mathcal{X}^n \to l^\infty(\mathbb{R})$ なる確率変数で，F-ブラウニアン橋と呼ばれる関数 $\mathbb{G}_F : \mathcal{X}^\infty \to l^\infty(\mathbb{R})$ が定める分布に弱収束する．Banach 空間値確率変数の共分散は定義していないが（いわば無限の成分を持つ行列 $\mathbb{R}^2 \to \mathbb{R}$），各組 $(x_k, x_l) \in \mathbb{R}^2$ に対して $\mathrm{Cov}[\mathbb{G}_F(x_k), \mathbb{G}_F(x_l)] = F(x_k \wedge x_l) - F(x_k)F(x_l)$ が成り立つ．

4.9　Poisson の少数の法則

記法 4.9.1. これより，単位時間内に電話がかかってくる回数や事故が起こる回数などは，Poisson 分布によって近似するのが筋が良いことがわかる．

記法 4.9.2. 点列 $\{p_n\} \subset (0,1)$ は $\lim_{n\to\infty} np_n = \lambda > 0$ を満たすとし，これが定める n 回の独立同試行 $(S^n := 2^n, \mathcal{F}, P), (\mathrm{pr}_n)_* P(\{1\}) = p_n$ の列を考える．これは，n が増加するにつれて，1 が出る確率は小さくなっていくが，n 回が終わってみたときに 1 が出る回数の平均は λ で変わらないようになっている．

定理 4.9.3. $Z_n : S^n \to \mathbb{Z}$ を S^n 上の 1-ノルムとする（1 が出る回数）．このとき，Z_n は $\mathrm{Pois}(\lambda)$ に法則収束する：$\forall_{l\in\mathbb{N}}\ \lim_{n\to\infty} P(Z_n = l) = e^{-\lambda}\dfrac{\lambda^l}{l!}$.

4.10 大偏差原理

無限次元空間における Laplace 原理

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．大数の法則から漏れた部分 a（＝偏差 $|a-m|$ の大きいもの）の確率は 0 に収束し $P(X_n=a)\xrightarrow{n\to\infty}0$，中心極限定理により指数減衰 $P(X_n\ge a)=\int_{z\ge a}p_n^{X_n}(z)e^{-nI(z)}dz=e^{-I(a)}$（第 2 項あやしい）をするのであるが，そのときの係数 $I(z)$ は，「大数の法則が指定する集中点に一番近い事象」すなわち「起こりにくい事象の中で最も起こりやすい事象」が支配するという原理である．これは Laplace の原理の無限次元版だとみなすと筋が良い．

歴史 4.10.1.

(1) 1929 に Khinchin が Bernoulli 列について扱う．
(2) 1938 に Cramer が $\exists_{t>0}\ E[e^{t|X_1|}]<\infty$ の条件の下で一般化．
(3) 1966 に Schilder が確率過程＝汎関数型の大偏差原理を定立．
(4) 1970s に Donsker-Varadhan の理論が生まれる．

4.10.1 定義

記法 4.10.2. X を可分完備距離空間とし，同時にある σ-代数 $\mathcal{F}$ により可測空間でもあるとする．

定義 4.10.3 (large deviations technique). $(X,\mathcal{F})$ 上の確率測度の列 $(\mu_n)_{n\in\mathbb{N}}$ が大偏差原理をみたすとは，ある下半連続関数 $I:X\to\mathbb{R}_+$ が存在して，任意の可測集合 $\Gamma\in\mathcal{F}$ に対して

$$-\inf_{x\in\Gamma^\circ}I(x)\le\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(\Gamma)\le\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(\Gamma)\le-\inf_{x\in\overline{\Gamma}}I(x)$$

が成り立つことをいう．このとき，I を **rate 関数**という．

定義 4.10.4 (good rate function). I が**良いレート関数**であるとは，任意の $l\ge0$ に対して，等位集合 $I^{-1}((-\infty,l])=\{x\in X\mid I(x)\le l\}$ がコンパクトであることをいう．

4.10.2 Laplace の原理

「大きなパラメータを持つ指数関数の積分の漸近挙動は，被積分関数の最大値付近からの寄与だけで決まる」という経験則である．

例 4.10.5. 有界閉区間 $[a,b]$ 上の連続関数 $f,g>0$ について，

$$\int_a^b e^{nf(x)}g(x)dx\approx e^{n\max_{x\in[a,b]}f(x)}\quad(n\to\infty)$$

が成り立つ．ただし，$F(n)\approx G(n):\Leftrightarrow\dfrac{\log F(n)}{\log G(n)}\xrightarrow{n\to\infty}1$ と定めた．

注 4.10.6. この結果を非有界な区間に一般化しようとすると，種々の技術的問題が生じる．が，同様の結果は成り立つことが多い．そこで「原理」と呼ばれている．

例 4.10.7. （対数の比を考えることで）定数倍を無視した弱い形の Stirling の公式 $n!=\int_0^\infty x^ne^{-x}dx\approx n^ne^{-n}\ (n\to\infty)$ を Laplace の原理から説明する．積分変換によって

$$n^n\int_0^\infty y^ne^{-ny}dy=n^n\int_0^\infty ne^{n\log y-y}dy$$

と書き直せる．すると，$\max_{y\in\mathbb{R}_+}(\log y-y)=-1$ と $\log y-y\xrightarrow{y\to 0}-\infty, \log y-y\xrightarrow{\infty}-\infty$ より，最大値 e^{-n} だけが積分に寄与することが予想される．

4.10.3　Cramer の理論

$m:=E[X_1]<\infty$ に対して，$A\in\mathcal{F}$ が $d(m,A)>0$ を満たせば，$\lim_{n\to\infty}P(Y_n\in A)=0$ であるが，このときの収束の速さは指数的に減衰する．

定理 4.10.8. 独立同分布を持つ確率変数列 (X_i) の積率母関数の定義域 $D_\nu:=\left\{t\in\mathbb{R}\mid M(t):=E[e^{tX_1}]<\infty\right\}$ は，0 を内点として持つとする．このとき，$S_n:=\sum_{i=1}^n X_i$ とおく．

(1) $\forall_{a>E[X_1]}\ \lim_{n\to\infty}\frac{1}{n}\log P(S_n\ge an)=-I(a)$．ただし，$I(z):=\sup_{t\in\mathbb{R}}(zt-\log M(t))$ をキュムラント母関数 $\log M$ の Legendre 変換とした．

(2) I は $\mathbb{R}$ 上下半連続な凸関数であり，$\lim_{z\to\pm\infty}I(z)=\infty, \forall_{z\in\mathbb{R}}\ I(z)\ge 0=I(E[X_1])$ を満たす．

要諦 4.10.9. $E[X_1]<a$ について $A:=[a,\infty)$ とおくと，

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}S_n\in A\right)=-\inf_{z\in A}I(z)$$

と書き換えられる．すなわち，事象 $\frac{S_n}{n}\in[a,\infty)$ が成り立つとき，$n\to\infty$ の目で見ると，殆ど a の近くの値を取るような事象によって実現されている」ということである．

4.10.4　Schilder の理論

確率過程に関する大偏差原理

$I(\phi)$ を，連続関数 ϕ の持つエネルギーだとすると，「起こりにくい事象の中では最も起こりやすい事象＝エネルギーが最小になる事象が起こる」という大偏差原理は，物理現象としても極めて自然な現象であることがわかる．

4.10.5　Varadhan の理論

大偏差原理は，Laplace の原理が有界線型汎関数の列に対しても成り立つための十分条件を与えると捉えると，その指数関数への注目と，位相のことばを用いた定義が自然に思える．

補題 4.10.10. $(X,\mathcal{F})$ 上の確率測度の列 (μ_n) が，I を良いレート関数として大偏差原理を満たすとする．このとき，X 上の有界連続関数 $f\in C_b(X)$ について，

$$\int e^{nf(x)}\mu_n(dx)\approx\exp\left(n\sup_{x\in X}|f(x)-I(x)|\right)\quad(n\to\infty)$$

第 5 章

マルチンゲール

確率変数の族を確率過程という．個々の確率変数の値域となる位相空間を状態空間という．確率変数族には独立性の概念が拡張できたが，これは応用上自然ではない．遥かに緩いクラスとして，マルチンゲールを定義する．1930 年代に，独立確率変数の和の理論を整備する過程で豊かに育った Kolmogorov のアイデアを一般化する試みの中で，Levy がマルチンゲールの概念を発明し，Doob が理論を立てた．Brown 運動も確率積分もマルチンゲールになる．

連続確率過程のマルチンゲールは，離散化したあとに適当な連続極限を取ることで，離散の場合の議論に帰着させることが出来る．

解析学に可測関数，連続関数，解析関数というようなクラスがあるように，確率論にもマルチンゲール，加法過程，Markov 過程，定常過程などのクラスがある．解析学に指数関数，Bessel 関数などの特殊関数があるように，確率論にも Weiner 過程，Poisson 過程というような特殊過程がある．ただし，分類の指導方針が全く違う．確率論の指導原理は独立性であってきた．

5.1 関数空間 C と D

5.1.1 ポーランド空間

補題 5.1.1. ポーランド空間（完備可分距離空間）の可算積はポーランド空間である．

定理 5.1.2. S 上のポーランド位相 τ と Hausdorff 位相 τ_1 を考える．$\tau_1\subset\tau$ ならば，これらが定める Borel σ-代数は一致する：$\mathcal{B}_\tau(S)=\mathcal{B}_{\tau_1}(S)$.

5.1.2 Kolmogorov σ-代数

定義 5.1.3. 集合 T 上の関数の空間 $\mathcal{F}\subset\mathrm{Map}(T,\mathbb{R})$ における σ-代数を考える．任意の $t\in T$ に対して，射影 $\mathrm{ev}_t:\mathcal{F}\to\mathbb{R}$ が可測となるような $\mathcal{F}$ 上の σ 加法族 $\mathcal{B}$ の中で最小のものを $\mathcal{B}_K(\mathcal{F})$ で表し，$\mathcal{F}$ 上の **Kolmogorov σ-加法族**という．

補題 5.1.4. $\mathcal{B}_K(\mathcal{F})=\bigvee_{t\in T}\pi_t^{-1}(\mathcal{B}^1)$ である．すなわち，Kolmogorov σ-代数は，$\{\pi_t^{-1}(\mathcal{B}^1)\}_{t\in T}$ が束 $P(\mathcal{F})$ の中でなす下限となる．

5.1.3 D 空間

関数解析では考察の対称となったことはないようであるが，確率論では C と同様に重要である．

記法 5.1.5. $I\subset\mathbb{R}$ を開区間とする．

定義 5.1.6. $f:I\to\mathbb{R}$ が**第 1 種不連続**または **cadlag** または右連続であるとは，I 上の各点で，右連続かつ有限な左極限が存在することを言う．これらの全体を $D(I)$ で表す．

5.2　確率過程に関する一般事項

5.3　情報と情報増大系

情報は σ-部分代数で，データは確率変数で表すとしたら，2 つの構造が何らかの意味で整合して居る必要がある．これを適合的という．

5.3.1　適合的な情報系

定義 5.3.1 (filtration / reference family)**.** $\mathcal{F}$ の部分 σ-加法族の広義単調増大列 $(\mathcal{F}_n)$ を，**フィルトレーション**または**（増加）情報系**という：$0 \le s \le t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t$.

定義 5.3.2 (adapted, predictable)**.** $F = (F_t)_{t\in T}$ 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．

例 5.3.3 (canonical filtration)**.** 確率過程 (X_n) に対して，$\mathcal{F}_n := \sigma[X_1, \cdots, X_n]$ で定まる列 $(\mathcal{F}_n)$ は適合的な情報系である．これを**自然な増加情報系**という．

5.3.2　情報系の連続性

定義 5.3.4.

(1) 情報系 $(\mathcal{F}_t)$ が**右連続**であるとは，$\mathcal{F}_{t+} := \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}$ としたとき，$\forall_{t\ge 0}\ \mathcal{F}_t = \mathcal{F}_{t+}$ が成り立つことをいう．
(2) 確率過程 (X_t) が**右連続**であるとは，任意の $t \in T$ に対して $X_t \in D(\Omega)$ が cadlag であることをいう．

例 5.3.5. 確率過程 (X_n) に対して，$\mathcal{F}_t := \cap_{\epsilon>0}\sigma[X_s \mid s \le t+\epsilon]$ と定めると，右連続で適合的な σ-部分代数となる．

5.4　停止時

確率過程 (X_n) を調べる時，時刻 $\mathbb{N}$ をランダムに抽出して調べたいことがある．ランダムな時刻のうち，特に振る舞いが良いクラスを用意する．

5.4.1　定義と例

定義 5.4.1 (Markov time / stopping time)**.**

離散　$\overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$-値確率変数 $\tau : (\Omega, \mathcal{F}) \to (\overline{\mathbb{N}}, \mathcal{B}(\overline{\mathbb{N}}))$[†1] が $(\mathcal{F}_n)$-**Markov 時刻**または**停止時刻**であるとは，

$$\forall_{n\in\mathbb{N}} \quad \{\tau \le n\} := \{\omega \in \Omega \mid \tau(\omega) \le n\} \in \mathcal{F}_n$$

を満たすことをいう．

連続　$\tau : \Omega \to [0, \infty]$ が $(\mathcal{F}_t)$-Markov 時刻であるとは，$\forall_{t\ge 0}\ \{\tau \le t\} := \{\omega \in \Omega \mid \tau(\omega) \le t\} \in \mathcal{F}_t$ が成り立つことを言う．

要諦 5.4.2. ランダムな時刻 $\tau : \Omega \to \overline{\mathbb{N}}$ が，時刻 n の前であるかあとであるかという事実だけは，$\mathcal{F}_n$ によって判断出来るとき，これを停止時刻というのである．

[†1] $\mathcal{B}(\overline{\mathbb{N}}) = P(\overline{\mathbb{N}})$ に注意

補題 5.4.3 (離散の場合の特徴付け). $\tau : \Omega \to \overline{\mathbb{N}}$ について，次の 2 条件は同値.

(1) τ は Markov 時刻である.

(2) $\forall_{n \in \mathbb{N}}\ \{\tau = n\} \in \mathcal{F}_n$ である.

補題 5.4.4 (連続の場合の特徴付け). $\tau : \Omega \to [0, \infty]$ について，次の 2 条件は同値.

(1) τ は Markov 時刻である.

(2) $\forall_{t \geq 0}\ \{\tau < t\} \in \mathcal{F}_t$.

(3) $\forall_{t \geq 0}\ \{\tau > t\} \in \mathcal{F}_t$.

(4) $\forall_{t \geq 0}\ \{\tau \geq t\} \in \mathcal{F}_t$.

例 5.4.5 (離散の例).

(1) 定値関数は Markov 時刻である.

(2) **到達時刻** (first hitting time) とは，$(\mathcal{F}_n)$-適合確率過程 (X_n) 対して，任意の事象 $A \in \mathcal{B}(\mathbb{R})$ に対し，

$$\tau_A(\omega) := \min\left\{n \in \overline{\mathbb{N}} \mid X_n(\omega) \in A\right\}$$

で定まる時刻である．ただし，$\min \emptyset = \infty$ とする．$\{\tau_A \leq n\} = \bigcup_{i \in [n]} \{X_i \in A\}$ より，Markov 時刻である.

(3) ある一定額 α 以上を賭けたら即座に賭けを中止すると決めているとき，この時刻は Markov 時刻である.

(4) **最終脱出時刻** (last exit time) とは，$(\mathcal{F}_n)$-適合確率過程 (X_n) 対して，任意の事象 $A \in \mathcal{B}(\mathbb{R})$ に対し，

$$\sigma_A(\omega) := \max\left\{n \in \overline{\mathbb{N}} \mid X_n(\omega) \in A\right\} + 1$$

とすると，これは確率変数ではあるが，Markov 時刻にはならない.「これが最後か？」を判定するには，さらに先の情報が必要だからである.

例 5.4.6. (X_t) の集合 $A \subset \mathbb{R}^d$ への到達時刻は

$$\tau_A(\omega) := \inf\{t > 0 \mid X_t(\omega) \in A\}$$

で定める．A が開または閉であるとき，τ_A は Markov 時刻になる.

5.4.2　構成

定理 5.4.7 (Markov 時刻の構成). τ, σ を Markov 時刻とする．このとき，$\tau \vee \sigma, \tau \wedge \sigma, \tau + \sigma$ はいずれも Markov 時刻である.

5.4.3　情報量

定義 5.4.8. τ を $(\mathcal{F}_n)$-Markov 時刻とする．τ が定める $\mathcal{F}$ の部分 σ-代数

$$\mathcal{F}_\tau := \{A \in \mathcal{F} \mid \forall_{n \in \mathbb{N}}\ A \cap \{\tau \leq n\} \in \mathcal{F}_n\}$$

を τ **時までの情報量**という.

要諦 5.4.9. $A \in \mathcal{F}_\tau$ であるとは，τ が n 時以前に起こっているという追加情報もつけた事象 $A \cap \{\tau \leq n\}$ は $\mathcal{F}_n$ によって判断可能であることをいう．τ が定数 m のとき，$\mathcal{F}_\tau = \mathcal{F}_m$ となる.

定理 5.4.10. τ, σ を Markov 時刻とする．次が成り立つ.

(1) $\mathcal{F}_\tau$ は σ-代数である.

(2) τ は $\mathcal{F}_\tau$-可測である.

(3) $\tau \leq \sigma$ ならば $\mathcal{F}_\tau \subset \mathcal{F}_\sigma$.

5.5 離散時変数のマルチンゲール

5.5.1 定義

定義 5.5.1 (martingale, submartingale). 確率過程 (X_n) が**情報系** $(\mathcal{F}_n)$ **についてマルチンゲール**であるとは，次の 3 条件が成り立つことをいう：

(1) $(\mathcal{F}_n)$-適合的である：$\forall_{n\in\mathbb{N}}\ X_n : \mathcal{F}_n$ measurable.
(2) 可積分列である：$\forall_{n\in\mathbb{N}}\ E[|X_n|] < \infty$.
(3) martingale 性：$\forall_{n\in\mathbb{N}}\ E[X_{n+1}|\mathcal{F}_n] = X_n$ a.s.

(3) の代わりに $\forall_{n\in\mathbb{N}}\ E[X_{n+1}|\mathcal{F}_n] \ge X_n$ a.s. が成り立つとき，**劣マルチンゲール**であるといい，$\forall_{n\in\mathbb{N}}\ E[X_{n+1}|\mathcal{F}_n] \le X_n$ a.s. が成り立つとき，**優マルチンゲール**であるという．

要諦 5.5.2. (3) は $\forall_{A\in\mathcal{F}_n}\ E[X_{n+1}1_A] = E[X_n 1_A]$ と同値．これが成り立つならば，繰り返し期待値の法則より，$m > n \ge 1 \Rightarrow E[X_m|\mathcal{F}_n] = X_n$ a.s. であり，特に，$E[X_n]$ は n に依らず一定である．また，(X_n) が劣マルチンゲールであることと，$(-X_n)$ が優マルチンゲールであることとは同値.

例 5.5.3.

(1) $\{Z_n\} \subset \mathcal{L}^1(\Omega)$ を期待値 0 かつ独立な確率変数列とし，$\mathcal{F}_n$ として自然な情報系を取る．和 $X_n := \sum_{k=1}^n Z_k$ はマルチンゲールである．$E[X_{n+1}|\mathcal{F}_n] = E[X_n + Z_{n+1}|\mathcal{F}_n] = E[X_n|\mathcal{F}_n] + E[Z_{n+1}|\mathcal{F}_n]$ であるが，第 1 項は X_n は $\mathcal{F}_n$ 可測であるから，$E[X_n|\mathcal{F}_n] = X_n$．また，$\sigma(Z_{n+1})$ と $\mathcal{F}_n$ は独立だから 3.5.2，$E[Z_{n+1}|\mathcal{F}_n] = E[Z_{n+1}] = 0$ a.s.．これは Kolmogorov の不等式 4.1.1 ですでにあった消息である．
(2) $(\mathcal{F}_n)$ を情報系とし，$X \in \mathcal{L}^1(\Omega)$ を可積分確率変数とする．$X_n := E[X|\mathcal{F}_n]$ とおけば，(X_n) はマルチンゲールである．実際，$E[X_{n+1}|\mathcal{F}_n] = E[E[X|\mathcal{F}_{n+1}]|\mathcal{F}_n] \overset{\text{a.s.}}{=} E[X|\mathcal{F}_n] = X_n$.

(1) の状況は公平な賭けなどの意味論を持つ．コイントスをして，表なら $+x$ 円，裏なら $-x$ 円の賭けで，所持金を X_n とすると，これはマルチンゲールである．

補題 5.5.4 (劣マルチンゲール性の保存).

(1) $\psi : \mathbb{R} \to \mathbb{R}$ は下に凸，(X_n) をマルチンゲールとする．このとき，$\forall_{n\in\mathbb{N}}\ E[|\psi(X_n)|] < \infty$ ならば，$(\psi(X_n))$ は劣マルチンゲールである．特に，ある $p \ge 1$ に関して $E[|X_n|^p] < \infty$ ならば，$(|X_n|^p)$ は劣マルチンゲールである．
(2) 下に凸な関数 $\psi : \mathbb{R} \to \mathbb{R}$ はさらに広義単調増加であるならば，(X_n) が劣マルチンゲールの場合でも，$(\psi(X_n))$ は劣マルチンゲールになる．

5.5.2 Doob 分解

定理 5.5.5 (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 $M = (M_n)$ と可予測な広義増加過程 $A = (A_n)$，すなわち，$0 = A_0 \le A_1 \le \cdots, A_n \in L^1(\mathcal{F}_{n-1})$ $(n = 1, 2, \cdots)$[†2]を満たす列 (A_n) とに一意的に分解される：$X_n = M_n + A_n$ a.s..

5.5.3 Doob の任意抽出定理

2 つのランダム関数 X, τ の交錯を考える．

定義 5.5.6. $(\mathcal{F}_n)$-適合な確率過程 (X_n) と，$\mathbb{N}$-値 Markov 時刻 τ に対して，$X_\tau : \Omega \to \mathbb{R}$ を $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$ で定める．

---

†2 マルチンゲールに対して $A_n \in L^1(\mathcal{F}_{n-1})$ とは，$\mathcal{F}_{n-1}$-可測の意味しかない．

補題 5.5.7. $X_\tau : \Omega \to \mathbb{R}$ は

(1) $\mathcal{F}$-可測である．
(2) $\mathcal{F}_\tau$-可測である．

［証明］．

(1) X_τ は可測関数の合成 $X \circ (\tau, \mathrm{id}) : \Omega \to \mathbb{N} \times \Omega \to \mathbb{R}$ であるため．

■

定理 5.5.8 (有界停止時刻によるマルチンゲール性の保存)**.** (X_n) を劣マルチンゲール，τ, σ を $\tau \le \sigma$ を満たす有界な Markov 時刻とする．このとき，X_τ, X_σ は共に可積分で，$E[X_\sigma|\mathcal{F}_\tau] \ge X_\tau$ a.s.. (X_n) がマルチンゲールであるとき，等号成立．

系 5.5.9 (optional sampling theorem)**.** (X_n) を $(\mathcal{F}_n)$-劣マルチンゲール，(τ_k) を有界な $(\mathcal{F}_n)$-マルコフ時刻の広義単調増加列とする．このとき，$Y_k := X_{\tau_k}$ は $(\mathcal{F}_{\tau_k})$-劣マルチンゲールである．

5.5.4　Doob の不等式

劣マルチンゲールに対しては，$\max\limits_{1 \le k \le n} X_K$ に関する評価を，X_n のみを用いて与えられる．一般の確率過程では決して成り立たない．この背後には Kolmogorov の不等式 4.1.1 がある．

定理 5.5.10 (Doob inequality)**.** (X_n) を劣マルチンゲールとする．このとき，$X_n^+ := X \vee 0$ とすると，任意の $a > 0$ について，次が成り立つ．

(1)
$$P\left(\max_{1 \le k \le n} X_k \ge a\right) \le \frac{1}{a} E\left[X_n 1_{\{\max_{1 \le k \le n} X_k \ge a\}}\right] \le \frac{1}{a} E[X_n^+]$$

(2)
$$P\left(\min_{1 \le k \le n} X_k \le -a\right) \le \frac{1}{a} E[X_n - X_1] - \frac{1}{a} E\left[X_n 1_{\{\min_{1 \le k \le n} X_k \le -a\}}\right] \le \frac{1}{a} E[X_n^+] - \frac{1}{a} E[X_1].$$

系 5.5.11. $\{M_n\} \subset \mathcal{L}^p(\Omega)$ を $p \ge 1$ 乗可積分なマルチンゲールとする．このとき，任意の $a > 0$ に対して，

$$P\left(\max_{1 \le k \le n} |M_k| \ge a\right) \le \frac{1}{a^p} E[|M_n|^p].$$

5.5.5　劣マルチンゲールの収束定理

劣マルチンゲールの正部分の期待値が「有界」ならば，X_n は概収束極限を持つ．その証明では，劣マルチンゲールの上向き横断回数の評価が肝要になる．

5.5.5.1　martingale 変換

定義 5.5.12 (martingale transformation)**.** 可予測な過程 (H_n) と $(\mathcal{F}_n)$-適合的な過程 (X_n) に対して，新たな確率過程 $(X'_n) := ((H \cdot X)_n)$ を次のように定める

$$X'_n = (H \cdot X)_n := \begin{cases} \displaystyle\sum_{k=2}^n H_k (X_k - X_{k-1}), & n \ge 2, \\ 0, & n = 1. \end{cases}$$

$(H \cdot X)_n$ を X_n の**マルチンゲール変換**という．連続時間の場合は，確率積分 $\displaystyle\int_0^t H dX$ となる．

要諦 5.5.13. (H_n) は戦略を表し，(X_n) は $\mathbb{Z}$ 上のランダムウォークとすれば，これによる変換 $(H \cdot X)_n$ は n 時に所持している利益分の金額となる．

例 5.5.14. 倍賭けの戦略は，次のように表せる．

$$H_n := \begin{cases} 2H_{n-1}, & Z_{n-1} = -1, \\ 1, & Z_{n-1} = 1. \end{cases}$$

定理 5.5.15. 可予測な確率過程 (H_n) は有界な列とする：$\forall_{n\in\mathbb{N}} \sup_{\omega\in\Omega}|H_n(\omega)| < \infty$．このとき，次が成り立つ．

(1) (X_n) がマルチンゲールならば，$(X'_n) = ((H\cdot X)_n)$ もマルチンゲールである．
(2) (X_n) が劣マルチンゲールで，(H_n) が非負ならば，$(X'_n) = ((H\cdot X)_n)$ も劣マルチンゲールである．

5.5.5.2　上渡回数定理

マルチンゲールは，（少なくとも期待値については）単調に増加する傾向があり，いつまでも区間 $[a,b]$ 付近にとどまってはいないか，概収束をする．

定義 5.5.16 (upcrossing number).

(1) 実数 $a < b$ について，列 (σ_j) を次のように定めると，Markov 時刻の狭義単調増加列となる：

$$\sigma_1 := \min\{n \geq 1 \mid X_n \leq a\}, \qquad \sigma_2 := \min\{n > \sigma_1 \mid X_n \geq b\},$$
$$\sigma_{2k+1} := \min\{n > \sigma_{2k} \mid X_n \leq a\}, \qquad \sigma_{2k+2} := \min\{n > \sigma_{2k+1} \mid X_n \geq b\}.$$

(2) Markov 時刻の狭義単調増加列に対して，$U_n := \max_{k\in\mathbb{N}\mid\sigma_{2k}\leq n}$ と定めると，$\mathbb{N}$-値確率変数の列となる．成分 U_n を，**時刻 n までの $a \nearrow b$ 間の上向き横断回数**という．

定理 5.5.17. (X_n) が劣マルチンゲールならば，

$$E[U_n] \leq \frac{1}{b-a}E[(X_n-a)^+].$$

5.5.5.3　劣マルチンゲールの収束定理

劣マルチンゲールに対しても，有界列は収束することに対応する結果が成り立つ．

定理 5.5.18. 劣マルチンゲール (X_n) は $\sup_{n\in\mathbb{N}} E[X_n^+] < \infty$ を満たすとする．このとき，ある確率変数 X が存在して $X = \lim_{n\to\infty} X_n$ a.s. かつ可積分 $X \in \mathcal{L}^1(\Omega)$ である．

要諦 5.5.19. 劣マルチンゲールに対して，有界性条件 $\sup_{n\in\mathbb{N}} E[X_n^+] < \infty$ は，平均の一様有界性 $\sup_{n\in\mathbb{N}} E[|X_n|] < \infty$ に同値．

5.5.6　積率不等式

Doob の不等式 5.5.10 を，マルチンゲールの p 次のモーメントに関する評価式に書き直せる．

5.5.6.1　Doob の不等式の一般化

定理 5.5.20. $p > 1$ について，$\{M_n\} \subset \mathcal{L}^p(\Omega)$ を p 乗可積分なマルチンゲールとする．このとき，

$$E\left[\max_{1\leq k\leq n}|M_k|^p\right] \leq \left(\frac{p}{p-1}\right)^p E[|M_n|^p].$$

5.5.6.2　Burkholder の不等式

記法 5.5.21. $\{M_n\} \subset \mathcal{L}^2(\Omega)$ を，$M_0 = 0$ を初項とする 2 乗可積分なマルチンゲールとする．このとき，$\forall_{n\in\mathbb{N}} E[M_n] = 0$ である．

定義 5.5.22. $p \geq 1$ について，マルチンゲール (M_n) の p **次変分**または p **次変動**とは，次で定まる実数列 $([M]_n)$ をいう：

$$[M]_n := \sum_{k=1}^{n} |M_k - M_{k-1}|^p.$$

特に $p = 1$ のとき，**変分**あるいは**全変動**という．

命題 5.5.23. 2 次変分 $[M]_n$ は，次の 2 条件をみたす：

(1) $(M_n^2 - [M]_n)$ はマルチンゲールである．
(2) $([M]_n)$ は増加過程である：$0 = [M]_0 \leq [M]_1 \leq \cdots$．

注 5.5.24. (M_n^2) は劣マルチンゲールだから，Doob 分解 $M_n^2 = N_n + A_n$ を持つ．このとき，$(M_n^2 - A_n)$ はマルチンゲールであるが，(A_n) も命題の 2 条件を満たす．(A_n) も (M_n) の 2 次変分と呼び，$(\langle M \rangle_n)$ で表す．$(\langle M \rangle_n)$ は可予測でもあるが，一般に $([M]_n)$ はそうではない．明確な区別が必要である．一方で，連続マルチンゲールにおいては，２つの概念は 1 つに退化する．

定理 5.5.25 (Burkholder-Davis-Gundy)**.** (M_n) を $M_0 = 0$ を満たす p 乗可積分なマルチンゲールとする．このとき，次が成り立つ：

$$\forall_{p \geq 1}\ \exists_{c_p, C_p > 0} \quad c_p E\left[[M]_n^{p/2}\right] \leq E\left[\max_{1 \leq k \leq n} |M_k|^p\right] \leq C_p E\left[[M]_n^{p/2}\right].$$

要諦 5.5.26. 右辺は，Doob の不等式の $E[|M_n|^p]$ を $E[[M]_n^{p/2}]$ で置き換えたものになっている．$p = 2$ のとき両者は一致するが，応用上は 2 次変分の方が計算しやすいことが多い．

5.6　連続時変数のマルチンゲール

5.6.1　定義

離散の場合のマルチンゲールは可積分性を仮定していたが，その全貌は右連続性である．

定義 5.6.1. 右連続な確率過程 (X_t) が $(\mathcal{F}_t)$-マルチンゲールであるとは，次の 3 条件を満たすことをいう．

(1) $(\mathcal{F}_t)$-適合である：$\forall_{t \geq 0}\ X_t$ は $\mathcal{F}_t$-可測．
(2) 可積分である：$\forall_{t \geq 0}\ X_t \in \mathcal{L}^1(\Omega, \mathcal{F})$．
(3) $\forall_{0 \leq s \leq t}\ E[X_t | \mathcal{F}_s] = X_s$ a.s.

条件 (3) の代わりに $\forall_{0 \leq s \leq t}\ E[X_t | \mathcal{F}_s] \geq X_s$ a.s. をみたすとき，$(\mathcal{F}_t)$-劣マルチンゲールという．

5.6.2 Doob の不等式

5.6.3 Doob の任意抽出定理

5.6.4 劣マルチンゲールの収束定理

5.6.5 Doob-Meyer 分解

5.6.6 Burkholder の不等式

5.7 Gauss 系

5.8 統計推測への応用

回帰モデル $X_i = f(X_{i-1},\cdots,X_{i-p}) + \epsilon_i$ 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 $\{Yt_0,\cdots,Y_{t_n}\}$ に基づいて未知パラメータ α_1,α_2,β 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[†3]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，

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

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．

[†3] 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．

第 6 章

Markov 過程

互いに独立な試行の列（確率変数の列）の，マルチンゲールとは別の方向への一般化を考える．独立性は一切の過去の履歴に依らないが，Markov 性は，現在の状態のみに依存する性質を指す．
Brown 運動は状態空間，時間パラメータのいずれも連続な場合であり，Poisson 過程は状態空間は離散的である例である．状態空間が離散的な場合，Markov 連鎖ともいう．またここで偏微分方程式との関係から，確率過程の一般化も自然に出現する．添字集合を多様体 M とした確率過程 $\Omega \times M \to \mathbb{R}^n$ を**確率場**という．このとき，時間概念が空間に置き換わっている．x_{n+1} の x_n への依存の仕方は経時変化しないという，時間的一様性の仮定をおいて議論する．すると，Markov 過程は推移作用素を定めることで分布が決まる．これは大数の法則を一般化する．また，推移作用素になり得る作用素は放物型偏微分方程式によって特徴付けられる．

6.1 Kolmogorov の拡張定理

Hopf の拡張定理の一般化である．

定理 6.1.1. 確率空間列 $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率測度列 (μ_n) が次の一貫性条件をみたすとき，$(\mathbb{R}^\mathbb{N}, \mathcal{B}(\mathbb{R}^\mathbb{N}))$ 上の確率測度 μ であって $\forall_{A \in \mathcal{B}(\mathbb{R}^n)}\ \mu(A \times \mathbb{R}^\mathbb{N}) = \mu_n(\mathcal{A})$ を満たすものが一意的に存在する．ただし，$A \times \mathbb{R}^\mathbb{N} = \left\{(\omega_n) \in \mathbb{R}^\mathbb{N} \mid (\omega_1, \cdots, \omega_n) \in A\right\}$ とした．なお，$\mathbb{R}^\mathbb{N}$ には直積位相を考える．

(consistency) $\forall_{n \in \mathbb{N}}\ \forall_{A \in \mathcal{B}(\mathbb{R}^n)}\ \mu_{n+1}(A \times \mathbb{R}) = \mu_n(A)$.

特に，この一貫性条件は $\mathbb{R}^\mathbb{N}$ 上の測度に延長できるための必要十分条件である．これは $\mathbb{R}$ を一般の完備可分空間としても成り立つ．

6.2 離散時間の Markov 連鎖

記法 6.2.1 (state space). I を可算集合とし，これを**状態空間**とする．見本過程は列 $\mathbb{N} \to I$ となり，経時的に I 上を動き回りことになる．

6.2.1 確率行列

歴史 6.2.2. Markov が Markov 連鎖と確率行列を発明した．言語分析やカードシャッフルの問題に用いるつもりであったが，たちまち他の分野でも有用だと解った．確率行列の概念は Kolmogorov に引き継がれることとなる．実際，量子状態を表す演算子も，行列表示を持つこととなる．

記法 6.2.3.

(1) $\mathbf{1}$ はすべての成分が 1 であるような縦ベクトルも表す．
(2) δ_i で，i 成分のみが 1 でそれ以外が 0 であるようなベクトルを表す．

定義 6.2.4 (stochastic matrix)**.**

(1) I 上の**確率ベクトル** $(\nu_i)_{i\in I}$ とは，I 上の確率質量関数 $I \to [0,1]$ をいう．

(2) （可算個の成分を持ち得る）行列 $\mathbb{P} = (p_{ij})_{i,j\in I}$ が **(右) 確率的**であるとは，各 i 行ベクトル $(p_{ij})_{j\in I}$ がそれぞれ I 上の確率ベクトルを定めることをいう．意味論として，成分 p_{ij} は，現状態 i から次の時刻 j に遷移する確率を定める．

補題 6.2.5 (確率行列の特徴付け)**.** 行列 $\mathbb{P}$ について，次の 3 条件は同値．

(1) $\mathbb{P}$ は確率的である．

(2) 各行の和が 1 で非負な作用素 (積分なので)：$f \geq 0 \Rightarrow \mathbb{P}f \geq 0$ かつ $\mathbb{P}\mathbf{1} = \mathbf{1}$.

(3) 横ベクトル ν が確率ベクトルならば，$\nu\mathbb{P}$ も確率ベクトルである．

補題 6.2.6 (確率行列は群をなす？)**.** 確率行列 $\mathbb{P} = (p_{ij}), \mathbb{P}' = (p'_{ij})$ の積は確率行列である．

6.2.2　Markov 連鎖の定義と構成

定義 6.2.7 (transition matrix, Markov chain)**.** I-値確率変数列 $\{X_n\} \subset \mathrm{Meas}(\Omega, I)$ が，**初期分布 ν，遷移行列 $\mathbb{P}$ を持つ空間 I 上の Markov 連鎖**であるとは，次が成り立つことをいう：

$$\forall_{n\in\mathbb{N}}\ \forall_{i_0,\cdots,i_n\in I}\ P(X_0 = i_0, \cdots, X_n = i_n) = \nu_{i_0}p_{i_0i_1}\cdots p_{i_{n-1}i_n}.$$

命題 6.2.8. 適当な確率空間の上に，初期分布 ν と遷移行列 $\mathbb{P}$ をもち，殆ど至る所 I 値な Markov 連鎖が存在する．

［証明］． I は可算だから単射 $I \hookrightarrow \mathbb{N}$ が存在する．以降，$I \hookrightarrow \mathbb{N} \hookrightarrow \mathbb{R}$ として，$\mathbb{R}$ の部分集合と同一視する．

構成　各 $n \in \mathbb{N}$ に対して，$(\mathbb{R}^{n+1}, \mathcal{B}(\mathbb{R}^{n+1}))$ 上の測度 P_{n+1} を

$$P_{n+1}(A) := \sum_{(i_0,\cdots,i_n)\in I^{n+1}\cap A} \nu_{i_0}p_{i_0i_1}\cdots p_{i_{n-1}i_n} \quad (A \in \mathcal{B}(\mathbb{R}^{n+1}))$$

とすると，これはたしかに確率測度である．

一貫性　任意の $A \in \mathcal{B}(\mathbb{R}^{n+1})$ について，

$$P_{n+2}(A \times \mathbb{R}) = \sum_{(i_0,\cdots,i_n)\in I^{n+1}\cap A} \nu_{i_0}p_{i_0i_1}\cdots p_{i_{n-1}i_n}\left(\sum_{i_{n+1}\in I} p_{i_ni_{n+1}}\right) = P_{n+1}(A).$$

検証　Kolmogorov の拡張定理 6.1.1 より，$(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ 上の確率測度 P であって，$P(A \times \mathbb{R}^{\mathbb{N}}) = P_{n+1}(A)$ を満たすものがただ一つ存在する．この空間上の実数値確率変数列 (X_n) を，$X_n(\omega) = \omega_n$ $(\omega = (\omega_0, \cdots) \in \mathbb{R}^{\mathbb{N}})$ と定めれば，これは殆ど至る所 I-値の，求める Markov 過程である．

■

例 6.2.9 (i.i.d. は Markov 過程)**.** $\mathbb{P} = (p_{ij})$ の行ベクトルが i に依らずすべて同じであるとき，(X_n) は独立同試行に従う確率変数列となる．

6.2.3　Markov 性

定義 6.2.10 (Markov property)**.** 過去の軌跡を $A_{i_0,\cdots,i_n} := \{\omega \in \Omega \mid X_0(\omega) = i_0, \cdots, X_n(\omega) = i_n\}$ とし，$P(A_{i_0,\cdots,i_n}) > 0$ ならば，$\forall_{i_{n+1}\in I}\ P[X_{n+1} = i_{n+1}|A_{i_0,\cdots,i_n}] = p_{i_ni_{n+1}}$ が成り立つことは定義からすぐにわかる．これを **Markov 性**という．

記法 6.2.11. $\mathcal{F}_n := \sigma[X_0, \cdots, X_n]$ を，Markov 連鎖の定める自然な増加情報系とする．

定理 6.2.12. (X_n) は初期分布 ν，遷移行列 $\mathbb{P}$ の Markov 連鎖とする．$m \in \mathbb{N}, i \in I$ は $P[X_m = i] > 0$ を満たすとする．このとき，条件付き確率測度 $P[-|X_m = i]$ の下で，次の 2 条件が成り立つ．

(1) $(X_{m+n})_{n\in\mathbb{N}}$ は初期分布 δ_i，遷移行列 $\mathbb{P}$ の Markov 連鎖である．

(2) $(X_{m+n})_{n\in\mathbb{N}}$ は $\mathcal{F}_m$ と独立である．

補題 6.2.13. 確率行列の n 乗の成分を $\mathbb{P}^n =: (p_{ij}^{(n)})$ と表すこととする．このとき，

$$\forall_{m\in\mathbb{N}}\ \forall_{i\in I}\ P(X_m = i) > 0 \Rightarrow P[X_{m+n} = j|X_m = i] = p_{ij}^{(n)}.$$

この成分を n **ステップ遷移確率**という．

6.2.4 強 Markov 性

いかなる時点においても，現在の状況にしか依存しない性質を Markov 性と言うのであった．さらに，時間をランダムに定めても，現状にしか依存しないはずである．これを強 Markov 性という．

定理 6.2.14. (X_n) を初期分布 ν，遷移確率 $\mathbb{P}$ の Markov 連鎖とし，$i \in I$ は $P[\tau < \infty, X_\tau = i] > 0$ とする．このとき，条件付き確率 $P[-|\tau < \infty, X_\tau = i]$ の下で，次の 2 条件が成り立つ．

(1) $(X_{\tau+n})_{n\in\mathbb{N}}$ は初期分布 δ_i，遷移行列 $\mathbb{P}$ を持つ Markov 過程である．
(2) $(X_{\tau+n})_{n\in\mathbb{N}}$ は $\mathcal{F}_\tau$ と独立である．

6.3 到達確率と差分作用素

差分は前進 $\Delta f(x) := f(x+1) - f(x)$ と後退 $\nabla f(x) := f(x) - f(x-1)$ の 2 つが考えられる．これが連続になると確率微分方程式となるのだ．

6.3.1 到達確率と特徴付け

定義 6.3.1 (hitting / absorption probability)**.** (X_n) を Markov 過程とする．

(1) 集合 $A \subset I$ に対して，A への到達時刻とは，$\tau_A := \min\{n \in \mathbb{N} \mid X_n \in A\}$ として定まる可測関数 $\Omega \to \overline{\mathbb{N}}$ であった 5.4.5.
(2) 初期分布 δ_i を持つ Markov 過程に関する確率を P_i で表す．$a_i := P_i[\tau_A < \infty]$ を**到達確率**または**吸収確率**という．
(3) I 上の確率測度 Q_i を，$i \in I$ からの遷移確率 $(p_{ij})_{j\in I}$ が定めるものとして，任意の $i \in I$ に対して Q_i-可積分な関数 $f : I \to \mathbb{R}$ に対する作用素 $\mathcal{L} : \cap_{i\in I} L^1(I, Q_i) \to \mathrm{Map}(I, \mathbb{R})$ を，$\mathcal{L}f(i) := \sum_{j\in I} p_{ij} f(j) - f(i)$ と定め，**差分作用素**という．

定理 6.3.2 (到達確率の特徴付け)**.** 到達確率 $(a_i)_{i\in I}$ は，方程式系

$$\forall_{i\in I\setminus A}\ \mathcal{L}a(i) = 0, \quad \forall_{i\in A}\ a_i = 0$$

の最小の非負解である．後者は前者の**境界条件**という．

要諦 6.3.3. 差分方程式を書き直すと，$i \in I \setminus A$ に関して $a_i = \sum_{j\in I} p_{ij} a_j$ となり，i からの遷移確率に関する，到達確率の平均になる．

6.3.2 Markov 過程の定める martingale

一般の Markov 過程について，これを martingale 理論の問題に還元することが可能である．これは全く確率微分方程式特有の構造である．

定理 6.3.4. (X_n) を Markov 過程，$f \in l^\infty(\mathbb{R})$ を有界関数とする．

$$Y_n := f(X_n) - f(X_0) - \sum_{k=0}^{n-1} \mathcal{L}f(X_k)$$

によって定まる過程 (Y_n) は $(\mathcal{F}_n)$-マルチンゲールである．

6.4 有限状態空間上の Markov 連鎖

$|I| < \infty$ の場合について，理論の広がりを見る．

6.4.1 不変分布とエルゴード性

確率行列 $\mathbb{P}$ の，確率分布の空間 $P(I)$ への作用を考えると，不動点が存在する．

定義 6.4.1 (ergodic, irreducible, aperiodic)**.** Markov 連鎖 $((X_n), I, \mathbb{P})$ について，

(1) $\mathbb{P}$ が**エルゴード的**であるとは，$\exists_{n_0\in\mathbb{N}}\ \mathbb{P}^{n_0} > 0$ が成り立つことをいう．
(2) $\mathbb{P}$ が**既約**であるとは，$\forall_{i,j\in I}\ \exists_{n_1\in\mathbb{N}}\ p_{ij}^{(n_1)} > 0$ を満たすことをいう．
(3) 状態 $i \in I$ が**非周期的**であるとは，$\exists_{n_2\in\mathbb{N}}\ \forall_{n\geq n_2}\ p_{ii}^{(n)} > 0$ を満たすことをいう．

補題 6.4.2 (エルゴード性の特徴付け)**.** Markov 連鎖 $((X_n), I, \mathbb{P})$ が $|I| < \infty$ を満たすとき，次の 3 条件は同値.

(1) $\mathbb{P}$ はエルゴード的である．
(2) $\mathbb{P}$ は既約で，すべての状態 $i \in I$ は非周期的である．
(3) $\mathbb{P}$ は既約で，ある状態 $i \in I$ は非周期的である．

例 6.4.3.

(1) 円周の N 等分点上のランダムウォークは，N が奇数ならばエルゴード的であるが，偶数ならば既約であっても非周期的にはならない．
(2) $p_{ii} = 1$ を満たす $i \in I$ を trap という．これがある Markov 過程はエルゴード的でない．

定理 6.4.4 (有限状態 Markov 過程のエルゴード定理)**.** Markov 連鎖 $((X_n), I, \mathbb{P})$ が $|I| < \infty$ を満たし，$\mathbb{P}$ はエルゴード的であるとする．このとき，(1) を満たす I 上の確率分布 π が一意的に存在する．この π は (2),(3) も満たす．

(1) 定常性：$\pi\mathbb{P} = \pi$.
(2) 極限分布：$\forall_{i,j\in I}\ \lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$.
(3) 混合性：(2) の収束は指数関数的である：$\exists_{C>0}\ \exists_{0<\lambda<1}\ \forall_{n\in\mathbb{N}}\ \forall_{i,j\in I}\ |p_{ij}^{(n)} - \pi_j| \leq C\lambda^n$.

この分布 π を**不変分布**または**定常分布**という．

6.4.2 大数の法則

Markov 連鎖がエルゴード的ならば，独立性の代わりになり，大数の法則が成り立つ．

定理 6.4.5 (大数の弱法則)**.** Markov 連鎖 $((X_n), I, \mathbb{P})$ が $|I| < \infty$ を満たし，$\mathbb{P}$ はエルゴード的であるとする．π を不変分布とすると，関数 $f : I \to \mathbb{R}$ について，

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{p} E^\pi[f].$$

定義 6.4.6 (number of visit)**.** $i \in I$ に関して，$f(j) := 1_{\{j=i\}}$ と定めると，$\sum_{k=1}^{n} f(X_k)$ とは時刻 n までの i への訪問回数 $\tau_i^{(n)}$ を表す．滞在時間ともいう．

系 6.4.7. $\frac{\tau_i^{(n)}}{n} \xrightarrow{p} \pi_i$.

定義 6.4.8 (stationarity)**.**

(1) Markov 連鎖 $(X_n)_{n\in\mathbb{N}}$ が**定常的**であるとは，$(X_n)_{n\in\mathbb{N}}$ と $(X_{n+1})_{n\in\mathbb{N}}$ との分布が等しいことをいう．
(2) 初期分布を π とするエルゴード的な Markov 連鎖を**定常 Markov 連鎖**という．

定理 6.4.9 (高次元化)**.** Markov 連鎖 $((X_n), I, \mathbb{P})$ が $|I| < \infty$ を満たし，$\mathbb{P}$ はエルゴード的であるとする．π を不変分布とすると，関数 $f : I^l \to \mathbb{R}$ $(l \geq 1)$ について，

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k, \cdots, X_{k+l-1}) \xrightarrow{p} E^{\pi}[f]$$

ただし，E^{π} は定常 Markov 連鎖 $(\overline{X}_n)_{n\in[l]}$ に関する期待値である．

6.5　正方格子上のランダムウォーク

次に非有限な状態空間を持つ Markov 過程の例を見る．代表的なものが，$\mathbb{Z}^d$ 上のランダムウォークである．

定義 6.5.1 (random walk)**.** Markov 過程 $((X_n), \mathbb{Z}^d, \mathbb{P})$ の遷移行列 $\mathbb{P}$ が，時間一様性に加えて空間一様性 $\forall_{x,y,z\in\mathbb{Z}^d}\ p_{xy} = p_{x+z,y+z}$ を満たすとき，これを**酔歩**という．

議論 6.5.2 (加法過程としての構成)**.** Markov 過程としての一貫性に訴えずとも，空間的一様性に注目すれば，初期分布 ν を持つ Z_0 と，分布 p を持つ $Z_1, Z_2, \cdots$ とが独立であるとき，$X_n := \sum_{k=0}^{n} Z_k$ とすればこれは酔歩である．

6.5.1　再帰性と非再帰性

平均的な一歩 $E[Z_1]$ が零ベクトルでない場合，酔歩は非再帰的である．また，再帰的であることと無限回 0 を踏むことは同値である．

記法 6.5.3. $\nu = \delta_0$，あるいは，$Z_0 = 0$ から始まる酔歩を考える．

$$A_n := \{X_n = 0, \forall_{1\leq k\leq n-1}\ X_k \neq 0\}$$
$$q := \sum_{n=1}^{\infty} P(A_n) = P[\{\exists_{n\in\mathbb{N}}\ X_n = 0\}]$$

とする．

定義 6.5.4 (recurrent)**.** 酔歩が $q = 1$ を満たすとき**再帰的**であるという．

補題 6.5.5. 酔歩について，次の 2 条件は同値．

(1) 再帰的である．
(2) $\sum_{n=1}^{\infty} P[X_n = 0] = \infty$.

定理 6.5.6 (非再帰性の十分条件)**.** $R := \max\left\{|z| \in \mathbb{R} \mid z \in \mathbb{Z}^d, p_{0z} > 0\right\} < \infty$ と仮定し，$m := \sum_{z\in\mathbb{Z}^d} p_{0z} z = E[Z_1] \in \mathbb{R}^d$ とおく．$m \neq 0$ のとき，酔歩は非再帰的である．

6.5.2 単純ランダムウォークの再帰性と非再帰性

単純酔歩では，2d 個の隣点にのみ，そして等確率に移動可能とする．$E[Z_1]=0$ なので，これだけで再帰性は判定できない．

定義 6.5.7 (simple random walk)**.** 遷移確率が

$$p_z:=\begin{cases}\frac{1}{2d}, & \lvert z\rvert=1,\\ 0, & \lvert z\rvert\neq 1\end{cases}$$

となる酔歩を**単純酔歩**という．

定理 6.5.8 (Polya)**.** d 次元の単純酔歩は，$d=1,2$ のとき再帰的であり，$d\geq 3$ のとき再帰的でない．

6.6 連続時間 Markov 過程

6.6.1 Chapman-Kolmogorov 方程式

熱核の半群性

Markov 過程の発展は，確率行列の積で表された．この連続化は，ある発展条件を満たすことである．この放物型偏微分方程式を Chapman-Kolmogorov 方程式という．これを解いて推移確率とし，Kolmogorov の拡張定理に基づけば拡散過程が構成できる．Kolmogorov は初期から物理学への応用を見据えて，多様体の言葉で論じていた．
この方法は Hormander が取ったように，偏微分方程式への迂回でもある．直接的に確率微分方程式に基づいて Brown 運動を「変形する」という確率論的手法を立てたのが伊藤清である．

議論 6.6.1. 離散集合 I 上の遷移行列 $\mathbb{P}$ が満たす規則は次のようにかける．

(1) $\forall_{i\in I}\ \sum_{j\in I}p_{ij}^{(n)}=1$.
(2) $\forall_{i,j\in I}\ \forall_{n,m\in\mathbb{N}}\ \sum_{k\in I}p_{ik}^{(n)}p_{kj}^{(m)}=p_{ij}^{(n+m)}$.

I を一般のポーランド空間，$\mathbb{N}$ を $\mathbb{R}_+$ へ，遷移行列を遷移作用素へ一般化したい．

(1) $\forall_{s\in\mathbb{R}_+}\ \forall_{x\in S}\ p(s,x,-)\in P(S)$．時刻 0 に x から出発する Markov 過程の，時刻 s での位置の分布．
(2) $\forall_{s,t\in\mathbb{R}_+}\ \forall_{x,z\in S}\ \int_S p(s,x,dy)p(t,y,dz)=p(s+t,x,dz)$．または，$\forall_{A\in\mathfrak{B}(S)}\ \int_S p(s,x,dy)p(t,y,A)=p(s+t,x,A)$.

こうして，行列積は積分に一般化される．(2) を時間一様な Chapman-Kolmogorov の等式という．これは，時刻 0 に x から初めて，$s+t$ に A に至るまでの時刻 s での経由地 $y\in S$ について積分しても等しくなる，という意味を持つ．

6.7 加法過程

見本道が殆ど至る所 cadlag である過程を D-過程という．見本過程 $\mathbb{R}_+\to\mathrm{Meas}(\Omega,\mathbb{R});s\mapsto X_s$ が確率連続な加法過程で D-過程でもあるものを，Levy 過程という．任意の確率連続な加法過程は Levy 過程に同等であり，Levy 過程の構造は解明済みである．
大雑把にいえば，連続な加法過程は Gauss 型，すなわちブラウン運動に限り，非連続的な加法過程は Poisson 型に限る．任意の Levy 過程は，ドリフト付きの Brown 運動と Levy のジャンプ過程との和に分解できる．

6.7.1　定義と例

例 6.7.1. $\mathbb{Z}^d$-酔歩は次の性質を満たす：任意の長さ $k \geq 2$ の部分列 $(n_j)_{j\in[k]}$ について，**増分**の過程 $(X_{n_j} - X_{n_{j-1}})_{j\in[k]}$ は独立.

定義 6.7.2 (increment, additive process)**.** 次の性質を満たす，初期分布 δ_0 な $\mathbb{R}^d$-値確率過程 $(X_s)_{s\in\mathbb{R}_+}$，すなわち，空間的一様かつ時間一様な Markov 過程を**加法過程**という.

(3) 任意の長さ $k \geq 2$ の狭義単調増加非負実数列 $(s_j)_{j\in[k]}, s_0 = 0$ が定める増分列 $(X_{s_j} - X_{s_{j-1}})_{j\in[k]}$ は独立.

このとき，S の原点を 0 とするとその転移確率は一様に $p(s,-) := p(s,0,-)$ と表わせ，Chapman-Kolmogorov の等式も

$$\forall_{s,t\in\mathbb{R}_+}\ \forall_{x,z\in S}\quad \int_S p(s,dy-x)p(t,dz-y) = p(s+t,dz-x)$$

と表せる.

定理 6.7.3 (Gauss 型と Poisson 型 Levy 過程)**.** X を Levy 過程とする.

(1) X がさらに概連続過程であれば，増分 $X_b - X_a$ $(b > a)$ は Gauss 分布に従う.
(2) X がさらに殆ど至る所飛躍 1 で増加する階段関数を見本過程に持つならば，増分 $X_b - X_a$ $(b > a)$ は Poisson 分布に従う.

6.7.2　Levy-Ito 分解

定理 6.7.4 (Levy 過程の分解定理 (Levy-Ito decomposition))**.** X を Levy 過程とする．$\Gamma = \mathbb{R}_+ \times (\mathbb{R}\setminus\{0\})$ 上の固有な Poisson 配置 N と，これと独立な Gauss 型 Levy 過程 G が存在して，

$$X(t,\omega) = G(t,\omega) + \lim_{n\to\infty}\left[\int_{s\leq t}\int_{|u|>1/n}(uN(dsdu,\omega) - \phi(u)n(dsdu))\right]$$

と表せる.

定義 6.7.5 (Levy-Khintchine triplet)**.** Levy 過程 X の連続部分 G を用いて，$m(t) := E[G(t)], v(t) := \mathrm{Var}[G(t)]$ は有限確定する．また，m は連続関数，v は連続な単調増加関数，$m(0) = v(0) = 0$ を満たす．また，n を Poisson 配置 $N = N_X$ の平均測度，すなわち Γ 上の測度で，次を満たす：

$$\forall_{t\in\mathbb{R}_+}\quad n(\{t\}\times(\mathbb{R}\setminus\{0\})) = 0,\qquad \int_{s\leq t}\int_{|u|>0}(u^2\wedge 1)n(dsdu) < \infty.$$

組 (n,m,v) を，Levy 過程 X の**特性量**または **Levy-Khintchine 組**という.

定理 6.7.6. 上記の条件を満たす (n,m,v) に対して，これを特性量として持つ Levy 過程が存在して，法則同等を除いて一意的である.

定義 6.7.7. $X_t - X_s$ $(t > s)$ の確率法則が $t - s$ の値のみに関係するような Levy 過程を，**時間的に一様な Levy 過程**という．このとき，特性値は次のように表せる.

$$m_X(t) = mt,\quad v_X(t) = vt,\quad n_X(dtdu) = dt\cdot n(du).$$

6.8　Brown 運動

連続な加法過程は，必然的に Gauss 型である．これを Brown 運動という．ドリフトはないものをまずは見る.

6.8.1　定義

熱方程式 $u_t(x,t)=u_{xx}(x,t), u(x,0)=0$ の基本解

$$H(x,t)=\frac{1}{\sqrt{4\pi t}}\exp\left(-\frac{x^2}{4t}\right)$$

は熱核または熱方程式の初期値問題の Green 関数と呼ばれ，初期条件 $u(x,0)=f(x)$ に関する解は

$$u(x,t)=H*f=\int_{\mathbb{R}}H(x-y,t)f(y)dy$$

と表される．熱の拡散と確率の拡散，エントロピーの概念は深いどこかでつながっているのであろうか．

定義 6.8.1. 確率空間 $(\Omega,\mathcal{F},P)$ 上の実数値確率過程 $(B_t)_{t\in\mathbb{R}_+}$ が **Brown 運動**であるとは，次の 3 条件をみたすことをいう：

(1) $B_0=0$ a.s.
(2) 任意の見本道 $B_t(\omega):\mathbb{R}_+\to\mathbb{R}$ は連続：$B_t\in W$.
(3) 任意の長さ $n\in\mathbb{N}_{>0}$ の $\mathbb{R}_+$ の狭義増加列 $(t_j)_{j\in[n]}, t_j=0$ が定める増分の組 $(B_{t_j}-B_{t_{j-1}})_{j\in[n]}$ は互いに独立に Gauss 分布 $N(0,t_j-t_{j-1})$ に従う.

要諦 6.8.2. 実は (3) のうち増分の正規性はなくても従うことは，Levy 過程の理論による.

定理 6.8.3 (Brown 運動の存在)**.** ある確率空間 $(\Omega,\mathcal{F},P)$ が存在して，その上に Brown 運動が存在する.

6.8.2　Wiener 測度

古典的 Wiener 空間 W_0 は 0 から始まる連続な見本過程 $R_+\to\mathbb{R}$ 全体の空間で，Banach 空間となる．Brown 運動はここに確率測度を押し出し (見本道のばらつき)，Brown 運動は，関数解析的には W_0 上の確率測度の 1 つと同一視出来る.

記法 6.8.4 (classical Wiener space)**.** 次の言葉を使えば，Brown 運動とは Wiener 空間に値を持つ確率変数 $\Omega\to W_0$ であって，カリー化 $\mathbb{R}_+\to\mathrm{Meas}(\Omega,\mathbb{R})$ は任意の有限成分について独立な Gauss 分布を定めるものをいう.

(1) $W=W^1:=C(\mathbb{R}_+)$ を連続な見本道の空間とする.
(2) $W_0:=\{w\in W\mid w_0=0\}$ とする．これを**古典的 Wiener 空間**という.

それぞれの空間には一様ノルムは入れられないので，広義一様収束位相を考え，Borel 集合族によって可測空間とみなす．$\mathbb{R}_+$ なので Banach 空間ではない.

定義 6.8.5 (Wiener measure (23))**.** $(W_0,\mathcal{B}(W_0))$ 上の射影の族 $(B_t)_{t\in\mathbb{R}_+}, B_t(\omega):=\mathrm{pr}_t(\omega)=\omega_t$ が Brown 運動になるような確率測度 P を **Wiener 測度**という.

補題 6.8.6. Wiener 測度は一意的に存在する.

［証明］.

存在　Brown 運動の存在 6.8.3 による．ある空間 $(\Omega,\mathcal{F},P)$ 上の Brown 運動 $B:\Omega\to W$ を取る．これによる像測度 P^B は $P^B(W_0)=1$ を満たすから，W_0 への制限を取れば，これが Wiener 測度である.

一意性　W_0 の柱状集合全体 $\mathcal{C}$2.8.2 上では一意的である．$\mathcal{C}$ は π-系・情報族であり，$\mathcal{B}(W_0)=\sigma(\mathcal{C})$ を満たすため，一意に延長される.

■

6.8.3 特性値

補題 6.8.7 (積率). B_t の奇数次の積率は消えてきて，偶数次の積率は

$$E[B_t^2] = t, \quad E[B_t^4] = 3t^2, \quad E[B_t^6] = 15t^3, \quad , E[B_t^{2n+1}] = (2n+1)(2n-1)\cdots 3\cdot 1 t^{2n+1}.$$

補題 6.8.8 (共分散). $\forall_{t,s\in\mathbb{R}_+}\ E[B_tB_s] = t \wedge s$.

6.8.4 独立増分性

加法過程としての独立増分性は，martingale 問題に繋がる．

記法 6.8.9. ブラウン運動 $(B_t)_{t\in\mathbb{R}_+}$ の自然な情報系を $\mathcal{F}_t^B := \sigma[B_s; s \le t]$ と表す．

命題 6.8.10. $0 \le s < t$ に関して，$B_t - B_s$ は $\mathcal{F}_s^B$ と独立．

系 6.8.11. Brown 運動 $(B_t)_{t\in\mathbb{R}_+}$ は $(\mathcal{F}_t^B)$ に関して martingale である．その 2 次変分は $\langle B\rangle_t = t$ で与えられる．

6.8.5 可微分性

定理 6.8.12 (Paley-Wiener-Zygmond). $B_t(\omega)$ は ω-a.e. に対して，t について至る所微分不可能である．

定理 6.8.13 (modulus of continuity). 1/2-Holder 連続性よりやや悪い連続度を持つ．

$$\limsup_{t_2-t_1=\epsilon\searrow 0, 0\le t_1<t_2\le 1} \frac{|B_{t_2} - B_{t_1}|}{\sqrt{2\epsilon\log(1/\epsilon)}} = 1 \quad \text{a.s.}$$

定理 6.8.14 (重複対数の法則).

$$\limsup_{t\searrow 0} \frac{B_t}{\sqrt{2t\log\log(1/t)}} = 1 \quad \text{a.s.}$$

6.9 Poisson 過程

D-過程であるが連続過程ではなく，見本過程が至る所ジャンプしているとき，これは Poisson 型 Levy 過程である．ここまでいかずとも，少しドリフト 0 分散 0 の Brown 運動を混ぜて，ジャンプを持つ加法過程で特に基本的な Poisson 過程を見る．

6.9.1 定義

$\mathbb{R}_+$ 上に強さ λ の Poisson 点過程を考える．その総数を整数で切ったものを Poisson 過程と呼ぼう．

定義 6.9.1 (Poisson process). $(\Omega, \mathcal{F}, P)$ 上の $\mathbb{Z}_+$-値確率過程 $(N_t)_{t\in\mathbb{R}_+}$ が**パラメータ $\lambda > 0$ を持つ Poisson 過程**であるとは，次の条件を満たすことをいう：

(1) $N_0 = 0$ a.s.
(2) 任意の見本道 $N_t(\omega)$ は右連続かつ単調増加である．
(3) 任意の長さ $n \in \mathbb{N}_{>0}$ の $\mathbb{R}_+$ の狭義増加列 $(t_j)_{j\in[n]}, t_0 = 0$ が定める増分の組 $(N_{t_j} - N_{t_{j-1}})_{j\in[n]}$ は独立で，それぞれパラメータ $\lambda(t_j - t_{j-1})$ を持つ Poisson 分布に従う．

議論 6.9.2 (Poisson 過程の構成). Kolmogorov の拡張定理により存在は保証されるが，次のように構成できる．独立にパラメータ λ の指数分布 9.4.5 に従う $\mathbb{R}_+$-値確率変数列 (S_j) を取る：$P[S_j \geq t] = e^{-\lambda t}$．なお，パラメータ $\lambda > 0$ の指数分布密度関数は

$$p(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x)$$

と表せる．このとき

$$Z_0 = 0, \quad Z_k = \sum_{j=1}^{k} S_j$$

とするとこれは強さ λ の Poisson 点過程であり，$t \in \mathbb{R}_+$ を超えた Z_k の数の過程

$$N_t := \max\{k \in \mathbb{N} \mid Z_k \leq t\} \in \overline{\mathbb{Z}}_+$$

は Poisson 過程となり，$P[N_t \in \mathbb{Z}_+] = 1$.

6.9.2　独立増分性

命題 6.9.3. $0 \leq s < t$ について，$N_t - N_s$ は $\mathcal{F}_s^N = \sigma[N_u; u \leq s]$ と独立である．

系 6.9.4. $(N_t - \lambda t)_{t \geq 0}$ は $(\mathcal{F}_t^N)$ に関してマルチンゲールである．

6.10　無限分解可能分布

6.10.1　定義と特徴付け

確率変数 X が，任意の $n \in \mathbb{N}$ に対して，ある分布 μ_n が存在してそれに従う独立同分布変数 n 個の和で表せるとき，これを無限可解であるという．これは半群 $P(\mathbb{R})$9.5.15 の言葉を用いて定義できる．

定義 6.10.1. 1 次元の分布 μ が**無限可解**であるとは，分布族 $\{\mu_t\}_{t\in\mathbb{R}_+}$ が存在して，μ を乗法単位元として $\mathbb{R}_+$ と同型な連続半群となることをいう：

(1) $\mathbb{R}_+ \to P(\mathbb{R}); t \mapsto \mu_t$ は弱位相に関して連続.
(2) $\forall_{t,s\in\mathbb{R}_+}\ \mu_t * \mu_s = \mu_{t+s}$.
(3) $\mu_0 = \delta_0$.
(4) $\mu_1 = \mu$.

補題 6.10.2. 次の 4 条件は同値.

(1) μ は無限可解である．
(2) $\forall_{n\in\mathbb{N}}\ \mu = \mu_n * \mu_n * \cdots * \mu_n$ と表せる．
(3) 任意の $\epsilon > 0$ に対し，分解 $\mu = \mu_1 * \cdots * \mu_n$ であって，Levy 距離に関して $d_L(\mu_i, \delta) < \epsilon$ を満たすものが存在する．
(4) $\mu_n = \mu_{n1} * \cdots * \mu_{nm(n)}$, $\max_{k\in[m(n)]} d_L(\mu_{nk}, \delta) \to 0$ を満たす列 (μ_n) が存在して，$\mu_n \to \mu$.

例 6.10.3. 畳み込みについて閉じている分布族 9.5.17 はみな無限可解な分布の例である．

例 6.10.4. 時間的に一様な Levy 過程 X は，任意の $t \in \mathbb{R}_+$ について X_t は無限可解である．また逆に，連続半群 (μ_t) に対して，これが定める一様な Levy 過程が存在する．

6.10.2　Levy 分解

無限可解分布，連続半群 $\{\mu_t\}_{t\in\mathbb{R}_+}$，時間的に一様な Levy 過程（の法則同値類）$X$ の間に，次の全単射対応がある：

$$\mu \xrightarrow{\mu=\mu_1} \{\mu_t\} \xrightarrow{\mu_t=P^{X(t)}} X$$

定理 6.10.5 (無限可解分布の Levy 分解定理). 任意の無限可解分布 μ は，特性関数 $\mathcal{F}\mu(z)$ が $\mathcal{F}\mu(z)=e^{\psi(z)}$ なる形になる．ただし，

$$\psi(z)=imz-\frac{v}{2}z^2+\int_{|u|>0}(e^{izu}-1-i\phi(u)z)n(du),\quad m\in\mathbb{R},v\geq 0,\int_{|u|>0}(u^2\wedge 1)n(du)<\infty.$$

定義 6.10.6. Poisson 分布，Cauchy 分布も Levy 分解を定め，これらに対応する一様 Levy 過程を Poisson 過程，Cauchy 過程という．

6.10.3　複合 Poisson 過程

定義 6.10.7. $m=0,v=0$ のときの $\psi(z)=\int_{|u>0|}(e^{izu}-1)n(du)$ と表せる μ のクラスを，**複合 Poisson 分布**という．これに対応する過程を**複合 Poisson 過程**という．

補題 6.10.8. 複合 Poisson 分布 μ は，ν^{*n} を Poisson 分布 p_λ によって加重平均を取ったものである：

$$\mu=\sum_{n=1}^{\infty}e^{-\lambda}\frac{\lambda^n}{n!}\nu^{*n}.$$

6.11　1 次元拡散過程

記法 6.11.1. $W:=C(\mathbb{R}_+)$ 上に確率測度の族 $(P_x)_{x\in\mathbb{R}}$ を考える．射影を $X_t:=\mathrm{pr}_t:W\to\mathbb{R};\omega\mapsto\omega(t)$ $(t\in\mathbb{R}_+)$ と表すと，この見本過程 $\mathbb{R}_+\to\mathrm{Map}(W,\mathbb{R})$ は任意の $x\in\mathbb{R}$ について (W,P_x) 上連続．$\mathcal{B}_t:=\sigma[X_s;s\leq t]$ とする．

定義 6.11.2 (diffusion process). 確率過程 $(X_t)_{t\in\mathbb{R}_+}$ が $P_x[X_0=x]=1$ と次の条件を満たすとき，$\mathcal{M}:=(\mathcal{M}_x)_{x\in\mathbb{R}},\mathcal{M}_x:=\{X_t(\omega)\in\mathbb{R}\mid t\in\mathbb{R}_+,\omega\in(C,P_x)\}$ を**一様な連続強 Markov 過程**または**拡散過程**という．

(強 Markov 性) $x\in\mathbb{R}$ と有限な (B_t)-Markov 時刻 $\tau:W\to\mathbb{R}_+$，$E\in\mathcal{B}^1(\mathbb{R}^1)$ について，$P_x[X_{\tau+t}\in E\mid\mathcal{B}_\tau]=P_x[X_t\in E]|_{x=X_\tau}$．

定義 6.11.3 (regular point, regular diffusion process). 拡散過程 $(\mathcal{M}_x)_{x\in\mathbb{R}}$ について，

(1) x は $\mathcal{M}$ の**正則点**であるとは，$P_x[\exists_{t\in\mathbb{R}_+}X_t>x]>0,P_x[\exists_{t\in\mathbb{R}_+}X_t<x]>0$ が成り立つことをいう．
(2) 任意の $x\in\mathbb{R}$ が正則点であるとき，$\mathcal{M}$ は**正則**であるという．

第 7 章

高次元確率論

ここではこれ以上具体論に深入りせず，高次元確率論と確率過程論を深掘りする．これらはいずれも $\Omega \to \mathrm{Meas}(\mathbb{R}_+, \mathbb{R})$ とみると，Banach 空間値確率変数として統一的に見れる．また，平均や分散などの母数も Banach 空間の点となる．実際，確率分布族を Banach 空間で添字付ける見方は有効な手法である．

第 8 章

離散確率分布

確率測度 $P(X)$ の具体例を，X が可算集合 $|\mathcal{X}| \le \infty$ 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、これらの母数は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」$\overline{X}$ は母集団の「平均」母数 μ の推定量である。

記法 8.0.1. 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，測度の畳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 $P(X) := \{P \in M(X) \mid \|\mu\| \le 1, P(1) = 1\} \subset (C(X))^*$ 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．

8.1 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 $\mathrm{id}_{\mathbb{N}}$ の確率測度による積分を平均といい，一般の確率変数 $X : \mathbb{N} \to \mathcal{X}$ 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 8.1.1 (確率分布の特徴付け)**.** こうして $X : \Omega \to \mathcal{X}$ 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) (non-negative) $\forall_{U \subset \mathcal{X}} \displaystyle\int_U dP^X \ge 0.$

(2) (normalized) $\displaystyle\int_X dP^X = 1.$

要諦 8.1.2 (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 8.1.3 (expectation value)**.** 確率空間 $(\Omega,\mathcal{F},P)$，終確率空間 $(\mathcal{X},\mathcal{B},P^X)$ について，確率変数 $f:\Omega\to\mathcal{X}$ の事象 $B\in\mathcal{B}$ 上での期待値とは，

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

注 8.1.4. 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 8.1.5 (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 $X:\Omega\to\mathcal{X}$ 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，$P(X\in A)=\int_A Xd\mu=\int_{X^*A}dP$ を満たすものである．

$$\begin{tikzcd} {[0,1]} & \\ (\Omega,\mathcal{F}) \arrow[u, "P"] \arrow[r, "X"'] & (\mathcal{X},\mathcal{B}) \arrow[ul, "X_*P"'] \end{tikzcd}$$

要諦 8.1.6. 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 $f:=\dfrac{dX_*P}{d\mu}$ と構成できる．

8.2　分布の特性値

8.2.1　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 $\mu:\mathcal{F}\to\mathbb{R}$ 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 8.2.1 (discrete probability function, probability mass function)**.**

(1) 可算集合 $\mathcal{X}$ 上の確率分布 $P^X:\mathcal{X}\to\mathbb{R}$ を**離散分布**という．

(2) 離散分布 $\nu:\mathcal{X}\to\mathbb{R}$ 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 8.2.2. $\mathcal{X}$ を可算集合として，次の標準的な対応が存在する．

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

[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．

8.2.2　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 8.2.3 (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 $\mu = (p_x)_{x\in\mathcal{X}} : \mathcal{X} \to \mathbb{R}$ 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 $\mu := \sum_{x\in\mathcal{X}} xp_x$ 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 $\dfrac{\mu_3}{\mu_2^{3/2}}$ を**歪度**という．[†3]

(7) 確率分布 $\dfrac{\mu_4}{\mu_2^2}$ を**尖度**という．[†4]

8.3　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．

8.3.1　特性関数と確率母関数

定義 8.3.1 (characteristic function, probability generating function).

(1) 実数上の確率変数 $\varphi(u) := \sum_{x\in\mathcal{X}} e^{iux} p_x : \mathbb{R} \to \mathbb{C}$ 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 8.3.2 (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) $\sigma^2 = \alpha_2 - \mu^2 = \sum_{x\in\mathcal{X}} x^2 p_x - \left(\sum_{x\in\mathcal{X}} xp_x\right)^2$.

(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価である．e なぞ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)

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

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

(3) 一般に，項別微分可能ならば $g^{(r)}(z)=\sum_{x=0}^\infty x(x-1)\cdots(x-r+1)p_xz^{x-r}$ であるが，いま $\mathcal{X}=\mathbb{Z}$ 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 $g'(1)=\mu$，$g''(1)=\sum_{x\in\mathcal{X}}x(x-1)p_x$ であるから，(1) より，$\sigma^2=\underbrace{g''(1)+g'(1)}_{=\sum(x(x-1)+x)p_x}-(g'(1))^2$

■

8.3.2　分位点関数

定義 8.3.3 (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，$F_L^{-1}(u):=\inf\{x\in\mathbb{R}\mid F(x)\ge u\}, F_R^{-1}(u):=\sup\{x\in\mathbb{R}\mid P(X\ge x)\ge 1-u\}$ と定める．

補題 8.3.4.

(1) F_L^{-1} は左連続である．

(2) F_R^{-1} は右連続である．

定義 8.3.5. 分布関数 $F(x)$ において，$x\to\pm\infty$ 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．

8.3.3　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 8.3.6 (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 $f(z)=\sum_{n=0}^\infty f_nz^n$

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 8.3.7 (母関数の応用)**.** 関手 Seq → 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 8.3.8 (moment)**.** 一般に，関数 $f : \mathbb{R} \to \mathbb{R}$ の，$x = c$ を中心とす rn 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．

8.4　一次元離散分布の例

8.4.1　デルタ分布

Shanonn 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 $H : \mathrm{Meas}(\mathcal{X}, \mathbb{R}) \to [0, \log \alpha]$ であるが，これが最小になるときの確率分布である．

定義 8.4.1 (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 $\epsilon_a(A) := 1_{\{B \in \mathcal{A} | a \in B\}}$ $(A \in \mathcal{A})$ 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]

8.4.2　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 8.4.2 (order statistic, empirical distribution function, resampling from sample, bootstrap method)**.** $X_1, \cdots, X_n \sim F$, 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 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ 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 $F_n(x) := \frac{1}{n}|\{i \in [n] \mid X_i \le x\}|$ 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 8.4.3. 大数の強法則によって，$P\left[\lim_{n\to\infty} |F_n(x) - F(x)| = 0\right] = 1$ が従うが，より強い「一様大数の法則」ともいうべき結果がある．

[†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 8.4.4 (Glivenko-Cantelli (1933)).

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

8.4.3 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 8.4.5 (Rademacher distribution, Rademacher / Steinhaus variable).

(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 $\epsilon_1,\cdots,\epsilon_n$ に対して，$X:=\sum_{i=1}^n\epsilon_i$ と定めた $X:\mathbb{R}^n\to\mathbb{R}$ を **Rademacher 過程**という．

(3) 複素数値 Rademacher 変数は Steinhaus 変数とも呼ばれ，$P(a<\arg(\epsilon)<b)=\frac{1}{2\pi}(b-a)$ で定まる．

注 8.4.6. 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 8.4.7 (Van Zuijlen's bound).

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

8.4.4 離散一様分布

整数 $1,2,\cdots,N$ 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 8.4.8 (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 $U:\mathrm{FinSet}\to\mathcal{M}(\mathcal{X})$ を**離散一様分布**という．

要諦 8.4.9. Shanonn 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 $H:\mathrm{Meas}(\mathcal{X},\mathbb{R})\to[0,\log\alpha]$ であるが，これが最大になるときの確率分布である．

8.4.5 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 8.4.10 (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** $B(n,p):\mathbb{N}\times[0,1]\to\mathcal{M}(n+1)$ 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 $b:n+1\to[0,1]$ は $b(x;n,p)=\binom{n}{x}p^xq^{n-x}$ と表される．$\mathrm{Bernoulli}(1,p):=B(1,p)$ を Bernoulli 分布という．

命題 8.4.11 (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)

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

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

$$\begin{aligned}
g(z) &= \sum_{x=0}^{n} b(x;n,p) z^x \\
&= \sum_{x=0}^{n} \binom{n}{x} p^x q^{1-x} z^x = (pz+q)^n
\end{aligned}$$

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 8.3.2 から，

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

■

要諦 8.4.12. 確率母関数が，二項展開の式に一致することから，ここに予想だにしなかった抜け道がある．

8.4.6 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．

8.4.6.1 **極限分布としての** Poisson **分布**

定義 8.4.13 (Poisson distribution (1838))**.** 可測空間 $(\mathbb{N}, P(\mathbb{N}))$ 上の **Poisson 分布** $\mathrm{Pois}(\lambda):(0,\infty)\to\mathcal{M}(\mathbb{N})$ とは，確率質量関数 $p(x;\lambda) = \dfrac{\lambda^x}{x!}e^{-\lambda}:\mathbb{N}\to[0,1]$ 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 8.4.14 (Poisson's limit theorem)**.** 減衰する確率 $p(n) := \dfrac{\lambda}{n}$ についての二項分布 $B(n,p(n))$ は，$\lim_{n\to\infty} b(x;n,p) = \dfrac{\lambda^x}{x!}e^{-\lambda}$ $(\forall_{x\in\mathbb{N}})$ を満たす．

[証明] .

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

$$= \frac{\lambda^x}{x!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\left(1+\frac{-\lambda}{n}\right)^{n/(-\lambda)\cdot(-\lambda)}\left(1-\frac{\lambda}{n}\right)^{-x}$$
$$\xrightarrow{n\to\infty} \frac{\lambda^x}{x!}e^{-\lambda}$$

この収束は一様収束である. ■

命題 8.4.15 (Poisson 分布の確率母関数). Poisson 分布 $\mathrm{Pois}(\lambda)$ 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 8.4.16. 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、統計学者のボ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。ボ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。

8.4.6.2　ポアソン過程

Poisson 点過程／ Poisson 点場において，ある有界領域内の点の数は Poisson 分布に従う確率変数となる．このことから名付けられたのが Poisson 点過程で，Poisson 自体はこれについて研究していない．

定義 8.4.17 (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 8.4.18. Poisson 過程は存在する．

命題 8.4.19. Poisson 過程 $(P_k(t))_{k\in\mathbb{N}}$ は $\mathbb{N}$ 上の Poisson 分布 $\mathrm{Pois}(\lambda t)$ 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，

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

[†10] 確率過程の独立はまだ定義していない．

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，$P_1(0) = \lambda t e^{-\lambda t}$．$k = 2$ とすると，

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

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

■

［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 $\mathrm{Pois}(\lambda t)$ 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．■

8.4.7　負の二項分布

定義 8.4.20 (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 3.5.12 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** $\mathrm{NB}(k,p) : \mathbb{Z}_+ \times [0,1] \to \mathcal{M}(\mathbb{N})$ 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** $G(p) := \mathrm{NB}(1,p)$ といい，確率関数は $p(x;1,p) = pq^x \ (x \in \mathbb{Z}_+)$ と表せる．

(3) 負の二項分布 $NB(k,p)$ は $k \in (0,\infty)$ 上に延長できる．

命題 8.4.21 (負の二項分布の確率母関数・平均・分散).

(1) 負の二項分布 $\mathrm{NB}(b,p)$ の確率母関数は $g(z) = p^k(1-qz)^{-k} = \left(\dfrac{p}{1-qz}\right)^k$ で表せる．

(2) 負の二項分布 $\mathrm{NB}(b,p)$ の平均は $\alpha_1 = \dfrac{kq}{p}$，分散は $\mu_2 = \dfrac{kq}{p^2}$ である．

［証明］．

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

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

$\left.\dfrac{(-k)(-k-1)\cdots(-k-x+1)}{x!}\right|_{x=0} = 1$ としたことに注意．よって，$\mathrm{NB}(k,p)$ 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 8.3.2(3) より，

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

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

■

命題 8.4.22 (Poisson 近似)**.** 平均 $\lambda := \dfrac{kq}{p} \in (0,\infty)$ を一定にして，成功数を表す母数を $k\to\infty$ とすると（このとき失敗率 q は極めて小さくなる），Poisson 分布 $\mathrm{Pois}(\lambda)$ に収束する：$NB(k,p)\to\mathrm{Pois}(\lambda)$ $(k\to\infty)$.

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

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

実際,

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

■

注 8.4.23 (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.

8.4.8　Katz 族

定義 8.4.24 (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 8.4.25.

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

命題 8.4.26 (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，$p=\mathrm{Pois}(b)$.
(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

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

と表せる.

8.4.9　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 8.4.27 (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 $p(x;N,n,r)=\dfrac{\binom{n}{x}\binom{N-n}{r-x}}{\binom{N}{r}}$ が，$X:=\{x\in\mathbb{N}\mid\max\{0,r-N+n\}\le x\le\min\{r,n\}\}$ 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 8.4.28.

(1) 平均は $\alpha_1=r\dfrac{n}{N}$.

(2) 分散は $\mu_2=r\left(\dfrac{N-r}{N-1}\right)\left(\dfrac{n}{N}\right)\left(1-\dfrac{n}{N}\right)$.

例 8.4.29 (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，$\widetilde{N}:=\dfrac{nr}{x}$ 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．

8.4.10　負の超幾何分布

8.4.11　対数分布

定義 8.4.30. $\mathbb{N}$ 上の分布・**対数分布** $\mathrm{Log}\,(\theta):=(p_x)_{x\in\mathbb{N}}$ $(\theta\in(0,1))$ とは，確率関数

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

で定まる分布をいう．

命題 8.4.31. $\mathrm{Log}\,(\theta)$ の確率母関数は $g(z)=\dfrac{\log(1-\theta z)}{\log(1-\theta)}$ である．

8.4.12　Ord 族

定義 8.4.32 (Ord family)**.** $\mathbb{Z}$ 上の分布であって，確率関数が差分方程式

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

を満たすもの全体を **Ord 族**という．

例 8.4.33. Katz 族と，超幾何分布，負の超幾何分布は Ord 族である．

8.5　多次元離散分布の例

8.5.1　確率変数の積

定義 8.5.1 (marginal distribution, joint / simultaneous distribution)**.** $I:=\{1,\cdots,d\}$ について，確率変数 $X_j:\Omega\to\mathcal{X}_j$ $(j\in I)$ の積 $X:=(X_j)_{j\in I}:\Omega\to\mathcal{X}:=\prod_{j=1}^{d}\mathcal{X}_j$ 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 8.5.2 (covariance, correlation coefficient)**.** 2 つの実確率変数 $X:\Omega \to \mathcal{X}, Y:\Omega \to \mathcal{Y}$ について，

(1) $\mathrm{Cov}[X,Y] := \mathcal{E}[(X-\mathcal{E}[X])(Y-\mathcal{E}[Y])]$ を**共分散**と呼ぶ．
(2) $\rho(X,Y) := \dfrac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}}$ を**相関係数**と呼ぶ．

補題 8.5.3.

(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) $E[X^2] < \infty, E[Y^2] < \infty, \mathrm{Var}(X), \mathrm{Var}(Y) > 0$ のとき，相関係数は存在し，$\mathrm{Im}\,\rho \subset [-1,1]$ を満たす．

8.5.2　多項分布

二項分布は多項分布の周辺分布であった．

定義 8.5.4 (multinomial distribution)**.** $I = \{E_1, \cdots, E_k\}$ とし，

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

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

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

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** $\mathrm{Mult}(n,p)$ という．

命題 8.5.5. $g_I(z_I; p) = (p_1 z_1 + \cdots + p_k z_k)^n$ が確率母関数である．

[証明] . 多項定理より． ■

命題 8.5.6. $\mathrm{Cov}[X_i, X_j] = \begin{cases} np_i(1-p_i) & i = j \\ -np_i p_j & i \neq j \end{cases}$

8.5.3　2 変量 Poisson 分布

定義 8.5.7 (bivariate Poisson distribution)**.** 確率変数 U_1, U_2, U_3 が独立で $U_i \sim \mathrm{Pois}(\lambda_i)$ 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 $(X_1, X_2) : \Omega \to \mathbb{N}^2$ が定める分布を **2 変量 Poisson 分布**といい，$\mathrm{BPois}(\lambda_1, \lambda_2, \lambda_3)$ で表す．周辺分布は $X_1 \sim \mathrm{Pois}(\lambda_1 + \lambda_3), X_2 \sim \mathrm{Pois}(\lambda_2 + \lambda_3)$ である．

命題 8.5.8 (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) $\mathrm{Cov}[X_1, X_2] = \lambda_3$.

8.5.4　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，この形での一般化を考える．

定義 8.5.9. $k > 0, P_i > 0$ をパラメータとする**負の多項分布** $NM(k, (P_1, \cdots, P_d))$ とは，$Q := 1 + \sum_{i=1}^{d} P_i$ とするとき，確率母関数

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

が定める $\mathbb{N}^d$ 上の分布を言う．

命題 8.5.10 (確率母関数の微分からわかること)**.**

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

第 9 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．

9.1 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．

9.1.1 積分の定義

命題 9.1.1. 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］． $\forall_{x\in\mathbb{R}}\ |x|^p \le 1 + |x|^q$ である．これと，Lebesgue の優収束定理より． ■

命題 9.1.2 (Markov の不等式)**.** 非負可測関数 X と任意の $p > 0, q \ge 0, \epsilon > 0$ について，

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

［証明］． $\epsilon^p 1_{\{X\ge\epsilon\}} X^q \le 1_{\{X\ge\epsilon\}} X^{p+q} \le X^{p+q}$ と，積分の単調性より． ■

9.1.2 期待値の定義

定義 9.1.3 (expectation / expected value / mean (value))**.** 確率変数 $X : \Omega \to \mathbb{R}$ 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 9.1.4 (r-th moment, r-th central moment, r-th absolute moment)**.** 確率変数 $X : \Omega \to \mathbb{R}$ 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，$\mathrm{Var}[X]$ で表す．
(5) $\sqrt{\mathrm{Var}[X]}$ を**標準偏差**と呼ぶ．

命題 9.1.5 (分散の性質)**.** 2 乗可積分な実確率変数 $X : \Omega \to \mathbb{R}$ について，

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

9.1.3　期待値不等式

記法 9.1.6. 積分論の記号を，期待値に関しても流用する．

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

定理 9.1.7 (測度論における結果)**.** X, Y を確率変数とする．

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

9.2　分布の特性値

測度論による確率の特徴量の理解

$(\mathbb{R}, \mathcal{B}_1)$ 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？

9.2.1　確率密度関数

定義 9.2.1 (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 9.2.2. 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 9.2.3. 任意の可測関数 $g : \mathbb{R} \to \mathbb{R}$ 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．

9.2.2　平均値と積率

定義 9.2.4 (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 9.2.5. $(\mathbb{R}, \mathcal{B}_1)$ 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 9.2.6 (変数変換公式:well-definedness). $(\mathcal{X}, \mathcal{A})$ を可測空間とし，X を $\mathcal{X}$-値確率変数とする．このとき，

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

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 9.2.7 (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 9.2.8 (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.

9.2.3 共分散と相関

定義 9.2.9 (covariance, correlation coefficient).

(1) X, Y, XY が P-可積分のとき，

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

を X, Y の共分散という．
(2) $X, Y \in L^2$ で $\mathrm{Var}[X]\mathrm{Var}[Y] \neq 0$ のとき，$\rho(X, Y) = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}}$ 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き，**共分散行列**は

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

で定まる $r \times c$ 行列をいう．$\mathrm{Cov}[X, Y] = \mathrm{Cov}[Y, X]^\perp$ が成り立つ.
(4) $\mathrm{Var}[X] := \mathrm{Cov}[X, X]$ で定まる $r \times r$ 行列を，**分散共分散行列**という．
(5) $\mathrm{Corr}[X] = (\rho(X_i, X_j))_{i,j\in[n]}$ 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 9.2.10. $X, Y \in L^2$ は十分条件である．

命題 9.2.11. $X, Y, Z \in L^2$ について，

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

命題 9.2.12 (Pearson の不等式). $X \in L^1$ について，$\mathrm{Var}[X] > 0$ 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 9.2.13 (共分散行列の双線型性).

(1) 可積分確率変数 $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ の積 X_iY_j も可積分とする．このとき，

$$\forall_{a_1,\cdots,a_m,b_1,\cdots,b_n\in\mathbb{R}} \quad \mathrm{Cov}\left(\sum_{i=1}^m a_iX_i, \sum_{j=1}^n b_jY_j\right) = \sum_{i=1}^m\sum_{j=1}^n a_ib_j\mathrm{Cov}()X_i, Y_j)$$

(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 $X_i, Y_j, X_iY_j \in L^1$ 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

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

注 9.2.14. 行列 M に関して，$|M| = (\mathrm{Tr}\ (MM^\perp))^{1/2}$ 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 9.2.15 (和の分散)**.** $X_1, \cdots, X_n \in L^2$ について，

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

命題 9.2.16 (相関係数の値域)**.** $\mu_X = E[X], \mu_Y = E[Y], \sigma_X = \sqrt{\mathrm{Var}[X]}, \sigma_Y = \sqrt{\mathrm{Var}[Y]}$ とする．

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

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

[証明] . $\sigma_X\sigma_Y \ne 0$ のとき，

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

より． ■

命題 9.2.17 (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，

$$E(X^TAY) = \mathrm{Tr}\ (\mathrm{Cov}(X, Y)^\perp A) + E(X)^\perp AE(Y)$$

9.2.4 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 9.2.18. 分散共分散行列 $\mathrm{Var}[X]$ は半正定値である．

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

注 9.2.19 (退化した多次元確率変数)**.** $\mathrm{Var}[X]$ が正定値でないとすると，$\exists_{u\in\mathbb{R}^d\setminus\{0\}}\ u^\perp\mathrm{Var}[X]u = E([u^\perp(X - E(X))]^2) = 0$ 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．

9.3 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．Fourier 解析により，関数空間における議論に変換することで，古典的な中心極限定理や分布の高次近似の結果をもたらすことが出来る．
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．

9.3.1 特性関数と分布

特性関数と分布の対応は，緩増加超関数の空間上の Fourier 逆変換 $\mathcal{F}^{-1}:\mathcal{S}'(\mathbb{R})\xrightarrow{\sim}\mathcal{S}'(\mathbb{R})$ の $P(\mathbb{R})$ への制限が与える．

定義 9.3.1 (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 $\varphi:\mathbb{R}\to\mathbb{C}$ 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 9.3.2 (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，すなわち，a,b が F の連続点ならば，

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

定理 9.3.3 (一意性定理)**.** 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．すなわち，任意の $\mu_1,\mu_2\in\mathcal{P}^d$ について，次は同値．

(1) $\mu_1=\mu_2$.
(2) $\varphi_{\mu_1}=\varphi_{\mu_2}$.

[証明]． $d=1$ の場合について，2 通りの方法で示す．

反転公式による証明　$\varphi_\mu=\varphi_{\widetilde{\mu}}$ とする．反転公式 9.3.2 により，任意の (端点が $F_\mu,F_{\widetilde{\mu}}$ の連続点である) 開区間の測度は φ が一意に定める：$F_\mu(b)-F_\mu(a)=F_{\widetilde{\mu}}(b)-F_{\widetilde{\mu}}(a)$．$b\in\mathbb{R}$ が $F_\mu,F_{\widetilde{\mu}}$ 両方の連続点であるとき，$F_\mu(b)=F_{\widetilde{\mu}}(b)$．分布関数は右連続で，不連続点は高々可算個だから，これは $F_\mu=F_{\widetilde{\mu}}$ を含意する．分布関数の一意性より，$\mu=\widetilde{\mu}$.

緩増加超関数の空間上の Fourier 変換による証明　まず，$P(\mathbb{R})\subset\mathcal{S}'(\mathbb{R})$ を示す．$\mu\in P(\mathbb{R})$ に対して，$\langle\mu,-\rangle:\mathcal{S}(\mathbb{R})\to\mathbb{R}$ は有界線型汎関数であることを示す．$|\langle\mu,f\rangle|\le\left|\int_{\mathbb{R}}f(x)\mu(dx)\right|\le\|f\|_\infty$．このとき，$\mu\in\mathcal{S}'(\mathbb{R})$ の Fourier 逆変換が φ_μ である．$\mathcal{S}'(\mathbb{R})$ 上の Fourier 変換は同型対応である．

■

系 9.3.4. L^2-確率変数 $X\in\mathcal{L}^2$ がある $\alpha>0$ について $\varphi_X(u)=\exp(-|u|^\alpha)$ ならば，$\alpha=2$ で，$X\sim N(0,2)$ である．

9.3.2 特性関数の特徴付け

命題 9.3.5. 関数 $\varphi:\mathbb{R}\to\mathbb{C}$ が特性関数であるならば，次の 3 条件が成り立つ．

(1) $\varphi(0)=1$.
(2) 一様連続である．
(3) 正定値である：$\forall_{n\in\mathbb{N},\xi_j\in\mathbb{R},z_j\in\mathbb{C}}\ \sum_{j,k=1}^{n}\varphi(\xi_j-\xi_k)z_j\overline{z}_k\ge 0$.

定理 9.3.6 (Bochner). 関数 $\varphi:\mathbb{R}\to\mathbb{C}$ が次の 3 条件を満たすならば，特性関数である．特に，命題の 3 条件を満たすならば，特性関数である．

(1) $\xi=0$ で連続.
(2) $\varphi(0)=1$.
(3) 正定値である.

9.3.3　分布の滑らかさと特性関数

特性関数は，分布の滑らかさに関する情報を次のような形で保持している．

命題 9.3.7. $\nu\in\mathcal{P}^d$ とする．ある $m\in\mathbb{Z}_+$ に関して

$$\int|u|^m|\varphi_F(u)|du<\infty$$

が成り立つならば，ν は C^m-級密度関数

$$f(x)=\frac{1}{(2\pi)^d}\int e^{-iu\cdot x}\varphi_F(u)du$$

をもち，$\forall_{k\le m}\ f^{(k)}\in C_0(\mathbb{R}^d)$ が成り立つ.

9.3.4　特性関数と分布の収束の対応

例 9.3.8.

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

は一様分布 $U(0,1)$ に弱収束する．実際，積分の定義より，

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

定理 9.3.9. 確率測度 ν_n の特性関数を φ_n とする．(1)⇒(2) である．各点収束極限 φ が原点で連続であるとき，(2)⇒(1) でもある．

(1) ある確率測度 ν について，$\nu_n\Rightarrow\nu$ である.
(2) φ_n が関数 φ に各点収束する.

系 9.3.10 (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) $\nu_n\to\nu$.
(2) $\forall_{u\in\mathbb{R}}\ \varphi_n(u)\to\varphi(u)$.

［証明］.

(1)⇒(2)　定義より自明.
(2)⇒(1)　定理による.

■

9.3.5　特性関数の Taylor 展開

命題 9.3.11. $|\beta_r|<\infty$ のとき，特性関数 φ は C^r 級で,

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

特に，$\varphi^{(r)}(0)=i^r\alpha_r$.

系 9.3.12 (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 9.3.13. 分布 ν 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 9.3.14. $\beta_r < \infty$ とする．このとき，φ は $u=0$ の周りで展開

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

を持つ．

9.3.6　第 2 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 9.3.15. $\beta_r < \infty$ とする．このとき，関数 $\psi(u) := \log\varphi(u)$ は次の形の展開を持つ：

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

定義 9.3.16 (cuumulant)**.** 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 $\kappa_j = \frac{1}{i^n}\frac{\partial^n}{\partial t^n}\psi(u)\Big|_{u=0}$ を，分布 ν 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 9.3.17 (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) $\kappa_4 = \alpha_4 - 3\alpha_2^2 - 4\alpha_1\alpha_3 + 12\alpha_1^2\alpha_2 - 6\alpha_1^4 = \mu_4 - 3\mu_2^2$.

9.3.7　Laplace 変換

$\mathbb{R}_+$ 上の分布に限っては，特性関数と同等に Laplace 変換も扱う．

記法 9.3.18.

$$\mathcal{P}_+ := \{\mu \in \mathcal{P}(\mathbb{R}) \mid \operatorname{supp}\mu \subset [0,\infty)\}$$

と表す．

定義 9.3.19. $F \in \mathcal{P}_+$ に対して，

$$\mathcal{L}_F(u) := \int_0^\infty e^{-ux}F(dx)$$

で定まる関数 $\mathcal{L}_F : \mathbb{R}_+ \to \mathbb{R}_+$ を，F の **Laplace 変換**という．

定理 9.3.20 (一意性定理)**.** $F, G \in \mathcal{P}_+$ について，次の 2 条件は同値．

(1) $F = G$.
(2) $\mathcal{L}_F = \mathcal{L}_G$.

9.3.8　分布関数

$\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$ を満たす右連続な広義単調増加関数と，これが定める Radon 積分とは一対一対応する．

定義 9.3.21 (distribution function).

(1) $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$ を満たす右連続な広義単調増加関数を**分布関数**という．
(2) 実確率変数 X に対して，$F^X(x) := P[X \le x]$ により定まる関数 $F^X : \mathbb{R} \to [0,1]$ を**（累積）分布関数**という．
(3) $(\mathbb{R}, \mathcal{B}_1)$ 上の確率測度 ν に対して，$F_\nu(x) := \nu((-\infty, x]) = \int_{-\infty}^{x} f(y)dy$ により定まる関数 $F^X : \mathbb{R} \to [0,1]$ を**（累積）分布関数**という．

命題 9.3.22 (一意性定理)**.** $\nu_1 = \nu_2$ と $F_{\nu_1} = F_{\nu_2}$ は同値.

命題 9.3.23. $\mathbb{R}$ 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) $\nu_n \to \nu$.
(2) F の任意の連続点 $x \in \mathbb{R}$ について，$F_n(x) \to F(x)$.

9.3.9　積率母関数

積率母関数が $u = 0$ の近傍で定義されるとき，これにより確率分布が一意に決定される．理論的な欠点は常に存在するとは限らないところで，漸近展開などで用いられるのは特性関数である．存在するならば特性関数に対して $\varphi(u) = \mathfrak{M}(e^{iu})$ なる関係があり，解析接続を通じて一意性定理が述べられる．
積率母関数の対数をキュムラント母関数という．一方で，特性関数の対数は第 2 キュムラント母関数という．

定義 9.3.24 (moment generating function)**.** $(\mathbb{R}, \mathcal{B}_1)$ 上の確率測度 ν に対して，$M_\nu(t) := \int_R e^{tx}\nu(dx)$ により定まる関数 $\mathbb{R} \nrightarrow \mathbb{R}$ を**積率母関数**という．定義域を

$$D_\nu := \left\{ t \in \mathbb{R}^d \mid \int e^{t\cdot x} F(dx) < \infty \right\}$$

で表す．

補題 9.3.25. $\mathfrak{M}_\nu$ が原点の近傍で存在するならば，そこで滑らかであり，$\forall_{n\in\mathbb{Z}^d_+}\ \partial^n \mathfrak{M}_\nu(0) = \alpha_n$．特に，

(1) $\alpha_1 = \partial\mathfrak{M}_\nu(0) = \partial(\log \mathfrak{M}_\nu)(0) = \partial\mathfrak{C}_\nu(0)$.
(2) $\mu_2 = \partial^2(\log \mathfrak{M}_\nu)(0) = \partial^2\mathfrak{C}_\nu(0)$.

定理 9.3.26 (一意性定理)**.**

(1) $\nu \in \mathcal{P}^d$ とする．D_ν が $0 \in \mathbb{R}^d$ の近傍ならば，ある整数 ϵ が存在して，積率母関数 $\mathfrak{M}_\nu$ は $\mathcal{R}_\epsilon := ((-\epsilon, \epsilon) \times i\mathbb{R})^d$ 上の解析関数 $\mathfrak{M}^\dagger_\nu$ に解析接続され，$\varphi_\nu(-) = \mathfrak{M}^\dagger_\nu(i-)$ が成り立つ．
(2) 正則関数 $\mathfrak{M}^\dagger_\nu$ は，$z \in \mathcal{D}_\epsilon := \{\zeta \in \mathbb{C} \mid |\zeta| < \epsilon\}^d$ 上において，絶対収束級数

$$\mathfrak{M}^\dagger_\nu(z) = \sum_{k=1}^{\infty} \frac{1}{k!} \int (z \cdot x)^k \nu(dx)$$

なる展開を持つ．
(3) $\nu_1, \nu_2 \in \mathcal{P}^d$ のとき，原点のある開近傍 $U \subset D_{\nu_1} \cap D_{\nu_2}$ 上で $\mathfrak{M}_{\nu_1} = \mathfrak{M}_{\nu_2}$ ならば，$\nu_1 = \nu_2$ が成り立つ．

9.3.10 キュムラント母関数

定義 9.3.27. $\nu \in \mathcal{P}^d$ に対して，$\mathfrak{C}_\nu(t) := \log \mathfrak{M}_F(t)$ で定まる D_ν 上の関数を，F の**キュムラント母関数**という．

例 9.3.28 (キュムラント母関数は極めて少ししか変わらなくても，密度関数は見かけ上全く異なる例)**.**

9.3.11 積率問題

積率問題の解の存在は，Hahn-Banach の定理の応用である．ここでは確率的な文脈であるが，物理的にも，電荷密度 ρ に関する条件として読める．

問題 9.3.29 (moment problem)**.** 実数列 $\alpha : \mathbb{N} \to \mathbb{R}$ に対して，これを積率の列に持つ分布が一意に定まるか？

命題 9.3.30. 有限個への制限 $\alpha|_{[N]}$ について，これを満たす積率問題の解は存在する．

定理 9.3.31 (存在の必要十分条件)**.** 区間 $[a, b]$ 上の分布であって，積率 α を持つものを考える．

(1) 積率問題は解を持つ．

(2) $\exists_{c>0}\ \forall_{N\in\mathbb{N}}\ \forall_{a_k\in\mathbb{R}}\ \left|\sum_{k=0}^N a_k\alpha_k\right| \le c \max_{a\le x\le b}\left|\sum_{k=0}^N a_k x^k\right|$.

定理 9.3.32 (一意性の十分条件)**.** $(\alpha_n)_{n\in\mathbb{N}}$ が $\exists_{t_0>0}\ \sum_{n=1}^\infty \frac{|\alpha_n|}{n!} t_0^n < \infty$ を満たすならば，α を積率とする分布はただ一つである．

例 9.3.33. 3 次以上のキュムラントがすべて 0 になる確率分布は，正規分布に限る．

9.3.12 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．

9.4 一次元連続分布の例

9.4.1 一様分布

定義 9.4.1 (uniform distribution)**.** $(\mathcal{X}, \mathcal{B}_1)$ 上の**一様分布** $U(a, b)$ $(a < b \in \mathbb{R})$ とは，確率密度関数 f

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

が定める確率分布をいう．

命題 9.4.2.

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

■

9.4.2 Gamma 分布・指数分布・カイ 2 乗分布

9.4.2.1 Gamma 分布の性質

定義 9.4.3 (gamma distribution)**.** $(\mathcal{X},\mathcal{B}_1)$ 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 9.4.4.

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

(2) $\alpha_1 = \frac{\nu}{\alpha}$.

(3) $\mu_2 = \frac{\nu}{\alpha^2}$.

(4) $\gamma_1 = \frac{2}{\sqrt{\nu}}, \gamma_2 = 3+\frac{6}{\nu}$.

[証明] .

(1)

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

でこの後 path を考えるらしい．

■

9.4.2.2 指数分布

カイ 2 乗分布は線型理論の騎手で，Pearson による．

定義 9.4.5 (exponential distribution, chi-square distribution)**.**

(1) $\mathrm{Exp}\,(\gamma) := G(\gamma,1)$ 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]

補題 9.4.6 (確率密度関数)**.**

(1) 指数分布の確率密度関数は，

$$g(x;\gamma,1) = \frac{1}{\Gamma(1)}\gamma^1 x^0 e^{-\gamma x}1_{x>0} = \frac{\gamma}{e^{\gamma x}}1_{x>0}.$$

(2)

補題 9.4.7. 指数分布 X_λ の特性関数は

$$E[e^{iuX_\lambda}] = \frac{\lambda}{\lambda-iu}.$$

[†2] 統計推測で重要なクラスである．

9.4.3 正規分布

定義 9.4.8 (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 9.4.9.

(1) $\displaystyle\int_{\mathbb{R}} \phi(x; \mu, \sigma^2) dx = 1.$

(2) $\varphi(u) = \exp\left(i\mu u - \dfrac{1}{2}\sigma^2 u^2\right).$

(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 9.3.12,

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

■

定義 9.4.10 (standard normal distribution)**.** $N(0, 1)$ を**標準正規分布**という．

補題 9.4.11.

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

［証明］.

$$P\left(\frac{X-\mu}{\sigma} \leq Z\right) = P\left(X \leq \mu + \sigma Z^{\dagger 4}\right)$$
$$= \int_{-\infty}^{\mu+\sigma Z} \phi(x; \mu, \sigma^2) dx.$$

■

注 9.4.12. 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．

9.4.4 Beta 分布

定義 9.4.13 (beta distribution)**.** $(\mathcal{X}, \mathcal{B}_1)$ 上の**（第１種）ベータ分布** $\mathrm{Beta}(\alpha, \beta) = B_E(\alpha, \beta)$ $(\alpha, \beta \in \mathbb{R}_{>0})$ とは，確率密度関数

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

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

[†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．

命題 9.4.14.

(1) $\alpha_1 = \dfrac{\alpha}{\alpha+\beta}$.

(2) $\mu_2 = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

(3) $\gamma_1 = \dfrac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$.

(4) $\gamma_2 = \dfrac{3(\alpha+\beta+1)(\alpha^2\beta+\alpha\beta^2+2\alpha^2-2\alpha\beta+2\beta^2)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$.

［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 9.4.15. 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．

9.4.5　Cauchy 分布

定義 9.4.16 (Cauuchy 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 9.4.17.

(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 9.4.18. 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．

9.4.6　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 9.4.19 (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 9.4.20.

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

(2) $\alpha_1 = \alpha\Gamma\left(\dfrac{\nu+1}{\nu}\right)$.

(3) $\mu_2 = \alpha^2 \left\{ \Gamma\left(\frac{\nu+2}{\nu}\right) - \Gamma\left(\frac{\nu+1}{\nu}\right)^2 \right\}$.

命題 9.4.21. $X \sim \mathrm{Exp}\left(\frac{1}{\alpha^\nu}\right)$ ならば，$X^{\frac{1}{\nu}} \sim W(\nu, \alpha)$.

9.4.7 対数正規分布

定義 9.4.22 (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 9.4.23.

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

命題 9.4.24.

(1) $\alpha_1 = \exp\left(\mu + \frac{\sigma^2}{2}\right)$.

(2) $\mu_2 = \left(e^{\sigma^2} - 1\right) \exp(2\mu + \sigma^2)$.

(3) $\gamma_1 = (\omega + 2)\sqrt{\omega - 1}$．ただし，$\omega := \exp(\sigma^2)$ とした．

(4) $\gamma_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3$.

9.4.8 logistic 分布

定義 9.4.25 (logistic distribution). $(\mathcal{X}, \mathcal{B}_1)$ 上の**ロジスティック分布** $\mathrm{Log}\,(\mu, \sigma)$ $(\mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0})$ 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 9.4.26.

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

［証明］．

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

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 9.4.27.

(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) $\alpha_1=\mu,\mathfrak{m}_2=\dfrac{\pi^2\sigma^2}{3},\mu_4=\dfrac{7\pi^4\sigma^4}{15}$.
(3) $\gamma_1=0,\gamma_2=4.2$.
(4) $\varphi(u)=e^{i\mu u}\dfrac{\pi\sigma u}{\sinh(\pi\sigma u)}$.

［証明］.

(1) 奇数の時は対称性より．
(2) $\zeta(2)=\dfrac{\pi^2}{6},\zeta(4)=\dfrac{\pi^4}{90}$ 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．

■

9.4.9　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 9.4.28. $(\mathcal{X},\mathcal{B}_1)$ 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

$$F(x)=1-\left(\frac{b}{x}\right)^a1_{x\ge b}$$

確率密度関数を

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

とする分布をいう．

命題 9.4.29.

(1) $\alpha_1=ab(a-1)^{-1}$ $(a>1)$.
(2) $\mu_2=ab^2(a-1)^{-2}(a-2)^{-1}$ $(a>2)$.
(3) $\gamma_1=2\dfrac{a+1}{a-3}\sqrt{\dfrac{a-2}{a}}$ $(a>3)$.
(4) $\gamma_2=\dfrac{3(a-2)(3a^2+a+2)}{a(a-3)(a-4)}$ $(a>4)$.

9.4.10　逆正規分布

ドリフト付きの Brown 運動の到達時間の分布として現れる．キュムラント母関数 $\log M$ が，正規分布のキュムラント母関数の逆数になっていることから．

定義 9.4.30 (inverse Gaussian distribution / Wald distribution)**.** 確率密度関数

$$f(x;\delta,\gamma)=1_{(0,\infty)}(x)\frac{\delta e^{\gamma\delta}}{\sqrt{2\pi}}x^{-3/2}\exp\left(-\frac{1}{2}\left(\gamma^2x+\frac{\delta^2}{x}\right)\right)$$

で定まる確率分布 $\mathrm{IG}(\delta,\gamma):\mathbb{R}_{>0}\times\mathbb{R}_{\ge0}\to P(\mathbb{R})$ を**逆正規分布**または **Wald 分布**という．

補題 9.4.31.

(1) $\gamma > 0$ のとき，確率密度関数は

$$f(x) = 1_{(0,\infty)}(x)\left(\frac{\delta^2}{2\pi}\right)^{1/2} x^{-3/2}\exp\left(-\frac{\delta^2(x-\delta\gamma^{-1})^2}{2(\delta\gamma^{-1})^2 x}\right)$$

とも表せる．

(2) $\gamma = 0$ のとき，確率密度関数は，$c := \delta^2$ と定めて，

$$f(x;c) = 1_{(0,\infty)}(x)\sqrt{\frac{c}{2\pi}}x^{-3/2}\exp\left(-\frac{c}{2x}\right)$$

と表せる．これが定める分布を **Levy 分布**といい，$\mathrm{Levy}(c) := \mathrm{IG}(c^{1/2}, 0)$ と表す．

(3) $\mathrm{IG}(\delta,\gamma)$ の積率母関数は，

$$\begin{aligned} M(s) &= \int_0^\infty e^{sx} f(x)dx \\ &= \exp\left(\delta\left(\gamma - \sqrt{\gamma^2 - 2s}\right)\right) && s \le \frac{\gamma^2}{2}, \gamma \ge 0 \\ &= \exp\left(\gamma\delta\left(1 - \sqrt{1 - \frac{2s}{\gamma^2}}\right)\right) && s \le \frac{\gamma^2}{2}, \gamma > 0. \end{aligned}$$

(4) $\mathrm{IG}(\delta,\gamma)$ の特性関数は $\varphi(u) = \exp\left(\gamma\delta\left(1-\sqrt{1-\frac{2iu}{\gamma^2}}\right)\right)$ となる．

(5) キュムラント母関数は $\log\varphi(u) = \frac{\delta}{\gamma}iu + \sum_{r=2}^\infty (2r-3)(2r-5)\cdots 1\cdot\frac{\delta}{\gamma^{2r-1}}\cdot\frac{(iu)^r}{r!}$ で，キュムラントは

$$\kappa_1 = \alpha_1 = \frac{\delta}{\gamma}, \qquad \kappa_r = (2r-3)(2r-5)\cdots 1\cdot\frac{\delta}{\gamma^{2r-1}}$$

(6) Levy 分布 $\mathrm{Levy}(c)$ の Laplace 変換は

$$\int_0^\infty e^{-\lambda x} f(x)dx = e^{-\sqrt{c\lambda}} \quad (\lambda \in \mathbb{R}_+)$$

である．

9.4.11　逆 Gamma 分布

命題 9.4.32. $X \sim \Gamma(n/2, n/2)$ $(n \in \mathbb{N})$ とし，確率変数 $Y := 1/X$ を考えると，この確率密度関数は

$$\begin{aligned} f_Y(y) &= f_X(1/y)y^{-2} \\ &= \frac{(n/2)^{n/2}}{\Gamma(n/2)}(1/y)^{n/2-1}\exp\left(-\frac{n/2}{y}\right)y^{-2} \\ &= \frac{(n/2)^{n/2}}{\Gamma(n/2)}y^{-n/2-1}\exp\left(-\frac{n/2}{y}\right) && y > 0 \end{aligned}$$

と表せる．

［証明］． $y > 0$ のとき $P[Y \le y] = P[X \ge 1/y]$ である． ■

9.4.12　Laplace 分布

定義 9.4.33. 密度関数

$$f(x) = \frac{1}{2}e^{-|x|}$$

が定める分布を**標準 Laplace 分布（両側指数分布）**という．

補題 9.4.34. (1) 特性関数は $\varphi(u) = (1+u^2)^{-1}$.

(2) X, Y, X', Y' が独立に標準正規分布 $N(0,1)$ に従うとする．このとき，$V := XY + X'Y'$ は密度 f を持つ Laplace 分布に従う．

9.4.13　Marcenko-Pastur 分布

MP$(\lambda, 1)$ 分布は Wishart 行列の固有値の標本スペクトル分布の漸近分布として導出された．

定義 9.4.35. パラメータ $(\lambda, \sigma^2) \in \mathbb{R}^2_{>0}$ が定める定数 $a := \sigma^2(1-\sqrt{\lambda})^2, b := \sigma^2(1+\sqrt{\lambda})^2$ について，確率密度関数

$$f(x;\lambda,\sigma^2) := \frac{1}{2\pi\sigma^2}\frac{\sqrt{(b-x)(x-a)}}{\lambda x}1_{[a,b]}(x) + 1_{[1,\infty)}(\lambda)\left(1-\frac{1}{\lambda}\right)\delta_0(x)$$

が定める分布 $\mathrm{MP}(\lambda,\sigma^2) : \mathbb{R}^2_{>0} \to P(\mathbb{R})$ を **Marcenko-Pastur 分布**という．

補題 9.4.36. $X \sim \mathrm{MP}(\lambda,\sigma^2)$ ならば，$\sigma^{-2}X \sim \mathrm{MP}(\lambda) := \mathrm{MP}(\lambda, 1)$.

命題 9.4.37. $X_{ij} \sim \text{i.i.d.} N(0,1)$ $(i \in [d], j \in [n])$ を成分とする $d \times n$ ランダム行列 X に対し，$n^{-1}XX^\perp$ の固有値を $0 \le \lambda_1 \le \cdots \le \lambda_d$ とする．

(1) 標本スペクトル分布は $\mu_{d,n}(dx) := \frac{1}{d}\sum_{i=1}^d \delta_{\lambda_i}(dx)$ で与えられる．

(2) $d = d_n$ が $\frac{d_n}{n} \to \lambda \in (0,\infty)$ を満たすとき，$\mu_{d_n,n} \xrightarrow{d} \mathrm{MP}(\lambda)$ を満たす．

9.4.14　Pearson 系

定義 9.4.38. $\mathbb{R}$ 上の確率密度関数 p であって，微分方程式

$$\frac{d}{dx}\log p(x) = \frac{(x-\lambda)-a}{b_2(x-\lambda)^2 + b_1(x-\lambda) + b_0}$$

を満たす

$$p(x) = C\exp\left(\int \frac{(x-\lambda)-a}{b_0 + b_1(x-\lambda) + b_2(x-\lambda)^2}dx\right)$$

の形をした分布を **Pearson 系**という．以降，$\lambda = 0$ とする．

(1) I 型分布とは，$p(x) = C\left(1-\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2}$ $(x \in (a_1, a_2))$ の形のものをいう．ただし，$a_1 < 0 < a_2, m_1, m_2 > -1$.

例 9.4.39.

(1) Beta 分布は I 型 Pearson 分布である．

(2)

9.4.15　Poisson 分布確率変数の和差

定義 9.4.40. 独立な $Y \sim \mathrm{Pois}(\alpha), Z \sim \mathrm{Pois}(\beta)$ に対して，

(1) $X = Y + 2Z$ の分布を，パラメータ $(\alpha,\beta) \in \mathbb{R}^2_{>0}$ の **Hermite 分布**とよび，$\mathrm{Hermite}(\alpha,\beta)$ で表す．

(2) $X = Y - Z$ の分布を，パラメータ $(\alpha,\beta) \in \mathbb{R}^2_{>0}$ の **Skellam 分布**とよび，$\mathrm{Skellam}(\alpha,\beta)$ で表す．

補題 9.4.41.

(1) Hermite 分布の特性関数は，$\varphi(u) = \exp\left(\alpha(e^{iu}-1) + \beta(e^{2iu}-1)\right)$.

(2) Hermite 分布の確率母関数は $g(z) = \exp\left(\alpha(z-1) + \beta(z^2-1)\right)$.

(3) Skellam 関数の特性関数は $\varphi(u) = \exp\left(\alpha(e^{iu}-1) + \beta(e^{-iu}-1)\right)$.

9.5 確率変数の独立性と期待値

有限個の確率変数 $X_1, \cdots, X_n$ の独立性とは，これが定める $(X_1, \cdots, X_n)$ の結合分布が，各 X_i の分布 P^{X_i} が定める直積速度 $P^{X_1} \times \cdots \times P^{X_n}$ に一致することをいう．本質的に「直積」とみなせるものを独立性という．

9.5.1 独立な確率変数の積の期待値

記法 9.5.1. $X_j : (\Omega, \mathcal{F}, P) \to (\mathcal{X}_j, \mathcal{B}_j)$ を確率変数とする．

命題 9.5.2. $n \geq 2$ を整数とする．可測関数 $f_j : \mathcal{X}_j \to \overline{\mathbb{R}}$ について，$f_j(X_j) \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ とする．このとき，$X_1, \cdots, X_n$ が独立ならば，積 $\prod_{i=1}^n f_i(X_i)$ も可積分で，

$$E\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n E[f_i(X_i)]$$

を満たす．

注 9.5.3. 複素数値可測関数にも一般化出来る．

系 9.5.4 (独立性の十分条件)**.** 任意の有界可測関数 f_i について

$$E\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n E[f_i(X_i)]$$

が成り立つならば，$X_1, \cdots, X_n$ は独立である．

9.5.2 確率変数の独立性と共分散

命題 9.5.5. $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ を独立とする．このとき，$\mathrm{Cov}[X, Y] = 0$.

9.5.3 確率変数の独立性と特性関数

平均作用素 E の積に対する振る舞いによって，独立性の十分条件を与えることが出来た．特性関数によっても特徴付けることが出来る．

記法 9.5.6. 多変量確率変数 Y の特性関数を φ_Y と表す．X_j $(j \in [n])$ を d_j 次元確率変数とする．$X := (X_1, \cdots, X_n)$ は $d := \sum_{j=1}^n d_j$ 次元確率変数である．X, X_j の特性関数は

$$\varphi_X(u) = E\left[e^{i\sum_{j=1}^n u_j \cdot X_j}\right] \; (u \in \mathbb{R}^d), \qquad \varphi_{X_j}(u_j) = E\left[e^{iu_j \cdot X_j}\right] \; (u_j \in \mathbb{R}^{d_j})$$

となる．

定理 9.5.7 (Kac)**.** 確率変数列 $X_1, \cdots, X_n$ について，次の 2 条件は同値．

(1) $X_1, \cdots, X_n$ は独立である．

(2) $\varphi_X = \varphi_{X_1} \otimes \cdots \otimes \varphi_{X_n} := \prod_{j=1}^n \varphi_{X_j}(u_j)$ である．

系 9.5.8. d 次元確率変数の列 $X_1, \cdots, X_n$ が独立ならば，

$$\varphi_{\sum_{j=1}^n X_j}(u) = \prod_{j=1}^n \varphi_{X_j}(u) \; (u \in \mathbb{R}^d)$$

9.5.4　独立同分布に従う確率変数列

記法 9.5.9. $X_1,\cdots,X_n$ が $\mathbb{R}^d$ に値を取る i.i.d. であって，$E[X_1]=\mu, \mathrm{Var}[X_1]=\Sigma$ であることを，$X_1,\cdots,X_n \sim \text{i.i.d.}(\mu,\Sigma)$ で表す．

9.5.5　独立確率変数の和と畳み込み

確率変数の和は，分布の畳み込みに対応し，特性関数のテンソル積に対応する．

系 9.5.10. d 次元確率変数 $X_1,\cdots,X_n$ が独立であれば，$X := X_1+\cdots+X_n$ について

$$\forall_{u\in\mathbb{R}^d}\quad \varphi_X(u)=\prod_{j=1}^n \varphi_{X_j}(u).$$

例 9.5.11. Hermite 分布と Skellam 分布**??**.

記法 9.5.12. $A-x := \left\{y\in\mathbb{R}^d \mid y+x\in A\right\}$ と表す．

定義 9.5.13 (convolution)**.** $\nu_1,\nu_2\in\mathcal{P}(\mathbb{R}^d)$ に対して，

$$\nu(A) := \int_{\mathbb{R}^d} \nu_1(A-x)\nu_2(dx)$$

で定まる確率測度 ν を $\nu_1*\nu_2$ で表す．

補題 9.5.14.

(1) $\nu_1*\nu_2(A)=\nu_2*\nu_1(A)=\displaystyle\int_{\mathbb{R}^{2d}} 1_A(x_1+x_2)\nu_1(dx_1)\nu_2(dx_2)$.
(2) 独立な確率変数 X_1,X_2 が $P^{X_1}=\nu_1, P^{X_2}=\nu$ を満たすとき，$P^{X_1+X_2}=\nu_1*\nu_2$.
(3) $\varphi_{\nu_1*\nu_2}=\varphi_{\nu_1}\otimes\varphi_{\nu_2}$.

(3) は ν_1,ν_2 の独立性を特徴付けない点に注意．

補題 9.5.15. $P(\mathbb{R}^d)$ は $*$ を積として，単位元 δ_0 を持つ可換な半群をなす．

定義 9.5.16 (reproducing property)**.** ある分布族について，畳み込みについて閉じていることを**分布族の再生性**という．

例 9.5.17.

(1) $G(\alpha,\nu_1)*G(\alpha,\nu_2)=G(\alpha,\nu_1+\nu_2)$.
(2) $N(\mu_1,\Sigma_1)*N(\mu_2,\Sigma_2)=N(\mu_1+\mu_2,\Sigma_1+\Sigma_2)$.
(3) $C(\mu_1,\sigma_1)*C(\mu_2,\sigma_2)=C(\mu_1+\mu_2,\sigma_1+\sigma_2)$.
(4) $B(n_1,p)*B(n_2,p)=B(n_1+n_2,p)$.
(5) $\mathrm{Pois}(\lambda_1)*\mathrm{Pois}(\lambda_2)=\mathrm{Pois}(\lambda_1+\lambda_2)$.

9.6　多次元分布の扱い

9.6.1　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 9.6.1 (joint / simultaneous distribution, marginal distribution)**.** $X=(X_1,\cdots,X_d):\Omega\to\mathbb{R}^d$ を確率変数とする．

(1) $(\mathbb{R}^d, \mathcal{B}_d)$ 上の確率分布 P^X を $X_1, \cdots, X_d$ 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．$P^{(X_{i_1},\cdots,X_{i_k})}$ も，$(X_{i_1}, \cdots, X_{i_k})$ の周辺分布という．

例 9.6.2 (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 $(X_1, \cdots, X_k)$ の分布を k **項分布** $M(n; p_1, \cdots, p_k)$ 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．

9.6.2　共分散

定義 9.6.3 (covariance, correlation coefficient)**.** 2 乗可積分実確率変数 X, Y について，

(1) $\mathrm{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])]$ を**共分散**という．
(2) $\rho(X, Y) := \dfrac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}}$ を**相関係数**という．

命題 9.6.4. 2 乗可積分実確率変数 X, Y, Z について，

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

例 9.6.5 (多項分布の共分散)**.** $X = (X_1, \cdots, X_k) \sim M(n; p_1, \cdots, p_k)$ とする．

(1) $E[X_i] = np_i$.
(2) $\mathrm{Var}[X_i] = np_i(1 - p_i)$.
(3) $\mathrm{Cov}[X_{i_1}, X_{i_2}] = -np_{i_1}p_{i_2}$ $(i_1 \neq i_2 \in [k])$.

[証明] . 多項定理より，

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

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

(1) 両辺の $(u_1, \cdots, u_k) = (0, \cdots, 0)$ における u_i 偏微分係数より，$np_i = \sum_{x_1,\cdots,x_k} {}^{*}x_iP^X[\{(x_1, \cdots, x_k)\}] = E[X_i]$.
(2) 同様に u_i の 2 階微分を考えて，$E[X_i^2] = n(n-1)p_i^2 + np_i$ を得る．よって，$\mathrm{Var}[X_i] = E[X_i^2] - (E[X_i])^2 = np_i(1 - p_i)$.
(3) $i_1 \neq i_2$ に関して，順に u_{i_1}, u_{i_2} での偏微分を考えることにより，$E[X_{i_1}X_{i_2}] = n(n-1)p_{i_1}p_{i_2}$．よって，共分散公式 9.6.4 より，

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

■

要諦 9.6.6. 証明が技巧的すぎる，多項定理に指数関数を代入する．

9.6.3　共分散行列

定義 9.6.7 (covariance matrix, variance-covariance matrix)**.** $X : \Omega \to \mathbb{R}^d$ 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,

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

を X, Y の**共分散行列**という.

(3) $\mathrm{Cov}[X, X]$ を X の**分散共分散行列**または**分散行列**と呼ぶ.

注 9.6.8. 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。[†5]

記法 9.6.9 (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 $E : M_{ij}(\mathbb{R}) \to M_{ij}(\mathbb{R}); E[M] = [E[M_{ij}]]$ と拡張すると，平均ベクトルは $E[M]$，共分散行列は $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])^T]$ と表せる.

命題 9.6.10.

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

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

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

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

命題 9.6.11 (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,

$$E[X^T GX] = \mathrm{Tr}\,(G\mathrm{Var}[X]) + E[X]^T GE[X].$$

9.6.4 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 9.6.12. $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$ に対して，$\partial_i := \dfrac{\partial}{\partial x_i}$ とし，$n := (n_1, \cdots, n_d) \in \mathbb{Z}_+^d$ に対して,

$$|n| := n_1 + \cdots + n_d, \qquad n! := n_1! \cdots n_d!,$$
$$x^n := x_1^{n_1} \cdots x_d^{n_d}, \qquad \partial^n := \partial_1^{n_1} \cdots \partial_d^{n_d}.$$

ただし，$x_j^0 = 1, \partial^0 = 1$ とよむ.

定義 9.6.13 (moment, central moment)**.** d 次元確率変数 $X = (X_1, \cdots, X_d)$ 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 9.6.14. $(\mathbb{R}^d, \mathcal{B}_d)$ 上の確率測度 ν に対して，可積分性の仮定の下で,

(1) $\mu := \left(\displaystyle\int_{\mathbb{R}^d} x_i \nu(dx)\right) \in \mathbb{R}^d$ を**平均（ベクトル）**という.

(2) $\alpha_n := \displaystyle\int_{\mathbb{R}^d} x^n \nu(dx)$ を n **次の積率**という.

[†5] https://ja.wikipedia.org/wiki/分散共分散行列

(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．

9.6.5　キュムラントの定義

定義 9.6.15. 第 2 キュムラント母関数によるキュムラントの定義

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

は，n を多重指数 $\mathbf{n}$ に読み替えることで，$\kappa_{\mathbf{n}} := i^{-|\mathbf{n}|}\partial^{\mathbf{n}}\psi(0)$ そのまま拡張される．

記法 9.6.16. $(\partial_{u_i})_0$ を，$u := (u_i) = 0$ における u_i-偏微分係数を表す．

議論 9.6.17 (キュムラントのテンソル表現). d 次元確率分布 ν について，$\alpha_1, \cdots, \alpha_r \in [d]$ に対して

$$\lambda^{\alpha_1\cdots\alpha_r} := (-i)^r (\partial_{u_{\alpha_1}})_0 \cdots (\partial_{u_{\alpha_r}})_0 \log\varphi(u)$$

と定める．これも ν の r 次のキュムラントといい，行列 $(\lambda^{\alpha_1\cdots\alpha_r})_{\alpha_1,\cdots,\alpha_r\in[d]}$ は対称テンソルを定める．

要諦 9.6.18. $\mathbf{r} := (r_1, \cdots, r_d) \in \mathbb{Z}_+^d$ が $r_k := |\{j \in [r] \mid \alpha_j = k\}|$ を満たすとする．このとき，

$$\lambda^{\alpha_1\cdots\alpha_r} = \kappa_{\mathbf{r}}$$

9.7　多次元連続分布

1 次元の絶対連続分布の概念を容易に d 次元に拡張できる．

9.7.1　多次元の確率密度関数

Radon-Nikodym の定理より，$\mathbb{R}^d$ 上の絶対連続分布と確率密度関数は一対一対応する．

記法 9.7.1. $\mathbb{R}^d$ の矩形 $\prod_{i=1}^{d}(a_i, b_i]$ の全体を $\mathcal{E}^d$ と表す．

定義 9.7.2. $\mathbb{R}^d$ 上の確率分布 ν が**絶対連続分布**であるとは，積分可能な関数 $f : \mathbb{R} \to \mathbb{R}_+$ が存在して

$$\forall_{A\in\mathcal{E}^d} \quad \nu(A) = \int_A f(x)dx$$

と表せることをいう．f を ν の**確率密度関数**という．

9.7.2　確率密度関数と独立性

独立性と特性関数について成り立つ Kac の定理 9.5.7 と同様の状況が，確率密度関数についても成り立つ．

命題 9.7.3 (独立性の確率密度関数による特徴付け). P^X が確率密度関数 p を持つとする．このとき，次の 2 条件は同値.

(1) $X_1, \cdots, X_n$ は独立.

(2) $p = \bigotimes_{i=1}^{n} p_i$ a.e. .

ただし，p_i は X_i の周辺密度関数とした．

9.7.3　畳み込み

絶対連続分布の畳み込みは，その確率密度関数の畳み込みに対応する．

9.7.4　変数変換と確率密度関数

例 9.7.4 (独立正規分布の比は Cauchy 分布)**.** $Y_1, Y_2 \sim N(0,1)$ を独立とする．$X := \dfrac{Y_1}{Y_2}$ は Cauchy 分布にしたがう．

例 9.7.5 (Box-Muller transform)**.** $Y_1, Y_2 \sim U(0,1)$ を独立とする．

$$\begin{cases} X_1 = \sqrt{-2\log Y_1}\cos(2\pi Y_2) \\ X_2 = \sqrt{-2\log Y_2}\sin(2\pi Y_2) \end{cases}$$

と定めると，これは独立な標準正規確率変数となる．

9.7.5　多変量正規分布

9.7.5.1　定義と特性値

定義 9.7.6. $\mu \in \mathbb{R}^d$，$\Sigma \in \mathrm{GL}_d(\mathbb{R})$ を正定値実対称行列として，$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上の測度

$$\mu_{\mu,\Sigma}(dx) = \phi(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}(\det\Sigma)^{1/2}}\exp\left(-\frac{1}{2}(x-m)^\perp \cdot \Sigma^{-1}(x-m)\right)dx$$

を，平均値 μ，共分散行列 Σ に関する d **次元正規分布**といい，$N_d(0,\Sigma)$ で表す．

命題 9.7.7. d 次元正規分布 $N_d(0,\Sigma)$ の特性関数は

$$\varphi(u) = \exp\left(i\mu\cdot u - \frac{1}{2}u^\top\Sigma u\right)$$

系 9.7.8. キュムラント母関数は $\psi(u) = \log\varphi(u) = i\mu\cdot u - \dfrac{1}{2}u^\top\Sigma u$ である．特に，平均ベクトルは μ，分散共分散行列が Σ で，３次以上のキュムラントはすべて消えている．

9.7.5.2　成分間の独立性の特徴付け

命題 9.7.9 (多次元正規確率変数の成分間の独立性)**.** $Y_1 := (X_1,\cdots,X_{k_1})^\top,\cdots,Y_l = (X_{k_{l-1}},\cdots,X_d)^\top$ $(l \ge 2)$ のように，X を l 個の確率変数 $Y_1,\cdots,Y_l$ に分ける．この分割に対して，Σ のブロック $\Sigma_{a,b} := \mathrm{Cov}[Y_a, Y_b]$ $(a,b \in [l])$ を考える．

(1) $Y_1,\cdots,Y_l$ は独立．

(2) $\forall_{a,b\in[l]}\ a \ne b \Rightarrow \Sigma_{a,b} = O$.

9.7.5.3　半正定値行列への拡張

任意の半正定値行列は分散共分散行列とみなすことが出来る．だから，それに対応した正規分布が存在するはずである．これは，次のように定義する．

定義 9.7.10. $\Sigma \in M_d(\mathbb{R})$ を半正定値行列とする（特に，非退化の可能性がある）．列 (Σ_n) を $\Sigma_n := \Sigma + n^{-1}I_n$ で定めるとこれは正定値行列の列で，$\Sigma_n \xrightarrow{n\to\infty} \Sigma$．対応する正規分布 $\nu_n := N_d(\mu,\Sigma_n)$ の特性関数の列は

$$\varphi_{\nu_n}(u) = \exp\left(iu^\top\mu - \frac{1}{2}u^\top\Sigma_n u\right) \xrightarrow{n\to\infty} \varphi(u) = \exp\left(iu^\top\mu - \frac{1}{2}u^\top\Sigma u\right)$$

と各点収束し，原点で連続である．よって，Bochner の定理 9.3.6 より，これはある確率分布 ν の特性関数である．この分布を，d 変量正規分布 $N_d(\mu,\Sigma)$ と呼ぶ．

9.7.5.4　正規分布に従うことの特徴付け

命題 9.7.11. $X:\Omega\to\mathbb{R}^d$ を確率変数，$\mu\in\mathbb{R}^d$ を平均ベクトル，$\Sigma\in M_d(\mathbb{R})$ を半正定値行列とする．このとき，次の 3 条件は同値．

(1) $X\sim N_d(\mu,\Sigma)$.

(2) $\forall_{u\in\mathbb{R}^d}\; u^\top X\sim N(u^\top\mu,u^\top\Sigma u)$.

(3) $\forall_{A\in M_{nd}(\mathbb{R}),b\in\mathbb{R}^n}\; AX+b\sim N_n(A\mu+b,A\Sigma A^\top)$.

系 9.7.12 (ユニタリ変換は成分の独立性を変えない). $X\sim N_d(\mu,\sigma^2I_d),U\in M_d(\mathbb{R})$ を直交行列とする．このとき，$UX\sim N_d(U\mu,\sigma^2I_d)$ である．特に，UX の成分は再び独立である．

9.7.6　Dirichlet **分布**

ベータ分布の高次元化である．

定義 9.7.13. $\Delta_p:=\left\{(x_1,\cdots,x_p)\in\mathbb{R}^p_{>0}\mid\sum_{i=1}^p x_i<1\right\}$ $(p\in\mathbb{N})$ について，

$$f(x_1,\cdots,x_p)=\frac{\Gamma(\nu)}{\prod_{i=1}^{p+1}\Gamma(\nu_i)}\left(\prod_{i=1}^d x_i^{\nu_i-1}\right)\left(1-\sum_{i=1}^p x_i\right)^{\nu_{p+1}-1}1_{\Delta_p}(x_1,\cdots,x_p)\qquad \nu_i>0,\nu:=\sum_{i=1}^{p+1}\nu_i$$

が定める分布を **Dirichlet 分布** $\mathrm{Dirichlet}(\nu_1,\cdots,\nu_{p+1}):\mathbb{R}^{p+1}_{>0}\to P(\mathbb{R}^p)$ という．

補題 9.7.14. $(X_1,\cdots,X_p)\sim\mathrm{Dirichlet}(\nu_1,\cdots,\nu_{p+1})$ とする．

$$\forall_{j\in[p]}\;X_j\sim\mathrm{Beta}\left(\nu_j,\sum_{i\ne j}\nu_i\right)$$

9.8　指数型分布族

ここから，分布の性質の議論から，分布族の性質を考えることへ，視点を移し，「確率分布へのパラメータの入り方」を議論する．

9.8.1　定義と例

指数型分布族は，有限個の x の関数 $T_i\in\mathrm{Meas}(\mathcal{X},\mathbb{R})$ の 1 次式の指数関数の $\mathrm{Meas}(\mathcal{X},\mathbb{R})$-倍で表せる Radon-Nikodym 微分を持つクラスである．このクラスに対しては，

定義 9.8.1. $\mathcal{P}=\{P_\theta\}_{\theta\in\Theta}\subset P(\mathcal{X})$ を参照測度 ν に対する絶対連続確率分布の族とし，p_θ をその Radon-Nikodym 微分とする．

$$p_\theta(x)=g(x)\exp\left(\sum_{i=1}^m a_i(\theta)T_i(x)-\psi(\theta)\right)$$

と表せるとき，$\mathcal{P}$ を**指数型分布族**という．

9.8.2　標準指数型分布族

定義 9.8.2 (密度関数による指数型分布族の生成). 参照測度 ν に関する密度関数 g を考える：$g \geq 0, \int g(x)\nu(dx) = 1$.

$$\Theta_0(g) := \left\{\theta \in \mathbb{R}^p \;\middle|\; \int e^{\theta \cdot x} g(x)\nu(dx) < \infty\right\}$$

とおくと，$\{0\} \in \Theta_0(g) \neq \emptyset$ である．g のキュムラント母関数を $\psi(\theta) := \log\left(\int e^{\theta \cdot x} g(x)\nu(dx)\right)$ と表すとき，密度関数

$$f(x|\theta) := g(x)\exp\left(x \cdot \theta - \psi(\theta)\right) \quad \theta \in \Theta_0(g)$$

を，$g(x)\nu(dx)$ **で生成される** p **次の自然指数型分布族**という．

補題 9.8.3. $\Theta_0(g)$ は凸集合である．これを**自然パラメータ空間**という．

9.8.3　共役事前分布

Howard Raiffa と Robert Schlaifer によるベイジアン決定理論で作られた概念である．共役というのは，事後分布が，事前分布の代数的な閉式で表せることを指している（したがって，数値積分が必要なく，解析的に計算可能）．

定理 9.8.4 (Bayes の定理（密度版）). 確率変数 (θ, X) は，参照測度 $\mu(dx)\nu(d\theta)$ に関する同時密度関数 $f(x,\theta)\mu(dx)\nu(d\theta)$ をもつとし，$f_X(x), \pi(\theta), f(x|\theta)$ をそれぞれ，X, θ の周辺密度と，θ を与えた下での x の条件付き密度とする．このとき，x を与えた下での θ の条件付き密度は，

$$\forall_{x \in \mathfrak{X}}\ f_X(x) > 0 \Rightarrow f(\theta|x) = \frac{\pi(\theta)f(x|\theta)}{\int_\Theta \pi(\theta)f(x|\theta)\nu(d\theta)} \quad \nu\text{-a.e.}$$

と表せる．このことを，$f(\theta|x) \propto \pi(\theta)f(x|\theta)$ と表す．

定義 9.8.5 (conjugate prior). 事後分布 $g(-|x)$ が事前分布 $\pi(-)$ と同じ分布型であるとき，π を θ の**共役事前分布**といい，$f(x|\theta)$ をその尤度関数という．

9.9　指数分散モデル

9.9.1　構成

記法 9.9.1. 一次元確率分布 $\mu(dx)$ が生成する 1 次元自然指数型分布族

$$\exp\left(x\theta - \psi(\theta)\right)\mu(dx), \quad \theta \in \Theta_0$$

を考える．ただし，ψ は μ の (第 2) キュムラント母関数とした．

$$\Theta_0 := \left\{\theta \in \mathbb{R} \mid \int e^{\theta x}\nu(dx) < \infty\right\}$$

は $\Theta_0^\circ \neq \emptyset$ を満たすとする．μ の積率母関数 $\mathfrak{M}$ が存在するとし，集合

$$\Lambda := \left\{\lambda > 0 \mid \mathfrak{M}(\theta)^\lambda \text{はある確率分布} \mu_\lambda \text{の積率母関数}\right\}$$

を考える．

補題 9.9.2 (exponential dispersion model).

(1) 任意の $\lambda \in \Lambda$ について，$\lambda\psi(\theta)$ はある分布 μ_λ のキュムラント母関数であり，$\exp(x\theta - \lambda\psi(\theta))\mu_\lambda(dx)$ $(\theta \in \Theta_0)$ は分布を定める．

(2) (1) の分布に従う確率変数 X について，$Y := \lambda^{-1}X$ の分布は $\widetilde{\mu}_\lambda := (\lambda^{-1})_*\mu_\lambda$ について次を満たす：

$$\forall_{B\in\mathcal{B}^1}\quad P^Y(B) = \int_B \exp(\lambda(y\theta - \psi(\theta)))\widetilde{\mu}_\lambda(dy), \quad \theta \in \Theta_0, \lambda \in \Lambda.$$

(2) が定める分布 $(P^Y)_{\lambda\in\Lambda}$ を μ または ψ が定める**指数分散モデル**といい，$\mathrm{EDM}(\theta,\lambda)$ で表す．

注 9.9.3. 定義から $\mathbb{N} \subset \Lambda$ である．

補題 9.9.4. $Y \sim \mathrm{EDM}(\theta,\lambda)$ とする．

(1) Y のキュムラント母関数は $\kappa_Y(u) = \lambda\left(\psi\left(\theta + \dfrac{u}{\lambda}\right) - \psi(\theta)\right)$ である．特に，$E[Y] = \psi'(\theta)$，$\mathrm{Var}[Y] = \lambda^{-1}\psi''(\theta)$.
(2) 平均値関数を $\tau(\theta) := \psi'(\theta) : \Theta^\circ \to \mathbb{R}$ で，その値域を $D := \tau(\Theta^\circ)$ で表す．このとき，τ は単射である．
(3) $V(\mu) := \tau'(\tau^{-1}(\mu)) : D \to \mathbb{R}_+$ について，$\mathrm{Var}[Y] = \lambda^{-1}V(\mu)$ と表せる．

この $V(\mu)$ を**単位分散関数**という．

注 9.9.5. $\lambda = 1$ でない限り，$V(\mu)$ そのものは分散ではない．これより，$\mathrm{EDM}(\theta,\lambda)$ を $\mathrm{EDM}(\mu,\lambda)$ で表す．

補題 9.9.6. 定数 $w_i \geq 0, w := \sum_{i=1}^n w_i > 0$ と，独立確率変数列 $Y_i \sim \mathrm{EDM}(\mu,\lambda w_i)$ について，

$$\frac{1}{w}\sum_{i=1}^n w_iY_i \sim \mathrm{EDM}(\mu,\lambda w)$$

9.9.2 Tweedie 分布族

定義 9.9.7. $\mathrm{EDM}(\mu,\lambda)$ のうち，単位分散関数が $V(\mu) = \mu^p$ $(p \in \mathbb{R}\setminus(0,1))$ と表せる分布族を，**Tweedie 分布族**といい，記号 $\mathrm{Tw}_p(\mu,\lambda)$ で表す．

例 9.9.8. $p = 0$ ならば正規分布族，$p = 1$ ならば Poisson 分布，$p = 2$ ならば Gamma 分布，$p = 3$ ならば逆正規分布となる．

命題 9.9.9. $\mathrm{EDM}(\mu,\lambda)$ について，$1 \in D, V(1) = 1$ とし，さらにある関数 $s : (0,\infty) \times \Lambda^{-1} \to \Lambda^{-1}$ で以下を満たすものが存在すると仮定する：$Y \sim \mathrm{EDM}(\mu,\lambda) \Rightarrow [\forall_{c>0}\, cY \sim \mathrm{EDM}(c\mu, 1/s(c,\lambda))]$．このとき，以下が成り立つ．

(1) $\exists_{p\in\mathbb{R}}\ Y \sim \mathrm{Tw}_p(\mu,\lambda)$.
(2) $s(c,\lambda) = c^{2-p}/\lambda$．すなわち，$c\mathrm{Tw}_p(\mu,\lambda) = \mathrm{Tw}_p(c\mu,\lambda/c^{2-p})$.
(3) $p = 0$ ならば $D = \mathbb{R}$．また，$p \neq 0$ ならば $D = (0,\infty)$．特に，$\mathcal{L}(Y)$ は無限分解可能である．

9.10 コピュラ

相関係数は，２つの確率変数に対して，実数 $[-1,1]$ を対応させるが，もっと表現力豊かに，変量間の従属性を，多次元分布によって直接的に表現することを考える．[a]
これはなんだか，周辺分布の張り合わせとして多次元分布を表現するホモトピー論のようである．逆に，コピュラなく，直積によって周辺分布を張り合わせた場合が，独立性である．

[a] この単語は元々音楽や言語学で使われていたが、統計学の用語として用いたのは、1959 年にスクラー (Abe Sklar) がパリ大学統計学会誌 (the Statistical Institute of the University of Paris) で発表したのが最初である。

記法 9.10.1. $I := [0,1]$ とする．

9.10.1 定義と例

定義 9.10.2 (copula). 次の条件を満たす関数 $C : I^2 \to I$ を**接合関数**という：

(1) $C(-,0)=C(0,-)=0$.
(2) $C(-,1)=C(1,-)=\mathrm{id}_I$.
(3) $\forall_{u_1,u_2,v_1,v_2\in I}\ u_1\le v_1, u_2\le v_2 \Rightarrow C(v_1,v_2)-C(u_1,v_2)-C(v_1,u_2)+C(u_1,u_2)\ge 0$.

例 9.10.3 (凸関数と product copula). C^2 級関数 C が $\partial_{u_1}\partial_{u_2}C(u_1,u_2)\ge 0$ をみたすとき，コピュラである．特に，$C^{\Pi}(u_1,u_2):=u_1u_2$ を**積コピュラ**という．

例 9.10.4 (二次元の分布関数が定めるコピュラ).
(1) 区間 $(0,1)$ 上の一様分布に従う 2 つの確率変数 $X_1,X_2\sim U((0,1))$ の分布関数 F はコピュラである．
(2)

9.10.2 Sklar **の定理**

多次元分布は，周辺分布関数とコピュラによって定まる．

定理 9.10.5 (2 次元の場合 (Sklar 1959)). (X_1,X_2) の結合分布関数を F，それぞれの周辺分布関数を F_1,F_2 とする．このとき，コピュラ C が存在して，

$$\forall_{x_1,x_2\in\mathbb{R}}\quad F(x_1,x_2)=C(F_1(x_1),F_2(x_2))$$

と表せる．F_1,F_2 が連続のとき，C は一意である．

要諦 9.10.6. 逆関数 F_1^{-1},F_2^{-1} が存在するときは，これを用いて $C(u_1,u_2)=F(F_1^{-1}(u_1),F_2^{-1}(u_2))$ とコピュラを構成できる．

9.10.3 Archimedes **コピュラ**

参考文献

[1] Vladimir Voevodsky "Notes on categorical probability"

[2] "Convergence of Probability Measures"

[3] "Weak Convergence of Measures"

[4] 伊藤清『確率論』

[5] Kolmogorov 『確率論の基礎概念』

[6] 吉田朋広『数理統計学』(朝倉書店，2006)

[7] 竹村彰道『現代数理統計学』（学術図書，2020）

[8] 久保川達也『現代数理統計学の基礎』（共立出版，2017）

[9] 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）

[10] 丸山徹 - 確率測度の*弱収束 https://core.ac.uk/download/pdf/145720102.pdf