

# 目次

第 1 章	モデルの枠組み	6
1.1	確率の数理構造の解明の歴史	6
1.1.1	確率とその解釈	6
1.1.2	事象・試行の集合論的構造	7
1.1.3	分割と代数の構造	8
1.1.4	確率分布	9
1.1.5	確率測度の経時発展	9
1.1.6	力学系の公理	10
1.1.7	量子確率空間の定式化	11
1.1.8	エントロピー	11
1.1.9	確率変数	11
1.2	確率空間とその構成	12
1.2.1	確率空間の圏	12
1.2.2	確率空間の射	12
1.2.3	確率空間の同型	13
1.2.4	直積測度の構成	14
1.2.5	積空間の性質	15
1.2.6	非可算積の扱い	16
1.3	確率測度論	17
1.3.1	事象の極限	17
1.3.2	Dynkin 族定理	17
1.3.3	単調族定理	18
1.3.4	確率測度の正則性	19
1.3.5	Bonferroni の不等式	19
1.4	独立性	19
1.4.1	分割の独立性	19
1.4.2	代数の独立性	20
1.4.3	確率変数の独立性	20
1.4.4	独立性の遺伝	21
1.4.5	Borel-Cantelli の補題	21
1.5	分布の扱い	23
1.5.1	分布関数	23
1.5.2	Lebesgue 分解	23
1.5.3	畳み込み	24
1.5.4	確率測度の Fourier 逆変換	24
1.5.5	特性関数の Taylor 展開	25
1.5.6	キュムラント	25
1.6	確率変数の空間	26

1.6.1	$L$ の構造	26
1.6.2	$L^p$ の構造	26
1.7	$L$ 上の汎関数	27
1.7.1	期待値と独立性	27
1.7.2	積率	27
1.7.3	$\mathbb{R}$ 上の分散	27
1.7.4	$\mathbb{R}$ 上の共分散	28
1.7.5	可分 Hilbert 空間上の平均	28
1.7.6	可分 Hilbert 空間上の分散	28
1.7.7	Chebyshev の不等式	29
1.7.8	Jensen の不等式	30
1.8	条件付け	30
1.8.1	条件付き期待値の定義	31
1.8.2	射影としての条件付き期待値	32
1.8.3	条件付き期待値の延長	33
1.8.4	条件付き期待値の代数的性質	33
1.8.5	確率変数に関する条件付け	34
1.8.6	可算生成の場合の条件付き確率	34
1.8.7	正則条件付き確率の定義	35
1.8.8	確率変数による正則条件付き確率	35
1.8.9	確率測度の分解	36
1.8.10	条件付き独立性	36
1.8.11	偏相関係数	37
1.9	独立確率変数列	37
1.9.1	Kolmogorov の 0-1 法則	37
1.9.2	Kolmogorov の不等式	38
1.9.3	Ottaviani の不等式	39
1.9.4	Hewitt-Savage の 0-1 法則	39
第 2 章	確率測度の空間	40
2.1	測度の表現定理	40
2.1.1	Radon 積分とこれを定める Radon 測度	40
2.1.2	$C$ の双対空間と符号付き測度	41
2.1.3	$C_0$ の双対空間	41
2.1.4	$C_b$ の双対空間	41
2.1.5	$L^\infty$ の双対空間	42
2.1.6	測度の収束の定義 5 種	42
2.2	$P(S)$ の研究	43
2.2.1	弱位相の定義	43
2.2.2	Alexandorff の特徴付け	44
2.2.3	弱収束の十分条件	46
2.2.4	双対空間の稠密部分集合	46
2.2.5	漠位相による特徴付け	46
2.2.6	弱位相の距離化	47
2.2.7	概収束列の構成	47
2.2.8	Fourier 変換の連続性	48
2.2.9	コンパクトになるとき	48
2.3	一様緊密性	48

2.3.1	定義	48
2.3.2	相対点列コンパクト性による特徴付け	49
2.3.3	一様緊密性の十分条件	50
2.4	測度の畳み込み	50
2.4.1	畳み込みと不等式	50
2.4.2	確率変数との関係	50
2.4.3	特徴量の関手性	51
2.4.4	独立確率変数の和と畳み込み	51
2.4.5	分布族の再生性	51
2.4.6	安定型	52
2.5	測度の正則性	52
2.5.1	内部正則性	52
2.5.2	緊密性	52
2.6	Donsker の定理	52
2.7	有限測度のノルム収束	53
2.8	分布関数の収束	53
2.8.1	弱収束の定義	53
2.8.2	相対コンパクト	53
第 3 章	確率変数の空間	55
3.1	原子	55
3.1.1	原子の定義	55
3.1.2	有限測度空間の分割	55
3.2	弱位相	56
3.2.1	弱収束列はノルム有界	56
3.3	一様可積分性	56
3.3.1	定義	56
3.3.2	積分の族の同程度絶対連続性としての特徴付け	57
3.3.3	弱相対点列コンパクト性との同値性	58
3.3.4	一様可積分性の十分条件	58
3.3.5	一様可積分列の概収束	59
3.3.6	一様可積分列の確率収束	60
3.3.7	分布収束列は一様可積分	60
3.4	$L$ の位相まとめ	61
3.4.1	確率収束の十分条件	61
3.4.2	標準確率空間上の収束の様子	62
3.4.3	確率収束の必要条件	62
3.4.4	確率収束列の概収束部分列の存在	63
3.4.5	確率収束の特徴付け	63
3.4.6	確率収束を定める距離	64
3.4.7	分布収束列の修正概収束部分列の構成	64
3.4.8	$L^1$ -弱相対コンパクトと一様可積分性は同値	64
3.4.9	概収束を定める位相	65
3.5	確率収束	65
3.5.1	Lebesgue の優収束定理の一般化	65
3.5.2	連続写像定理	65
3.5.3	積の収束	65
3.5.4	同じ極限に収束するための条件	66

3.5.5	漸近同等性と確率有界性	66
3.6	分布収束	67
3.6.1	Lebesgue の優収束定理の一般化	67
3.6.2	弱収束の十分条件	67
3.6.3	連続写像定理	67
3.6.4	積の収束	68
3.6.5	弱収束の遺伝と独立性	68
3.6.6	積空間上の弱収束	68
3.6.7	デルタ法	69
3.7	安定収束	69
3.7.1	定義と例	69
3.7.2	安定収束極限定理	70
第 4 章	独立性	71
4.1	条件付き期待値	71
4.1.1	動機	71
4.1.2	定義	72
4.1.3	性質	72
4.1.4	可測写像を与えたもとでの条件付き期待値	73
4.1.5	正則条件付き確率	74
4.2	確率変数の独立性	75
第 5 章	独立確率変数列	76
5.1	独立確率変数列の和の振る舞い	76
5.2	概収束するための条件	76
5.3	独立確率変数の和の収束同等性	78
5.4	概発散の描像	79
5.5	大数の法則	80
5.5.1	本質的要件	80
5.5.2	種々の十分条件	81
5.5.3	弱法則の十分条件	82
5.5.4	例	83
5.6	中心極限定理	83
5.6.1	指数関数の Taylor 展開に関する補題	83
5.6.2	Lindeberg の条件	84
5.6.3	種々の十分条件	86
5.7	重複対数の法則	86
5.7.1	ルートより少し速い収束	86
5.7.2	大数の法則の精緻化	87
5.7.3	重複対数の法則	87
5.8	大数の法則的現象	87
5.8.1	積分	87
5.8.2	Weierstrass の多項式近似	87
5.8.3	Borel の正規数定理	89
5.8.4	Maxwell 分布	89
5.8.5	熱力学的極限	89
5.9	中心極限定理的現象	90
5.9.1	極限計算	90

5.9.2	Gauss の誤差論	90
5.9.3	Poisson の少数の法則	90
5.9.4	Weyl の一様分布定理	91
5.10	経験分布論	91
5.10.1	大数の法則の精緻化	91
5.10.2	中心極限定理の精緻化	91
5.11	大偏差原理	91
5.11.1	定義	92
5.11.2	Laplace の原理	92
5.11.3	Cramer の理論	93
5.11.4	Schilder の理論	93
5.11.5	Varadhan の理論	94
5.12	集中不等式	94
5.12.1	Berry-Esseen の不等式	94
5.12.2	Chebyshev の不等式	94
5.12.3	Hoeffding の不等式	95
5.12.4	Chernoff の不等式	95
5.12.5	確率グラフ	96
5.12.6	劣 Gauss 分布	96
5.12.7	Orlicz 空間	97
5.12.8	一般化 Hoeffding 不等式	97
5.12.9	劣指数分布	97
5.12.10	Bernstein の不等式	97
第 6 章	離散確率過程に対する極限定理	98
6.1	非独立過程の大数の法則	98
6.1.1	$L^2$ -有界な直交系の列	98
6.1.2	乗法系	98
6.1.3	エルゴード定理	99
6.2	McLeish の中心極限定理	99
6.3	第 1 逆正弦法則	100
6.3.1	投票でのリード	100
6.3.2	第 1 逆正弦法則	101
6.4	Donsker の定理	101
第 7 章	参考文献	102
参考文献		103
参考文献		104

## 第 1 章

# モデルの枠組み

確率論とは有界測度論に他ならず、これを解析するには面積・体積・時間の概念同様、極限の概念によって数学モデルに落とし込んでのち、極限定理と漸近展開によって解析するのが数理である。

決定的な予測をすることが不可能な程度に、不完全な情報しか持たない現象をモデルする技法を考える。20 世紀に identification の問題と確率論の問題とが分離され、前者は頻度論的確率解釈、後者は確率の公理が用いられる。後者の範疇では量子論のように原理的に不可知であるか、意思決定のように複雑性やノイズの大きさによるものかは、形式的には関係ない。

**試行** とは標本空間の分割で、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。

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

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

**エントロピー** とは、系で考える全ての試行の期待驚愕度の上限である。

### 1.1 確率の数理構造の解明の歴史

The beginning definitions in any field of mathematics are always misleading, and the basic definitions of probability are perhaps the most misleading of all. [Rota, 2001]

#### 1.1.1 確率とその解釈

種々の物理模型と同様、identification の問題が控えている。これと確率論を分離したことは、歴史的には随分最近の仕事であった。現状、確率論は数学分野として発展することに成功し、現実問題との identification の立場を取ることでそれは統計になる [Kotz et al., 2005] Frequency Interpretation.

**注 1.1.1** (確率とその identification の問題の分離). 標本とは "Merkmal" であり、label のことである。そして [von Mises, 1919] が標本空間として想定しているのは無限列全体の集合、すなわち  $n^{\mathbb{N}}$  である。なお、von Mises のいう確率の定義にはもう 1 つあるが、こちらは「ランダム性」の定義に当たり、統計模型の適用の際に生じる identification の問題に関わる。この点は Kolmogorov 自身も次のように指摘している。

... that the basis for the applicability of the results of the mathematical theory of probability to real 'random phenomena' must depend on some form of the frequency concept of probability, the unavoidable nature of which has been established by von Mises in a spirited manner. [Kolmogorov, 1963]

最も興味深いのは、ランダム性の定義を得るために Kolmogorov はアルゴリズムの計算複雑性の研究を行い、"there cannot be a very large number of simple algorithms"と結論付けたことである。von Mises と Kolmogorov の定式化はライバルのように扱われることが多いが、本当に対極的なのは、統計の下で確率を捉えたか、確率の先に統計を捉えたかの違いでしかなく、結局 identification の問題は未解決のままである。

**注 1.1.2** (Interpretation of probability). 識別可能性の問題は、「統計モデルの識別可能性」として、計量経済学など多くの分野で表面化して、[Kotz et al., 2005] でもその項しかない。数学基礎論の問題としては議論は広くないようで、Interpretation of probability と呼ぶようだ。現状、 $P[A]$  は、 $A, \Omega \setminus A$  からなる Bernoulli 試行の頻度極限であり、ある種の算譜に関して極限が普遍のものとして、現実世界との identification を理解できそうだ。これは実験を前提としたものの見方であり、意思決定を前提とする場合は Bayes' interpretation が優先しそうである。

**定義 1.1.3** (structure, observationally equivalent, identified [Kotz et al., 2005] Identification Problems).  $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$  を統計的実験、 $X \in L(\Omega)$  を可観測量とする。

- (1) 2つの構造  $(\Omega, \mathcal{F}, P_{\theta_1}), (\Omega, \mathcal{F}, P_{\theta_2})$  が観測上等価であるとは、 $X^{P_{\theta_1}} = X^{P_{\theta_2}}$  が成り立つことをいう。
- (2) 任意の  $\theta \in \Theta$  が、自分自身以外に観測上等価なものを持たないとき、統計的実験  $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$  は識別可能であるという。
- (3) 可観測量  $X$  が識別不可能であるとは、真値  $\theta_0 \in \Theta$  が他に観測上等価な元を  $\Theta$  内に持たないことをいう。

### 1.1.2 事象・試行の集合論的構造

#### 「集合の公理」のもう一つの見方

事象を空間の中で捉えた「標本空間 (Merkmalraum)」の概念は [von Mises, 1919] による。

We shall not speak of probability until a collective has been defined. [von Mises and Geiringer, 1957]

するとその後ここに測度・位相の構造を載せるという自然な発展を辿った。この視点の転換は、集合の正則性公理や選択公理などを、事象のためにも流用することと捉えられる。非可測集合の存在は選択公理と同値であるというのは、人類の認知論の一つの歴史的メルクマールに到達しているということでもある。

**記法 1.1.4** (mutually exclusive). 集合の和・差の演算を、さらに細かく区別する。

- (1) 事象  $A, B$  が排反であるとは、集合として互いに素であることをいう。この時の和事象を  $A + B$  で表し直和という。  
 $P(A + B) = P(A) + P(B), P(A \cup B) = P(A) + P(B) - P(A \cap B)$  と書き分ける。
- (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。

**要諦 1.1.5.** Zermelo は Poincare の再帰定理にもとづいて、recurrence paradox としてエントロピー増大の原理の Boltzman による定式化である  $H$  定理を批判した。ここに確率論の基礎論的・認識論的な難しさが見て取れる。再帰的関数は基礎論の話題になるのだ。

**定義 1.1.6** (sample space, sample, event). 可測空間  $(\Omega, \mathcal{F})$  を標本空間または状態空間という。

- (1)  $\Omega$  の元  $\omega \in \Omega$  を標本 (点) または状態という。
- (2)  $\mathcal{F}$  の元を事象と呼ぶ。<sup>†1</sup> 事象としての  $\Omega \in \mathcal{F}$  を全事象という。

**例 1.1.7** (statistical ensemble, path space). Boltzman の統計力学、Poincare と天体力学、Feynman と量子力学などでも考えられている。

- (1) 統計力学における巨視系など、施行を繰り返した空間  $E^{\mathbb{N}}$  が理想的な標本空間の形態である。統計力学においてはこのようにして得られる確率空間を特に統計的集団という。この上の確率密度関数 (に対応する対象) を分配関数という。統計力学

<sup>†1</sup> 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。



や量子場の理論で用いられる概念であるが、Yang-Mills 理論などの場の理論としては初等的なものでも、数学的な基礎づけは出来ていない。

- (2) 古典力学は経路空間上の作用汎関数の微分の理論であり、量子力学は経路空間上の作用汎関数の測度に関する積分の理論である。よって大域的な情報が必要になり、微分幾何学や層の理論に非常に接近してくる。そこで、見本道の空間上の積分が問題になる。そして時間発展は、見本道の空間の保測変換、すなわち Hilbert 空間上の Hermite 作用素となる。
- (3) 経路空間を一般化して、状態空間  $E$  を底空間とし、このファイバー束  $\iota: F \rightarrow E, \pi: E \rightarrow M$  を考える。 $F$  を内部空間とし、場の値域にあたり、 $M$  は場の定義域にあたる。こうして確率空間は切断の全体  $\Gamma(M, E)$  とし、これを場の空間ともいう。
- (4) 統計力学における相空間  $M \subset \mathbb{R}^3$  上の  $r$  粒子系において、粒子の位置が  $\mathbb{R}^r$  の一様分布に従うとき、**Maxwell-Boltzman の統計が成立する**というが、これが成り立つ実際の物理系はない。一方で、全ての区別できない  $r$  粒子の付値全体の集合上の一様分布に従うとき、**Bose-Einstein の統計に従う**という。光子など、相互作用のない Bose 粒子系で成立する。さらに、2 以上の粒子が同じ区画に入らないものの中で、区別できない付値の全体の集合上の一様分布に従うとき、**Fermi-Dirac の統計が成り立つ**といい、電子・中性子・陽子について成立する。
- (5) 連続体無限個の確率変数の族を**確率場**という。

□

### 1.1.3 分割と代数の構造

全ての閉  $\sigma$ -代数はある観測の結果とみなせる。ここでも有向集合の構造が見つかる。位相も、観測の一種と考えられるのかもしれない。

**定理 1.1.8 (すべての閉  $\sigma$ -代数はある観測の結果とみなせる).** 任意の閉  $\sigma$ -加法族  $\mathcal{F}$  に対して、ある実確率変数  $X \in \mathcal{L}(\Omega)$  が存在して、 $\mathcal{F} = \mathcal{F}[X]$  が成り立つ。

**要諦 1.1.9.**  $\sigma$ -代数とは、観測の結果のモデルである。観測が細くなればなるほど、知識は増えていく。逆に、 $\sigma$ -代数以外の言葉で、状態空間  $(X, \mathcal{A}, \mu)$  の様子を観察する手立ては人類にはない。

**定理 1.1.10 (確率変数が定める分割).**  $X, Y \in \mathcal{L}(\Omega)$  について、 $X < Y$  a.s.  $:\Leftrightarrow [\exists \varphi \in \text{Map}(\Omega, \Omega) \ X = \varphi \circ Y \text{ a.s.}]$  と表すと、これは同値類  $\sim$  とその上の順序を定め、

- (1)  $Y < X$  a.s.  $\Leftrightarrow \mathcal{F}[Y] \subset \mathcal{F}[X]$ .
- (2)  $Y \sim X$  a.s.  $\Leftrightarrow \mathcal{F}[Y] = \mathcal{F}[X]$ .

が成り立ち、この順序は可算な上限と下限を備える束の構造を持つ。これは、可算個の確率変数の結合変数が定める分割として自然に現れる構造である。全く同様に、 $\sigma$ -代数の包含関係は完備な束を定める。<sup>†2</sup>

**定義 1.1.11 .**

- (1)  $2 := \{A \in \mathcal{F} \mid P[A] \in \{0, 1\}\}$  は  $\sigma$ -加法族である。
- (2)  $\sigma$ -加法族  $\mathcal{G}$  が分割  $\Delta$  に従属するとは、 $\mathcal{G}$  が定める分割 ( $\mathcal{G}$  を含む最小の分割)  $\mathcal{G}_\Delta$  に対して、 $\Delta \subset \mathcal{G} \subset \mathcal{G}_\Delta$  を満たすことをいう。
- (3) 2つの  $\sigma$ -加法族  $\mathcal{G}_1, \mathcal{G}_2$  が**同等**であるとは、 $\mathcal{G}_1 \vee 2 = \mathcal{G}_2 \vee 2$  を満たすことをいう。これは、任意の元に対して、もう一方の  $\sigma$ -代数の元であって、対称差が零であるようなものが存在することをいう。

<sup>†2</sup> 伊藤 [?] はより細かい分割を定める方を**母**と呼んでいる。



### 1.1.4 確率分布

標本空間に有限測度を導入して確率のモデルとすることは、Borel (1909) の大数の強法則の証明、Wiener (1920-24) の Brown 運動の研究で成功を収め、Kolmogorov (1933) が公理にまとめた。確率空間とは知識の空間であり、数理物理の相空間も標本空間である。このように空間に関する知識のとき、標本空間はファイバー束と見た方が良い。一方で確率過程のように時間発展に関する知識のとき、標本空間はフィルター付きと見たほうが良い。

**公理 1.1.12 (probability).**  $(\Omega, \mathcal{F})$  を標本空間とする。この上の有限な測度  $P$  であって、 $P(\Omega) = 1$  と正規化されたものを**確率**という。測度空間  $(\Omega, \mathcal{F}, P)$  を**確率空間**という。

**定義 1.1.13 (random variable / observable, probability distribution / law).**  $(\Omega, \mathcal{F}), (\mathcal{X}, \mathcal{A})$  を標本空間とする。

- (1) 可測空間の射  $X \in \mathcal{L}(\Omega; \mathcal{X})$  を  $\mathcal{X}$ -**値確率変数**または**可観測量**という。
- (2) これが  $(\mathcal{X}, \mathcal{A})$  に押し出す像測度  $X_*P =: P^X$  を**確率分布**または**法則**という。
- (3) 像  $\text{Im } X$  も同様の記法  $\Omega^X$  で表し、これを  $X$  の**標本空間**という。

**定義 1.1.14 (discrete / continuous distribution).** 分布が

- (1) **離散型**とは、台が可算な分布をいう。すなわち、原子の和として得られる分布をいう。
- (2) **連続型**とは、分布関数 1.5.1 が連続であることをいう。すなわち、原子を持たない分布をいう。
- (3) さらに Lebesgue 測度に対して絶対連続な分布と、そうでない特異連続な分布がある。台が非可算でも Lebesgue 測度が 0 であることがあり得るためである。

**例 1.1.15 (時空間のモデルの例).**

- (1) 統計力学における相空間  $M \subset \mathbb{R}^3$  上の  $r$  粒子系において、粒子の位置が  $\mathbb{R}^r$  の一様分布に従うとき、**Maxwell-Boltzman の統計が成立する**というが、これが成り立つ実際の物理系はない。一方で、全ての区別できない  $r$  粒子の付値全体の集合上の一様分布に従うとき、**Bose-Einstein の統計に従う**という。光子など、相互作用のない Bose 粒子系で成立する。さらに、2 以上の粒子が同じ区画に入らないものの中で、区別できない付値の全体の集合上の一様分布に従うとき、**Fermi-Dirac の統計が成り立つ**といい、電子・中性子・陽子について成立する。

□

### 1.1.5 確率測度の経時発展

#### 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 1.1.16 (conditional probability).**

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

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

を、事象  $A$  の下での  $B$  の条件付き確率という。  $P_A(B)$  とも表す。<sup>†3</sup>  $P(A) = 0$  の時はこの値は任意に定めることで、2 変数

<sup>†3</sup>  $i: A \hookrightarrow \Omega$  についての引き戻し測度である？

測度  $P(\cdot | \cdot) : \mathcal{F} \times \mathcal{F} \rightarrow [0, 1]$  が定まる.

(2)  $P(\cdot | A)$  は再び  $\Omega$  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.

**命題 1.1.17 (law of total probability : 全確率の分解).**  $(H_i)_{i \in I}$  を互いに背反で  $\Omega = \sum_{i \in I} H_i$  を満たす事象の族とする.

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

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

$$\begin{aligned} P(A) &= P(\cup_{i \in I} A \cap H_i) \\ &= \sum_{i \in I} P(A \cap H_i) && \text{加法性} \\ &= \sum_{i \in I} P(A | H_i) P(H_i) && \text{定義 1.1.16} \end{aligned}$$

**定理 1.1.18 (Bayes).**  $(H_i)_{i \in I}$  を互いに背反で  $\Omega = \prod_{i \in I} H_i$  を満たす事象の族とする.  $P(A) > 0$  ならば,

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

**[証明].**

$$\begin{aligned} P(H_i | A) &= \frac{P(H_i \cap A)}{P(A)} && \text{条件付き確率の定義 1.1.16} \\ &= \frac{P(A | H_i) P(H_i)}{\sum_{j \in I} P(A | H_j) P(H_j)} && \text{分母は定義 1.1.16, 分子は全確率の分解則 1.1.17} \end{aligned}$$

**要諦 1.1.19.** 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.<sup>†4</sup>  $I = 1$  の場合は,

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

となる.

### 1.1.6 力学系の公理

力学系では自己射を考え, その自然なフィルトレーションとを考える.

**定義 1.1.20.** 確率空間  $(X, \mathcal{A}, \mu)$  について,

- (1) 測度を保存する自己写像  $\varphi : X \rightarrow X, \varphi_* \mu = \mu$  との組  $(X, \mathcal{A}, \mu, \varphi)$  を**力学系**という.
- (2) 可観測量の空間上の作用素  $T_\varphi : \mathcal{L}(X) \rightarrow \mathcal{L}(X); f \mapsto f \circ \varphi$  を **Koopman 作用素**という.

**要諦 1.1.21 (flow).** 自己射  $(\varphi_t)$  は通常 Hamilton の原理などの物理法則から定まり, 1-パラメータ変換群をなす. これを**流れ**ともいう.

**定理 1.1.22 (Poincare recurrence theorem).** 力学系  $(X, \mathcal{A}, \mu, \varphi)$  について, 次が成り立つ:

$$\forall E \in \mathcal{A} \quad \mu(\{x \in E \mid \varphi^n(x) \notin E \text{ f.e.}\}) = 0.$$

<sup>†4</sup> In quantum mechanics, the collapse of the wavefunction may be seen as a generalization of Bayes's Rule to quantum probability theory. This is key to the Bayesian interpretation of quantum mechanics.

### 1.1.7 量子確率空間の定式化

状態を可観測量  $f: X \rightarrow \mathbb{R}$  を通じて観測するというモデルである力学系の公理の考え方を取り入れて、確率空間の公理との交差点が量子確率空間とみなせる。

**定義 1.1.23 (quantum probability space, state on a  $C^*$ -algebra).** 位相を加えた確率空間  $(\Omega, \mathcal{F}, P)$  の、 $X$  の Gelfand 双対となる  $C^*$ -代数  $\mathcal{A}$  と、 $P$  のその上への双対となる状態  $\langle - \rangle: \mathcal{A} \rightarrow \mathbb{C}$  との組を量子確率空間という。一般に、単位的  $C^*$ -代数  $\mathcal{A}$  の状態とは、正な線型汎関数  $\rho: \mathcal{A} \rightarrow \mathbb{C}$  であって  $\rho(1) = 1$  を満たすものをいう。

### 1.1.8 エントロピー

#### エントロピーとは確率不変量である

知識の限界のモデルである確率空間には保存量が定義でき、系の時間発展を見るエルゴード理論や通信の限界を考える情報理論で用いられる。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。<sup>a</sup>

<sup>a</sup> Entropy is a measure of disorder, given by the amount of information necessary to precisely specify the state of a system. <https://ncatlab.org/nlab/show/entropy>

**定義 1.1.24 (surprisal / self-information, information function, Shannon entropy).**  $(X, \mathcal{A}, \mu)$  を確率空間とし、 $f: X \rightarrow \mathbb{R}$  を可観測量、 $\mathcal{P} \subset \mathcal{A}$  を  $f$  が定める  $X$  の分割とする。

- (1)  $\mu!(A) := -\log \mu(A)$  で定まる測度  $\mu!: \mathcal{A} \rightarrow \overline{\mathbb{R}_+}$  を驚愕度という。
- (2) 分割  $\mathcal{P}$  に対して、各元  $x \in X$  にその属する  $A \in \mathcal{P}$  の驚愕度に対応させる関数  $I_{\mathcal{P}}: X \rightarrow \mathbb{R}$  を  $\mathcal{P}$  の情報量関数という：

$$I_{\mathcal{P}} := - \int_{A \in \mathcal{P}} 1_A \log \mu(A).$$

- (3) この積分  $H(\mathcal{P}) := \int_X I_{\mathcal{P}} d\mu$  を分割  $\mathcal{P}$  のエントロピーという。

**定義 1.1.25 (Kolmogorov-Sinai entropy).** 力学系  $(X, \mathcal{A}, \mu, T)$  について、 $\mathbb{P}$  を  $X$  の有限な分割  $\mathcal{P} \subset \mathcal{A}$  のなす有向集合とする。  $\mathcal{P} \in \mathbb{P}$  の像

$$h(T) := \sup_{\mathcal{P} \in \mathbb{P}} \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathcal{P} \vee T_* \mathcal{P} \vee \dots \vee T_*^{n-1} \mathcal{P})$$

を力学系  $(X, \mathcal{A}, \mu, T)$  のエントロピーという。これは力学系の同型類において等しい値を取る。

**歴史 1.1.26.** Khinchin が Shannon のエントロピーが保測写像  $T$  や一般の  $\sigma$ -代数にも、ネットの極限のことばで定義出来ることを発見した。力学系の同型類の研究において、Kolmogorov はエントロピーが不変量になることを発見した。Sinai と Ornstein はこれを応用して Bernoulli 力学系に対する同型問題を解決した (Ornstein の同型定理 1970)。

### 1.1.9 確率変数

The basic object of study in probability is the random variable and I will argue that it should be treated as a basic construct, like spaces, groups and functions, and it is artificial and unnatural to define it in terms of measure theory. [Mumford, 2000]

## 1.2 確率空間とその構成

現状確率空間は測度論によって理解されるが、制限をつけなければ極めて込み入った状況になる。これが不自然だと疑問を投げかける立場は [Mumford, 2000] にも見られる。そこでよりふるまいの良いサブクラスに限定することになるが、そこでの圏論的ふるまいを特徴づける試み (synthetic approach) が現在進んでいる。最先端は monad と valuation の理論だと見受けられる。<sup>a</sup>

<sup>a</sup> 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>

### 1.2.1 確率空間の圏

確率論は確率変数を調べる学問だとすれば、圏論的には  $\mathbf{Prob}$  上の層のなす圏の研究である。

**定義 1.2.1** (standard Borel space [Simpson, 2017]). 確率空間  $(\Omega, \mathcal{F}, P)$  が次を満たすとき、**Polish 確率空間**または**標準 Borel 空間**であると呼ぼう。

- (1)  $\Omega$  は Polish 空間である。
- (2)  $\mathcal{F} = \mathcal{B}(\Omega)$  はその位相に関する Borel  $\sigma$ -代数である。

Polish 確率空間と Borel 可測な保測写像のなす圏を  $\mathbf{Prob}$  で表す。

**命題 1.2.2** ([Simpson, 2017] Prop.1).  $L(-, A) : \Omega \mapsto L(\Omega; A)$  は関手  $L : \mathbf{Prob}^{\text{op}} \rightarrow \mathbf{Set}$  を定める。特に、 $L(-, A)$  は前層である。

### 1.2.2 確率空間の射

**定理 1.2.3** (可測性の特徴付け [Landkov, 1972]).  $X : \Omega \rightarrow \mathbb{R}^m, f : \Omega \rightarrow \mathbb{R}$  とし、 $\sigma[X]$  を  $\xi$  が  $\Omega$  上に引き戻す  $\sigma$ -代数とする。このとき、次の2条件は同値：

- (1)  $f$  は  $\sigma[X]$ -可測。
- (2) ある可測関数  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$  が存在して、 $f = \varphi(X)$  である。

**[証明].**

(2) $\Rightarrow$ (1) 任意の  $a \in \mathbb{R}$  について、

$$\{f \geq a\} = \{\varphi(X) \geq a\} = \{X \in \varphi^{-1}([a, \infty))\} \in \sigma[X].$$

(1) $\Rightarrow$ (2)  $f \geq 0$  の場合について示せば十分。

■

**命題 1.2.4** (積写像の可測性).  $X \in L(\Omega; \mathbb{R}^d), \varphi$  を Borel 可測とする。

- (1)  $\varphi(X) \in L(\Omega)$ .
- (2)  $X$  が可測であることは、各  $X_i$  ( $i \in [d]$ ) が可測であることに同値。
- (3)  $\sigma[X] = \sigma[X_1, \dots, X_d]$  である。

**[証明].**

(1) 1

(2)  $\Rightarrow$  任意の  $A \in \mathcal{B}(\mathbb{R})$  に対して、

$$X_i^{-1}(A) = X^{-1}(\mathbb{R}^{i-1} \times A \times \mathbb{R}^{d-i}) \in \sigma[X].$$

が成り立つためである。

⇐ まず、矩形集合  $A_1 \times \cdots \times A_d$  ( $A_i \in \mathcal{G}(\mathbb{R})$ ) の全体は  $\mathcal{G}(\mathbb{R}^d)$  を生成する。これに注意すれば、

$$X^{-1}(A_1 \times \cdots \times A_d) = \bigcap_{i=1}^d X_i^{-1}(A_i)$$

から従う。

(3) の証明 任意の  $A \in \mathcal{G}(\mathbb{R})$  に対して、

$$X_i^{-1}(A) = X^{-1}(\mathbb{R}^{i-1} \times A \times \mathbb{R}^{d-i}) \in \sigma[X].$$

が成り立つから、 $\sigma[X_i] \subset \sigma[X]$  が任意の  $i \in [n]$  について成り立つ。

⊂ の証明 任意の  $A_i \in \mathcal{G}(\mathbb{R})$  に対して、

$$X^{-1}(A_1 \times \cdots \times A_d) = \bigcap_{i=1}^d X_i^{-1}(A_i) \in \sigma[X_1, \dots, X_d].$$

であるため。

■

### 1.2.3 確率空間の同型

#### 議論領域を $\mathbb{R}$ 上の絶対連続分布と離散分布に限る

Meas は極めてひどい性質を持つので議論領域を「区間  $I \subset \mathbb{R}$  上の Lebesgue 測度空間」と「原子からなる空間」とその合併とに制限し、これを標準確率空間または **Lebesgue 空間** という。完備な Borel 空間は標準確率空間をなす。

- (1)  $(\mathbb{R}, \mathcal{G}(\mathbb{R}))$  上の完備確率測度と同型な空間を標準確率空間という。
- (2) 原子を持たない標準確率空間は  $([0, 1], \overline{\mathcal{G}([0, 1])}, m)$  と同型である。
- (3) 可分完全確率空間は、 $\mathbb{R}$  の部分空間上の Radon 確率測度と狭義同型である。

可分完全空間は一見さらに制限が強いが、必要条件に「台集合が連続体濃度以下である」ことが含まれるのみであり、台集合が連続体濃度以下の場合には標準確率空間であることに同値。

**定義 1.2.5.** 確率空間  $(\Omega_1, \mathcal{F}_1, P_1), (\Omega_2, \mathcal{F}_2, P_2)$  について、

- (1) **強準同型**とは、全射な可測写像  $\varphi: \Omega_1 \rightarrow \Omega_2$  であって、測度の押し出しでもあるものをいう： $\varphi_* P_1 \sim P_2$ 。
- (2) **準同型**または**商写像**とは、それぞれの充満集合  $\Omega'_1, \Omega'_2$  とその間の強準同型  $\varphi': \Omega'_1 \rightarrow \Omega'_2$  が存在することをいう。
- (3) 確率空間の**強同型**とは、単射でもあり、可測な逆を持つ強準同型をいう。すなわち、 $\sigma$ -加法族を押し出し  $\varphi^*(\mathcal{F}_1) = \mathcal{F}_2$ 、かつ、測度を押し出すことをいう。
- (4) 確率空間の**同型**とは、それぞれの充満集合  $\Omega'_1, \Omega'_2$  とその間の全単射写像  $\varphi: \Omega'_1 \rightarrow \Omega'_2$  が存在して次を満たすことをいう：
  - (a) 同値な測度を押し出す： $\varphi_* P_1 = P_1 \circ \varphi^* = P_2$ 。
  - (b) 相対測度について可測： $\varphi^{-1}(\mathcal{F}_2 \cap \Omega'_2) = \mathcal{F}_1 \cap \Omega'_1$ 。
 すなわち、それぞれの充満集合  $\Omega'_1, \Omega'_2$  が存在して、 $(\Omega'_i, \mathcal{F}_i \cap \Omega'_i, P_i|_{\mathcal{F}_i \cap \Omega'_i})$  が互いに強同型であることをいう。
- (5) ある  $\mathbb{R}$  上の完備 Borel 確率空間  $(\mathbb{R}, \mathcal{G}(\mathbb{R}), \mu)$  と同型な完備確率空間  $(\Omega, \mathcal{F}, P)$  を**標準確率空間**という。

**観察 1.2.6** (Meas の特徴). 圏論的には極めてひどい性質を持つ。特に、Meas は閉圏 (cartesian closed category) ではない。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)。<sup>15</sup> □

**歴史 1.2.7.** これを逃れるための概念の模索は von Neumann (1932) から始まり、Rokhlin による Lebesgue 空間 (1940) として結晶した。Kolmogorov と Gnedenko による完全性もその 1 つであるが、本質的には同じである。

<sup>15</sup> これは Ord, Top, Diff と同根で、これらの圏は空間の圏  $\mathcal{C}$  から代数の圏  $\mathcal{D}$  へのファイバー  $U: \mathcal{C} \rightarrow \mathcal{D}$  として捉えられ、**位相的な圏**という。すると、Borel  $\sigma$ -加法族  $\mathcal{G}(S)$  は、関手  $\mathcal{G}: \text{Top} \rightarrow \text{Meas}\sigma\text{-Alg}$  として定まる。

**記法 1.2.8** (伊藤 [伊藤清, 1991] の用語のまとめ).

- (1) 任意の測度空間  $(X, \mathcal{G}, \mu)$  に対して存在する完備化  $\bar{\mu}$  を, **Lebesgue 拡大**と呼ぶ.
- (2) 確率空間に位相の情報を加えることを考える. 定義域が Borel 集合族と一致するとき, 確率測度を Borel といい, その完備化として得られる測度を**正則**というが, ここでは完備 Borel の語を採用する. 完備 Borel 確率測度を**分布**といい,  $P(X)$  で表すこととする.
- (3) 像測度を  $Pf^{-1}, fP$  で表す.
- (4) 測度の内部正則性を  $K$ -正則ともいう.

**定義 1.2.9** (**perfect (Kolmogorov), separating family, properly separable**).

- (1) 完備な確率測度  $\mu$  が**完全**であるとは, 任意の可測関数  $f: S \rightarrow \mathbb{R}$  に沿った押出が完備 Borel 確率測度になることをいう:  $f_*\mu \in P(\mathbb{R})$ .  $\mu$  が完備 Borel 測度なら, 完全である.
- (2) 可算集合  $\mathcal{G} = \{B_n\} \subset \mathcal{F}$  が**分離族**であるとは,

$$\forall (C_n)_{n \in \mathbb{N}} \in \prod_{n \in \mathbb{N}} \{B_n, B_n^c\} \quad |\cap_{n \in \mathbb{N}} C_n| \leq 1.$$

- (3) 完備確率測度  $\mu$  について,  $\text{Dom}(\mu)$  が可算な分離族によって生成されるとき,  $\mu$  を**真に可分**という.

**定理 1.2.10** (可分完全性は抽象 Lebesgue 空間よりもさらに狭いクラスである).

- (1) 標準確率測度は完全で, さらに連続体濃度以下ならば可分でもある.
- (2) 完全な測度の可測写像による押し出しは再び完全である.
- (3) 完備可分距離空間上の完備 Borel 確率測度は
  - (a) 標準的である. すなわち, ある完備確率測度  $\mu \in P(\mathbb{R})$  が存在して,  $(\mathbb{R}, \overline{\mathcal{B}(\mathbb{R})}, \mu)$  に同型である.
  - (b) 原子的でないならば, Lebesgue 空間である. すなわち,  $([0, 1], \overline{\mathcal{B}([0, 1])}, m)$  に同型である.
  - (c) 可分完全である.
- (4) 完備な確率空間  $(S, \mu)$  について, 次の 2 条件は同値.
  - (a)  $(S, \mu)$  は可分完全である.
  - (b)  $\mathbb{R}^1$  のある部分集合とその上の Radon 確率測度が存在して, これと強同型である.

## 1.2.4 直積測度の構成

**定義 1.2.11** (直積  $\sigma$ -代数 / Kolmogorov の  $\sigma$ -加法族).  $(\Omega_n, \mathcal{F}_n, P_n)$  を確率空間の列とする.  $\Omega := \prod_{n \in \mathbb{N}} \Omega_n$  とし,  $\text{pr}_n: \Omega \rightarrow \Omega_n$  を射影とする. 射影の系  $(\text{pr}_n)$  が  $\Omega$  上に定める  $\sigma$ -加法族

$$\bigoplus_{n \in \mathbb{N}} \mathcal{F}_n := \sigma \left[ \left\{ E_1 \times \cdots \times E_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots \mid E_i \in \mathcal{F}_i, n \in \mathbb{N} \right\} \right].$$

を, **積  $\sigma$ -加法族**, または **Kolmogorov の  $\sigma$ -加法族**という.

**要諦 1.2.12** (cylinder set). 射影  $\text{pr}_n^\infty := (\text{pr}_1, \dots, \text{pr}_n)$  による  $\sigma$ -加法族  $S_1 \times \cdots \times S_n$  の元の逆像  $q_i^*(A_i)$  ( $A_i \in S_i$ )

$$C_A^{(n)} = \{\omega \in \Omega \mid (\omega_1, \dots, \omega_n) \in A\} \quad A \in S_1 \times \cdots \times S_n$$

を**柱状集合**という. 例えば,  $\text{pr}_1: \mathbb{R}^2 \rightarrow \mathbb{R}$  の場合は帯領域である. なお, 柱状集合の全体  $\mathcal{C}$  は有限加法族で, これが生成する  $\sigma$ -加法族が Kolmogorov  $\sigma$ -加法族である.

**定理 1.2.13** (直積測度の一意性).  $(\Omega_n, \mathcal{F}_n, P_n)$  を確率空間の列とする. 積可測空間  $\left( \prod_{n \in \mathbb{N}} \Omega_n, \bigoplus_{n \in \mathbb{N}} \mathcal{F}_n \right)$  上の確率測度  $P$  で, 次を満たすものはただ一つである:

$$P[E_1 \times \cdots \times E_n \times \Omega_{n+1} \times \cdots] = \prod_{i=1}^n P_i[E_i], \quad E_i \in \mathcal{F}_i, n \in \mathbb{N}.$$

これを  $P =: \bigoplus_{n \in \mathbb{N}} P_n$  と表す.



記法 1.2.14. 射影を

$$\begin{array}{ccc} \text{pr}_m^n : \prod_{i=1}^n \mathbb{R}^d & \longrightarrow & \prod_{i=1}^m \mathbb{R}^d \\ \downarrow \Psi & & \downarrow \Psi \\ (x_i)_{i \in [n]} & \longmapsto & (x_i)_{i \in [m]} \end{array}$$

と定める. ただし,  $m < n \in \overline{\mathbb{N}}$ .

**定理 1.2.15 (Kolmogorov の拡張定理).** 確率測度の系  $P_n \in P\left(\prod_{i=1}^n \mathbb{R}^d\right)$  について, 次は同値:

(1)  $\left(\prod_{i=1}^{\infty} \mathbb{R}^d, \mathcal{B}\left(\prod_{i=1}^{\infty} \mathbb{R}^d\right)\right)$  上の確率測度  $P$  が存在して,

$$P[(\text{pr}_n^{\infty})^{-1}(A_n)] = P_n[A_n], \quad A_n \in \mathcal{B}\left(\prod_{i=1}^n \mathbb{R}^d\right), n \in \mathbb{N}.$$

(2) 系  $(P_n)$  は次の一貫性条件を満たす:

$$P_n[(\text{pr}_m^n)^{-1}(A_m)] = P_m[A_m], \quad A_m \in \mathcal{B}\left(\prod_{i=1}^m \mathbb{R}^d\right), n < m \in \mathbb{N}.$$

また,  $\mathbb{R}^d$  を一般の Polish 空間  $S$  としても成立するし,  $\Omega := (\mathbb{R}^d)^{\mathbb{R}_+}$  についても成立する.

### 1.2.5 積空間の性質

#### 圏論的な振る舞いの良さの回復

完備可分距離空間上の完備な Borel 確率空間については, 位相構造と測度構造が両立しており, 積  $\sigma$ -加法族は積位相の  $\sigma$ -加法族になる. より一般に, 標準確率空間の積空間は再び標準になる.

**定理 1.2.16 (完備可分空間の直積  $\sigma$ -代数は, 直積位相の Borel  $\sigma$ -代数).**  $\Omega_n$  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  $\sigma$ -加法族  $\mathcal{F}$  は,  $\Omega$  の直積位相が生成する Borel 集合族に一致する.

**定理 1.2.17 (正則確率空間の可算直積).**

- (1) 完備可分距離空間の列  $(S_n)$  の直積  $S := \prod_{n \in \mathbb{N}} S_n$  は完備可分である.
- (2)  $S_n$  上の完備な Borel 確率測度の列  $(P_n)$  の直積の完備化  $\overline{\prod_{n \in \mathbb{N}} P_n}$  は  $S$  上の完備な Borel 確率測度である.
- (3) 標準確率測度の列  $(P_n)$  の積の完備化  $\overline{\prod_{n \in \mathbb{N}} P_n}$  も標準確率測度になる.

**命題 1.2.18 (積写像の可測性).**

- (1) 実数値可測関数として,  $X_1 \in \mathcal{L}(\Omega; S_1), X_2 \in \mathcal{L}(\Omega; S_2)$  と,  $(X_1, X_2) \in \mathcal{L}(\Omega; S_1 \times S_2)$  は同値.
- (2) しかし,  $S_1 \times S_2$  が可分でない限り,  $X_1, X_2$  が確率変数であっても,  $(X_1, X_2)$  も確率変数とは限らない. すなわち,  $(X_1, X_2)$  が押し出す測度が確率測度であるならば,  $(X_1, X_2)$  も確率変数である.

**定理 1.2.19 (非可算無限直積の Set 上の特徴付け).**  $T$  を集合とし,  $[T] := \{S \subset T \mid |S| < \infty\}$  とおくと, 直積集合  $\mathbb{R}^T$  はネット  $(\mathbb{R}^S)_{S \in [T]}$  の射影極限でもある. すなわち, 任意の集合  $A \in \text{Set}$  と写像の系  $(f_S : A \rightarrow \mathbb{R}^S)_{S \in [T]}$  について, 次を可換にするただ一つの写像  $h : \mathbb{R}^T \rightarrow A$  が存在する:

$$\begin{array}{ccc} \mathbb{R}^T & \xleftarrow{f_S} & A \\ \text{pr}_S \downarrow & \swarrow h & \\ \mathbb{R}^S & & \end{array}$$



## 1.2.6 非可算積の扱い

**命題 1.2.20.**  $S := C([0, 1]; \mathbb{R}^d)$  に一様位相を与えて距離空間と見る. Borel 集合族  $\mathcal{B}(S)$  は集合族

$$\mathcal{G} := \{C(t_1, \dots, t_n; A_1, \dots, A_n) \in P(S) \mid n \in \mathbb{N}, 0 \leq t_1 < \dots < t_n \leq 1, A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^d)\}$$

で生成される  $\sigma$ -代数  $\sigma(\mathcal{G})$  に等しい.

**[証明].**

$$\sigma(\mathcal{G}) \subset \mathcal{B}(S)$$

任意の  $n \in \mathbb{N}, 0 \leq t_1 < \dots < t_n \leq 1, A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^d)$  に対して,

$$C(t_1, \dots, t_n; A_1, \dots, A_n) = \bigcap_{i \in [n]} \text{pr}_{t_i}^{-1}(A_i)$$

と表せる. ただし,  $\text{pr}_{t_i} : S \rightarrow \mathbb{R}^d$  を,  $\omega \in S$  を  $\omega(t_i) \in \mathbb{R}^d$  に写す写像とした. すると, 各線型汎関数  $\text{pr}_{t_i}$  は連続であるから,  $A_i$  が全て開集合であるとき,  $C$  も開集合となる. 次に  $A_i \in \mathcal{B}(\mathbb{R}^d)$  が一般 (開集合または閉集合の可算和・積) のときも,  $C$  は開集合または閉集合の可算和・積で表せる.

$$\mathcal{B}(S) \subset \sigma(\mathcal{G})$$

一様収束の位相は, ノルム  $\|\omega\| = \sup_{t \in [0, 1]} |\omega(t)|$  が定める位相に等しい. これについての開球

$$B(\omega, \epsilon) = \{\omega_1 \in S \mid \|\omega_1 - \omega\| < \epsilon\} \stackrel{\text{open}}{\subset} S$$

は,  $\omega \in S$  が連続関数であることに注意すれば,  $B(\omega, \epsilon) = \bigcup_{t \in \mathbb{Q} \cap [0, 1]} C(t, B(\omega(t), \epsilon))$  と表せる. 開球  $B$  は一様収束の位相の基底をなすから, この議論から  $\mathcal{B}(S) \subset \sigma(\mathcal{G})$  が従う. ■

**命題 1.2.21.**  $X = \{X_t\}_{t \in [0, 1]} \subset L(\Omega)$  を族とする.

(1)

$$\sigma[X] = \{A \in \mathcal{F} \mid A \in \sigma[X_t \mid t \in I], I \subset [0, 1] \text{ は可算}\}.$$

(2) 任意の  $F : \Omega \rightarrow \mathbb{R}$  について, 次は同値:

(a)  $\sigma[X]$ -可測である.

(b) Borel 可測関数  $\varphi \in L(\mathbb{R}^{\mathbb{N}})$  と列  $\{t_i\} \subset [0, 1]$  が存在して,

$$F(\omega) = \varphi(X_{t_1}(\omega), \dots, X_{t_n}(\omega), \dots).$$

(c) Borel 可測関数  $\varphi \in L(C([0, 1]))$  が存在して,

$$\varphi(X) = F.$$

(3)

$$\{\omega \in \Omega \mid t \mapsto X_t \text{ は連続}\} \notin \sigma[X].$$

(4)  $X \in C(\Omega)$  のとき, 転置  $X_\bullet : \Omega \rightarrow C([0, 1])$  は可測である.

## 1.3 確率測度論

### 1.3.1 事象の極限

単調列の極限として必ず定まる  $\limsup, \liminf : \mathcal{F}^{\mathbb{N}} \rightarrow \mathcal{F}$  を用いて, Darboux の方法で事象列  $\{A_n\} \subset \mathcal{F}$  に極限を定める. すると, 確率測度  $\mu : \mathcal{F} \rightarrow [0, 1]$  は, この収束列を保存する写像として特徴付けられる.

**定義 1.3.1** (上極限事象, 下極限事象, 極限事象). 事象列  $(A_n) : \mathbb{N} \rightarrow \mathcal{F}$  に対して, 次の事象が定義できる.

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

**命題 1.3.2** (測度の連続性と劣加法性). 可測空間  $(\Omega, \mathcal{F})$  上の確率測度  $\mu : \mathcal{F} \rightarrow [0, 1]$  について, 次が成り立つ.

- (1) 単調増加列  $(A_n)$  に対して,  $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\bigcup_{n=1}^{\infty} A_n)$ .
- (2) 単調減少列  $(A_n)$  に対して,  $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\bigcap_{n=1}^{\infty} A_n)$ .
- (3) (Boole's inequality / subadditivity) 測度  $\mu$  に対して,  $\mu(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mu(A_i)$  ( $n \in \mathbb{N}$ ).
- (4) 一般の収束列  $(A_n) \rightarrow A$  に対して,  $P[A_n] \rightarrow P[A]$ .

**[証明]**. (4) は Fatou の補題により,

$$P[A] \leq \liminf_{n \rightarrow \infty} P[A_n] \leq \limsup_{n \rightarrow \infty} P[A_n] \leq P[A].$$

■

**補題 1.3.3** (概収束するならば確率収束する).  $\limsup, \liminf : \mathcal{F}^{\mathbb{N}} \rightarrow \mathcal{F}$  は  $^c$  に関して可換.

- (1)  $\limsup A_n^c = (\liminf A_n)^c$ .
- (2)  $\liminf A_n^c = (\limsup A_n)^c$ .

特に, 概収束するなら確率収束する.

**要諦 1.3.4.**  $A_n := \{|X_n - X| \geq \epsilon\}$  とすると,

- (1)  $\forall \epsilon > 0 \quad P[(\limsup A_n)^c] = P[\liminf A_n^c] = 1$  が概収束の定義であり,
- (2)  $\forall \epsilon > 0 \quad P[A_n^c] \xrightarrow{n \rightarrow \infty} 1$  が確率収束の条件である.

すなわち, 事象列  $(A_n)$  が,  $P$  を通じて 0 に収束することが確率収束で, 事象列自体が零集合に下限収束することが概収束である. よって,

- (1) Fatou の補題より,  $(0 \leq) \limsup_{n \rightarrow \infty} P[A_n] \leq P[\limsup A_n]$  であるから, 概収束するなら確率収束する.
- (2) Borel-Cantelli の補題により, 概収束するための十分条件の 1 つが  $\sum_{n=1}^{\infty} P[A_n] < \infty$  である.

### 1.3.2 Dynkin 族定理

位相空間論のような基底論は展開出来ないが, 話を Dynkin 族に限ると一意性定理を得る. 乗法族は Dynkin 族の  $\sigma$ -生成に関する基底となる.

**定義 1.3.5** ( $\pi$ -system, Dynkin family /  $\lambda$ -system).

- (1) 集合族  $\mathcal{G} \subset P(X)$  が乗法族または  $\pi$ -系であるとは、共通部分  $\cap$  について閉じていることをいう。
- (2) 集合族  $\mathcal{G} \subset P(X)$  が Dynkin 族または  $\lambda$ -系であるとは、<sup>†6</sup> 次の3条件を満たすことをいう。
- (a) 乗法単位元:  $X \in \mathcal{G}$ .
  - (b) 差集合:  $\forall A, B \in \mathcal{G} \quad A \subset B \Rightarrow B \setminus A \in \mathcal{G}$ .
  - (c) 単調生成:  $\forall (A_i)_{i \in \mathbb{N}}: \mathbb{N} \rightarrow \mathcal{G} \quad [\forall i \neq j \in \mathbb{N} \quad A_i \cap A_j = \emptyset] \Rightarrow \sum_{i \in \mathbb{N}} A_i \in \mathcal{G}$ .
- この条件は、(2) の下で  $\mathcal{G} \supset \{A_n\} \nearrow A \Rightarrow A \in \mathcal{G}$  と同値。

例 1.3.6.  $\mathbb{R}$  上の  $\{(-\infty, a]\}_{a \in \mathbb{R}}$  や、 $\mathbb{R}^d$  上の

$$\mathcal{G} = \left\{ \prod_{i=1}^d (a_i, b_i] \subset \mathbb{R}^d \mid a_i \leq b_i \in \mathbb{R} \right\}.$$

は乗法族である。 □

**補題 1.3.7 (Dynkin 族は合併について閉じている).** 任意の集合  $X$  上の Dynkin 族の族  $(A_\lambda)_{\lambda \in \Lambda}$  について、 $\cap_{\lambda \in \Lambda} A_\lambda$  も Dynkin 族である。

**[証明].** 任意の  $\lambda \in \Lambda$  について  $X \in A_\lambda$  より、 $X \in \cap_{\lambda \in \Lambda} A_\lambda$ .  $A \subset B$  を満たす  $A, B \in \cap_{\lambda \in \Lambda} A_\lambda$  と  $\cap_{\lambda \in \Lambda} A_\lambda$  の族  $(A_i)_{i \in \mathbb{N}}$  を任意にとると、任意の  $\lambda \in \Lambda$  について  $B \setminus A \in A_\lambda$  かつ  $\sum_{i \in \mathbb{N}} A_i \in A_\lambda$ . よって、 $B \setminus A, \sum_{i \in \mathbb{N}} A_i \in \cap_{\lambda \in \Lambda} A_\lambda$ . ■

**定理 1.3.8 ( $\pi$ - $\lambda$  定理).**  $\mathcal{P}$  を  $\pi$ -系、 $\mathcal{L}$  を  $\lambda$ -系とする。

- (1) 任意の集合族  $\mathcal{A} \subset \mathcal{F}$  について、次は同値：
  - (a)  $\mathcal{A}$  は  $\sigma$ -代数.
  - (b)  $\mathcal{A}$  は乗法族かつ Dynkin 族である.
- (2)  $\mathcal{P} \subset \mathcal{L} \Rightarrow \sigma[\mathcal{P}] \subset \mathcal{L}$ .
- (3) 特に、 $\sigma[\mathcal{P}]$  は  $\mathcal{P}$  を含む最小の  $\lambda$ -系として特徴付けられる。

### 1.3.3 単調族定理

**定義 1.3.9 (monotone class).**  $\mathcal{M} \subset P(\Omega)$  が単調族であるとは、任意の単調列の極限について閉じていることをいう。

**定理 1.3.10 (単調族定理).** 有限加法族  $\mathcal{G}$  の単調族閉包と  $\sigma$ -閉包は等しい:  $\mathcal{M}[\mathcal{G}] = \sigma[\mathcal{G}]$ .

**系 1.3.11 (Borel 集合族の生成核).**

$$\mathcal{A} := \left\{ \prod_{i=1}^d (a_i, b_i] \subset \mathbb{R}^d \mid a_i \leq b_i \in \overline{\mathbb{R}} \right\}$$

の  $\sigma$ -閉包は  $\mathcal{B}(\mathbb{R}^d)$  である。ただし、 $(a, a] = \emptyset, ([a, \infty) = (a, \infty)$  とみなす。

**定理 1.3.12 (関数に関する単調族定理).**  $\mathcal{F} \subset P(\Omega)$  を乗法系、 $\mathcal{T} \subset \text{Map}(\Omega, \mathbb{R})$  を有界関数の属とする。次を満たすならば、 $\mathcal{T}$  は  $\Omega$  上の  $\sigma[\mathcal{F}]$ -可測な有界関数の全体  $L_{\sigma[\mathcal{F}]}^\infty(\Omega)$  を含む。

- (1)  $\mathcal{T}$  は  $1 \in \mathcal{T}$  を満たす実線型空間.
- (2) 次の意味で  $\mathcal{F} \hookrightarrow \mathcal{T}$  である:  $\forall A \in \mathcal{F} \quad 1_A \in \mathcal{T}$ .
- (3) 任意の非負関数の単調増加列  $\{f_n\} \subset \mathcal{T}$  について、これが有界関数  $f$  に収束するならば  $f \in \mathcal{T}$  である。

**[証明].**

Step1

$$\mathcal{G} := \{A \in P(\Omega) \mid 1_A \in \mathcal{T}\}.$$

とすると、仮定 (2) より、 $\mathcal{F} \subset \mathcal{G}$ . いま、 $\mathcal{G}$  は Dynkin 族である。

Step2 よって Dynkin 族定理より、 $\sigma[\mathcal{F}] \subset \mathcal{G}$ . 特に、任意の  $A \in \sigma[\mathcal{F}]$  について、 $1_A \in \mathcal{T}$ . あとは仮定 (1),(3) から結論を得る。

<sup>†6</sup> Dynkin 自身は  $\lambda$ -系と呼んだ

### 1.3.4 確率測度の正則性

**系 1.3.13.**  $P \in \mathcal{P}(\mathbb{R}^d)$  を確率測度とする. 任意の  $A \in \mathcal{B}(\mathbb{R}^d)$  と任意の  $\epsilon > 0$  に対して, 次を満たすコンパクト集合  $K \subset^{\text{cpt}} \mathbb{R}^d$  と開集合  $G \subset^{\text{open}} \mathbb{R}^d$  の組が存在する:

- (1)  $K \subset A \subset G$ .
- (2)  $P[G \setminus A] \leq \epsilon, P[A \setminus K] \leq \epsilon$ .

**[証明].**

外部正則性について

$$\mathcal{G} := \left\{ A \in \mathcal{B}(\mathbb{R}^d) \mid \begin{array}{l} \text{任意の } \epsilon > 0 \text{ に対して } G \subset^{\text{open}} \mathbb{R}^d \\ \text{が存在して } P[G \setminus A] \leq \epsilon \end{array} \right\}.$$

と定めると,  $\mathcal{A} \subset \mathcal{G}$  かつ  $\mathcal{A}$  は単調族であるから, 系 1.3.11 より外部正則性に関する条件は従う.

内部正則性について 任意の  $\epsilon > 0$  を取ると, ある  $G \subset^{\text{open}} \mathbb{R}^d$  が存在して,  $A^c \subset G$  かつ  $P[G \setminus A^c] \leq \epsilon$  を満たす. よって,

$$P[A \cap (G^c)^c] = P[A \cap G] \leq \epsilon$$

を満たすから,  $F := G^c \subset A$  とすると閉集合で  $P[A \setminus F] \leq \epsilon$  を満たす. あとは  $F$  を有界に取ればよい.  $F \cap [-N, N]^d \nearrow F$  より, ある  $N \in \mathbb{N}$  が存在して,  $P[F \setminus (F \cap [-N, N]^d)] \leq \epsilon$  を満たす.

### 1.3.5 Bonferroni の不等式

**命題 1.3.14 (Bonferroni の不等式).**

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

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

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

## 1.4 独立性

### 1.4.1 分割の独立性

確率測度  $P$  は,  $\mathcal{F}$  上に対称的な  $n$  項関係を定める. 独立性は推移的ではないが,  $^c, \cup, \cap$  については保存されるから, 実は  $\sigma$ -代数上に定義されている.

**定義 1.4.1 (independent).** 集合族  $\{A_\lambda\}_{\lambda \in \Lambda} \subset \mathcal{F}$  が**独立**であるとは, 任意の  $n \in \mathbb{N}$  個の相異なる元  $A_{\lambda_1}, \dots, A_{\lambda_n}$  に対して,

$$P[A_{\lambda_1} \cap \dots \cap A_{\lambda_n}] = P[A_{\lambda_1}] \cdots P[A_{\lambda_n}]$$

が成り立つことをいう.

**命題 1.4.2 (事象の独立性の特徴付け).** 有限個の事象  $A_1, \dots, A_n$  について, 次の3条件は同値:

- (1)  $(2^n - 1 - n)$  次連立方程式  $A_1, \dots, A_n$  は独立である:  $\forall_{j \in [n]} \forall_{k_1 < \dots < k_j \in [n]} P[A_{k_1} \cap \dots \cap A_{k_j}] = P[A_{k_1}] \cdots P[A_{k_j}]$ ,
- (2)  $\forall_{(B_1, \dots, B_n) \in \prod_{j \in [n]} \{A_j, \Omega\}} P[B_1 \cap \dots \cap B_n] = P[B_1] \cdots P[B_n]$ .
- (3)  $\forall_{(B_1, \dots, B_n) \in \prod_{j \in [n]} \{A_j, A_j^c\}} P[B_1 \cap \dots \cap B_n] = P[B_1] \cdots P[B_n]$ .

**[証明].** いずれも,  $\sigma[A_1], \dots, \sigma[A_n]$  が独立であることに同値である.  $P[X \cap A_i] = P[X]P[A_i]$  や  $P[\emptyset \cap A_i] = P[\emptyset]P[A_i]$  は自明に成り立つことに注意すれば, (1) は  $\sigma[A_1], \dots, \sigma[A_n]$  ใดなる部分集合も独立であることを主張しており, (2) は  $\{A_i, \Omega, \emptyset\}$  は  $\sigma[A_i]$  を Dynkin 生成する乗法族であることから. ■

**補題 1.4.3 (独立性の遺伝).** 事象  $A_1, \dots, A_n$  が独立であるとする.

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

**補題 1.4.4 (独立事象に対する確率の関手性).** 列  $(A_i)_{i \in \mathbb{N}}$  が独立ならば,

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

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

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

## 1.4.2 代数の独立性

$\sigma$ -代数が独立  $\mathcal{G} \perp \mathcal{H}$  であるとは, 積空間への憧憬  $(\Omega, \mathcal{F}, P) \xrightarrow{\sim} (\Omega_1, \mathcal{G}, P_1) \otimes (\Omega_2, \mathcal{H}, P_2)$  が取れることに同値.  $\sigma$ -代数の独立性については, その部分乗法族について独立性を確かめればよい.  $\sigma$ -代数の独立性を通じて, 確率変数の独立性が定義される.

**定義 1.4.5.** 部分集合  $\mathcal{A}_1, \dots, \mathcal{A}_n \subset \mathcal{F}$  が独立であるとは, 任意の組  $(A_1, \dots, A_n) \in \prod_{j \in [n]} \mathcal{A}_n$  が互いに独立であることとする.

**定理 1.4.6.**  $\mathcal{A}_1, \dots, \mathcal{A}_n$  を互いに独立な乗法族とする. このとき,  $\sigma(\mathcal{A}_1), \dots, \sigma(\mathcal{A}_n)$  も独立である.

**系 1.4.7.** 独立な列  $\mathcal{G} := \{A_n\}_{n \in \mathbb{N}} \subset \mathcal{F}$  の可算な分割  $\mathcal{G} = \cup_{i \in \mathbb{N}} \mathcal{F}_i$  について,  $\mathcal{F}_1, \mathcal{F}_2, \dots$  は互いに独立である.

**命題 1.4.8 (測度の独立性の直積による特徴付け).** 確率変数族  $\{X_\lambda\}_{\lambda \in \Lambda} \subset L(\Omega)$  の独立性を, それらが引き起こす  $\sigma$ -代数族の独立性によって定める. このとき, 次の2条件は同値:

- (1)  $\{X_\lambda\}_{\lambda \in \Lambda}$  は独立.
- (2) 直積  $\sigma$ -代数  $\otimes \mathcal{F}_\lambda$  上の測度として,  $P_{(X_\lambda)_{\lambda \in \Lambda}} = \otimes_{\lambda \in \Lambda} P_{X_\lambda}$ .

**要諦 1.4.9.** おそらく, 独立な確率変数列が引き起こす測度は, ある直積確率空間と同型になる, というような理論である. 独立でない場合というのが, 集合論的な退化現象が起こる場合である (ランク落ちのような).

## 1.4.3 確率変数の独立性

**定理 1.4.10 (独立性の十分条件).**  $\{X_\lambda\} \subset L(\Omega; \mathbb{R}^d)$  と乗法族  $\mathcal{G} \subset \mathcal{B}(\mathbb{R}^d)$  とについて, 次は同値:

- (1)  $\{X_\lambda\}$  は独立.
- (2) 任意の有限部分集合  $\{\lambda_1, \dots, \lambda_n\} \subset \Lambda$  と  $I_1, \dots, I_n \in \mathcal{G}$  について,

$$P[X_{\lambda_1} \in I_1, \dots, X_{\lambda_n} \in I_n] = \prod_{i=1}^n P[X_i \in I_n].$$

**系 1.4.11.**  $\{X_i\}_{i \in [n]} \subset L(\Omega)$  について, 次は同値:

- (1)  $\{X_i\}_{i \in [n]}$  は独立.

(2) 同時分布関数は積で表される：任意の  $x_1, \dots, x_n \in \mathbb{R}$  について,

$$P[X_1 \leq x_1, \dots, X_n \leq x_n] = \prod_{i \in [n]} P[X_i \leq x_i].$$

(3) (存在するならば) 同時密度関数が積で表される： $f_{(X_1, \dots, X_n)} = f_{X_1} \cdots f_{X_n}$ .

(4) 分布が積で表される： $P^{(X_1, \dots, X_n)} = P^{X_1} \otimes \cdots \otimes P^{X_n}$ .

(5) 特性関数が積で表される： $\varphi_{(X_1, \dots, X_n)} = \varphi_{X_1} \cdots \varphi_{X_n}$ .

**[証明].**

(1) $\Leftrightarrow$ (2) 定理による.

(2) $\Leftrightarrow$ (3) 微分による.

(1) $\Leftrightarrow$ (4) 独立性の条件式

$$P[X_1 \in A_1, \dots, X_n \in A_n] = \prod_{i=1}^n P[X_i \in A_i]$$

とは,

$$P^{(X_1, \dots, X_n)}[A_1 \times \cdots \times A_n] = \prod_{i=1}^n P^{X_i}[A_i]$$

と書き換えられるためである. ■

#### 1.4.4 独立性の遺伝

**定理 1.4.12.**  $\{X_\lambda\} \subset L(\Omega; \mathbb{R}^d)$  は独立とする. 任意の分割  $\Lambda = \sqcup_{i \in I} \Lambda_i$  に対して,  $\{F_i := \sigma[X_\lambda | \lambda \in \Lambda_i]\}_{i \in I}$  は独立である.

**[証明].**

$$\mathcal{F}_i := \left\{ \bigcup_{k=1}^n X_{\lambda_k}^{-1}(E_k) \in \mathcal{F} \mid \lambda_1, \dots, \lambda_n \in \Lambda_i, n \in \mathbb{N}, E_k \in \mathcal{B}(\mathbb{R}^d) \right\}.$$

と定めるとこれは乗法系で,  $\sigma[\mathcal{F}_i] = \mathcal{F}_i$  を満たす. これについて, 任意の有限集合  $\{i_1, \dots, i_k\} \subset I$  と任意の  $A_1 \in \mathcal{F}_{i_1}, \dots, A_k \in \mathcal{F}_{i_k}$  について,

$$P[\cap_{i=1}^k A_i] = \prod_{i=1}^k P[A_i].$$

が成り立つから, 定理 1.4.10 より  $\{\mathcal{F}_i\}$  は独立. ■

**系 1.4.13.**  $\{X_i\}_{i \in [n]} \subset L(\Omega; \mathbb{R}^d)$  を独立とする. 任意の Borel 可測関数  $f, g$  について,

$$f(X_1, \dots, X_k) \perp\!\!\!\perp g(X_{k+1}, \dots, X_n).$$

#### 1.4.5 Borel-Cantelli の補題

有界測度論に話を限ると, 測度の和が収束する事象列の極限事象  $\limsup, \liminf$  はいずれも 0 に収束している必要がある. そして, 測度の和が発散するとき, 互いに独立ならば事象列の極限事象  $\limsup$  は確率 1 で起こる必要がある. 振動しても良いので,  $\liminf$  は確率 1 で起こる必要はない.

**命題 1.4.14 (集合に関する Fatou の補題).**  $\limsup : \mathcal{F}^{\mathbb{N}} \rightarrow \mathcal{F}$  は次を満たす.

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

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

**定理 1.4.15 (Borel-Cantelli lemma).**

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

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty \Rightarrow \mu(\limsup_{n \rightarrow \infty} A_n) = 0.$$

特に,  $\mu(\liminf_{n \rightarrow \infty} A_n^c) = \mu(X)$ .

(2) 列  $(A_i) : \mathbb{N} \rightarrow \mathcal{F}$  が独立であるとき、確率測度  $P$  について、

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\limsup_{n \rightarrow \infty} A_n) = 1.$$

特に,  $P(\liminf_{n \rightarrow \infty} A_n^c) = 0$ .

**[証明].**

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

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

$$\mu(\limsup_{n \rightarrow \infty} A_n) = \mu(\cap_{m=1}^{\infty} B_m) = \lim_{m \rightarrow \infty} \mu(B_m) \quad \text{単調列の積 1.3.2(3)}$$

$$\leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mu(A_n) = 0. \quad \text{劣加法性 1.3.2(4)}$$

また、de Morgan の法則より、

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

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

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

もう一つの結論も de Morgan の定理から従う。 ■

**要諦 1.4.16** (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。

**系 1.4.17**. 事象列  $(A_n)$  が次の (1) と、(2)(a),(2)(b) のいずれかの 2 条件を満たすならば、 $P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$ ,  $\lim_{n \rightarrow \infty} P(A_n) = 0$  が成り立つ。

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

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

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



## 1.5 分布の扱い

### 1.5.1 分布関数

$((-\infty, x])_{x \in \mathbb{R}}$  も  $((-\infty, x))_{x \in \mathbb{R}}$  も  $\mathcal{G}(\mathbb{R})$  を生成するから、この上の値 (これを分布関数という) が決まれば分布  $\mu \in P(\mathbb{R})$  が決まる.

**定義 1.5.1 ((cumulative) distribution function).** 確率分布  $\mu \in P(\mathbb{R})$  または確率変数  $X \in L(\Omega)$  に対して、

$$F(x) := \mu((-\infty, x]) = P[X \leq x]$$

を分布関数という.

**補題 1.5.2 (分布関数の特徴付け).**  $(\mathbb{R}, \mathcal{B}_1(\mathbb{R}))$  上の確率測度  $\mu$  の分布関数を  $F(x) := \mu((-\infty, x])$  とする.

- (1)  $F$  は広義単調増加である.
- (2)  $\lim_{x \rightarrow \infty} F(x) = 1, \lim_{x \rightarrow -\infty} F(x) = 0$  をみたす有界関数である.
- (3) (一意化) 右半連続である.

**要諦 1.5.3.** 右半連続関数  $F: \mathbb{R} \rightarrow [0, 1]$  に対して、Stieltjes 積分  $\int (-) dF: C_c(\mathbb{R}) \rightarrow \mathbb{R}$ , またその延長として Lebesgue-Stieltjes 積分  $\int (-) dF: \mathcal{L}^1(\mathbb{R}) \rightarrow \mathbb{R}$  が定まる. すなわち、確率分布  $\mu$  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是一对一対応する.

**定理 1.5.4 (Radon-Riesz の表現定理の系).**

- (1)  $\mathbb{R}$  上の Borel 確率測度  $\int: C_c(\mathbb{R}) \rightarrow \mathbb{R}$  について、ある有界かつ右連続な単調増加関数  $F: \mathbb{R} \rightarrow [0, 1]$  が一意的存在して、これについての Stieltjes 積分と一致する.
- (2)  $F$  が補題の 3 条件を満たすとき、ある  $\mu \in P(\mathbb{R})$  が存在して、 $F = F_\mu$ .

### 1.5.2 Lebesgue 分解

まずは、分布関数に不連続点 (原子) があるかどうか、次に連続と言っても Lebesgue 測度に関して絶対連続か、の 2 段階で分類出来る.

**補題 1.5.5 (continuous / diffuse, atomic).** 分布関数  $F_\mu: \mathbb{R} \rightarrow [0, 1]$  について、次の 2 条件は同値:

- (1)  $F_\mu$  は連続である.
- (2)  $\mu$  は原子を持たない:  $\forall x \in \mathbb{R} \mu(\{x\}) = 0$ .

このとき、 $F$  は連続であるといい、これが成り立たないとき**原子的**であるという.

**[証明].**

(1)  $\Rightarrow$  (2)  $\mu$  が原子  $\{x\}$  を持つならば、数列  $\{x - 1/n\}_{n \in \mathbb{N}}$  について、 $F_\mu(x - 1/n) \xrightarrow{n \rightarrow \infty} \mu((-\infty, x)) < \mu((-\infty, x])$ .

(2)  $\Rightarrow$  (1)  $\mu$  は原子を持たないから、任意の  $x \in \mathbb{R}$  について、

$$0 = \mu(\{x\}) = \lim_{n \rightarrow \infty} \mu\left(\left(x - \frac{1}{n}, x + \frac{1}{n}\right]\right) = \lim_{n \rightarrow \infty} (F(x + 1/n) - F(x - 1/n)).$$

そこで、 $F$  の単調性に注意すれば、任意の  $\epsilon > 0$  について、ある  $n \in \mathbb{N}$  が存在して  $F(x + 1/n) - F(x - 1/n) < \epsilon$  であるから、 $|\delta| < 1/n$  を満たすように取れば、 $|F(x + \delta) - F(x)| < \epsilon$ .

**補題 1.5.6 (Lebesgue 分解).**

- (1) 任意の Radon 積分  $C_c(X) \rightarrow \mathbb{R}$  について, 連続と原子的部分との和で一意的に表せる.
- (2) 連続な Radon 測度  $\mu$  は, 任意の Radon 測度  $\nu$  に対して,  $\nu$ -絶対連続部分と  $\nu$ -特異部分とに a.e. の違いを除いて一意的に分解できる.

**定義 1.5.7 (discrete, absolute continuous, density, singular).** 分布関数  $F: \mathbb{R} \rightarrow [0, 1]$  について,

- (1)  $F$  が (純) 不連続または原子的であるとは, Lebesgue-Stieltjes 測度  $dF$  がデルタ測度の可算な凸結合で表せるときをいう.
- (2)  $F$  が原子的でないならば, 補題より  $F$  は関数として連続になる.
  - (a)  $F$  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:  $F(x) = \int_{-\infty}^x p(y)dy$ .
  - (b)  $F$  が特異であるとは,  $F$  は連続であるが, Lebesgue-Stieltjes 測度  $dF$  が Lebesgue 測度  $dx$  と互いに特異であるときをいう:  $\exists A \in \mathcal{G}_1(\mathbb{R}) \ dF(A) = 1 \wedge m(A) = 0$ .

ただし, 純不連続ならば特異である.

**定理 1.5.8 (Lebesgue decomposition).** 任意の分布関数  $F: \mathbb{R} \rightarrow \mathbb{R}$  は, 不連続分布関数  $F_1$ , 絶対連続分布関数  $F_2$ , 特異な連続分布関数  $F_3$  が一意的に存在して, これらの凸結合で表せる.

## 1.5.3 畳み込み

分布の畳み込みは, その分布を定める独立確率変数の和の演算に対応する.

**定義 1.5.9.**  $\mu_1, \mu_2 \in P(\mathbb{R})$  について,

$$\mu_1 * \mu_2(E) := \int_{\mathbb{R}} \mu_1(E - x) \mu_2(dx) = \int_{\mathbb{R}^2} 1_E(x + y) \mu_1(dx) \mu_2(dy)$$

**定理 1.5.10.** 任意の  $n \in \mathbb{N}$  と有界な Borel 可測関数  $f \in \mathcal{L}_b(\mathbb{R})$  に対して,

$$\int_{\mathbb{R}} f(x) \mu(dx) = \int_{\mathbb{R}^n} f(x_1 + x_2 + \cdots + x_n) \mu_1(dx_1) \cdots \mu_n(dx_n).$$

**命題 1.5.11 (測度の代数).**  $(P(\mathbb{R}), *)$  は交換的で結合的であるが, 超関数の世界を借りなければ単位的でない.

- (1)  $\mu_1 * \mu_2 = \mu_2 * \mu_1$ .
- (2)  $\mu_1 * (\mu_2 * \mu_3) = (\mu_1 * \mu_2) * \mu_3$ .

**定理 1.5.12 (Fourier 逆変換は代数の準同型である).**  $\mu, \nu \in P(\mathbb{R}^d)$  について, 畳み込みの特性関数は, 特性関数の積である:  $\varphi_{\mu * \nu} = \varphi_{\mu} \cdot \varphi_{\nu}$ . また,  $X_1, \dots, X_n \in L(\Omega)$  が独立ならば,  $\varphi_{X_1 + \cdots + X_n} = \varphi_{X_1} \cdots \varphi_{X_n}$ .

## 1.5.4 確率測度の Fourier 逆変換

特性関数の広義一様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. 確率測度が定める積分について,  $P(X) \hookrightarrow S'(X)$  と緩増加超関数の空間に埋め込むことが出来, 超関数に関する Fourier 変換と定義が一致し, これは位相線型空間  $S'(X)$  の自己同型を差玉得る.

**定義 1.5.13 (characteristic function).**  $\mu \in P(\mathbb{R}^n)$  の特性関数  $\varphi: \mathbb{R}^n \rightarrow \mathbb{C}$  とは, 緩増加分布としての Fourier 逆変換  $\mathcal{F}: S'_n \xrightarrow{\sim} S'_n$  による像

$$\varphi(u) := (\mu * e_{-u})(0) = \int_{\mathbb{R}^n} e^{iu \cdot x} \mu(dx)$$

をいう. 指標  $e_t(x) = e^{it \cdot x}$  との畳み込みであるためにこう呼ぶ.

**補題 1.5.14.**  $(\Omega, \mathcal{F}, P)$  を確率空間,  $\hat{f}$  は Fourier 逆変換とする.

- (1)  $\mu(dx) = f(x)dx$  であるとき,  $\hat{\mu} = \hat{f}$  が成り立つ.
- (2) 一般に Fourier 変換は  $\mathcal{F} : L^1(\Omega) \rightarrow C_0(\Omega)$  と定まっているノルム減少的な有界線型作用素であるが, 特性関数については, この値域は単位球面に収まる:  $\|\varphi\|_\infty = 1$ .
- (3) Fourier 変換の一般論から,  $\widehat{(-u)} = \overline{\hat{\mu}(u)}$  が成り立つ.

**定理 1.5.15 (一意性定理).**

- (1) Fourier 変換の制限  $\mathcal{F} : P(\mathbb{R}^n) \rightarrow \mathcal{F}(P(\mathbb{R}^n))$  は周期 4 の位相同型を定める. なお,  $P(\mathbb{R}^n)$  には弱収束の  $\sigma(M(\mathbb{R}^n), C_b(\mathbb{R}^n))$ -位相,  $\mathcal{F}(P(\mathbb{R}^n)) \subset C(\mathbb{R}^n; \mathbb{C})$  には広義一様収束の位相を入れた.
- (2) (Bochner-Khinchin) 像  $\mathcal{F}(P(\mathbb{R}^n))$  は,  $\mathbb{R}^n$  上の正の定符号関数の全体に等しい.

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

**定理 1.5.16 (群準同型).** Fourier 変換の制限  $\mathcal{F} : P(\mathbb{R}^n) \rightarrow \mathcal{F}(P(\mathbb{R}^n))$  は, 畳み込み  $*$  から各点積  $\cdot$  への群準同型を定める:  $\mathcal{F}(\mu_1 * \mu_2) = \varphi_1 \cdot \varphi_2$ .

**定理 1.5.17 (特性関数の特徴付け (Bochner)).** 複素関数  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$  が次の 2 条件を満たすならば, 特に正の定符号関数であるならば, ある 1 次元絶対連続分布  $\mu \in P(\mathbb{R})$  の特性関数である:

- (1)  $\varphi(0) = 1$  かつここで連続.
- (2) 正定値である.

### 1.5.5 特性関数の Taylor 展開

$\mathcal{F} : L^1(\mathbb{R}^d) \rightarrow C_0(\mathbb{R}^d; \mathbb{C})$  の制限は,  $\mathcal{F} : L^p(\mathbb{R}^d) \rightarrow C^p(\mathbb{R}^d)$  という様子になっている.

**定理 1.5.18 (滑らかさと絶対積率の存在 [?]).**  $\mu \in P(\mathbb{R})$  とその特性関数  $\varphi \in C_0(\mathbb{R}; \mathbb{C})$  について, 次の 2 条件は同値:

- (1)  $\varphi$  は原点において  $2r$  階連続微分可能である.
- (2)  $2r$  次の絶対積率が存在する:  $\beta_{2r} < \infty$ . すなわち,  $\mu$  の確率密度関数は  $L^{2r}(\Omega)$  の元.

このとき, 原点だけでなく  $\varphi \in C_0^{2n}(\mathbb{R}; \mathbb{C})$  でもある. また, (2) $\Rightarrow$ (1) の方向は, 一般の  $\beta_r$  ( $r \in \mathbb{N}$ ) についても成り立つ.

**定理 1.5.19 (特性関数の Taylor 展開).**  $\mu \in P(\mathbb{R})$  は  $\beta_r < \infty$  を満たすとする. このとき,  $\varphi \in UC_0(\mathbb{R}; \mathbb{C})$  は Maclauren 展開

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

を持つ.

[証明].  $\varphi$  の Maclauren 展開は  $u \in \mathbb{R}$  に関して広義一様収束する. したがって項別微分が可能で,

$$\begin{aligned} \varphi(u) &= \int e^{iux} f(x) dx \\ &= \int \left( 1 + iux + \frac{(iux)^2}{2!} + \cdots \right) f(x) dx \\ &= \int f(x) dx + iu \int x f(x) dx + (iu)^2 \int x^2 f(x) dx + \cdots = 1 + \alpha_1(iu) + \alpha_2 \frac{(iu)^2}{2!} + \cdots. \end{aligned}$$
■

### 1.5.6 キュムラント

**定理 1.5.20.** 第 2 キュムラント関数  $\psi := \log \circ \varphi$  は  $u = 0$  の近傍で次の Maclauren 展開を持つ:

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

このときの展開係数  $\kappa_j$  は,  $\varphi$  の展開係数  $\alpha_j$  に対して次の関係を持つ:

- (1)  $\kappa_1 = \alpha_1$ . これは平均ベクトルである.
- (2)  $\kappa_2 = \alpha_2 - \alpha_1^2$ . これは分散共分散行列である.
- (3)  $\kappa_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3$ .

[証明]. 広義一様収束性を確認したら, 合成関数の微分則によって従う. ■

**命題 1.5.21**. 一般に,  $1 \leq |n| \leq r$  を満たす多重指数  $n \in \mathbb{N}$  について  $\kappa_n$  はテンソルであり,

$$\kappa_n = \sum_{k=1}^{|n|} \sum_{n_1 + \dots + n_k = n \in \mathbb{N}^d} \frac{(-1)^{k-1}}{k} \binom{n}{n_1 \dots n_k} \prod_{m=1}^k \alpha_{n_m}.$$

$$\alpha_n = \sum_{k=1}^{|n|} \sum_{n_1 + \dots + n_k = n \in \mathbb{N}^d} \frac{1}{k!} \binom{n}{n_1 \dots n_k} \prod_{m=1}^k \kappa_{n_m}.$$

## 1.6 確率変数の空間

### 1.6.1 $L$ の構造

確率空間の射を確率変数といい, その全体を  $\mathcal{L}(\Omega; \mathcal{G})$  で表す. 特に  $\mathcal{G} = \mathbb{R}$  のときは  $\mathcal{L}(\Omega)$  と略記する.  $L(\Omega)$  も確率収束の位相について完備可分距離空間となる.

**定理 1.6.1**.  $L(\Omega)$  は確率収束の擬ノルム

$$\|X\|_0 = E \left[ \frac{|X|}{1 + |X|} \right]$$

について完備可分距離空間となる. 詳しくは, 次の通り:

- (1)  $\|\cdot\|_0$  は正定値性と三角不等式は満たすが, スカラー倍については  $-1$  の作用についてしか満たさない  $G$ -ノルムである.
- (2) この確率収束の定める距離について  $L(\Omega)$  は位相線型空間となる (スカラー倍と加法が連続になる). すなわち, 特に  $F$ -ノルムである.
- (3) この  $F$ -擬ノルムについて完備である. すなわち, 特に  $F$ -空間である.
- (4) この位相は局所凸である. すなわち, 特に Frechet 空間である.<sup>†7</sup>

**要諦 1.6.2**.  $L^p(\Omega) \subset L(\Omega)$  の  $p$ -ノルム位相は,  $L(\Omega)$  が定める相対位相よりも強い. これは Chebyshev 不等式  $P[|X|^p] \geq P[|X| \geq \epsilon] \epsilon^p$  から分かる.

**注 1.6.3**.  $L^p(\Omega)$  ( $0 < p < 1$ ) 空間の極限としても得られないか?

### 1.6.2 $L^p$ の構造

**定理 1.6.4** ( $L^\infty$  の小ささ). 確率空間上の  $L^p(P)$  ( $1 < p < \infty$ ) の閉部分空間が,  $L^\infty(P)$  にも含まれるならば有限次元である.

**定理 1.6.5** (Lyapunov の不等式). 確率空間において,  $0 < p < q$  のとき,  $\mathcal{L}^q(X) \subset \mathcal{L}^p(X)$  かつ  $\|\cdot\|_p \leq \|\cdot\|_q$  である.

[証明].  $r := q/p > 1$  とすると,  $|X|^p, 1$  についての Hölder の不等式より,

$$\| |X|^p \|_1 \leq \| |X|^p \|_r \|1\|_{r^*} \leq \| |X|^p \|_r.$$

これは,

$$\int_{\mathbb{R}} |x|^p P(dx) = \left( \int_{\mathbb{R}} |x|^{pr} P(dx) \right)^{1/r} = \left( \int_{\mathbb{R}} |x|^q P(dx) \right)^{p/q}$$

を意味するから,  $\|X\|_p \leq \|X\|_q$  を含意している. ■

<sup>†7</sup> この提示の仕方は高信 [高信敏, 2015] で初めてみた.

**定理 1.6.6.**  $\forall X \in L(\Omega) \quad \lim_{p \rightarrow \infty} \|X\|_{L^p(\Omega)} = \|X\|_{L^\infty(\Omega)}.$

## 1.7 L 上の汎関数

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

モデル上の汎関数  $\mathcal{P} \rightarrow \mathbb{R}$  を母数という. 標本空間  $\mathcal{X}^n$  上の関数  $\mathcal{X}^n \rightarrow \mathbb{R}$  をその推定量という.

また, 確率変数  $X$  の全体は場  $\mathfrak{F} := \Gamma(\Omega, E)$  と思え, その上の作用汎関数  $S: \mathfrak{F} \rightarrow \mathbb{C}$  の平均は分配関数と呼ばれる.

### 1.7.1 期待値と独立性

積分は正な線型汎関数だから,  $X \leq Y \Rightarrow E[X] \leq E[Y]$  である. 条件付き期待値とは射影  $L^2(\Omega) \rightarrow L^2_{\mathcal{G}}(\Omega)$  で,  $L^p$  ( $p \in [1, \infty]$ )-距離に関してノルム減少的である.

**定理 1.7.1.**  $\{X_i\} \subset L^1(\Omega)$  を独立とする. このとき  $X_1 \cdots X_n \in L^1(\Omega)$  でもあり,

$$E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n].$$

[証明]. まず  $\{X_i\}$  が非負単関数である場合に示す. ■

### 1.7.2 積率

1 次の積率  $\alpha_1(\mu)$  を平均といい, 2 次の中心積率  $\mu_2$  を分散という.

**定義 1.7.2.** 分布  $\mu \in P(\mathbb{R})$  に対して,

- (1)  $\alpha_p(\mu) = M_p(\mu) := \|\text{id}_{\mathbb{R}}\|_{L^p(\mu)}^p$  により定まる  $M_p: P(\mathbb{R}) \rightarrow \mathbb{R}_+$  を能率という.
- (2)  $\beta_p(\mu) := |M|_p(\mu) := \|\text{id}_{\mathbb{R}}\|_{L^p(\mu)}^p$  により定まる  $|M|_p: P(\mathbb{R}) \rightarrow \mathbb{R}_+$  を絶対能率という.

絶対積率は Holder の不等式より,  $p$  について単調増加する.

**系 1.7.3.** Chebyshev の不等式より,

$$\mu([-a, a]) \geq 1 - \frac{|M|_p(\mu)}{a^p}$$

### 1.7.3 $\mathbb{R}$ 上の分散

**定理 1.7.4.**  $\text{Var}: L^2(\Omega) \rightarrow \mathbb{R}_+$  は

- (1) 位置変数について不変  $\forall a \in \mathbb{R} \quad \text{Var}[X + a] = \text{Var}[X]$  で,
- (2) 2 次の斉次性  $\forall a \in \mathbb{R} \quad \text{Var}[aX] = a^2 \text{Var}[X]$  を持ち,
- (3) 独立確率変数の和に対して分解する  $\text{Var}[X_1 + \cdots + X_n] = \text{Var}[X_1] + \cdots + \text{Var}[X_n].$

**命題 1.7.5 (分散の特徴付け).**  $X \in L^2(\Omega)$  について,

$$\text{Var}[X] = \min_{c \in \mathbb{R}} E[(X - c)^2].$$

[証明].

$$E[(X - c)^2] = c^2 - 2cE[X] + E[X^2] = (c - E[X])^2 + E[X^2] - E[X]^2.$$

の最小値は  $E[X] = c$  のときの  $E[X^2] - E[X]^2 = \text{Var}[X]$  である. ■

1.7.4  $\mathbb{R}$  上の共分散

**定理 1.7.6.**  $\text{Cov} : L^2(\Omega) \times L^2(\Omega) \rightarrow \mathbb{R}$  について,

- (1) 対称性 :  $\text{Cov}[X, Y] = \text{Cov}[Y, X]$ .
- (2) 双線型性 :  $\text{Cov}[aX + bY, Z] = a\text{Cov}[X, Z] + b\text{Cov}[Y, Z]$ .
- (3) 定数で消える :  $\text{Cov}[X, 1] = 0$ . 特に,  $\text{Cov}[aX + b, Y] = a\text{Cov}[X, Y]$ .
- (4) 共分散公式 :  $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$ .

## 1.7.5 可分 Hilbert 空間上の平均

平均とはその確率測度に関する積分である。確率変数が可分な実 Hilbert 空間値になったとき、これは Pettis 積分に対応する。 $\mathbb{R}^n$  の分布は任意の線型汎関数による  $\mathbb{R}$  への一次元投影によって特徴付けられる (Cramer-Wold) ことは、弱可測性と Pettis 積分の結果と一致する。

**命題 1.7.7 (Pettis integral).**  $H$  を可分 Hilbert 空間,  $X : \Omega \rightarrow H$  を弱可測関数とする。このとき,  $\|X\| : \Omega \rightarrow \mathbb{R}$  が可積分ならば, ただ一つの元  $E[X] \in H$  が存在して,

$$\forall x \in H \quad (E[X]|x) = \int_H (X(\omega)|x) \mu(d\omega).$$

**定義 1.7.8 (平均).**  $H$  を可分 Hilbert 空間とし,  $\mu \in P(H)$  をその上の Borel 確率測度とする。  $\mu$  の平均  $m \in H$  とは, 恒等関数  $X = \text{id}_H$  の Pettis 積分とする:

$$\forall x \in H \quad \langle m, x \rangle = \int_H \langle x, y \rangle \mu(dy).$$

**要諦 1.7.9.** すなわち, 平均の各  $x$ -成分は,  $x$ -成分の平均に等しい。平均の定める線型汎関数  $F(x) = \int_H (x|y) \mu(dy) = (x|m)$  は, ランダムな  $y \in H$  と内積を取ったときの期待値は  $m$  と内積を取ることと等しい, という可換性を表していると思える。

## 1.7.6 可分 Hilbert 空間上の分散

$\mathbb{R}^n$  の分散は  $n \times n$  行列になるが, 要は  $B(H)$  の元である。分散公式は  $\text{Tr} Q = E[|x|^2] - |m|^2$  という形になる。共分散は内積とのアナロジーで考えると極めてわかりやすい。

$\text{Cov} : L^2(\Omega; \mathbb{R}^r) \times L^2(\Omega; \mathbb{R}^c) \rightarrow M_{r,c}(\mathbb{R})$  は  $\text{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])^\top]$  で定まり, まるで「中心化された, 内積の逆」のようなものである。

- (1) エルミート対称性 :  $\text{Cov}[X, Y] = \text{Cov}[Y, X]^\top$ .
- (2) 双線型性 :  $\text{Cov}[aX + bY, Z] = a\text{Cov}[X, Z] + b\text{Cov}[Y, Z]$ .
- (3) 2つ合わせる :  $\text{Cov}[AX, Y] = A\text{Cov}[X, Y]$ . よって  $\text{Cov}[X, BY] = \text{Cov}[X, Y]B^\top$ .
- (4) 定数で消える :  $\text{Cov}[X, I] = 0$ .
- (5) 共分散公式 :  $\text{Cov}[X, Y] = E[XY^\top] - E[X]E[Y]^\top$ .
- (6) 共分散公式を一般化すると, 一般の確率ベクトルの 2 次形式の平均についての表示を得る:

$$E[X^\top A Y] = \text{Tr}(\text{Cov}[X, Y]^\top A) + E[X]^\top A E[Y].$$

$\text{Var}[X] := \text{Cov}[X, X] : L^2(\Omega; \mathbb{R}^r) \rightarrow M_r(\mathbb{R})$  を分散共分散行列という。半正定値で (非負の固有値をもち), 直交系となる固有ベクトル系を持つ。固有値は固有ベクトルの Rayleigh 商として表せる:  $\lambda_i = \frac{\|A v_i\|^2}{\|v_i\|^2} \geq 0$ .

**定義 1.7.10.** 双線型写像  $G : H \times H \rightarrow \mathbb{R}$  を表現する有界作用素  $Q \in B(H)$  を分散という:

$$G(x, y) = \int_H (x|z - m)(y|z - m) \mu(dz) = (Qx|y).$$

**命題 1.7.11** .  $E[|x|^2] < \infty$  とする.

- (1) 共分散  $Q$  は正作用素 (半正定値) であり, かつ, 対称である:  $(Qx|y) = (x|Qy)$ .
- (2) 跡が 2 次の中心化モーメントに等しく,  $\text{Tr}Q < \infty$  である (普段見る分散とはこれである).
- (3) また 2 次のモーメントは  $E[|x|^2] = \text{Tr}Q + |m|^2$  と表わせ, これを共分散公式という.
- (4)  $Q$  はコンパクト作用素である.

平均は Pettis 積分に直感的な理解を与え, 分散は跡に直感的な理解を与える.

### 1.7.7 Chebyshev の不等式

分布の集中性に対して, 線型な評価を与える.

**定理 1.7.12** .  $\psi \in \mathcal{L}(\mathbb{R})_+$  を非負な Borel 可測関数とする.

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

[証明].

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

が成り立つ. 両辺の期待値を取れば良い. ■

**系 1.7.13** .

- (1) (Markov) 特に  $\psi(x) = |x|^p$  ( $p > 0$ ),  $A := \{|x| \geq a\}$  とすると, 次が成り立つ:

$$\forall a > 0 \quad P[|X| \geq a] \leq \frac{1}{a^p} E[|X|^p].$$

- (2) (Chebyshev)  $X \in L^2(\Omega)$  とし, この標準偏差を  $\sigma(X) := \sqrt{\text{Var}[X]}$  と表す.

$$\forall a > 0 \quad P[|X - EX| \geq a] \leq \frac{\sigma^2}{a^2}.$$

- (3) (Bienaymé)  $f: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  を単調増加関数,  $X, f(X) \in L^1(\Omega)$  とする.

$$\forall a > 0 \quad P[|X| \geq a] \leq \frac{1}{f(a)} E[f(|X|)].$$

[証明].

- (1)

$$E[|X|^p] \geq E[|X|^p 1_{|X| \geq a}] \geq a^p P[|X| \geq a].$$

- (2)  $\sigma(X) = 0$  のとき  $\sigma(X) = 0 \Leftrightarrow X(\omega) = EX$  a.s. であるから, 上の不等式は当然成り立つ.

$\sigma(X) \neq 0$  のとき 求める事象を  $A := \{\omega \in \Omega \mid |X(\omega) - EX| > a\sigma(X)\}$  とおくと,

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

と評価できる.

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

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

**要諦 1.7.14** (平均値から標準偏差の  $a$  倍以上離れる確率は  $\frac{1}{a^2}$  以下である. ). まさかそんなに当然な評価の変形だったのか.



## 1.7.8 Jensen の不等式

**定義 1.7.15 (convex function).** 凸集合  $C \subset \mathbb{R}^p$  上の, エピグラフ (上部分) が凸集合であるような関数  $f: C \rightarrow \mathbb{R}$  は凸であるという. すなわち, 任意の  $[x, y] \subset C$  について,

$$\forall t \in [0, 1] \quad \varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y).$$

**命題 1.7.16 (Jensen の不等式の一般化).**  $(a, b)$  ( $a < b \in \mathbb{R}$ ) 上の凸関数  $\varphi: (a, b) \rightarrow \mathbb{R}$  について, 次が成り立つ:

- (1)  $\varphi$  は連続で, 特に局所 Lipschitz である.  
 (2) 凸関数の特徴付け: ある可算個の  $\{(p_\alpha, q_\alpha)\}_{\alpha \in A} \subset \mathbb{R}^2$  が存在して,

$$\varphi(x) = \sup_{\alpha \in A} (p_\alpha x + q_\alpha), \quad x \in (a, b).$$

- (3) 任意の  $X \in L^1(\Omega; [a, b])$  に対して,  $E[\varphi(X)] \in \overline{\mathbb{R}}$  は確定し,

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

**[証明].**

- (1) Step1 任意の  $x < y < z \in (a, b)$  に対して,

$$\frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(z) - \varphi(y)}{z - y}.$$

が成り立つ. 実際, この式を同値変形していくと,

$$(\varphi(y) - \varphi(x))(z - y) \leq (\varphi(z) - \varphi(y))(y - x).$$

$$\varphi(y)(z - x) \leq \varphi(x)(z - y) + \varphi(z)(y - x).$$

$$\varphi(y) \leq \underbrace{\frac{z - y}{z - x}}_{=:t} \varphi(x) + \underbrace{\frac{y - x}{z - x}}_{=:1-t} \varphi(z).$$

であるが, いま

$$tx + (1-t)z = \frac{zx - yx + yz - xz}{z - x} = y$$

より, これは成立.

Step2 よって, 任意の  $x_0 \in (a, b)$  に対して, 任意の

$$p_0 \in \left[ \sup_{a < x \leq x_0} \frac{\varphi(x_0) - \varphi(x)}{x_0 - x}, \inf_{x_0 \leq x < b} \frac{\varphi(x) - \varphi(x_0)}{x - x_0} \right] \neq \emptyset.$$

について,

$$\varphi(x) \geq p_0(x - x_0) + \varphi(x_0) = p_0x + \underbrace{(\varphi(x_0) - p_0x_0)}_{=:q_0}.$$

で,  $x = x_0$  にて等号が成立する. よって, 任意の  $[c, d] \subset (a, b)$  について,

$$\|\varphi|_{[c,d]}\| := \max \left\{ \inf_{c < x} \frac{\varphi(x) - \varphi(c)}{x - c}, \sup_{x < d} \frac{\varphi(d) - \varphi(x)}{d - x} \right\}$$

を Lipschitz 係数として Lipschitz 連続である.

- (2) (1) の Step2 より, すべての  $x \in (a, b) \cap \mathbb{Q}$  にて  $(p_0, q_0)$  を考えればよい.  
 (3) まず  $E[X] \in (a, b)$  より  $\varphi(E[X])$  が定まっていることに注意すれば, 次のように計算できる:

$$\begin{aligned} \varphi(E[X]) &= \sup_{\alpha \in A} (p_\alpha E[X] + q_\alpha) = \sup_{\alpha \in A} (E[p_\alpha X + q_\alpha]) \\ &\leq \sup_{\alpha \in A} E[\sup_{\alpha \in A} (p_\alpha X + q_\alpha)] = \sup_{\alpha \in A} E[\varphi(X)] = E[\varphi(X)]. \end{aligned}$$

■

## 1.8 条件付け

$\sigma$ -代数が系に対する情報だと解釈出来た。これに対する解析法として最も重要なものが「条件付け」である。

### 1.8.1 条件付き期待値の定義

**定義 1.8.1 (conditional expectation, conditional probability, regular).**  $(\Omega, \mathcal{F}, P)$  を確率空間とし,  $\mathcal{G}$  を  $\mathcal{F}$  の部分  $\sigma$ -代数とする. 可積分確率変数  $X \in L^1(\Omega)$  について, 次の 2 条件を満たす,  $P$ -零集合を除いて一意な確率変数を条件付き期待値といい,  $E[X|\mathcal{G}]$  で表す.

- (1)  $\mathcal{G}$ -可測でもある  $P$ -可積分確率変数である.
- (2) 任意の  $\mathcal{G}$ -可測集合  $B \in \mathcal{G}$  上では  $X$  と期待値が同じ確率変数になる:  $\forall B \in \mathcal{G} \quad E[X1_B] = E[E[X|\mathcal{G}]1_B]$ .<sup>†8</sup>

**命題 1.8.2.**  $E[-|\mathcal{G}] : L^1(\Omega) \rightarrow L^1(\Omega)$  はノルム減少的で正な線型汎作用素である. すなわち,

- (1)  $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$  a.s.
- (2)  $X \leq Y$  a.s.  $\Rightarrow E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$  a.s.
- (3)  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  を凸関数とする.  $\varphi(X) \in L^1(\Omega)$  ならば,  $\varphi(E[X|\mathcal{G}]) \leq E[\varphi(X)|\mathcal{G}]$  a.s.
- (4)  $|E[X|\mathcal{G}]| \leq E[|X||\mathcal{G}]$  a.s.

いずれも  $L^1_{\mathcal{G}}(\Omega)$  上の等式・不等式であることに注意.

**[証明].**

- (2)  $X' := E[X|\mathcal{G}]$  とおく.  $A_n := \{X' \leq 1/n\} \in \mathcal{G}$  について条件付き期待値の定義から

$$0 \leq E[X, A_n] = E[X', A_n] \leq -\frac{1}{n}P[A_n].$$

より,  $P[A_n] = 0$  が必要. これより,

$$P[X' < 0] = P[\cup_{n=1}^{\infty} A_n] = 0.$$

が解る.

- (3)  $X \in L^1(\Omega)$  は単関数で,

$$X = \sum_{i=1}^N a_i 1_{A_i}, \quad \Omega = \sqcup_{i=1}^N A_i.$$

と表せるとき, (2) から,  $p_i := E[1_{A_i}|\mathcal{G}] \geq 0$  a.s. で,

$$\begin{aligned} \sum_{i=1}^N p_i &= E\left[\sum_{i=1}^N 1_{A_i} \middle| \mathcal{G}\right] \text{ a.s.} \\ &= E[1|\mathcal{G}] = 1 \text{ a.s.} \end{aligned}$$

より, 凸性から,

$$\begin{aligned} \varphi(E[X|\mathcal{G}]) &= \varphi\left(E\left[\sum_{i=1}^N a_i 1_{A_i} \middle| \mathcal{G}\right]\right) \\ &= \varphi\left(\sum_{i=1}^N a_i p_i\right) \text{ a.s.} \\ &\leq \sum_{i=1}^N \varphi(a_i) p_i \text{ a.s.} \end{aligned}$$

一方で,

$$E[\varphi(X)|\mathcal{G}] = E\left[\sum_{i=1}^N \varphi(a_i) 1_{A_i} \middle| \mathcal{G}\right]$$

<sup>†8</sup> これは 2 段階に分けて積分していると見れる.

$$\begin{aligned}
&= \sum_{i=1}^N \varphi(a_i) E[1_{A_i} | \mathcal{G}] \quad \text{a.s.} \\
&= \sum_{i=1}^N \varphi(a_i) p_i.
\end{aligned}$$

から, 結論を得る.

(4)  $\varphi(x) = |x|$  と取ればよい.

**系 1.8.3 (条件付き期待値の連続性).** 任意の  $\sigma$ -代数  $\mathcal{G} \subset \mathcal{F}$  について,

- (1) 条件付き期待値  $E[-|\mathcal{G}] : L^p(\Omega) \rightarrow L^p_{\mathcal{G}}(\Omega)$  は任意の  $p \in [1, \infty]$  についてノルム減少的.
- (2)  $X_n \rightarrow X$  in  $L^p(\Omega)$  ならば,  $E[X_n | \mathcal{G}] \rightarrow E[X | \mathcal{G}]$  in  $L^p(\Omega)$  である.

**[証明].** 凸関数  $f(x) = |x|^p$  について,  $|E_{\mathcal{G}}[X]|^p \leq E_{\mathcal{G}}[|X|^p]$  より,

$$E[|E_{\mathcal{G}}[X]|^p] \leq E[E_{\mathcal{G}}[|X|^p]] = E[|X|^p] < \infty$$

だから, たしかに像は  $L^p$  に入り,  $\|E_{\mathcal{G}}[X]\|_p \leq \|X\|_p$  である. 全射性は, 任意の  $X \in \mathcal{L}^p_{\mathcal{G}}$  について,  $E_{\mathcal{G}}[X] = X$  より.

**命題 1.8.4 (Lebesgue の優収束定理).**  $X_n \rightarrow X$  a.s. かつ  $\{X_n\} \subset L^1(\Omega)$  は可積分な優関数を持つとする. このとき,  $\{E[X_n | \mathcal{G}]\} \subset L^1_{\mathcal{G}}(\Omega)$  は  $E[X | \mathcal{G}]$  に概収束かつ  $L^1$ -収束する.

**要諦 1.8.5.** 条件付き期待値には, 引き続き Markov の不等式, Jensen の不等式, Holder の不等式が成り立つ.

## 1.8.2 射影としての条件付き期待値

**定理 1.8.6 (条件付き期待値の射影としての特徴付け).** 部分  $\sigma$ -代数  $\mathcal{G} \subset \mathcal{F}$  と  $Y \in L^2(\Omega)$  を考える. 任意の  $\hat{Y}_{\mathcal{G}} \in L^2_{\mathcal{G}}(\Omega)$  について, 次は同値:

- (1)  $\|Y - \hat{Y}_{\mathcal{G}}\|_{L^2(\Omega)} = \inf_{Y' \in L^2_{\mathcal{G}}(\Omega)} \|Y - Y'\|_{L^2(\Omega)}.$
- (2)  $\forall Z \in L^2_{\mathcal{G}}(\Omega) \quad E[ZY] = E[Z\hat{Y}_{\mathcal{G}}].$

**[証明].**

(1) $\Rightarrow$ (2) 任意の  $Z \in L^2(\mathcal{G})$  を取る. すると任意の  $a \in \mathbb{R}$  について,  $\hat{Y}_{\mathcal{G}} + aZ \in L^2_{\mathcal{G}}(\Omega)$  であるから, 最小性より

$$\begin{aligned}
\|Y - (\hat{Y}_{\mathcal{G}} + aZ)\|_2^2 - \|Y - \hat{Y}_{\mathcal{G}}\|_2^2 &= a^2 E[Z^2] - 2a E[Z(Y - \hat{Y}_{\mathcal{G}})] \\
&= E[Z^2] \left( a - \frac{E[Z(Y - \hat{Y}_{\mathcal{G}})]}{E[Z^2]} \right)^2 - \frac{EZ[Y - \hat{Y}_{\mathcal{G}}]^2}{E[Z^2]} \geq 0.
\end{aligned}$$

左辺は  $a = \frac{E[Z(Y - \hat{Y}_{\mathcal{G}})]}{E[Z^2]}$  のとき, 最小値  $-\frac{E[Z(Y - \hat{Y}_{\mathcal{G}})]^2}{E[Z^2]} \leq 0$  を取るから, 特に  $E[Z(Y - \hat{Y}_{\mathcal{G}})] = 0$  が必要である.

(2) $\Rightarrow$ (1)  $E[Y\hat{Y}_{\mathcal{G}}] = E[\hat{Y}_{\mathcal{G}}^2]$  に注意すれば, 任意の  $Z \in L^2_{\mathcal{G}}(\Omega)$  について,

$$\begin{aligned}
\|Y - Z\|_2^2 &= E[Y^2 - 2YZ + Z^2] = E[Y^2] - 2E[\hat{Y}_{\mathcal{G}}Z] + E[Z^2] \\
&= E[Y^2] - 2E[Y\hat{Y}_{\mathcal{G}}] + 2E[\hat{Y}_{\mathcal{G}}^2] - 2[\hat{Y}_{\mathcal{G}}Z] + E[Z^2] \\
&= \|Y - \hat{Y}_{\mathcal{G}}\|_2^2 + \|\hat{Y}_{\mathcal{G}} - Z\|_2^2 \geq \|Y - \hat{Y}_{\mathcal{G}}\|_2^2.
\end{aligned}$$

**系 1.8.7 .** 条件付き期待値  $E[-|\mathcal{G}]$  は

- (1) 線型である.
- (2) 正値性を持つ.

(3) Pythagoras の関係:  $\|Y\|_2^2 = \|Y - E[Y|\mathcal{G}]\|_2^2 + \|E[Y|\mathcal{G}]\|_2^2$ .

(3) は次と同値で, **全分散の公式**ともいう:

$$\text{Var}[Y] = \text{Var}[Y - E[Y|\mathcal{G}]] + \text{Var}[E[Y|\mathcal{G}]], \quad \text{Var}[Y|\mathcal{G}] := E[(Y - E[Y|\mathcal{G}])^2|\mathcal{G}].$$

### 1.8.3 条件付き期待値の延長

**定理 1.8.8**. 部分  $\sigma$ -代数  $\mathcal{G} \subset \mathcal{F}$  について,

(1) 任意の  $Y \in L(\Omega)_+$  に対して, 次を満たす  $E[Y|\mathcal{G}] \in L_{\mathcal{G}}(\Omega)_+$  が唯一存在する:

$$\forall Z \in L_{\mathcal{G}}(\Omega)_+ \quad E[ZY] = E[ZE[Y|\mathcal{G}]].$$

特に,  $Y \in L^2(\Omega)$  ならば,  $E[Y|\mathcal{G}]$  は  $L_{\mathcal{G}}^2(\Omega)$  への射影である.

(2) 任意の  $Y \in L^1(\Omega)$  に対して, 次を満たす  $E[Y|\mathcal{G}] \in L_{\mathcal{G}}^1(\Omega)$  が唯一存在する:

$$\forall Z \in L_{\mathcal{G}}^{\infty}(\Omega) \quad E[ZY] = E[ZE[Y|\mathcal{G}]].$$

特に,  $Y \in L^2(\Omega)$  ならば,  $E[Y|\mathcal{G}]$  は  $L_{\mathcal{G}}^2(\Omega)$  への射影である.

### 1.8.4 条件付き期待値の代数的性質

条件付き期待値の延長定理 1.8.8 の条件  $Z \in L_{\mathcal{G}}^{\infty}(\Omega)$  を満たすには, その単関数近似と単調収束定理より,  $1_A$  ( $A \in \mathcal{G}$ ) のみについて要請すれば十分, ということである.

**命題 1.8.9** (可測関数の取り出し).  $X \in L^1(\Omega), Y \in L_{\mathcal{G}}^1(\Omega)$  について,

(1)  $XY \in L^1(\Omega)$  ならば,  $E[XY|\mathcal{G}] = YE[X|\mathcal{G}]$  a.s.

(2) 特に,  $E[Y|\mathcal{G}] = Y$  a.s.

**命題 1.8.10** (独立確率変数に対する性質).  $X \in L^1(\Omega)$  は  $\mathcal{G}$  と独立とする.

(1)  $E[X|\mathcal{G}] = E[X]$  a.s.

(2) 特に,  $E[X|2] = E[X]$  a.s..

**命題 1.8.11** (tower property).  $\mathcal{G}_1 \subset \mathcal{G}_2$  ならば,  $E_{\mathcal{G}_1} = E_{\mathcal{G}_1} \circ E_{\mathcal{G}_2}$ . すなわち,

$$E[X|\mathcal{G}_1] = E[E[X|\mathcal{G}_2]|\mathcal{G}_1] \text{ a.s.}$$

**[証明]**. 右辺を  $Z$  とおく. 任意の  $A \in \mathcal{G}_1$  について,  $A \in \mathcal{G}_2$  でもあるから,

$$E[Z, A] = E[E[X|\mathcal{G}_2], A] = E[X, A].$$

■

**命題 1.8.12** (条件付き期待値の相等).  $\mathcal{G}, \mathcal{H} \subset \mathcal{F}, X, Y \in L^1(\Omega)$  について, ある  $A \in \mathcal{G} \cap \mathcal{H}$  が存在して,

[A1]  $\{A \cap G\}_{G \in \mathcal{G}} = \{A \cap H\}_{H \in \mathcal{H}}$ .

[A2]  $X = Y$  on  $A$ .

を満たすとする. このとき,

$$E[X|\mathcal{G}] = E[Y|\mathcal{H}] \text{ on } A.$$

## 1.8.5 確率変数に関する条件付け

$\sigma[X]$  を与えた下での条件付き期待値と、確率変数  $X$  を与えた下での条件付き期待値とは等しい。

**補題 1.8.13.**  $S$ -値確率変数  $X$  と実確率変数  $Y$  について、次は同値：

- (1)  $Y$  は  $\sigma[X]$ -可測.
- (2) ある Borel 関数  $f: S \rightarrow \mathbb{R}$  が存在して、 $Y = f(X)$  である.

**定義 1.8.14.** 確率変数  $X \in L(\Omega; S)$  による  $Y \in L^1(\Omega)$  の条件付き期待値は、次を満たす可測関数  $E[Y|X = -]: S \rightarrow \mathbb{R}$  とする：

$$\forall B \in \mathcal{B}(S) \quad \int_{X^{-1}(B)} Y(\omega) P(d\omega) = \int_B E[Y|X = x] P^X(dx).$$

**命題 1.8.15.**  $\sigma[X]$  を与えた下での条件付き期待値と、確率変数  $X$  を与えた下での条件付き期待値とは等しい：

$$E[Y|X](\omega) = E[Y|X = x]|_{x=X(\omega)} \quad \text{a.s.}$$

## 1.8.6 可算生成の場合の条件付き確率

一口に条件付き確率と言っても、事象を与えた下でのものと、 $\sigma$ -代数を与えた下でのものとは違う。後者は各  $\omega \in \Omega$  に対して、それが属する  $A_i \in \Delta$  を対応させ、この事象を与えた下での条件付き確率  $P[-|A_i](\omega)$  を返す写像となっている。

**定義 1.8.16.**  $\Omega$  の可算分割  $\Delta = \{A_i\}_{i \in \mathbb{N}} \subset \mathcal{F} \setminus \mathcal{N}$  に対して、事象  $A_i$  の下での  $B$  の条件付き確率を

$$P[B|A_i] := \frac{P[B \cap A_i]}{P[A_i]}, \quad (B \in \mathcal{F})$$

と定める。これは、 $A_i$  が起こったと知った際の  $B$  の確率であり、確率空間 (全事象) の移行  $\Omega \mapsto A_i$  を意味する。

**要諦 1.8.17** (事象を与えた下での条件付き確率の意味)。これを、

$$P[B|\sigma[A_i]](\omega) := \frac{P[B \cap A_i]}{P[A_i]} 1_{A_i}(\omega) + \left(1 - \frac{P[B \cap A_i]}{P[A_i]}\right) 1_{A_i^c}(\omega), \quad (\omega \in \Omega)$$

の  $A_i$  上での値と見る。

**命題 1.8.18.**  $\Omega$  の可算分割  $\Delta = \{A_i\}_{i \in \mathbb{N}} \subset \mathcal{F} \setminus \mathcal{N}$  に対して、 $\sigma$ -代数  $\sigma[\Delta]$  が与えられた下での事象  $B \in \mathcal{F}$  の条件付き確率は、

$$P[B|\Delta](\omega) = \sum_{i \in \mathbb{N}} P[B|A_i] 1_{A_i}(\omega) = \sum_{i \in \mathbb{N}} \frac{P[B \cap A_i]}{P[A_i]} 1_{A_i}(\omega), \quad (\omega \in \Omega)$$

に等しい。

**要諦 1.8.19** ( $\sigma$ -代数を与えた下での条件付き確率の意味)。

- (1) これは確率変数としては各  $A_i \in \Delta$  上定値であり、任意の  $B \in \mathcal{F}$  に対して、各  $\omega \in A_i$  上での  $A_i$  内での存在感/含有割合  $\frac{P[B \cap A_i]}{P[A_i]}$  を返す。
- (2)  $\omega \in A_i$  を固定すると、単に事象  $A_i$  を与えた下での条件付き確率  $P[-|A_i](\omega)$  である。

## 1.8.7 正則条件付き確率の定義

本来は  $\mu_\Delta(-, -) : \Delta \times \mathcal{F} \rightarrow [0, 1]$  として,  $\mu_\Delta(G_i, A) = P[A|G_i]$  としたいが, 一般の  $\sigma$ -代数  $\mathcal{G}$  に対して, これを生成する分割  $\Delta$  の元  $G_i \in \Delta$  を正確に指定する代わりに, 元  $\omega \in \Omega$  を通して行うので,  $(\Omega, \mathcal{G}) \times \mathcal{F}$  上の関数, ということにして定義する. この点による可測性の問題が生じて議論が煩雑さを持つ. これをケアするのが, 条件付き期待値についての知識 (2) である.  $\mathcal{G}$  の元  $B \in \mathcal{G}$  上での振る舞いを規定することで達成する.

観察 1.8.20 (問題の所在).

- (1)  $P[A|\mathcal{G}] := E[1_A|\mathcal{G}]$  ( $A \in \mathcal{F}$ ) によって定まる  $L^1_{\mathcal{G}}(\Omega)$ -確率変数の同値類への対応  $P[-|\mathcal{G}] : \mathcal{F} \rightarrow L^1_{\mathcal{G}}(\Omega)$  を  $\mathcal{G}$  の定める条件付き確率と呼びたい.
- (2)  $\omega \in \Omega$  を固定した際に確率測度を定めるとは限らない. これが確率測度を定める (ような修正が存在する) とき, 正則条件付き確率という.
- (3)  $\mathcal{G}$  が  $\Omega$  の可算な分割から生成される場合, 任意の  $\omega \in \Omega$  に対して  $A \mapsto P[A|\mathcal{G}](\omega)$  は確率測度を定める.
- (4) 問題の所在: 任意の互いに素な可測集合列  $\{F_n\} \subset \mathcal{F}$  について, 条件付き期待値の線形性と単調収束定理より,

$$P\left[\sum F_n \middle| \mathcal{G}\right] = E\left[\sum 1_{F_n} \middle| \mathcal{G}\right] = \sum E[1_{F_n}|\mathcal{G}] = \sum P[F_n|\mathcal{G}] \text{ a.s.}$$

が成り立つが, このときの零集合

$$\mathcal{N} := \left\{ \omega \in \Omega \mid P\left[\sum F_n \middle| \mathcal{G}\right] \neq \sum P[F_n|\mathcal{G}] \right\}$$

が, 任意の (おそらく非可算無限個ある) 互いに素な可測集合列  $\{F_n\} \subset \mathcal{F}$  について, 一様に零集合を取れるとは限らないが, 「標準確率空間」については気にしなくてよい.

□

定義 1.8.21.  $\mu_{\mathcal{G}}(-, -) : \Omega \times \mathcal{F} \rightarrow \overline{\mathbb{R}_+}$  を,  $\mathcal{G}$  を与えた下での正則条件付き確率という:

- (1) 任意の  $A \in \mathcal{F}$  について,  $\mu_{\mathcal{G}}(-, A)$  は  $\mathcal{G}$ -可測.
- (2) 任意の  $A \in \mathcal{F}, B \in \mathcal{G}$  に対して,  $P[A \cap B] = \int_B \mu_{\mathcal{G}}(\omega, A) P(d\omega)$ .
- (3) 任意の  $\omega \in \Omega$  について,  $\mu_{\mathcal{G}}(\omega, -)$  は  $(\Omega, \mathcal{F})$  上の確率測度.

要諦 1.8.22. (1)/(2) は単に条件付き期待値としての要件であり, (3) が測度を定めるための追加の要件である.

- (1)  $\mu_{\mathcal{G}}(-, A) : \Omega \rightarrow [0, 1]$  は,  $\omega \in \Omega$  という根元事象が起こったという事実を  $\mathcal{G}$  の粒度でしか知らなかった場合の  $A$  の条件付き確率を表す.
- (2)  $\mu_{\mathcal{G}}(\omega, -) : \mathcal{F} \rightarrow [0, 1]$  は,  $A \in \mathcal{F}$  の条件付き確率を, 各  $\omega \in \Omega$  に対応する  $\mathcal{G}$  の元で条件付けた際の条件付き確率である.

定理 1.8.23 (正則条件付き確率の存在). 標準 Borel 空間  $(S, \mathcal{B}(S), P)$  について, 任意の  $\mathcal{G} \subset \mathcal{F}$  に対して, ある零集合  $N \in \mathcal{G}$  が存在して,  $A \mapsto \mu_{\mathcal{G}}(\omega, A)$  は  $N^c$  上で一意に定まる.

## 1.8.8 確率変数による正則条件付き確率

記号としては  $\mu_{\sigma[X]} = \mu^{Y|X}$  と対応する.

定義 1.8.24.  $X \in L(\Omega; S_X), Y \in L(\Omega; S_Y)$  について, 次を満たす  $\mu^{Y|X}(-, -) : S_X \times \mathcal{B}(S_Y) \rightarrow \overline{\mathbb{R}_+}$  を  $X$  を与えた下での正則条件付き分布という:

- (1) 任意の  $A \in \mathcal{B}(S_Y)$  について,  $\mu^{Y|X}(-, A)$  は  $\mathcal{B}(S_X)$ -可測.
- (2) 任意の  $A \in \mathcal{B}(S_Y), B \in \mathcal{B}(S_X)$  に対して,  $P[A \cap B] = \int_B \mu^{Y|X}(x, A) P^X(dx)$ .
- (3) 任意の  $x \in S_X$  について,  $\mu^{Y|X}(x, -)$  は  $(S_Y, \mathcal{B}(S_Y))$  上の確率測度.

特に,  $X \perp\!\!\!\perp Y$  のとき,  $\mu^{Y|X}(x, dy) = P^Y(dy)$ .

**要諦 1.8.25.** 条件 (2) は, 各  $B \in \mathcal{G}(S_X)$  上での積分を通じて見れば,  $\mu^{Y|X}(-, A)$  は, 素朴な意味での条件付き確率  $P[A|\sigma[X]]$  を模倣していることを意味する.

**定理 1.8.26.** 完備可分距離空間  $S_X$  上の確率空間  $(S_X, \mathcal{G}(S_X), P)$  には,  $Y$  を与えた下での正則条件付き確率  $\mu^{Y|X}$  が存在して, ある零集合  $N \in \mathcal{G}(S_X)$  が存在して,  $A \mapsto \mu^{Y|X}(\omega, A)$  は  $N^c$  上で一意に定まる.

### 1.8.9 確率測度の分解

**定理 1.8.27.** 完備可分距離空間  $S_X$  上の確率空間  $(S_X, \mathcal{G}(S_X), P)$  が  $\mathcal{G}(S_Y) = \mathcal{G}(S_X)$  を満たすとし,  $Z \in L_{\sigma[X]}(\Omega; S_Z)$ ,  $f \in L^1(S_Y \times S_Z)$  とする. このとき,

(1)

$$E[f(Y, Z)|X](\omega) = \int f(y, Z(\omega)) \mu^{Y|X}(X(\omega), dy) \quad \text{a.s.}$$

(2) 特に,  $X \perp\!\!\!\perp Y$  ならば,  $g(z) := E[f(Y, z)]$  に関して,

$$E[f(Y, Z)|X](\omega) = g(Z(\omega)) \quad \text{a.s.}$$

(3) 特に, ある可測関数  $h \in L(S_X; S_Z)$  に対して,

$$E[f(Y, Z)|X](\omega) = (g \circ h)(X(\omega)) \quad \text{a.s.}$$

**例 1.8.28** (条件付き密度関数).  $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$  が参照測度  $\mu \otimes \nu$  に関して密度  $f(x, y)$  を持つとする.  $X$  の周辺密度を  $f(x) := \int_{\mathbb{R}} f(x, y) \nu(dy)$  で表す.

$$f(y|x) := \begin{cases} \frac{f(x, y)}{f(x)} & f(x) > 0, \\ 0 & f(x) = 0. \end{cases}$$

と定めると,

$$E[Y|X = x] := \int_{\mathbb{R}} y f(y|x) \nu(dy) \quad P^X\text{-a.e. } x$$

は  $X$  を与えた下での  $Y$  の条件付き期待値の定義を満たす. □

### 1.8.10 条件付き独立性

独立性と条件付き独立性との間に包含関係はない.

**定義 1.8.29.**

(1) 事象  $A, B, C \in \mathcal{F}$ ,  $P[C] > 0$  について,  $A \perp\!\!\!\perp B | C$  であるとは,

$$P[A \cap B|C] = P[A|C]P[B|C]$$

を満たすことをいう. これは  $P[B|A \cap C] = P[B|C]$  に同値. これは「 $C$  の情報さえ与えられたならば,  $A$  による追加情報は  $B$  の条件付き確率を変えない」と見做せる.

(2)  $\mathcal{G} \subset \mathcal{F}$  を部分  $\sigma$ -代数とする.  $\mathcal{G}_1, \dots, \mathcal{G}_n$  は  $\mathcal{G}$ -条件付き独立であるとは,

$$\forall_{A_k \in \mathcal{G}_k} P \left[ \bigcap_{k \in [n]} A_k \middle| \mathcal{G} \right] = \prod_{k \in [n]} P[A_k | \mathcal{G}] \quad \text{a.s.}$$

を満たすことをいう.

(3) 各  $\mathcal{G}_\lambda \subset \mathcal{F}$  を部分  $\sigma$ -代数とする. 任意の有限部分集合  $\lambda_1, \dots, \lambda_m \in \Lambda$  について,  $\mathcal{G}_{\lambda_1}, \dots, \mathcal{G}_{\lambda_m}$  が  $\mathcal{G}$ -条件付き独立であるとき, 族  $(\mathcal{G}_\lambda)_{\lambda \in \Lambda}$  は  $\mathcal{G}$ -条件付き独立であるという.



特に、 $\mathcal{G} = 2$  のとき、 $\mathcal{G}$ -条件付き独立性は単なる独立性に一致する。

**定理 1.8.30** . 部分  $\sigma$ -代数  $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$  について、次は同値：

- (1)  $\mathcal{G} \perp \mathcal{H} \mid \mathcal{G}$ .
- (2)  $\forall H \in \mathcal{H} \quad P[H|\mathcal{G} \vee \mathcal{G}] = P[H|\mathcal{G}] \text{ a.s.}$

**定理 1.8.31 (条件付き独立性の連鎖律)**. 部分  $\sigma$ -代数  $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{H} \subset \mathcal{F}$  について、次は同値：

- (1)  $\mathcal{H} \perp (\mathcal{G}_1, \mathcal{G}_2, \dots) \mid \mathcal{G}$ .
- (2)  $\forall m \in \mathbb{N} \quad \mathcal{H} \perp \mathcal{G}_{m+1} \mid (\mathcal{G}, \mathcal{G}_1, \dots, \mathcal{G}_m)$ .

**系 1.8.32** . 部分  $\sigma$ -代数  $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2, \mathcal{H} \subset \mathcal{F}$  について、次は同値：

- (1)  $\mathcal{H} \perp (\mathcal{G}_1, \mathcal{G}_2) \mid \mathcal{G}$ .
- (2)  $\mathcal{H} \perp \mathcal{G}_1 \mid \mathcal{G} \wedge \mathcal{H} \perp \mathcal{G}_2 \mid (\mathcal{G}, \mathcal{G}_1)$ .

### 1.8.11 偏相関係数

線型予測子  $X$  に関して、 $Y_1 \perp Y_2 \mid \sigma[X]$  が達成されているかを調べる手法をグラフィカルモデリングと総称する。正規性の仮定の下で、条件付き相関を調べることが常套手段である。

**記法 1.8.33**. 偏相関係数  $\text{Corr}[Y_1, Y_2|X]$  はしばしば  $\rho_{Y_1, Y_2|X}$  で表される。

**定義 1.8.34** .  $Y_1, Y_2 \in L^2(\Omega), \mathcal{G} \subset \mathcal{F}$  について、

- (1)  $\epsilon_j := Y_j - E[Y_j|\mathcal{G}] \ (j = 1, 2)$ .
- (2)  $\text{Corr}(Y_1, Y_2|\mathcal{G}) := \text{Corr}(\epsilon_1, \epsilon_2) = \frac{\text{Cov}[\epsilon_1, \epsilon_2]}{\sqrt{\text{Var}[\epsilon_1]\text{Var}[\epsilon_2]}}$ .
- (3)  $\text{Var}[Y_j|\mathcal{G}] := E[(Y_j - E[Y_j|\mathcal{G}])^2|\mathcal{G}]$ .
- (4)  $\text{Cov}[Y_1, Y_2|\mathcal{G}] := E[(Y_1 - E[Y_1|\mathcal{G}])(Y_2 - E[Y_2|\mathcal{G}]|\mathcal{G})]$ .

**命題 1.8.35** .

$$\text{Corr}[Y_1, Y_2|\mathcal{G}] = \frac{E[\text{Cov}[Y_1, Y_2|\mathcal{G}]]}{\sqrt{E[\text{Var}[Y_1|\mathcal{G}]]E[\text{Var}[Y_2|\mathcal{G}]]}}.$$

**問題 1.8.36 (linear estimator)**.  $X \in L(\Omega; \mathbb{R}^d), \mathcal{G} := \sigma[X]$  とし、 $E[Y_j|X] = \alpha_j^\top X + b_j$  である場合を考える。すなわち、 $X$  が系列  $(Y_j)_{j=1,2}$  を線型に予測する。

## 1.9 独立確率変数列

### 1.9.1 Kolmogorov の 0-1 法則

独立な  $\sigma$ -代数の列  $(\mathcal{B}_n)$  に対して、極限  $\sigma$ -代数は  $\limsup \mathcal{B}_n = \bigvee_k \bigwedge_{n>k} \mathcal{B}_n \subset 2$  となる、Borel-Cantelli の補題の一般化とも見れる。

**定義 1.9.1** . 事象列  $\{A_n\} \subset \mathcal{F}$  について、 $\mathcal{T} := \bigcap_{n=1}^{\infty} \sigma[A_n, A_{n+1}, \dots]$  を末尾  $\sigma$ -代数といい、その元を末尾事象という。

**例 1.9.2**. 列  $\{X_n\} \subset L(\Omega)$  について、 $X_n, S_n, S_n/n$  が収束するという事象は末尾事象である 5.1.2. □

**定理 1.9.3 (Kolmogorov)**.  $\sigma$ -代数の列  $\mathcal{B}_1, \mathcal{B}_2, \dots$  が独立であるとする。このとき、

$$\limsup_{n \rightarrow \infty} \mathcal{B}_n = \bigvee_k \bigwedge_{n>k} \mathcal{B}_n \subset 2.$$

**[証明].** 定理の主張は,  $\mathcal{G} := \bigvee_k \bigwedge_{n>k} \mathcal{G}_n$  は自分自身と独立という状況と同値. いま,  $\mathcal{G}_1, \dots, \mathcal{G}_k, \bigvee_{n>k} \mathcal{G}_n$  は独立.  $\mathcal{G} \subset \bigvee_{n>k} \mathcal{G}_n$  より特に  $\mathcal{G}_1, \dots, \mathcal{G}_k, \mathcal{G}$  は独立. これが任意の  $k \in \mathbb{N}$  について成り立つから,  $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}$  も独立. したがって,  $\mathcal{G}, \mathcal{G}$  は独立.  $\mathcal{G} \subset \mathcal{G}$  より,  $\mathcal{G}$  はそれ自身と独立. ■

**系 1.9.4 (見本平均は殆ど確実な定数に概収束または概発散する).**

- (1) 独立な事象列  $\{A_n\} \subset \mathcal{F}$  について, 任意の末尾事象  $A \in \mathcal{T}$  は  $P[A] \in \{0, 1\}$  を満たす.
- (2)  $(X_n)$  を独立な確率変数列とする. 見本平均の概収束極限  $Y := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k$  が存在するとする. このとき,  $Y$  は殆ど確実に定数である.

**[証明].**

- (1)  $\mathcal{G}_k := \sigma[A_k]$  と取れば良い.
- (2)  $Y$  は  $\mathcal{T}$ -可測確率変数であるため.

## 1.9.2 Kolmogorov の不等式

### 独立確率変数の和の過程の最大値

Markov の不等式の右辺を分散の言葉で表せば, 任意の  $S_k$  について

$$\forall_{k \in [n]} P[|S_k| \geq a] \leq \frac{1}{a^2} E[|S_k|^2] = \frac{1}{a^2} \sum_{i=1}^k V_i \leq \frac{\text{Var}[S_n]}{a^2}$$

同様の評価が,  $\max_{k \in [n]} |S_n|$  についても出来る. これは  $(S_n)$  がマルチンゲールであるから, 最後の時点  $S_n$  にだけ注目すれば良いことによる.

**定理 1.9.5 (Kolmogorov).** 実確率変数列  $\{X_n\} \subset L^2(\Omega, \mathcal{F}, P)$  は独立で,  $E[X_n] = 0, V_n := \text{Var}[X_n] < \infty$  とする. このとき, 和の過程  $S_k := \sum_{i=1}^k X_i$  について,

$$\forall_{a>0} P[\max_{k \in [n]} |S_k| \geq a] \leq \frac{\text{Var}[S_n]}{a^2}.$$

**[証明].** 確率を求めたい事象を  $A$ , これらがどの段階で初めて  $a > 0$  を超えるかによって場合分けをして

$$A := \left\{ \omega \in \Omega \mid \max_{k \in [n]} |S_k| \geq a \right\} \quad A_k := \left\{ \omega \in \Omega \mid \forall_{i \in [k-1]} |S_i| < a \wedge |S_k| \geq a \right\}$$

とおくと,  $A = \sum_{k \in [n]} A_k$  が成り立ち,  $A_k \in \sigma[X_1, \dots, X_k]$ . いま,  $(S_n^2)$  は劣マルチンゲールで, Doob の不等式より特に  $\forall_{k \in [n-1]} E[Z_n^2, A_k] \geq E[Z_k^2, A_k]$  であるから,

$$\begin{aligned} P[A] &= \sum_{k \in [n]} P[A_k] \leq \sum_{k \in [n]} \frac{1}{a^2} E[S_k^2, A_k] && \because A_k \text{ 上では } a^2 \leq S_k^2 \\ &\leq \frac{1}{a^2} \sum_{k \in [n]} E[S_n^2, A_k] = \frac{1}{a^2} E[S_n^2, A] \\ &\leq \frac{1}{a^2} E[S_n^2] = \frac{1}{a^2} \sum_{i=1}^n V_i. \end{aligned}$$

**系 1.9.6 (一般の  $L^p$  の場合).** 実確率変数列  $\{X_n\} \subset L^p(\Omega)$  は独立で,  $E[X_n] = 0$  とする. このとき,

$$P[\max_{1 \leq i \leq n} |S_i| \geq \epsilon] \leq \frac{E[|S_n|]^p}{\epsilon^p} \quad \epsilon > 0.$$

### 1.9.3 Ottaviani の不等式

**定理 1.9.7.**  $X_1, \dots, X_n$  を独立,  $\alpha, \beta > 0$  とする. (1) $\Rightarrow$ (2) が成り立つ.

- (1)  $\forall_{k \in n} P[|X_{k+1} + \dots + X_n| \leq \alpha] \geq \beta$ .  
 (2)  $P[\max_{k \in [n]} |S_k| > 2\alpha] \leq \frac{1}{\beta} P[|S_n| > \alpha]$ .

[証明].

(a) (1) の整理をすると,  $A := \left\{ \max_{k \in [n]} |S_k| > 2\alpha \right\}$  をどの段階で達成するかによって

$$A_k := \{\forall_{1 \leq i \leq k-1} |S_i| \leq 2\alpha, |S_k| > 2\alpha\}$$

と場合分けをすると,  $\sum_{k=1}^n P[A_k] \leq \frac{1}{\beta} P[|S_n| > \alpha]$  を示せば良い.

(b) (2) の整理をすると,  $B_k := \{|S_n - S_k| \leq \alpha\}$  とするといずれも確率は  $\beta$  以上で,  $(A_i \cap B_i)$  も互いに素なままであり,  $A_i \cap B_i \subset \{|S_n| > \alpha\}$  を満たす.  $A_k \in \sigma[X_1, \dots, X_k], B_k \in \sigma[X_{k+1}, \dots, X_n]$  より互いに独立だから,  $P[A_k \cap B_k] = P[A_k]P[B_k] \geq \beta P[A_k]$ .

(c) 以上を併せると,

$$\sum_{k \in [n]} P[A_k] \leq \frac{1}{\beta} \sum_{k \in [n]} P[A_k \cap B_k] \leq \frac{1}{\beta} P[|S_n| > \alpha].$$

■

### 1.9.4 Hewitt-Savage の 0-1 法則

de Finetti や Savage らが Bayes 統計学の基礎を築く上で肝要になった定理である.

**定義 1.9.8 (exchangeable).**  $X_1, X_2, \dots \in \text{Meas}(\Omega, \mathbb{R})$  を確率変数列とする.

- (1)  $(X_n)$  が交換可能列であるとは,  $\forall_{n \in \mathbb{N}} \forall_{\sigma \in S_n} (X_1, \dots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \dots, X_{\sigma(n)})$  が成り立つことをいう. すなわち, 有限個の確率変数の置換について, 無限列  $(X_n)$  が分布同等であることをいう.  
 (2) 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

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

が成り立つことをいう.

**要諦 1.9.9.** Walter Shewhart (1924) による統計管理 (statistical control) と同等な概念である.

**定理 1.9.10.**

- (1) 交換可能な事象  $S \in \mathcal{F}$  の全体は部分  $\sigma$ -加法族  $\mathcal{L} \subset \mathcal{F}$  をなす.  
 (2) (Hewitt-Savage) 交換可能列  $(X_n)$  が独立同分布であったとき,  $\sigma$ -代数  $\mathcal{L}$  の元は零集合か充満集合である.  
 (3) (de Finetti 1) 交換可能列  $(X_n)$  を二項分布に従う確率変数の列とする. これらは,  $\mathcal{L}$  で条件付けると, 独立同分布列となる.  
 (4) (de Finetti 2)

**要諦 1.9.11.** 交換可能列と, 独立同分布列の凸結合とが同値であるというのが, de Finetti の定理に続く後世の研究結果である. 「凸結合」というのは, ある確率分布に対する積分だと見れば, これは交換可能列は予測分布と見れることだと解することが出来る. Hewitt-Savage (1955) による証明は Klein-Milman の定理による.

## 第 2 章

# 確率測度の空間

- (1) 完備距離空間上の Borel 確率測度は正則である。これを  $P(S) \subset M(S)$  と表すと、これは  $(M(S), C_b(S))$ -コンパクトになる。
- (2) 一般に Hausdorff 空間上の有限 Radon 測度は正則になり、この範囲で弱収束の理論が展開でき、 $X$  が完全正則であるとき  $\sigma(M(X), C_b(X))$ -位相に一致する (Topsoe 1970)。
- (3)  $P(S)$  の中で「一様に内部正則」という性質が  $(M(S), C_b(S))$ -相対コンパクト性を特徴付ける。なお、これは  $L^1(\Omega)$  上の一様可積分性に完全に対応する。
- (4) 一様緊密性の十分条件には、 $p > 1$  次の絶対積率について有界であることなどがある。
- (5) 緩増加超関数に対する Fourier 変換  $\mathcal{F} : P(\mathbb{R}^n) \rightarrow \mathbb{C}^{\mathbb{R}^n}$  は連続である。ただし、特性関数の列  $(\varphi_n)$  の各点収束極限  $\varphi$  が再び特性関数であるためには、 $\varphi$  の  $t = 0$  での連続性が必要。

## 2.1 測度の表現定理

### 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<sup>a</sup>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。

<sup>a</sup> 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]

### 定理 2.1.1 (Borel 測度の正則性についての結果)。

- (1) Hausdorff 空間  $X$  上の Radon 測度は有限ならば外部正則である。特に、正則な Borel 測度である。
- (2) 任意の完全正則空間（距離空間はこれを満たす） $X$  上の Borel 測度は正則である。特に、完備可分距離空間上の Borel 測度は正則な Radon 測度である。

### 2.1.1 Radon 積分とこれを定める Radon 測度

**補題 2.1.2.** 各点完備な位相線型束上の線型汎関数は、正ならば連続である。

**定理 2.1.3 (実数上の Radon 積分の Lebesgue-Stieltjes 積分としての表現).** 正な線型汎関数  $\int \in (C_c(\mathbb{R}))_+^*$  について、Lebesgue-Stieltjes 積分が定める対応

$$F : \{m \in \text{Map}(\mathbb{R}, \mathbb{R}) \mid m \text{ は単調増加}\} \rightarrow (C_c(\mathbb{R}))_+^*$$

は全射で、 $F$  が定める同値類は  $m$  を下半連続に取ることで代表元とすることができ、これによって集合の同型が引き起こされる。

**定理 2.1.4 (Radon-Riesz<sup>†1</sup>).**  $X$  を局所コンパクトハウスドルフ空間,  $(C_c(X))^*_+$  を正な線型汎関数のなす正錐,  $\text{RM}(X; [0, \infty])$  をその上の Radon 測度とする. 次の対応は全単射である:

$$\begin{array}{ccc} \text{RM}(X) & \xrightarrow{\quad} & (C_c(X))^*_+ \\ \downarrow \Psi & & \downarrow \Psi \\ \mu & \longmapsto & \mu(A) = \int_* [A] \end{array}$$

また, 位相同型でもある.<sup>†2</sup>

**歴史 2.1.5.** Riesz の元来の定理は次のものであるが, これと,  $[a, b]$  上の有界変動関数とその上の正則な Borel 測度との間には全単射が存在するために, この定理も Riesz の名前で呼ばれる.

**注 2.1.6.**  $\text{RM}(X; [0, \infty])$  の元は, 「外部正則かつ開集合上内部正則かつ局所有限な Borel 測度  $\mu^*$ 」と一対一対応し, この関係を, 「 $\mu$  は  $\mu^*$  に付随する本質的測度である」という.<sup>†3</sup>

## 2.1.2 $C$ の双対空間と符号付き測度

**定理 2.1.7.** 次の対応  $F$  は全射である:

$$F : \{m \in \text{Map}(\mathbb{R}; \mathbb{R}) \mid m \text{ は単調増加}\} \rightarrow (C(\mathbb{R}))_+.$$

**定理 2.1.8 (絶対連続関数の双対空間の Lebesgue 積分による特徴付け (Riesz 1909)??).**

$$C([a, b])^* \simeq_{\text{Set}} \{\rho \in \text{BV}([a, b])\}$$

## 2.1.3 $C_0$ の双対空間

**補題 2.1.9.**  $C_c(X)$  は  $C_0(X)$  上一様ノルムについて稠密である. 特に,  $C_c(X)^* = C_0(X)^*$ .

**定理 2.1.10 (Riesz-Markov (1938)).**  $X$  を局所コンパクトハウスドルフ空間,  $\text{RM}(X; \mathbb{R})$  をその上の符号付き Radon 測度の全変動ノルムに関する Banach 空間,  $(C_0(X))^*$  を  $C_0(X)$  上の有界線型汎関数の Banach 空間とする. このとき, 次の対応は等長同型  $(C_0(X))^* \simeq_{\text{Ban}} \text{RM}(X; \mathbb{R})$  を定める:

$$\begin{array}{ccc} (C_0(X))^* & \xrightarrow{\quad} & \text{RM}(X; \mathbb{R}) \\ \downarrow \Psi & & \downarrow \Psi \\ \Phi & \longmapsto & \Phi f = \mu[f] \quad (f \in C_0(X)) \end{array}$$

**注 2.1.11.** このとき Radon 測度は有限だから, 実際は正則な Borel 測度にもなっている. なお, 符号付き Radon 測度の代わりに複素 Radon 測度としても成立する.

**系 2.1.12 (実数上の Radon 測度は有界変動関数によって表現される).** 次の対応は等長同型  $(C_0(\mathbb{R}))^* \simeq_{\text{Ban}} \text{BV}(\mathbb{R})$  を定める:

$$\begin{array}{ccc} C_0(\mathbb{R}) & \xrightarrow{\quad} & \text{BV}(\mathbb{R}) \\ \downarrow \Psi & & \downarrow \Psi \\ \Phi & \longmapsto & \Phi(f) = \int_{\mathbb{R}} f(u) d\alpha(u). \end{array}$$

## 2.1.4 $C_b$ の双対空間

**定義 2.1.13 (Baire measure).** 位相空間  $X$  上の Baire  $\sigma$ -代数とは,  $C(X)$  を可測にする最小の  $\sigma$ -代数をいう. このとき,  $C_b(X)$  を可測にする最小の  $\sigma$ -代数としても特徴付けられる. Baire  $\sigma$ -代数上の測度を Baire 測度という.

<sup>†1</sup> Malliavin と Alexandorff も Riesz and Radon と呼んでいる.

<sup>†2</sup> <https://ncatlab.org/nlab/show/Riesz+representation+theorem>

<sup>†3</sup> [urlhttps://ncatlab.org/nlab/show/Radon+measure](https://ncatlab.org/nlab/show/Radon+measure)

**補題 2.1.14 .**

- (1) 距離空間上では Baire 集合族と Borel 集合族とは一致する.
- (2) Tychonoff 空間 (距離空間と正規空間はこれを満たす, コンパクトハウスドルフ空間は正規である) 上のタイトな Baire 測度は, Radon 測度に一意に延長する.

**定理 2.1.15 (Riesz-Markov-Kakutani (1941)).**  $X$  をコンパクトハウスドルフ空間,  $M_B(X)$  を符号付き Baire 測度が全変動ノルムについてなす Banach 空間とする. このとき, 次の対応は等長同型  $(C(X))^* \simeq_{\text{Ban}} M(X)$  を定める:

$$\begin{array}{ccc} (C(X))^* & \longrightarrow & M_B(X) \\ \Psi & & \Psi \\ \Phi & \longmapsto & \Phi f = \mu[f] \quad (f \in C(X)) \end{array}$$

**要諦 2.1.16.** したがって, Radon 測度の空間へは埋め込みを定める.

**定理 2.1.17 (Markov-Alexandorff (1940) ([?], Th'm IV.6.2)).**  $X$  を  $T_4$ -位相空間,  $M_f(X)$  をその上の有限加法的な符号付き Radon 測度が全変動ノルムについてなす空間とする. 次の対応は等長同型  $(C_b(X))^* \simeq_{\text{Ban}} M_f(X)$  を定める:

$$\begin{array}{ccc} (C_b(X))^* & \longrightarrow & M_f(X) \\ \Psi & & \Psi \\ \Phi & \longmapsto & \Phi(f) = \int_X f d\mu \end{array}$$

またこの同型は順序も保つ.

**歴史 2.1.18.** 初めて  $C_B(\mathbb{R})$  上の線型汎関数の空間を考察したのは Fichtenholz G. と Kantorovitch, L. (1934) であった. Alexandorff (1940,[?]) で charge と呼んでいるものは, 有限加法的な符号付き測度で, 有限加法的な集合関数である. これは Markov (38) の仕事の拡張だという. この論文では Riesz-Radon の拡張が試みられ, そこで charge の語を生み出している.<sup>†4</sup> してまさか  $T_4$ -空間を表す正規性は, この Riesz-Radon の研究から出てきたとは.

**注 2.1.19.** [?] で  $X = \mathbb{R}$  の場合は取り上げられているが, 少し差があり, 本当に  $X$  が正規な場合にも成り立つのか疑問が残る. どうやら Dunford-Schwartz Theorem IV 6.2 には載っているらしい. 他に参考になりそうなのは, Malliavin, この PDF など. さらに,  $(C_b(\mathbb{R}))^*$  は  $\mathbb{R}$  の Stone-Cech コンパクト化上の Radon 測度の空間でもあるらしい.<sup>†5</sup>

## 2.1.5 $L^\infty$ の双対空間

**定理 2.1.20 (Hildebrandt, T. H. (1934)).** 次の対応は等長同型  $(l^\infty(\mathbb{R}))^* \simeq_{\text{Ban}} M_f(\mathbb{R})$  を定める.

$$\begin{array}{ccc} (l^\infty(\mathbb{R}))^* & \longrightarrow & M_f(\mathbb{R}) \\ \Psi & & \Psi \\ \Phi & \longmapsto & \Phi(f) = \int_{\mathbb{R}} f d\mu \end{array}$$

**定理 2.1.21 ([?], Th'm IV.8.16)).**  $(\Omega, \mathcal{F}, \mu)$  を  $\sigma$ -有限な測度空間とする. このとき,  $M_f$  を  $\mathcal{F}$  上の有限加法的な符号付き測度の空間とすると,  $(L^\infty(\Omega))^* \simeq_{\text{Ban}} M_f(\Omega)$  が成り立つ.

## 2.1.6 測度の収束の定義 5 種

**定義 2.1.22 (narrow topology / topology étroite, wide topology (Bourbaki)).** 位相空間  $X$  上の可測空間  $(X, \mathcal{G}(X))$  上の符号付き Radon 測度のなす空間を  $M(X; \mathbb{F})$  ( $\mathbb{F} = \mathbb{R}, \mathbb{C}$ ) とする.

- (1)  $M(X)$  は「各点収束」について閉じていることが, Vitali-Hahn-Saks の定理により保証される.
- (2)  $M(X)$  は全変動ノルムについて Banach 空間となり, このノルムにの定める位相を強位相という.

<sup>†4</sup> to call it measure seemed to be inconvenient, therefore, lead by a natural physical analogy, I call it a charge.

<sup>†5</sup> <https://mathoverflow.net/questions/83593/dual-space-of-continuous-functions>

- (3)  $\sigma(M(X), (M(X))^*)$ -位相を**弱位相**という.
- (4)  $\sigma(M(X), C_b(X))$ -位相を**狭位相**または**弱位相**という.
- (5)  $\sigma(M(X), C_c(X))$ -位相を**広位相**または **$w^*$ -位相**または**漠位相**という.

**命題 2.1.23.**

- (1)  $X$  を局所コンパクトハウスドルフ空間とする. このとき,  $M(X)$  の広位相と狭位相とは,  $P(X) \subset M(X)$  上に同じ相対位相を定める.
- (2)  $X$  がコンパクトならば,  $M(X)$  の広位相と狭位相とは一致し, これについて  $P(X) \subset M(X)$  もコンパクトである.
- (3)  $X$  がコンパクトでないならば,  $P(X) \subset M(X)$  は広位相についてコンパクトではない.
- (4)  $X$  を局所コンパクトハウスドルフ空間とする. Radon 確率測度の列が一様にタイトならば, 漠位相についてコンパクトである (収束する部分列が存在し, その収束先は確率測度である).

**注 2.1.24.** 確率測度の弱収束を  $w^*$ -収束と呼びたいのならば, この事実が根拠になるが, 極めて混乱を招きやすい呼称だろう.

**例 2.1.25.**

- (1)  $\delta_n \rightarrow 0$  は漠収束の意味で成り立つ.
- (2)  $\frac{\delta_{1/n} + \delta_n}{2} \rightarrow \delta_0/2$  は漠収束の意味で成り立つ.

□

**命題 2.1.26.**

- (1)  $X$  が  $\sigma$ -コンパクトならば,  $w^*$ -位相は距離化可能である.
- (2) 任意の  $X$  について, 正錐  $M(X)_+$  は距離化可能である.
- (3)  $X$  が  $\sigma$ -コンパクトならば,  $M(X)$  の有界集合は弱相対コンパクトである.

## 2.2 $P(S)$ の研究

$C_b(S)$  を Banach 空間,  $M_f(S)$  をその双対空間 (有限加法的な符号付き Radon 測度),  $M(S)$  を完全加法的なものがなす部分空間,  $P(S) := \{\mu \in M(S) \mid \|\mu\| \leq 1, \mu(1) = 1\} \subset B^*$  とする.

**記法 2.2.1.**  $S$  で可分な距離空間を表す.

### 2.2.1 弱位相の定義

**定義 2.2.2 (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.

- (1)  $\sigma(M(X), C_b(X))$ -位相に関する収束を**弱収束**という.
- (2)  $M(X)$  の  $w^*$ -位相に関する収束を**漠収束**という.

**要諦 2.2.3.**  $C_b(X)$  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. 実際,  $(\delta_n)$  は弱収束しないが,  $0$  に漠収束し,  $0 \notin P(S)$ .

**補題 2.2.4 (弱収束理論の基本補題).**  $S$  を可分な距離空間,  $P, Q$  をその上の任意の確率測度とする.

- (1) (位相的正則性) 任意の  $B \in \mathcal{B}(S)$  に対して,  $P[B] = \sup_{F^{\text{closed}} \subset B} P[F] = \sup_{B \subset G^{\text{open}} \subset S} P[G]$ .
- (2)  $C_b(S)$  は  $P(S)$  上の分離族となる:

$$\left[ \forall f \in C_b(S) \int_S f(x)P(dx) = \int_S f(x)Q(dx) \right] \Rightarrow P = Q.$$



(3)  $S$  はさらに完備でもあるとする。このとき、 $S$  上の確率測度は緊密 (=内部正則) である。

(4)  $S$  はさらに完備でもあるとする。このとき、 $P[B] = \sup_{K \subset B} P[K]$ .

特に、弱収束の極限は一意的である。

[証明].

(1) 次のように  $\mathcal{A} \subset \mathcal{G}(S)$  を定めると、これは部分  $\sigma$ -代数になるから、 $\mathcal{G}(S) \subset \mathcal{A}$ 、すなわち、 $\mathcal{O}(S) \subset \mathcal{A}$  を示せば良い：

$$\mathcal{A} := \left\{ B \in \mathcal{G}(S) \mid P[B] = \sup_{F \subset B} P[F] = \sup_{B \subset G} P[G] \right\}.$$

(2) (1) より、任意の閉集合  $F \subset S$  に対して  $P[F] = Q[F]$  が成立することを示せば良い。  $C_b(S)$  は  $1_F$  を近似できるので、Lebesgue の優収束定理より判る。

(3)

■

## 2.2.2 Alexandorff の特徴付け

**定理 2.2.5 (Portmanteau 定理：弱収束の特徴付け (Alexandroff)).**  $\{\mu_n\} \subset P(\mathbb{R}^d), \mu \in P(\mathbb{R}^d)$  について、次の 6 条件は同値。これは分布関数を用いる条件 (5),(6) を除いて、一般の距離空間  $X$  とネット (あるいは可分距離空間と点列) についても成り立つ。

- (1)  $\mu_n \Rightarrow \mu$ .
- (2) 任意の開集合  $G \subset \mathbb{R}^d$  について、 $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G)$ .
- (3) 任意の閉集合  $F \subset \mathbb{R}^d$  について、 $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$ .
- (4)  $A \in \mathcal{G}(\mathbb{R}^d)$  について、 $\mu(\partial A) = 0 \Rightarrow \lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ . この条件を満たす集合  $A$  を  $\mu$ -連続集合という。<sup>†6</sup>
- (5)  $\mu_n, \mu$  が定める分布関数  $F_n, F$  について、 $\forall x \in \text{Cont}(F) \lim_{n \rightarrow \infty} F_n(x) = F(x)$ .
- (6) 任意の  $f \in C_c(\mathbb{R}^d)$  について、 $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu$ .

**歴史 2.2.6.** 測度の弱収束の理論は、A. D. Alexandorff と Prohorov による。凸解析と関数解析が肝要となる。Alexandroff 自身も、凸集合の幾何の研究が、自身の抽象測度論の研究の源泉になったと言っている。I don't know who invented such a nonsensical name for Alexandroff's theorem. [Bogachev, 2018]

[証明].

(1) $\Rightarrow$ (2) 任意の  $G \subset \mathbb{R}^d$  をとる。

Step1

$$F_k := \{x \in \mathbb{R}^d \mid d(x, G^c) \geq 1/k\}.$$

と定めると、これは  $G$  を内側から近似する閉集合の列で、 $G = \cup_{k \geq 1} F_k$  を満たす単調増加列である。ここで、 $\{\varphi_k\} \subset C_b(\mathbb{R}^d; [0, 1])$  は  $G$  上に台を持ち、 $F_k$  上で 1 な連続関数とすると、 $\varphi_k \nearrow 1_G$  を満たす。

Step2 任意の  $\epsilon > 0$  をとる。  $\varphi_k \nearrow 1_G$  より、ある  $k \in \mathbb{N}$  が存在して、

$$\int_{\mathbb{R}^d} \varphi_k d\mu \geq \mu(G) - \epsilon.$$

が成り立つ。いま  $\mu_n \Rightarrow \mu$  の仮定と  $\varphi_k \in C_b(\mathbb{R}^d)$  より、

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \varphi_k d\mu_n = \int_{\mathbb{R}^d} \varphi_k d\mu.$$

であるが、

$$\mu_n(G) \geq \int_{\mathbb{R}^d} \varphi_k d\mu_n, \quad n \in \mathbb{N}.$$

<sup>†6</sup> このような集合は体/代数をなす。

を満たしながらであるから,

$$\limsup_{n \rightarrow \infty} \mu_n(G) \geq \int_{\mathbb{R}^d} \varphi_k d\mu \geq \mu(G) - \epsilon.$$

(2)⇔(3) 任意の閉集合  $F \subset^{\text{closed}} \mathbb{R}^d$  をとると,  $F^c$  は開集合だから,

$$1 - \liminf_{n \rightarrow \infty} \mu_n(F^c) \leq 1 - \mu(F^c).$$

逆も辿れる.

(2),(3)⇒(4) 任意に  $\mu$ -連続集合  $A \in \mathcal{G}(\mathbb{R}^d)$  をとると,  $A^\circ, A, \bar{A}$  はいずれも測度が等しい. よって,

$$\limsup_{n \rightarrow \infty} \mu_n(A) \leq \limsup_{n \rightarrow \infty} \mu_n(\bar{A}) \leq \mu(\bar{A}) = \mu(A).$$

$$\liminf_{n \rightarrow \infty} \mu_n(A) \geq \liminf_{n \rightarrow \infty} \mu_n(A^\circ) \geq \mu(A^\circ) = \mu(A).$$

(4)⇒(5)  $x \in \mathbb{R}^d$  について,  $x \in \text{Cont}(F)$  は

$$\mu \left( \partial \left( \prod_{i=1}^d (-\infty, x_i] \right) \right) = 0.$$

と同値. よって, 仮定から従う.

(5)⇒(6)  $d = 1$  として示す. 任意の  $f \in C_c(\mathbb{R})$  と  $\text{supp } f \subset [-R, R]$  を  $\{-R, R\} \subset \text{Cont}(F)$  を満たすように取る.  $[-R, R]$  の分割  $\Delta := \{x_i\}_{i \in [N]} \subset \text{Cont}(F)$  に対して,

$$f_\Delta(x) := \begin{cases} f(x_0) & x \in [x_0, x_1], \\ f(x_{i-1}) & x \in (x_{i-1}, x_i] \quad (2 \leq i \leq N). \end{cases}$$

と定めると, 任意の  $\epsilon > 0$  に対して, 細かい分割  $\Delta = \{x_i\}_{i \in [N]}$  が存在して,

$$\max_{x \in [-R, R]} |f_\Delta(x) - f(x)| \leq \epsilon.$$

すると,

$$\int_{[-R, R]} f d\mu_n - \int_{[-R, R]} f d\mu = \int_{-R}^R (f - f_\Delta) d\mu_n + \int_{-R}^R f_\Delta (d\mu_n - d\mu) + \int_{-R}^R (f_\Delta - f) d\mu =: I_1 + I_2 + I_3.$$

のうち,  $|I_1| + |I_3| \leq 2\epsilon$  と評価できる. 最後に  $I_2$  は

$$I_2 = \sum_{i=1}^N f(x_{i-1}) \left( (F_n(x_i) - F_n(x_{i-1})) - (F(x_i) - F(x_{i-1})) \right)$$

と表せるが,  $\{x_i\} \subset \text{Cont}(F)$  より,  $\lim_{n \rightarrow \infty} I_2 = 0$ .

(6)⇒(Step1) 任意の  $\epsilon > 0$  に対して, ある  $R > 0$  が存在して,

$$\inf \{ \mu_n(B_R(0)) \wedge \mu(B_R(0)) \in [0, 1] \mid n \in \mathbb{N} \} \geq 1 - \epsilon.$$

Step2  $\varphi_R \in C_c(\mathbb{R}^d; [0, 1])$  を,  $B_R(0)$  上で 1 である  $B_{R+1}(0)$  上に台を持つ関数とすると, 任意の  $f \in C_b(\mathbb{R}^d)$  について,

$$\int_{\mathbb{R}^d} f (d\mu_n - d\mu) = \int_{\mathbb{R}^d} f \varphi_R (d\mu_n - d\mu) + \int_{\mathbb{R}^d} f (1 - \varphi_R) d\mu_n - \int_{\mathbb{R}^d} f (1 - \varphi_R) d\mu =: I_1 + I_2 + I_3.$$

と分解できる. まず  $\lim_{n \rightarrow \infty} I_1 = 0$  であり,  $|I_2|, |I_3| \leq \|f\|_\infty \epsilon$  が成り立つ. これより,

$$\limsup_{n \rightarrow \infty} \left| \int_{\mathbb{R}^d} f (d\mu_n - d\mu) \right| \leq 2\|f\|_\infty \epsilon.$$

**系 2.2.7 (弱収束極限の一意性).** 弱収束の極限は一意的である.

**[証明].**  $\mu_n \rightarrow \mu, \mu_n \rightarrow \nu$  とする.  $\mu, \nu$  の分布関数の不連続点の合併  $D_\mu \cup D_\nu$  は可算で, その上で分布関数は一致する.  $F, G$  は右連続だから,  $\mathbb{R}$  上一致することが従う. ■

**系 2.2.8 (離散分布の弱収束の特徴付け).**  $\{\mu_n\} \subset P(\mathbb{N})$  について,

(1)  $\mu_n \Rightarrow \mu$ .

(2) 任意の  $k \in \mathbb{N}$  において,  $\mu_n(\{k\}) \rightarrow \mu(\{k\})$ .

## 2.2.3 弱収束の十分条件

$F$  の連続点で  $F_n(x) \rightarrow F$  とは、分布  $\mu$  の連続点  $x$  について、 $\mu_n((-\infty, x]) \rightarrow \mu((-\infty, x])$  に同値。

**系 2.2.9** (密度が収束するならば弱収束する).  $\{\mu_n\} \subset P(\mathbb{R}^d)$  と  $\mu \in P(\mathbb{R}^d)$  は Lebesgue 測度  $l$  について絶対連続であるとする. このとき、密度関数の列  $(d\mu_n/dl)$  が  $d\mu/dl$  に殆ど至る所収束するならば、 $\mu_n \Rightarrow \mu$  である。

**[証明]**. 任意の  $G \subset \mathbb{R}^d$  について、Fatou の補題より、

$$\begin{aligned} \mu(G) &= \int_G \frac{d\mu}{dl} dx \\ &= \int_G \lim_{n \rightarrow \infty} \frac{d\mu_n}{dl} dx \\ &\leq \liminf_{n \rightarrow \infty} \int_G \frac{d\mu_n}{dl} dx = \liminf_{n \rightarrow \infty} \mu_n(G). \end{aligned}$$

■

**定理 2.2.10** (一致の定理が成り立つ部分 Borel 集合族上での一致).  $\mathcal{A}_P \subset \mathcal{B}(S)$  が次の 2 条件をみたすとき、 $\forall A \in \mathcal{A}_P$   $P_n(A) \rightarrow P(A)$  は  $P_n \Rightarrow P$  を含意する。

- (1)  $\mathcal{A}_P$  は乗法族である。
- (2) 任意の開集合が  $\mathcal{A}_P$  の元の可算和で表せる。

また、条件 (2) は、次のいずれと入れ替えても成立。

- (a)  $S$  が可分で  $\forall x \in S \ \forall \epsilon > 0 \ \exists A \in \mathcal{A}_P \ x \in A^\circ \subset A \subset B_\epsilon(x)$ .

## 2.2.4 双対空間の稠密部分集合

一般に、局所凸空間  $X$  について、 $Y \subset X^*$  が  $w^*$ -稠密であることは、 $X$  の分離族を定めることに同値 (Hahn-Banach の分離定理の系)。そして、 $C_b(\mathbb{R}^d)$  の稠密部分集合には  $C_c^\infty(\mathbb{R}^d)$  があるので、これを含む関数クラスはどれでも十分。

**定理 2.2.11** (有界 Lipschitz 連続関数は確率測度の分離族である). 任意の  $P, Q \in P(S)$  について、次の 2 条件は同値：

- (1)  $\forall f \in \text{Lip}_b(S) \ Pf = Qf$ .
- (2)  $P = Q$ .

**系 2.2.12** (弱収束の特徴付け).  $\{\mu_n\} \subset P(\mathbb{R}^d), \mu \in P(\mathbb{R}^d)$  について、次の 2 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) 任意の有界な Lipschitz 関数  $f \in \text{Lip}_b(\mathbb{R}^d)$  について、 $P_n f \rightarrow P f$ .
- (3) 任意の  $f \in C_c^\infty(\mathbb{R}^d)$  について、 $\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)]$ .

## 2.2.5 漠位相による特徴付け

弱収束と漠収束は同じ位相に関する収束である (一般に  $X$  が Hausdorff のときこれが成り立つ)。しかし、 $P(X) \subset M(X)$  を部分集合とみると、 $X$  がコンパクトでないならば決して漠位相についてコンパクトでない。すなわち、漠収束は極限が  $\mu(X) \leq 1$  を満たすことしか確約し得ない。

**命題 2.2.13** (漠収束の特徴付け). 有限測度の列  $\{\mu_n\} \subset M(\mathbb{R})_+, \mu \in M(\mathbb{R})_+$  について、次の 2 条件は同値。

- (1)  $\mu_n$  は  $\mu$  に漠収束する.  
 (2) 端点が  $\mu$  の連続点となるような区間  $(a, b]$  について,  $\mu_n(a, b] \rightarrow \mu(a, b]$ .

**命題 2.2.14 (Helly の選出定理の系).** 有界測度の列  $\{\mu_n\} \subset M(\mathbb{R})_+$  が全変動ノルムについて有界ならば:  $\sup_{n \in \mathbb{N}} \mu_n(\mathbb{R}) < \infty$ , 収束する部分列を含む.

**命題 2.2.15 (漠収束による特徴付け).**  $\{\mu_n\} \subset P(\mathbb{R}), \mu \in M(\mathbb{R})$  について, 次の2条件は同値.

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

一般の有限測度列  $\{\mu_n\} \subset M(\mathbb{R}; \mathbb{R}_+)$  については, 次が成り立つ:

- (1)  $(\mu_n)$  は  $\mu$  に弱収束し,  $\lim_{n \rightarrow \infty} \mu_n(\mathbb{R}) = \mu(\mathbb{R})$ .  
 (2)  $(\mu_n)$  は  $\mu$  に弱収束する.

## 2.2.6 弱位相の距離化

$S$  が Polish 空間であるとき,  $P(S)$  の弱位相は距離化可能である.

一般の Banach 空間について,  $X$  が可分であることと  $B^* \subset X^*$  が  $w^*$ -距離化可能であることは同値である.  $S$  が可分であるとき,  $C_c(S) \subset L^p(S)$  ( $1 \leq p < \infty$ ) も可分.  $P(S) \subset M(S)$  は  $M(S)$  の単位閉球の部分集合であり, 前双対  $C_c(S)$  が可分であるために, 距離化可能であり,  $M(S)$  も可分である.

**例 2.2.16 (Levy-Prokhorov 距離).**  $B_\epsilon(E) := \{x \in X \mid \rho(x, E) < \epsilon\}$  と表す.

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

は距離を定め, これが定める位相は弱位相に一致する. この  $X = \mathbb{R}$  の場合に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. □

**定理 2.2.17 (弱位相との関係).**  $\{P_n\} \subset P(S)$  について, (1) $\Rightarrow$ (2) が成り立つ.  $S$  が可分であるとき, (2) $\Rightarrow$ (1) も成り立つ.

- (1)  $\pi(P_n) \rightarrow \pi(P)$ .  
 (2)  $P_n \Rightarrow P$ .

**系 2.2.18 (Prokhorov metric).**  $S$  を完備可分距離空間とする.  $P(S)$  の弱位相は Prokhorov 距離  $\pi$  によって完備可分に距離付け可能で, 弱相対点列コンパクト性は  $\pi$  についての弱コンパクト性で特徴付けられる.

**例 2.2.19 (1-Wasserstein distance / Kantorovich-Rubinstein distance / Monge-Kantorovich distance).** さらに  $X$  を完備とする (すべて併せてポーランド空間とする). このとき, 次は  $P(X)$  の距離を与える:

$$W_1(\mu, \nu) = \sup \left\{ \left| \int \varphi d\mu - \int \varphi d\nu \right| \in \mathbb{R} \mid \varphi \in \text{Lip}(X; \mathbb{R}), L(\varphi) \leq 1 \right\}.$$

□

## 2.2.7 概収束列の構成

**定理 2.2.20 (Skorohod).**  $S$  を完備可分距離空間,  $\{P_n\} \in P(S)$  を  $P \in P(S)$  への弱収束列とする. このとき, ある確率空間  $(\Omega, \mathcal{F}, P)$  と確率変数列  $X_n: \Omega \rightarrow S$  が存在して, 次を満たす:

- (1)  $\forall_{n \in \mathbb{N}} P^{X_n} = P_n$ , かつ,  $P^X = P$ .  
 (2)  $X_n \xrightarrow{\text{a.s.}} X$ .

### 2.2.8 Fourier 変換の連続性

$P(X)$  上の弱位相  $(\sigma(P(X), C_b(X)))$ -位相) と,  $C(\mathbb{R}; \mathbb{C})$  上の各点収束の位相とについて, Fourier 変換は位相同型である (Glivenko の定理). が, その像は「確率分布の特性関数」については閉じていない. すなわち, 特性関数の各点収束極限は特性関数とは限らない. [?]

**定理 2.2.21 (Glivenko).**  $\{\mu_n\} \subset P(\mathbb{R}^n)$  について, 次の 2 条件は同値:

- (1)  $\{\mu_n\}$  は  $\mu$  に弱収束する.
- (2)  $\varphi_{\mu_n}$  は  $\varphi_\mu$  に各点収束する.

なお, (1) が成り立つとき, (2) は特に広義一様収束する.

**定理 2.2.22 (Levy の連続性定理).**  $\{\mu_n\} \subset P(\mathbb{R}^n)$  の特性関数の列  $(\varphi_{\mu_n})$  はある  $\varphi$  に各点収束し,  $\varphi$  は  $t = 0$  で連続とする. このとき, ある  $\mu \in P(\mathbb{R}^n)$  が存在して,  $\varphi = \varphi_\mu$  が成り立つ.

### 2.2.9 コンパクトになるとき

$X$  がコンパクトであるとき,  $P(X)$  は弱コンパクトになる. コンパクト距離空間  $X$  は第 2 可算, 特に可分でもあるので, このとき実は  $P(X)$  は距離化可能でもある.

**命題 2.2.23.** 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.  $C(X)$  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  $\lambda \in (0, 1)$  を任意にとると,  $\lambda\mu_1 + (1 - \lambda)\mu_2 \in P(X)$  がわかる.
- (2) 線型汎函数  $\text{ev}_1: M(X) \rightarrow \mathbb{F}; \mu \mapsto \mu(1)$  は  $w^*$ -位相について連続である  $M(X) = (C(X))^*$  の閉単位球  $B^*$  は  $w^*$ -コンパクト (Alaoglu) である.

■

## 2.3 一様緊密性

一様可積分性に対応する概念である

$S$  がコンパクトでない場合,  $P(S)$  上の  $(M(S), C_c(S))$ -位相と  $(M(S), C_b(S))$ -位相とは一致しなくなる. 後者の位相における相対コンパクト集合は, 緊密性によって特徴付けられる.

### 2.3.1 定義

**定義 2.3.1 (uniformly tight / tendue).** 確率測度の族  $\Gamma \subset P(X)$  が一様に緊密<sup>†7</sup>であるとは, 任意の  $\epsilon > 0$  に対して, あるコンパクト集合  $K \subset X$  が存在して,

$$\forall \mu \in \Gamma \quad \mu(K) \geq 1 - \epsilon.$$

<sup>†7</sup> 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.

が成り立つことをいう。

**注 2.3.2.** 距離空間上の任意の Borel 確率測度は正則であるから、特に内部正則 (=緊密) である。この際の評価が一様になりたつための条件である。一般の有限測度列  $\{\mu_n\} \subset M(X; \mathbb{R}_+)$  には、さらに有界性  $\sup_{\mu \in \Gamma} \mu(\mathbb{R}) < \infty$  を課す。

**補題 2.3.3.**  $\mu_n, \mu \in P(\mathbb{R}^d)$  について、

- (1) 有限集合  $\{\mu_1, \dots, \mu_n\}$  は一様に緊密である。
- (2)  $\mu_n \Rightarrow \mu$  ならば、 $\{\mu_n\}_{n \in \mathbb{N}}$  も一様に緊密である。

**[証明].**

- (1) 明らか。
- (2)  $\mu$  について、ある  $K \subset^{\text{cpt}} \mathbb{R}^d$  が存在して  $\mu(K) \geq 1 - \epsilon$ 。Portmanteau 定理の特徴付けの (2) より、任意の  $K$  の有界開近傍  $K \subset G \subset^{\text{open}} \mathbb{R}^d$  について、 $\mu_n(G) \geq \mu(G) \geq 1 - \epsilon$ 。

**例 2.3.4** (一様に緊密でない例).  $\{x_n\}, \{y_n\} \subset \mathbb{R}^d$  を  $x \in \mathbb{R}^d$  への収束列と発散列としたとき、

$$\mu_n := \frac{\delta_{x_n} + \delta_{y_n}}{2}$$

は一様に緊密ではない。

### 2.3.2 相対点列コンパクト性による特徴付け

$\sigma(M(X), C_b(X))$ -位相と  $\sigma(M(X), C_c(X))$ -位相との代表的な違いは、コンパクト集合の違いである。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。

**命題 2.3.5** (Prohorov [Billingsley, 1999] Th'm 5.1, 5.2).  $S$  を可分距離空間、 $\Lambda \subset P(S)$  を部分集合とする。(1) $\Rightarrow$ (2) が成り立ち、 $S$  が完備なとき逆も成り立つ。

- (1) 一様に緊密である。
- (2) 相対点列コンパクトである：任意の列は収束部分列を持つ。

**[証明].**  $S = \mathbb{R}^d$  として 2 つの同値性を示す。

(2) $\Rightarrow$ (1)

**系 2.3.6** (漠収束の特徴付け). 完備可分距離空間  $X$  上の確率測度列  $(\mu_n)$  について、次の 2 条件は同値。<sup>†8</sup>

- (1)  $\mu$  に弱収束する。
- (2)  $\mu$  に漠収束し、かつ  $(\mu_n)$  は一様に緊密である。

**[証明].** (1) $\Rightarrow$ (2) は明らかだから、(2) $\Rightarrow$ (1) を示す。漠収束の特徴付け 2.2.15 より、極限測度  $\mu$  が  $P(X)$  に属することを示せば十分である。略。

**系 2.3.7** (全有界性による特徴付け [?]). 実数上の分布の族  $\Lambda \subset P(\mathbb{R})$  について、次の 4 条件は同値。

- (1) 弱相対コンパクトである。
- (2)  $\lim_{a \rightarrow \infty} \inf_{\mu \in \Lambda} \mu[-a, a] = 1$ 。

<sup>†8</sup> <https://math.stackexchange.com/questions/313986/are-vague-convergence-and-weak-convergence-of-measures-both-weak-convergence>

(3) Prohorov の距離  $\pi$  に関して全有界である.

(4) 一様に緊密である.

**系 2.3.8.**  $\Lambda \subset P(\mathbb{R})$  が  $p$  次の絶対積率に対して有界  $\sup_{\mu \in \Lambda} \beta_p(\mu) < \infty$  ならば, 相対コンパクトである.

### 2.3.3 一様緊密性の十分条件

**命題 2.3.9.**  $\{X_n\} \subset L(\Omega)$  が

$$\sup_{n \in \mathbb{N}} E[|X_n|^p] < \infty.$$

を満たすとき,  $\{\mu_{X_n}\}_{n \in \mathbb{N}}$  は一様に緊密である.

**[証明].** Markov の不等式 1.7.13 より,

$$P[|X_n| \geq R] \leq \frac{E[|X_n|^p]}{R^p}.$$

## 2.4 測度の畳み込み

Lebesgue 空間  $L^1(\mathbb{R})$  は, 畳み込みを積として非単位的な Banach 代数をなす. 単位元は Dirac 関数に相当する. なお, 局所コンパクトハウスドルフ位相群  $G$  に対して,  $L^1(G)$  が単位的であることと,  $G$  が離散群であることが同値. ノルム代数  $L^1(\mu)$  の公理に劣乗法性が入るが, この結果を Young の不等式という. そして, Fourier 変換は Banach 代数の射であり, 畳み込みを関数積に移す.

### 2.4.1 畳み込みと不等式

**記法 2.4.1.** 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  $\mu$  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 2.4.2 (convolution).**

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

によって,  $\text{Meas}(\mathbb{R}^d, \mathbb{R})^2 \rightarrow \text{Meas}(\mathbb{R}^d, \mathbb{R})$  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 2.4.3 (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,

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

### 2.4.2 確率変数との関係

**命題 2.4.4 (独立和の分布は畳み込み).** 2つの確率変数  $X_1, X_2$  は像測度  $\mu, \nu$  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,

$$\begin{aligned} \mu * \nu(E) &= \iint_{\mathbb{R}} 1_E(x+y) P^{X_1}(dx) P^{X_2}(dy) \\ &= \iint_{\mathbb{R}} 1_E(x+y) P^X(dxdy) \\ &= P(X_1 + X_2 \in E) = P^X(E). \end{aligned}$$



### 2.4.3 特徴量の関手性

**命題 2.4.5.** 分布  $\mu, \nu \in P(\mathbb{R})$  に対して, 平均と分散は畳み込みを加法に写す:

- (1)  $\alpha_1(\mu * \nu) = \alpha_1(\mu) + \alpha_1(\nu).$
- (2)  $\mu_2(\mu * \nu) = \mu_2(\mu) + \mu_2(\nu).$

### 2.4.4 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 2.4.6.**  $d$  次元確率変数  $X_1, \dots, X_n$  が独立であれば,  $X := X_1 + \dots + X_n$  について

$$\forall u \in \mathbb{R}^d \quad \varphi_X(u) = \prod_{j=1}^n \varphi_{X_j}(u).$$

**例 2.4.7.** Hermite 分布と Skellam 分布??.

□

**記法 2.4.8.**  $A - x := \{y \in \mathbb{R}^d \mid y + x \in A\}$  と表す.

**定義 2.4.9 (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  $\nu$  を  $\nu_1 * \nu_2$  で表す.

**補題 2.4.10.**

- (1)  $\nu_1 * \nu_2(A) = \nu_2 * \nu_1(A) = \int_{\mathbb{R}^{2d}} 1_A(x_1 + x_2) \nu_1(dx_1) \nu_2(dx_2).$
- (2) 独立な確率変数  $X_1, X_2$  が  $P^{X_1} = \nu_1, P^{X_2} = \nu_2$  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  $\nu_1, \nu_2$  の独立性を特徴付けない点に注意.

**補題 2.4.11.**  $P(\mathbb{R}^d)$  は  $*$  を積として, 単位元  $\delta_0$  を持つ可換な半群をなす.

### 2.4.5 分布族の再生性

**定義 2.4.12 (reproducing property).**

- (1) 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.
- (2) ある分布  $\varphi \in UC_0(\mathbb{R}; \mathbb{C})$  について, その**分布型**とは, affine 変換による像  $\{F(ax + b)\}_{a>0, b \in \mathbb{R}}$  で定まる分布族をいう. 同じ分布型に属する特性関数  $\varphi_1, \varphi_2$  は条件  $\exists a>0 \quad \varphi_2(t) = e^{itb} \varphi_1(at)$  を満たす.

**命題 2.4.13.**  $\varphi \in UC_0(\mathbb{R}; \mathbb{C})$  が生成する分布型が安定であることは,

$$\forall a_1, a_2 > 0 \quad \exists a_3 > 0, b \in \mathbb{R} \quad \varphi(a_1 t) \varphi(a_2 t) = e^{itb} \varphi(a_3 t)$$

を満たすことに同値.

**例 2.4.14.**

- (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)  $C(\mu_1, \sigma_1) * C(\mu_2, \sigma_2) = C(\mu_1 + \mu_2, \sigma_1 + \sigma_2).$

$$(4) B(n_1, p) * B(n_2, p) = B(n_1 + n_2, p).$$

$$(5) \text{Pois}(\lambda_1) * \text{Pois}(\lambda_2) = \text{Pois}(\lambda_1 + \lambda_2).$$

□

### 2.4.6 安定型

実は、安定な対称分布は、Cauchy 分布と正規分布との一般化された分布族のクラスとして特定できる。

#### 補題 2.4.15.

- (1)  $\varphi_{(\alpha)}(t) = e^{-|t|^\alpha}$  ( $0 < \alpha \leq 2$ ) は特性関数である.
- (2) これに属する  $\varphi_{(\alpha)}$  は安定型である.
- (3)  $\varphi_{(\alpha)}$  は実数値であるから、特に対応する分布は対称である.

**定理 2.4.16.** 安定な対称分布の特性関数は、 $\varphi_{(\alpha)}(ct)$  ( $c > 0, 0 < \alpha \leq 2$ ) の形をしている.

## 2.5 測度の正則性

一般の位相空間  $S$  上で、どれほど  $\mathbb{R}^n$  と並行な議論が行えるかを考える.

### 2.5.1 内部正則性

**補題 2.5.1.**  $S$  を距離空間とし、 $(S, \mathcal{G}(S), P)$  をその上の Borel 確率空間とする. 任意の開集合  $F \overset{\text{closed}}{\subset} S$  に対して、開集合の列  $\{G_n\} \subset \mathcal{O}(S)$  が存在して、

$$F \subset G_n \quad \wedge \quad P[G_n \setminus F] \leq \frac{1}{n}, \quad n = 1, 2, \dots.$$

**定理 2.5.2 (距離空間上の確率測度の正則性).**  $S$  を距離空間とする.

- (1) 任意の  $A \in \mathcal{G}(S)$  と  $\epsilon > 0$  に対して、開集合  $G \overset{\text{open}}{\subset} S$  と閉集合  $F \overset{\text{closed}}{\subset} S$  とが存在して、

$$F \subset A \subset G \quad \wedge \quad P[G \setminus A] \leq \epsilon \quad \wedge \quad P[A \setminus F] \leq \epsilon.$$

を満たす.

- (2)  $S$  が完備可分であるとする.  $F$  はコンパクトにとれる.

### 2.5.2 緊密性

**定理 2.5.3 (緊密性の十分条件).**  $S$  が次のいずれかを満たすならば、任意の  $S$  上の確率測度は緊密である.

- (1)  $S$  は完備で可分.
- (2)  $S$  は  $\sigma$ -コンパクト (例えば可分な局所コンパクト空間).

## 2.6 Donsker の定理

$C$  上の弱収束の理論によって、中心極限定理を精緻化することを考える.

## 2.7 有限測度のノルム収束

Radon 電荷の Banach 空間  $M(\Omega)$  上でのノルム収束を、全変動収束という。その双対空間は  $C_0(\Omega, \mathbb{C})$  であるが (Riesz-Markov), この元に対する収束が弱収束である。確率測度の空間  $P(X)$  はもちろんノルム閉ではないが、凸である。

**補題 2.7.1** (確率測度の全変動距離の特徴付け).  $\mu, \nu \in P(\Omega, \mathcal{F}, P)$  について,

$$\|\mu - \nu\| = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.$$

**系 2.7.2 (Scheffe).**  $\{X_n\}, X \ll \mu$  とする.  $p_n \rightarrow p$   $\mu$ -a.e. が成り立つならば,  $X_n$  は  $X$  に全変動ノルムについて収束する:  $\|X_n - X\| \rightarrow 0$ . すなわち,  $\lim_{n \rightarrow \infty} F_n(x) = F(x)$  が一様収束する. 逆は成り立たない.

## 2.8 分布関数の収束

Euclid 空間の確率測度の弱収束理論のもう一つの展開方法

Portmanteau 定理 2.2.5 の特徴づけに、分布関数の言葉がある。むしろ、こちらを確率分布の弱収束として定義する事もありえる。これは、 $\mathbb{R}$  上の Borel 集合は  $(-\infty, x]$  が生成しており、 $x$  が  $F$  の連続点であることと  $P[\{x\}] = 0$  とは同値。これは分布と絶対連続関数との対応 2.1.3 の一般化ともみなせる。

### 2.8.1 弱収束の定義

**定義 2.8.1 (weak convergence of CDF).**  $F$  の不連続点??はたかだか有限個である。 $F$  の連続点  $\text{Cont}(F)$  での値が与えられたら、その不連続点での値は右連続性から確定する。単調右連続関数の列  $(M_n)$  と  $M$  について、 $M_n \Rightarrow M$  とは、

$$\forall x \in \text{Cont}(M) \quad M_n(x) \rightarrow M(x)$$

が成り立つことをいう。

**注 2.8.2** (緊密性に対応する概念). 実は、確率測度のときと違って、分布関数の弱収束極限は分布関数になるとは限らない。そこで、次の概念が必要となることもある：

$$M_n \Rightarrow M \text{ (proper)} \Leftrightarrow M_n(\infty) - M_n(-\infty) \rightarrow M(\infty) - M(-\infty).$$

真に弱収束するとき、その収束先の右連続化は分布関数になる。

### 2.8.2 相対コンパクト

確率測度の弱収束の相対コンパクト性に関する Prokhorov の定理??はこの定理から示される。やはり関数論的に扱いやすいのは分布関数である。

**定理 2.8.3 (Helly, E. selection theorem).** 分布関数  $(F_n)$  の列には、ある部分列  $(F_{n_k})$  と右連続な単調増加関数  $F$  とが存在して、 $F_{n_k} \Rightarrow F$ . 収束先  $F$  が分布関数になるとは限らないことに注意。

[証明].

**部分列の構成**  $\mathbb{Q} = \{\alpha_j\}$  と附番すると、 $\{F_n(\alpha_1)\} \subset \mathbb{R}$  は有界だから、収束する部分列  $\{F_{n_1}(\alpha_1)\}$  を持つ。さらにこの中から、 $\{F_{n_1}(\alpha_2)\}$  が収束するような部分列  $\{F_{n_2}\}$  が取れる。これを繰り返すことで、列の列  $\{\{F_{n_j}(\alpha_j)\}_{j \in \mathbb{N}}\}_{n_j \in \mathbb{N}}$  が取れるから、この対角線を選んで関数列  $\{F_{n_k}\}$  を取ると、これは  $\mathbb{Q}$  上で収束する。

**収束先の構成**  $G(x) := \lim_{n \rightarrow \infty} F_{n_k}(x)$  によって  $G: \mathbb{Q} \rightarrow [0, 1]$  が定まる. これに対して  $F(x) := \inf_{x < r \in \mathbb{Q}} G(r)$  とすると, これは単調かつ右連続な  $\mathbb{R}$  への延長である. 単調性は明らか. 右連続性は, 定義から  $\forall x \in \mathbb{R} \forall \epsilon > 0 \exists x < r G(r) < F(x) + \epsilon$  ということだから,  $\forall y \in [x, r) F(y) \leq G(r) < F(x) + \epsilon$  と併せると,  $F(y) - F(x) < \epsilon$  を満たすような  $\delta := y - x \geq 0$  が取れたことになる.

**弱収束の証明** 任意の  $x \in \text{Cont}(F)$  について, 任意の  $\epsilon > 0$  を取る. このとき,  $\exists y < x F(x) - \epsilon < F(y)$  かつ  $\exists x < s \in \mathbb{Q} G(s) < F(x) + \epsilon$  だから, 任意の  $r \in \mathbb{Q} \cap (y, x)$  について,

$$F(x) - \epsilon < G(r) \leq G(s) < F(x) + \epsilon.$$

これと単調性  $F_n(r) \leq F_n(x) \leq F_n(s)$  を併せると,  $n_k \rightarrow \infty$  のとき,  $F(x) - \epsilon < F_{n_k}(x) < F(x) + \epsilon$  を満たす.

■

## 第 3 章

# 確率変数の空間

弱収束の理論は関数解析の理論の系である。[?] の 120p から。確率測度の弱収束とは、 $\sigma(P(X), C_b(X))$ -位相に関する収束をいった。

- (1)  $L^1(\Omega)$  の弱相対コンパクト性は一様可積分性が特徴付ける。
- (2)  $(\Omega, \mathcal{F}, P)$  が有限であるとき、一様可積分性は、 $L^1$ -有界性と同程度絶対連続性とは特徴付ける。非原子的分布、特に絶対連続分布については後者のみで十分。
- (3) 十分条件は、可積分な優関数を持つこと (束構造について  $\sup_{n \in \mathbb{N}} |X_n| \leq Y \in L^1(\Omega)$ )、そして  $L^{1+\epsilon}$  ( $\epsilon > 0$ )-有界であることである。

### 3.1 原子

離散分布と連続分布を数学的に区別する消息である。

#### 3.1.1 原子の定義

**定義 3.1.1 (atom).**  $(X, \mathcal{A}, \mu)$  を測度空間とする。集合  $A \in \mathcal{A}$  が次の 2 条件を満たすとき、 $\mu$ -原子という：

- (1)  $\mu(E) \neq 0$ .
- (2)  $\forall B \in \mathcal{A} \ B \subset A \Rightarrow \mu(B) \in \{\mu(A), 0\}$ .

**補題 3.1.2.**  $A_1, A_2$  を原子とする。次のうちいずれか一方のみが成り立つ：

- (1)  $\mu(A_1 \cap A_2) = 0$ .
- (2)  $\mu(A_1 \triangle A_2) = 0$ .

**例 3.1.3.**  $\mathbb{R}^n$  上の Lebesgue 測度は原子を持たない。 □

#### 3.1.2 有限測度空間の分割

有限測度空間が原子を持たないならば、測度について「全有界」である。

**定理 3.1.4 (Saks).**  $(X, \mathcal{A}, \mu)$  を有限測度空間とする。任意の  $\epsilon > 0$  に対して、 $X$  の有限な分割  $\{E_i\}_{i \in [n]} \subset \mathcal{A}$  であって、任意の  $E_i$  について、次のいずれかが成り立つものが存在する：

- (1)  $\mu(E_i) > \epsilon$  を満たす原子である。
- (2)  $\mu(E_i) \leq \epsilon$  を満たす。

## 3.2 弱位相

$L(\Omega)$  の元の分布収束  $X_n \Rightarrow X$  は,  $L^1(X)$  上では弱位相に一致する. 特に, 分布収束する  $L^1(\Omega)$  の列は一様可積分である.

### 3.2.1 弱収束列はノルム有界

**定理 3.2.1 (ノルム空間の弱収束の必要条件).**  $X$  をノルム空間とし,  $\{x_n\} \subset X$  を点列,  $x \in X$  とする. (1) $\Rightarrow$ (2) が成り立つ.

- (1)  $\{x_n\}$  は  $x$  に弱収束する.
- (2)  $\{x_n\}$  はノルム有界で:  $\sup_{n \in \mathbb{N}} \|x_n\| < \infty$ ,  $\|x\| \leq \liminf_{n \rightarrow \infty} \|x_n\|$  を満たす.

## 3.3 一様可積分性

一様可積分性は  $L^1(\Omega)$  の部分集合が弱相対コンパクトであることに同値な条件である. 距離空間のコンパクト集合は有界であることに注意.

### 3.3.1 定義

一様可積分ならば可積分である, という状況が成り立つのは, 有限測度空間においてのみである. Lebesgue の優収束定理は, 可積分関数を上界に持って一様有界ならば積分が収束することをいうが, 有界測度空間では一様可積分性を用いてさらに強い結果が成り立つ. また一様可積分性は, 非原子的な測度空間では, 得られる不定積分の族が一様に絶対連続であることと特徴付けられる.

**観察 3.3.1.**  $f \in L^1(X)$  ならば,  $\|f\|1_{\{|f| \geq \alpha\}} \xrightarrow{\alpha \rightarrow \infty} 0 \in L^1(X)$  が成り立ち, 優関数  $|f|$  が存在するから Lebesgue の優収束定理より,

$$\lim_{\alpha \rightarrow \infty} \int_{\{|f| \geq \alpha\}} |f| d\mu = 0$$

も成り立つ. □

**定義 3.3.2.** 族  $\{f_\lambda\} \subset L^1(X)$  が一様可積分とは,

$$\lim_{\alpha \rightarrow \infty} \sup_{\lambda \in \Lambda} \int_{\{|f_\lambda| \geq \alpha\}} |f_\lambda| d\mu = 0$$

が成り立つことをいう.

**要諦 3.3.3.** 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は同値である.

**補題 3.3.4.** 族  $\mathcal{F} := \{f_\lambda\} \subset \mathcal{L}(X)$  について,

- (1)  $(X, \mathcal{A}, \mu)$  が有限のとき,  $\mathcal{F}$  が一様可積分ならば, 各  $f_\lambda$  は可積分である.
- (2)  $\mu(X) = \infty$  のとき,  $\mathcal{F}$  が例え一様に有界であっても可積分でない  $f_\lambda$  が存在し得る.

**[証明].**

- (1) 任意の  $a > 0$  について,

$$\int_X |f| d\mu = \int_{|f| < a} |f| d\mu + \int_{|f| \geq a} |f| d\mu \leq P[|f| < a]a + \int_{|f| \geq a} |f| d\mu$$

が成り立つが, 第1項は定数, 第2項は有界列である.

(2) 第1項が有限にならない。裾が長い関数を考えれば良い。

■

### 3.3.2 積分の族の同程度絶対連続性としての特徴付け

有限測度空間上の絶対連続分布について、 $\mathcal{F} \subset L^1(X)$  が一様可積分であることと同程度絶対連続な有界集合であることは同値。このクラスは、概収束するならば  $L^1$ -収束もする。

**定理 3.3.5 (一様可積分性の特徴付け).**  $(\Omega, \mathcal{F}, P)$  を有限測度空間とする。  $\{X_\lambda\}_{\lambda \in \Lambda} \subset L^1(\Omega)$  について、次の二条件は同値。ただし、(a) $\Rightarrow$ (b) と (b),(c) $\Rightarrow$ (1) に注意。

- (1)  $\{X_\lambda\}_{\lambda \in \Lambda}$  は一様可積分である。
- (2) (a)  $L^1$ -有界性:  $\{X_\lambda\}_{\lambda \in \Lambda}$  は  $L^1(\Omega)$  の有界集合である。
- (b)  $\sup_{\lambda \in \Lambda} P[|X_\lambda| > A] \xrightarrow{A \rightarrow \infty} 0$ .
- (c) 一様に絶対連続:  $P[B] \rightarrow 0 \Rightarrow \sup_{\lambda \in \Lambda} E[|X_\lambda|, B] \rightarrow 0$ .

[証明].

(1) $\Rightarrow$ (a)

$$\begin{aligned} \|X_\lambda\|_1 &\leq E[|X_\lambda|, |X_\lambda| > A] + E[|X_\lambda|, |X_\lambda| \leq A] \\ &\leq E[|X_\lambda|, |X_\lambda| > A] + A \end{aligned}$$

と評価出来るが、第一項は一様可積分性を仮定したとき、 $\lambda \in \Lambda$  に依らずに  $A$  を十分大きく取ることによって有界に出来る。

(a) $\Rightarrow$ (b)

$$\|X_\lambda\|_1 \geq E[|X_\lambda| 1_{|X_\lambda| > A}] \geq AP[|X_\lambda| > A]$$

の評価が成り立つから、 $A \rightarrow \infty$  のとき  $P[|X_\lambda| > A] \rightarrow 0$  が必要。

(1) $\Rightarrow$ (c) 任意の  $\epsilon > 0$  に対して、 $\sup_{\lambda \in \Lambda} E[|X_\lambda|, |X_\lambda| > A] < \epsilon$  を満たす十分大きな  $A$  と、 $P[B] < \epsilon/A$  を満たす  $B \in \mathcal{F}$  を取れば、

$$\begin{aligned} E[|X_\lambda| 1_B] &\leq E[|X_\lambda| 1_{\{|X_\lambda| > A\}}] + E[|X_\lambda| 1_{\{|X_\lambda| \leq A\} \cap B}] \\ &< \epsilon + AP[B] < 2\epsilon \end{aligned}$$

と評価出来る。

(b),(c) $\Rightarrow$ (1)  $B_\lambda := \{|X_\lambda| > A\}$  と置くと、(b) の仮定より、 $P[B_\lambda] \xrightarrow{A \rightarrow \infty} 0$ 。よって、(c) の仮定より、 $\forall \lambda' \in \Lambda \sup_{\lambda \in \Lambda} E[|X_\lambda|, B_{\lambda'}] \rightarrow 0$ 。

■

**要諦 3.3.6.** 一般のノルム空間上において、弱収束列はノルム有界なのであった。この必要条件に (c) を加えれば十分になるというのが、Lebesgue 空間  $L^1(\Omega)$  の特徴である。

**定理 3.3.7 ([一般の有限測度空間上の一様可積分性の特徴付け ([?], Ex 16.9)].**  $(X, A, \mu)$  を有限測度空間とする。  $\{f_n\} \subset \mathcal{L}(X)$  について、次の2条件は同値。

- (1)  $\{f_n\}$  は一様可積分。
- (2) (a)  $\left\{ \int f_n d\mu \right\}_{n \in \mathbb{N}} \subset \mathbb{R}$  は有界。
- (b) 定積分列  $\left\{ \int_{-}^{\cdot} |f_n| d\mu \right\}$  は一様に絶対連続:

$$\forall \epsilon > 0 \exists \delta > 0 \forall A \in \mathcal{A} \mu(A) < \delta \Rightarrow \sup_{n \in \mathbb{N}} \int_A f_n d\mu \leq \epsilon.$$

さらに、測度  $\mu$  が非原子的ならば、(b) $\Rightarrow$ (a) が成り立つ。



### 3.3.3 弱相対点列コンパクト性との同値性

まずは一様可積分性の意義を関数解析的に見てしまうと、 $L^1(\Omega)$  の弱相対コンパクト集合を特徴づける概念に他ならないのである。ここから、収束するならば、極限が可積分であることが分かる。

**定理 3.3.8 (Dunford-Pettis).**  $\{X_n\} \subset L^1(\Omega)$  について、次の2条件は同値：

- (1) 一様可積分である。
- (2) 弱相対点列コンパクトである。

**要諦 3.3.9.**  $L^1(\Omega)$  は回帰的でないので、ノルム閉単位球  $B \subset L^1(\Omega)$  は弱コンパクトではない。有界なだけでは足りず、同程度絶対連続であることが必要。また、 $L^1(\Omega)$  は第2可算とは限らないので、弱相対点列コンパクト性と弱相対コンパクト性とは一致するとは限らない。

**系 3.3.10 ( $L^1$ -有界集合の分布収束極限は可積分).** 一様可積分列  $\{X_n\}$  の分布収束極限  $X$  は可積分である。なお、 $\{X_n\}$  は  $L^1$ -有界ならば、弱収束極限  $X$  は可積分である。

**[証明].**  $X_n \Rightarrow X$  のとき、任意の  $A \in \mathbb{R}$  について、 $x \mapsto |x| \wedge A$  は有界連続関数であるから、 $E[|X_n| \wedge A] \rightarrow E[|X| \wedge A]$ 。よって Fatou の補題より、

$$E[|X| \wedge A] \leq \liminf_{n \rightarrow \infty} E[|X_n| \wedge A] \leq \liminf_{n \rightarrow \infty} E[|X_n|] < \infty.$$

■

### 3.3.4 一様可積分性の十分条件

**命題 3.3.11 (Lebesgue の優収束定理の前提条件は一様可積分性の十分条件).** 列  $\{X_n\} \subset L^1(\Omega)$  は可積分な優関数  $Y \in L^1(\Omega)$  を持つとする： $\sup_{n \in \mathbb{N}} |X_n| \leq Y$ 。このとき、 $\{X_n\}$  は一様可積分である。

**[証明].**

$$\sup_{n \in \mathbb{N}} E[X_n, |X_n| \geq \epsilon] \leq E[Y, |Y| \geq \epsilon] \xrightarrow{\epsilon \rightarrow \infty} 0.$$

■

**定理 3.3.12 (一様可積分性の特徴付け<sup>†1</sup>).**  $\{X_n\} \subset \mathcal{L}(X)$  に対して、次の2条件は同値：

- (1)  $(X_n)$  は一様可積分である。
- (2) 極限において線型よりも速い発散をする Borel 可測関数  $\psi \in \mathcal{L}(\mathbb{R})_+$ ,  $\lim_{|x| \rightarrow \infty} \frac{|x|}{\psi(x)} = 0$  が存在して  $\sup_{n \in \mathbb{N}} E[\psi(X_n)] < \infty$  を満たす

**系 3.3.13 (絶対積率が定める十分条件).** 列  $(f_n)$  は

$$\exists \epsilon > 0 \quad \beta_{1+\epsilon}^{1+\epsilon} = \sup_{n \in \mathbb{N}} \int |f_n|^{1+\epsilon} d\mu < \infty$$

を満たすならば一様可積分である。すなわち、部分空間  $L^p(X) \subset L^1(X)$  ( $p > 1$ ) に入っていれば一様可積分である。

**[証明].** 1.7.13 より、

$$P[|X_n| \geq R] \leq \frac{E[|X_n|^p]}{R^p}.$$

■

<sup>†1</sup> Dellacherie-Meyer, 2章 24p

## 3.3.5 一様可積分列の概収束

可積分関数の概収束列について、一様可積分性と  $L^1$ -収束性とは同値。また、一様可積分列は概収束するならば  $L^1$ -収束するが、逆は言えない。

**命題 3.3.14 (Lebesgue の収束定理の精緻化 (Vitali の収束定理)).**  $\{X_n\} \subset L^1(\Omega)$  は一様可積分とする。このとき、 $(X_n)$  が  $X$  に概収束するならば、 $L^1$ -収束もする。特に、 $X \in L^1(\Omega)$  で、 $\lim_{n \rightarrow \infty} E[X_n] = E[X]$  も成り立つ。

**[証明].**

- (1)  $X \in L^1(\Omega)$  であることは、Fatou の補題より  $E[|X|] \leq \liminf_{n \rightarrow \infty} E[|X_n|] < \infty$  であることから分かる。なお、 $(X_n)$  が一様可積分ならば  $L^1$ -有界である。
- (2) 任意の  $\lambda > 0$  に対して

$$E[|X_n - X|] = E[|X_n - X|1_{\{|X_n - X| < \lambda\}}] + E[|X_n - X|1_{\{|X_n - X| \geq \lambda\}}].$$

第1項は Lebesgue の優収束定理より 0 に収束し、第2項は  $(X_n - X)$  の一様可積分性より 0 に収束する。

- (3) 期待値の収束は、三角不等式  $|E[X_n] - E[X]| \leq E[|X_n - X|] \rightarrow 0$  による。

■

**系 3.3.15 (概収束する可積分関数列の一様可積分性の特徴付け).**  $\{X_n\} \subset L^1(X)$  は  $X \in L(X)$  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)  $X_n \rightarrow X \in L^1(X)$ 。特に  $X \in L^1(X)$   
 (3)  $\|X_n\|_{L^1(X)} \rightarrow \|X\|_{L^1(X)} < \infty$ 。

(2) $\Leftrightarrow$ (3) は Scheffe の補題と呼ばれる。

**[証明].**

- (1) $\Rightarrow$ (2) まず、 $L^1$ -収束列は  $L^1$ -有界であるから、あとは同程度絶対連続性を示せば良い。任意の  $\epsilon > 0$  を取ると、個々の不定積分は絶対連続だから、ある  $\delta_n$  が存在して  $P[B] < \delta_n \Rightarrow E[|X_n|, B] < \epsilon$ 。さらに、三角不等式より

$$E[|X_n|, B] \leq E[|X|, B] + \|X_n - X\|_1$$

であるから、 $\exists N \in \mathbb{N} \forall n \geq N P[B] < \delta \Rightarrow E[|X_n|, B] < 2\epsilon$  も分かる。よって、 $P[B] < \min(\delta, \delta_1, \dots, \delta_N)$  ととれば、 $\forall n \in \mathbb{N} E[|X_n|, B] < 2\epsilon$  と一様に評価出来る。

(2) $\Rightarrow$ (3)

$$\|E[|X_n|] - E[|X|]\| \leq E[||X_n| - |X||] \leq E[|X_n - X|].$$

- (3) $\Rightarrow$ (1) 任意の  $\lambda > 0$  について、 $P[|X| = \lambda] = 0$  ならば、Lebesgue の優収束定理から、

$$E[|X_n|1_{\{|X_n| \geq \lambda\}}] = E[|X_n|] - E[|X_n|1_{\{|X_n| < \lambda\}}] \rightarrow E[|X|] - E[|X|1_{\{|X| < \lambda\}}] = E[|X|, |X| \geq \lambda].$$

$P^X$  の原子は高々可算個だから、変わらず定積分の絶対連続性より、 $\lambda \rightarrow \infty$  のとき 0 に収束する。

■

## 3.3.6 一様可積分列の確率収束

実はいままでの議論で尽きている．一様可積分性の下で確率収束するとき， $L^1$ -収束もするのは，Lebesgue の優収束定理の一般化と見れる．

**補題 3.3.16** .  $X$  を位相空間， $\{x_n\} \subset X$  を点列とする．

- (1)  $x_n \rightarrow x$ .
- (2)  $\{x_n\}$  の任意の部分列が  $x$  に収束する部分列を持つ．

**[証明]** .  $x_n$  は  $x$  に収束しないとすると  $\exists U \in \mathcal{O}(x) \forall n \in \mathbb{N} \exists n_1 \geq n x_{n_1} \notin U$  . すると， $x_n$  の部分列であって， $U$  に終局しないものが構成出来る． ■

**定理 3.3.17** .  $\{X_n\} \subset L^1(\Omega)$  について，次の2条件は同値：

- (1)  $(X_n)$  は一様可積分で， $X$  に確率収束する．
- (2)  $(X_n)$  は  $X$  に  $L^1$ -収束する．

**[証明]** . (1) $\Rightarrow$ (2) を示せば良い． $\{X_n\}$  の任意の部分列は  $X$  に  $L^1$ -収束する部分列を持つことを示せば良い． $\{Y_n\} \subset \{X_n\}$  を部分列とすると，これも  $X$  に確率収束するから， $X$  に概収束する部分列  $\{Z_n\} \subset \{Y_n\}$  を持つ．これは一様可積分だから， $X$  には  $L^1$ -収束もする． ■

**定理 3.3.18 (Generalized Scheffe's lemma<sup>t2</sup>)** .  $\{X_n\} \subset L^1(\Omega)$  は  $X$  に確率収束するとする．このとき，次の3条件は同値：

- (1)  $\|X_n\|_1 \rightarrow \|X\|_1$ .
- (2)  $\{X_n\}$  は一様可積分．
- (3)  $X_n \xrightarrow{L^1} X$ .

## 3.3.7 分布収束列は一様可積分

## Vitali-Hahn-Saks の定理

$f_n$  が定める不定積分が，集合関数  $\mathcal{A} \rightarrow \mathbb{R}$  として各点収束するならば一様可積分で，極限も不定積分であることを論じる．これは  $f_n$  を密度関数としてもつ分布列  $\mu_n$  が各点収束するならば一様可積分であることを含意している．

**定理 3.3.19 (Vitali-Hahn-Saks)** .  $(S, \mathcal{M}, \mu)$  を  $\sigma$ -有限とする． $\{f_n\} \subset L^1(S)$  の不定積分の列  $\{F_n(E) := \mu[f_n 1_E]\}_{n \in \mathbb{N}} \subset \text{Map}(\mathcal{M}, \mathbb{R})$  が各点収束極限  $F : \mathcal{M} \rightarrow \mathbb{R}$  をもつならば， $\{F_n\}$  は一様に絶対連続であり， $F$  も  $\mu$  に関するある不定積分である．

**系 3.3.20 (分布収束列は一様可積分である [Billingsley, 1999]Th'm 3.6)** .  $X, \{X_n\} \subset L^1(S)$  を確率変数とする．分布収束  $X_n \Rightarrow X$ ，すなわち弱収束し，かつ  $\|X_n\|_{L^1(\Omega)} \rightarrow \|X\|_{L^1(\Omega)}$  ならば， $\{X_n\}$  は一様可積分．

**要諦 3.3.21** .  $\{X_n\}$  は分布収束する  $L^1$ -確率変数であるから，弱収束列である．これが弱コンパクトでもあるためには， $E[X_n] \rightarrow E[X]$  が必要． $L^1(S)_+$  は第2可算とは限らないので，点列コンパクトならばコンパクトとは言えないことに注意．

**系 3.3.22 (弱収束の Saks 流の特徴付け)** . 可積分関数列  $\{f_n\} \subset L^1(\Omega)$  について，次の2条件は同値：

- (1)  $\{f_n\}$  は弱収束する．
- (2) (a)  $\{f_n\} \subset L^1(\Omega)$  はノルム有界である．  
(b)  $\forall A \in \mathcal{F} \int_A f_n d\mu \in \mathbb{R}$  が存在する．

<sup>t2</sup> <https://math.stackexchange.com/questions/4401886/generalizing-scheffes-lemma-using-only-convergence-in-probability>

## 3.4 L の位相まとめ

### 3.4.1 確率収束の十分条件

測度収束とは、Egorov の定理が示唆する状況に他ならない。

**定義 3.4.1** .  $\{X_n\} \subset L(\Omega)$  が  $X \in L(\Omega)$  に確率収束するとは、任意の  $\epsilon > 0$  に対して、

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \epsilon] = 0$$

が成り立つことをいう。

**命題 3.4.2** .  $\{X_n\} \subset \mathcal{L}(\Omega)$  とする。

- (1)  $X$  に概収束するならば、確率収束する。
- (2)  $X$  に  $L^p$ -収束するならば、確率収束する。
- (3)  $X$  に  $L^\infty$ -収束するならば、概収束する。<sup>†3</sup>

**[証明]** .

- (1) (a) 事象の包含関係に注目する本質論  $A_{m,j} := \left\{ |X_m - X| \leq \frac{1}{j} \right\}$  とすると、 $\forall j \in \mathbb{N} \quad P[A_{m,j}] \xrightarrow{m \rightarrow \infty} 1$  を示せば良い。いま、

$$\left\{ \lim_{n \rightarrow \infty} X_n = X \right\} = \left\{ \forall j \in \mathbb{N} \exists n \in \mathbb{N} \forall m \geq n \quad |X_m - X| \leq \frac{1}{j} \right\} = \bigcap_{j \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{m=n}^{\infty} A_{m,j}$$

と表せるから、 $\left\{ \lim_{n \rightarrow \infty} X_n = X \right\} \subset \bigcup_{n \in \mathbb{N}} \bigcap_{m=n}^{\infty} A_{m,j}$  であり、 $P$  の連続性より、

$$1 = P \left[ \bigcup_{n \in \mathbb{N}} \bigcap_{m=n}^{\infty} A_{m,j} \right] = \lim_{n \rightarrow \infty} P \left[ \bigcap_{m=n}^{\infty} A_{m,j} \right] \leq \lim_{n \rightarrow \infty} P[A_{m,j}] = 1.$$

- (b) 補集合を評価する 集合  $A_n := \{|X_n - X| \geq \epsilon\}$  と定めると、 $\limsup_{n \rightarrow \infty} A_n$  が零集合であることを示せば、 $\lim_{n \rightarrow \infty} A_n$  も零集合であることが従い、これは確率収束を意味する。いま

$$\limsup_{n \rightarrow \infty} A_n = \{|X_n - X| \geq \epsilon \text{ i.o.}\} \subset \left\{ \lim_{n \rightarrow \infty} X_n \neq X \right\}.$$

で、右辺は零集合である。

- (c) Lebesgue の収束定理による方法  $1_{\{|X_n - X| > \epsilon\}} : \Omega \rightarrow \{0, 1\}$  を、 $\omega \mapsto |X_n(\omega) - X(\omega)| \mapsto 1_{(\epsilon, \infty)}(|X_n - X|)$  とよみかえることで、

$$P[|X_n - X| > \epsilon] = E[1_{\{|X_n - X| > \epsilon\}}] = E[1_{(\epsilon, \infty)}(|X_n - X|)] \xrightarrow{n \rightarrow \infty} 0.$$

- (d) Egorov の定理による方法  $X_n \rightarrow X$  a.s. のとき、 $\forall \delta > 0 \exists A \in \mathcal{F} \quad P[A] > 1 - \delta$  かつ  $X_n \xrightarrow{\text{uniformly}} X$  on  $A$ . よって、

$$\forall \epsilon > 0 \quad \forall \delta > 0 \quad \exists N \in \mathbb{N} \quad \forall n \geq N \quad P[|X_n - X| > \epsilon] \leq P[A^c] < \delta.$$

- (2) Markov の不等式より、 $\forall p > 0 \quad \forall \epsilon > 0 \quad P[|X| \geq \epsilon] \leq \frac{1}{\epsilon^p} E[|X|^p] \rightarrow 0.$

(3)

■

<sup>†3</sup> 高信 [高信敏, 2015]

### 3.4.2 標準確率空間上の収束の様子

例 3.4.3 (反例). 確率空間を  $([0, 1], \mathcal{B}([0, 1]), I)$  とする.

- (1)  $X_{nk}(\omega) := 1_{(\frac{k-1}{n}, \frac{k}{n})}(\omega)$  ( $k \in [n]$ ) を  $(n, k)$  に関する辞書式順序で並べたものは, 0 に確率収束し,  $L^\infty$ -収束するから  $L^p$ -収束もする ( $1 \leq p \leq \infty$ ). が, 概収束しない.
- (2)  $X_n(\omega) := n1_{(0, \frac{1}{n})}(\omega)$  とすると, 0 に確率収束し, 概収束もするが,  $L^1$ -収束はしない.
- (3)  $X_n$  を,  $1_{[0, 1/2]}$  と  $1_{[1/2, 1]}$  を交互に並べたものとする. すると, いずれも  $\{0, 1\}$  上に Rademacher 分布を押し出すために法則収束するが, 確率収束はしない.

□

例 3.4.4.

- (1)  $\{f_n\} \subset L^1([0, 1])$  であって,  $f_n \rightarrow 0$  in  $L^1([0, 1])$  であるが, 任意の点で 0 に収束しない例.
- (2)  $\{f_n\} \subset C([0, 1])$  であって,  $f_n \rightarrow 0$  であるが, 一様収束はせず, また  $L^1([0, 1])$ -収束もしない例.

□

[証明].

- (1) [伊藤清, 1991] にあるような,  $[0, 1]$  の区間の列で, 幅が縮小するが任意の点  $x \in [0, 1]$  を無限回訪れる列を取り, その定義関数は例になる.
- (2) 面積が同じ三角形であって, その  $[0, 1]$  上の底辺がどんどん縮小していくものをとればよい.

■

### 3.4.3 確率収束の必要条件

これは一般の可分な距離空間上で成り立つ.

命題 3.4.5.  $(X_n)$  が  $X$  に確率収束するならば, 弱収束する.

[証明].

Portmanteau 定理 2.2.5 による方法 任意に  $g \in C_c(\mathbb{R})$  を取ると, これは一様連続であるから, 任意の  $\epsilon > 0$  に対して  $\delta > 0$  が存在して,

$$|x - y| \leq \delta \Rightarrow |\varphi(x) - \varphi(y)| \leq \epsilon.$$

これに対して  $X_n \xrightarrow{P} X$  より, ある  $N \in \mathbb{N}$  が存在して  $\forall n \geq N \ P[|X_n - X| \geq \delta] \leq \epsilon$ . よって,

$$\begin{aligned} |E[g(X_n)] - E[g(X)]| &\leq E[|g(X_n) - g(X)|] \\ &\leq E[|g(X_n) - g(X)|1_{\{|X_n - X| \leq \delta\}}] + 2\|g\|_\infty P[|X_n - X| > \delta] \\ &\leq \epsilon + 2\|g\|_\infty \epsilon. \end{aligned}$$

特性関数と Glivenko の定理

一般化された Lebesgue の優収束定理 任意の  $f \in C_b(\mathbb{R})$  に対して,  $\{f(X_n)\}$  は可積分な優関数  $\|f\|_\infty$  を持つので, 特に一様可積分である. 連続写像定理より,  $f(X_n) \xrightarrow{P} f(X)$ . よって, 確率収束に関する Lebesgue の優収束定理から  $E[f(X_n)] \rightarrow E[f(X)]$ .

■

補題 3.4.6 (定数への収束は同値).  $\{P_n\} \subset P(S), a \in S$  について, 次の 2 条件は同値:

- (1)  $P_n \xrightarrow{P} a$ .
- (2)  $P_n \xrightarrow{w} a$ .

**[証明].**  $a = 0$  の場合について示せば十分. 一般に, 確率収束する確率変数は弱収束するから,  $X_n$  の分布がデルタ測度  $\delta_0$  に弱収束するとして,  $X_n$  の 0 への確率収束を示せば良い.  $X_n$  の分布関数を  $F_n$  で表すと, 仮定より  $F_n \Rightarrow F := 1_{[0, \infty)}$  である. いま, 任意の  $\epsilon > 0$  に対して,

$$P[|X_n| \geq \epsilon] = F_n(-\epsilon) + (1 - F_n(\epsilon)) + P^{X_n}[\{\epsilon\}].$$

である. まず,  $\delta_0$  の連続点は  $\mathbb{R} \setminus \{0\}$  であるから,  $\delta_0$  の分布関数を  $F$  とすると,  $F_n(-\epsilon) \rightarrow F(-\epsilon) = 0$  かつ  $1 - F_n(\epsilon) \rightarrow 1 - F(\epsilon) = 0$ . また,  $\delta_0(\{\epsilon\}) = \delta_0(\{0\}) = 0$  より,  $P^{X_n}[\{\epsilon\}] \rightarrow \delta_0(\{\epsilon\}) = 0$ . 以上を併せると, 任意の  $\epsilon > 0$  に対して  $P[|X_n| \geq \epsilon] \rightarrow 0$ . すなわち,  $X_n$  は 0 に確率収束する. ■

### 3.4.4 確率収束列の概収束部分列の存在

**命題 3.4.7.** 確率変数列  $\{X_n\} \subset \mathcal{L}(\Omega)$  について, 次の 2 条件は同値:

- (1)  $(X_n)$  は  $X$  に確率収束する.
- (2) 集合  $\{X_n\}$  の任意の列は,  $X$  に概収束する部分列を含む.

**[証明].**

(1)  $\Rightarrow$  (2)  $(X_n)$  を  $\{X_n\}$  の任意の部分列とする. ある部分列  $(Y_n)$  が存在して,  $P[|Y_n - X| > 2^{-n}] < 2^{-n}$  を満たす. というのも, 十分大きな  $N \in \mathbb{N}$  に対して,  $Y_n := X_{N+n}$  とすればよい. すると  $\sum_{n \in \mathbb{N}} P[|Y_n - X| > 2^{-n}] < 1$  であるから Borel-Cantelli の補題より,

$$P[|Y_n - X| > 2^{-n} \text{ i.o.}] = 0$$

だから, 概収束する.

(2)  $\Rightarrow$  (1) 背理法による. ある  $\epsilon > 0$  が存在して  $P[|X_n - X| > \epsilon]$  が 0 に収束しないならば,  $\sum_{n \in \mathbb{N}} P[|X_n - X| > \epsilon] = \infty$ . 実際, この級数が収束するならば, Cauchy 列でもあることから,  $P[|X_n - X| > \epsilon] \xrightarrow{n \rightarrow \infty} 0$  が従ってしまう. よって, Borel-Cantelli の補題より,  $P[|X_n - X| > \epsilon \text{ i.o.}] = 1$ . これは, 概収束しない部分列が選べることを意味する. ■

**系 3.4.8 (確率収束の連続写像定理).**  $X_n \xrightarrow{P} X \in \mathcal{L}(X)$  とする.

- (1)  $X_n \xrightarrow{P} Y \in \mathcal{L}(X)$  でもあるなら,  $X = Y$  a.s.
- (2) 連続関数  $f: \mathbb{R} \rightarrow \mathbb{R}$  は確率収束を保つ.

**[証明].**

- (1)  $(X_n)$  にはある部分列が存在して  $X$  に概収束し, さらにその部分列が存在して  $Y$  にも概収束する. よって,  $X = Y$  a.s.
- (2) 任意に部分列  $\{f(X_{n_k})\} \subset \{f(X_n)\}$  を取ると,  $X$  に概収束する部分列  $\{Y_n\} \subset \{X_{n_k}\}$  が取れる. これについて,  $f(Y_n)$  は  $X$  に概収束するから,  $\{f(Y_n)\}$  も  $f(X)$  に概収束する. ■

**要諦 3.4.9.**  $L^p$ -収束列が概収束部分列を持つことは有限測度空間上では自明だが, これは一般の Lebesgue 空間  $L^p(\Omega)$  上では言える.

**定理 3.4.10 (概収束部分列の存在).**  $\{f_n\} \subset L^p(\Omega)$  ( $\Omega \subset \mathbb{R}^n$ ) は  $f$  に  $L^p$ -収束するとする. このとき, ある部分列と  $h \in L^p(\Omega)$  が存在して,  $h$  を優関数として概収束する:  $\forall_{k \in \mathbb{N}} |f_{n_k}| \leq h$  a.e. かつ  $f_{n_k} \rightarrow f$  a.e.

### 3.4.5 確率収束の特徴付け

**命題 3.4.11.**  $\{X_n\} \subset L(\Omega)$  について, 次は同値:

- (1)  $X_n \xrightarrow{P} X$ .

- (2) ある  $R > 0$  が存在して,  $\lim_{n \rightarrow \infty} E[|X_n - X| \wedge R] = 0$ .  
 (3) 任意の  $R > 0$  に対して,  $\lim_{n \rightarrow \infty} E[|X_n - X| \wedge R] = 0$ .

### 3.4.6 確率収束を定める距離

**命題 3.4.12** (確率収束の特徴付け).  $\{X_n\} \subset \mathcal{L}(X)$  について, 次の2条件は同値:

- (1)  $\exists_{X \in \mathcal{L}(X)} X_n \xrightarrow{P} X$ .  
 (2)  $\forall_{\epsilon > 0} \exists_{N \in \mathbb{N}} \forall_{n, m \geq N} P[|X_m - X_n| > \epsilon] < \epsilon$ .

(2) の条件が成り立つとき, 確率的 **Cauchy 列** (fundamental in probability) という.

**定義 3.4.13**.  $\mathcal{L}(X)$  上に擬距離

$$d(X, Y) := \inf \{ \epsilon > 0 \mid P[|X - Y| \geq \epsilon] \leq \epsilon \}$$

を定めると,  $L(X)$  上の距離になる.

**命題 3.4.14**.

- (1)  $d$  について収束することと確率収束することとは同値.  
 (2) 距離空間  $(L(X), d)$  は完備可分である.

**要諦 3.4.15**. 次の距離も同値になる.

- (1)  $d_2(X, Y) := \inf \{ P[|X - Y| > \alpha] + \alpha \in \mathbb{R}_+ \mid \alpha > 0 \}$   
 (2)  $d_3(X, Y) := E[|X - Y| \wedge 1]$   
 (3)  $d_4(X, Y) := E \left[ \frac{|X - Y|}{1 + |X - Y|} \right]$

$$\forall_{\epsilon \in (0, 1)} \epsilon P[|X - Y| > \epsilon] \leq d_3(X, Y) \leq \epsilon + P[|X - Y| > \epsilon].$$

が成り立つ. いずれの距離も, ある可測関数  $f \in L(\Omega)$  から可積分関数を作り出す標準的な方法であり, 極めて自然な位相であることが理解できる. 例えば高信 [?] では  $\|X\|_0 := E \left[ \frac{|X|}{1 + |X|} \right]$  としている.

### 3.4.7 分布収束列の修正概収束部分列の構成

**定理 3.4.16** (Skorohod の概収束実現定理). 実確率変数列  $X_n$  は  $X$  に分布収束するとする. このとき, 確率空間  $(0, 1)$  上に同じ分布を誘導する実確率変数  $\tilde{X}_n, \tilde{X}$  が存在して,  $\tilde{X}_n \rightarrow \tilde{X}$  a.s. が成り立つ.

**定理 3.4.17** (一般の距離空間の場合). 距離空間  $(S, d)$  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  $S$ -値確率変数の列  $(X_n)$  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.

### 3.4.8 $L^1$ -弱相対コンパクトと一様可積分性は同値

$L^1(\mathbb{R})$  において, 弱相対コンパクトであることと一様可積分であることは同値.

**議論 3.4.18**.  $L(X)$  での各点収束と  $L^1(X)$  でのノルム収束との関係は全くの混沌に思える.  $\{X_n\} \subset L^1(X)$  が概収束しても  $X \in L^1(X)$  とは限らないし, ましてや  $L^1$ -収束するとは限らない. しかしこの状況は「一様可積分性」の仮定で一気に解決する.

また  $L^1$ -収束するからといって概収束するとは限らないことは, 積分論の

$$\int |f_n| dx \rightarrow 0 \Rightarrow f_n \rightarrow 0 \text{ a.e.}$$

からよく学んでいる.



### 3.4.9 概収束を定める位相

概収束を定める位相は濃すぎて距離付け可能ではない。  $L^p$ -位相と順序は定まらない。

**命題 3.4.19.** 概収束を定める位相は存在しない。

**[証明].** 仮に概収束を定める位相が存在したとすると、確率収束の特徴付け 3.3.17 により、その位相は確率収束の位相に一致する。これは矛盾。 ■

## 3.5 確率収束

確率収束の定義  $\forall M > 0 \ P[|X_n| \geq M] \xrightarrow{n \rightarrow \infty} 0$  について、 $n$  と  $M$  の役割を入れ替えて、 $n \in \mathbb{N}$  について一様に  $M \rightarrow \infty$  としたとき 0 に収束することが一様緊密性である。確率収束列、さらに法則収束列は一様に緊密である。

### 3.5.1 Lebesgue の優収束定理の一般化

この結果はさらに一般化される。

**定理 3.5.1.** 一様可積分な列  $\{X_n\} \subset L^1(X)$  は  $X \in L^1(X)$  に確率収束するとする（可積分な優関数を持つならばこれを満たす）。このとき、 $E[X_n] \rightarrow E[X]$ 。

**[証明].**  $\{E[X_n]\} \subset \mathbb{R}$  の任意の部分列  $\{E[X_{n_k}]\}$  に対して、 $E[X]$  に収束する部分列が存在することを示せば良い。確率収束の特徴付けより、ある部分列  $\{Y_n\} \subset \{X_{n_k}\}$  が存在して、 $X$  に概収束するから、一様可積分な概収束列に関する Lebesgue の優収束定理より、 $E[Y_n] \rightarrow E[X]$ 。 ■

### 3.5.2 連続写像定理

**命題 3.5.2 (確率収束に関する連続写像定理).**  $(S_1, d_1), (S_2, d_2)$  を可分距離空間、 $X_n, X$  を  $S_1$ -値確率変数とする。  $h : S_1 \rightarrow S_2$  は可測写像で、 $P[X \in C] = 1$  を満たす可測集合  $C \in \mathcal{G}(S_1)$  上で連続であるものとする。このとき、

$$X_n \xrightarrow{P} X \implies h(X_n) \xrightarrow{P} h(X).$$

**[証明].**  $\{h(X_n)\}$  の任意の部分列が、概収束する部分列を持つことを示せば良い。さらなる部分列  $\{Y_n\} \subset \{X_{n_k}\} \subset \{X_n\}$  が存在して、 $X$  に概収束する。このとき、

$$1 = P\left[\lim_{n \rightarrow \infty} d_1(Y_n, X) = 0\right] = P\left[\lim_{n \rightarrow \infty} d_1(Y_n, X) = 0, X \in C\right] \leq P\left[\lim_{n \rightarrow \infty} d_2(h(Y_n), h(X)) = 0\right] = 1.$$

**系 3.5.3.**  $X_n, Y_n, X, Y \in L(\Omega)$  について、 $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y$  とする。  $X_n Y_n \xrightarrow{P} XY$  でもある。

### 3.5.3 積の収束

**定理 3.5.4.**

- (1)  $X_n \xrightarrow{d} X$  かつ  $Y_n \xrightarrow{P} c$  ならば、 $(X_n, Y_n) \xrightarrow{d} (X, c)$ 。
- (2)  $X_n \xrightarrow{P} X$  かつ  $Y_n \xrightarrow{P} Y$  ならば、 $(X_n, Y_n) \xrightarrow{P} (X, Y)$ 。

**注 3.5.5.** 一方で、結合分布の弱収束は、各成分の弱収束より強い 3.6.10。これも、各成分の分布だけでは結合分布が定まらないことに起因する。

**系 3.5.6.**  $X_n, Y_n, X, Y \in L(\Omega)$  は  $X_n \xrightarrow{P} X$  かつ  $Y_n \rightarrow Y$  を満たすとする。このとき、 $X_n Y_n \xrightarrow{P} XY$ 。

**[証明].** 連続写像定理からすぐに従うが、直接も示せる。

直接の評価 いま、

$$\begin{aligned} |X_n Y_n - XY| &\leq |X_n Y_n - XY_n| + |XY_n - XY| \\ &= |Y_n| |X_n - X| + |X| |Y_n - Y|. \end{aligned}$$

と分解できるから、それぞれの項の確率収束を示せば十分。それに当たっては、 $|Y_n|, |X|$  の値の大きさ ( $M > 0$  より大きい小さいか) に応じて場合分けすることが肝要になる。

第1項 任意の  $M > 0$  について、

$$\begin{aligned} |Y_n| |X_n - X| \geq \epsilon &\Rightarrow |X_n - X| \geq \frac{\epsilon}{M} \quad \vee \quad |Y_n| \geq M \\ &\Rightarrow |X_n - X| \geq \frac{\epsilon}{M} \quad \vee \quad |Y| \geq \frac{M}{2} \quad \vee \quad |Y_n - Y| \geq \frac{M}{2}. \end{aligned}$$

が成り立つから、

$$P[|Y_n| |X_n - X| \geq \epsilon] \leq P[|X_n - X| \geq \epsilon/M] + P[|Y| \geq M/2] + P[|Y_n - Y| \geq M/2].$$

と評価できる。任意の  $\epsilon > 0$  に対して、 $M > 0$  を十分大きく取り、それに合せて  $n$  を十分大きくとると、右辺は任意に 0 に近づけることが出来る。

第2項 任意の  $M > 0$  について、

$$|X| |Y_n - Y| \geq \epsilon \Rightarrow |Y_n - Y| \geq \frac{\epsilon}{M} \quad \vee \quad |X| \geq M.$$

が成り立ち、

$$P[|X| |Y_n - Y| \geq \epsilon] \leq P[|Y_n - Y| \geq \epsilon/M] + P[|X| \geq M].$$

と評価でき、 $M > 0$  を十分大きく取り、それに合せて  $n \in \mathbb{N}$  を十分大きくとると右辺は 0 にいくらでも近づけることが出来る。

**特徴付けを用いる** 任意の部分列  $\{X_{n_k} Y_{n_k}\}_{k \in \mathbb{N}}$  を取る。仮定から、 $(X_{n_k})_{k \in \mathbb{N}}$  の部分列  $X_i$  であって  $X_i \xrightarrow{\text{a.s.}} X$  を満たすものと、 $(Y_{n_k})_{k \in \mathbb{N}}$  の部分列  $Y_i$  であって  $Y_i \xrightarrow{\text{a.s.}} Y$  を満たすものが存在するから、これについて  $X_i Y_i \xrightarrow{\text{a.s.}} XY$  が成り立つ。

■

### 3.5.4 同じ極限に収束するための条件

確率収束は距離が定めることを考えると、積に関して良い振る舞いをする。

**命題 3.5.7 (確率収束の遺伝).**  $(X_n, Y_n)$  を  $S \times S$  上の確率変数列とし、積距離を  $\rho : S \times S \rightarrow \mathbb{R}_+$  とする。  $X_n \Rightarrow X$  かつ  $\rho(X_n, Y_n) \Rightarrow 0$  ならば、 $Y_n \Rightarrow X$  である。

**定理 3.5.8.**  $(X_{un}, X_n)$  は二重列とし、 $X_{un} \Rightarrow Z_u \Rightarrow X$  を満たすとする。このとき、 $\forall \epsilon > 0 \lim_u \limsup_n P[\rho(X_{un}, X_n) \geq \epsilon] = 0$  ならば、 $X_n \Rightarrow X$  も成り立つ。

### 3.5.5 漸近同等性と確率有界性

一様に緊密な測度を押し出す確率変数列を**確率有界**ともいう。

**定義 3.5.9 (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。

**要諦 3.5.10.** これは、実確率変数  $X_n$  が押し出す測度の列  $\{P^{X_n}\}$  が一様に緊密であることを言っている。

**補題 3.5.11** (分布収束列は一様に緊密である).  $X_n \xrightarrow{d} X$  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].** 一様緊密性 2.3.9 にて議論した. また, 漠収束の特徴付け 2.2.15 から従う. ■

**補題 3.5.12** (一様に緊密な確率変数列の積も和も一様に緊密).  $(X_n), (Y_n)$  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 3.5.13.** 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,  $X_n = O_p(r_n) :\Leftrightarrow \frac{X_n}{r_n} = O_p(1)$ .

## 3.6 分布収束

### 3.6.1 Lebesgue の優収束定理の一般化

**定理 3.6.1.** 一様可積分な列  $\{X_n\} \subset \mathcal{L}(X)$  は  $X \in \mathcal{L}(X)$  に弱収束するとする. このとき  $X \in L^1(X)$  で,  $E[X_n] \rightarrow E[X]$ .

**定理 3.6.2.**  $X_n \Rightarrow X$  ならば,  $E[X] \leq \liminf_n E[X_n]$ .

**定理 3.6.3.**  $\{X_n\} \subset L^1(X)$  は一様可積分で  $X_n \Rightarrow X$  を満たすとする. このとき,  $X$  も可積分で  $E[X_n] \Rightarrow E[X]$ .

**定理 3.6.4** (優収束定理).  $Y \in L^1(\Omega)_+$  は  $\forall n \in \mathbb{N} \quad \forall t > 0 \quad P[|X_n| \geq t] \leq P[Y \geq t]$  を満たすとする. このとき,  $X_n \Rightarrow Y$  は  $E[X_n] \rightarrow E[X]$  を含意する.

### 3.6.2 弱収束の十分条件

**命題 3.6.5** (密度関数が概収束するならば弱収束する).  $\{X_n\}, X \subset \mathcal{L}(X)$  の分布は密度関数  $f_n, f$  を持つとする.  $f_n \rightarrow f$  a.s. ならば,  $X_n \Rightarrow X$ .

### 3.6.3 連続写像定理

**記法 3.6.6.**  $(S_i, d_i)$  ( $i = 1, 2$ ) を距離空間,  $\nu, \nu_n \in P(S_1)$  を確率測度とする.

**命題 3.6.7** (連続写像は弱連続作用素を引き起こす).  $T : S_1 \rightarrow S_2$  は連続であるとする. このとき,  $T_* : P(S_1) \rightarrow P(S_2)$  は確率測度の弱収束の位相について連続である. すなわち,  $\nu_n \Rightarrow \nu$  ならば,  $T_* \nu_n \Rightarrow T_* \nu$ .

**[証明].**  $T_* : C_b(S_2) \rightarrow C_b(S_1)$  が定まるため, 任意の  $f \in C_b(S_2)$  について引き戻して考えれば

$$E_{T_* \nu_n}[f] = E_{\nu_n}[f \circ T] \rightarrow E_{\nu}[f \circ T] = E_{T_* \nu}[f].$$

**系 3.6.8** (擬連続写像列に対する一般化).  $T, T_n : S_1 \rightarrow S_2$  は可測写像で, ある  $P^X$ -充満な可測集合  $C \in \mathcal{B}(S_1)$  が存在して, 条件

(C) 任意の  $x \in C$  に収束する任意の列  $\{x_n\} \subset S_1$  に対して,  $T_n(x_n) \xrightarrow{n \rightarrow \infty} T(x)$ .

が満たされるとする. このとき,  $\nu_n \Rightarrow \nu$  ならば,  $T_* \nu_n \Rightarrow T_* \nu$ .

**補題 3.6.9.** 点  $x \in C$  に対する条件 (C) は, 次の条件 (C') と同値である.

(C')  $\forall \epsilon > 0 \quad \exists (n_0, \delta) \in \mathbb{N} \times (0, \infty) \quad \forall n \geq n_0 \quad \forall x' \in U_\delta(x) \quad d_2(T(x), T_n(x')) > \epsilon$

## 3.6.4 積の収束

**定理 3.6.10 (Slutsky).**  $X, X_n, Y_n \in \mathcal{L}(\Omega, M_{nm}(\mathbb{R}))$  ( $n, m \geq 1$ ) を確率ベクトル・行列.  $c \in M_{nm}(\mathbb{R})$  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)  $Y_n X_n \xrightarrow{d} cX$ .
- (4)  $c$  が正則行列のとき,  $Y_n^{-1} X_n \xrightarrow{d} c^{-1} X$ .

**系 3.6.11.**  $X, X_n, Y_n$  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.

## 3.6.5 弱収束の遺伝と独立性

**命題 3.6.12.**  $X_n, Y_n, X, Y \in L(\Omega)$  について,  $X_n$  と  $Y_n$  は独立で,  $X_n \xrightarrow{d} X$  かつ  $Y_n \xrightarrow{d} Y$  とする. このとき, 次は同値:

- (1)  $(X_n, Y_n) \xrightarrow{d} (X, Y)$ .
- (2)  $X \perp\!\!\!\perp Y$ .

**例 3.6.13.**  $\{\xi_i\}_{i=1}^\infty$  を  $N(0, 1)$  の独立同分布列とし,

$$X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i, \quad Y_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^2 - 1).$$

と定めると,

- (1)  $X_n, Y_n$  は独立でない.
- (2) 法則収束極限  $(X_n, Y_n) \xrightarrow{d} (X, Y)$  が存在して,  $X \perp\!\!\!\perp Y$  である.

□

**[証明].**

- (1) 仮に独立とすると,

$$E[X_n^2 Y_n] = E[X_n^2] E[Y_n] = E[X_n^2] \cdot 0 = 0.$$

が必要であるが, これは直接の計算

$$\begin{aligned} E[X_n^2 Y_n] &= E \left[ \frac{1}{n^{3/2}} \sum_{i,j,k=1}^n \xi_i \xi_j (\xi_k^2 - 1) \right] \\ &= \frac{1}{n^{3/2}} \sum_{i=1}^n E[\xi_i^4 - \xi_i^2] = \frac{2}{\sqrt{n}}. \end{aligned}$$

に矛盾する.

- (2)

■

**命題 3.6.14.**  $X_n, Y_n, X, Y \in L(\Omega)$  について,  $X_n$  と  $Y_n$  は独立で,  $X_n \xrightarrow{a.s.} X$  かつ  $Y_n \xrightarrow{d} Y$  とする. このとき,  $X_n Y_n \xrightarrow{d} XY$  である.

## 3.6.6 積空間上の弱収束

**定理 3.6.15.**  $T := S_1 \times S_2$  が可分ならば, 次の 2 条件は同値:

- (1)  $P_n^1 \times P_n^2 \Rightarrow P^1 \times P^2$ .

$$(2) P_n^1 \Rightarrow P^1 \text{ かつ } P_n^2 \Rightarrow P^2.$$

**定理 3.6.16.**  $S \times S$  を可分とする  $X_n \Rightarrow X$  かつ  $Y_n \Rightarrow a$  ならば,  $(X_n, Y_n) \Rightarrow (X, a)$ .

**定理 3.6.17.**  $S \times S$  を可分で,  $X, Y$  を独立,  $Y \stackrel{d}{=} Z$  かつ  $\rho(Y_n Z) \Rightarrow 0$ ,  $\mathcal{F}_0 \subset \mathcal{F}$  を集合体とする. 次の2条件が成り立つならば,  $(X_n, Y_n) \Rightarrow (X, Y)$ .

- (1)  $\forall A \in \mathcal{P}(S) \forall E \in \mathcal{F}_0 P[X \in \partial A] = 0 \Rightarrow P[(X_n \in A) \cap E] \rightarrow P[X \in A]P[E].$
- (2)  $\forall n \in \mathbb{N} X_n \in \mathcal{L}_{\mathcal{F}_1}(\Omega).$

### 3.6.7 デルタ法

パラメータ  $\theta \in \mathbb{R}^k$  の推定量  $T_n: \mathcal{X}^n \rightarrow \mathbb{R}^k$  と, 関数  $\varphi: \mathbb{R}^k \rightarrow \mathbb{R}$  を考える.  $T_n \xrightarrow{P} \theta$  が成り立つとき,  $\varphi$  が  $\theta$  で連続ならば  $\varphi(T_n) \xrightarrow{P} \varphi(\theta)$  も成り立つ. では,  $\sqrt{n}(T_n - \theta) \xrightarrow{d} \varphi'(\theta)\sqrt{n}(T_n - \theta)$  も成り立つのであろうか?

**定理 3.6.18 (デルタ法).** 次の条件を仮定する.

- (D1) 関数  $f: \mathbb{R}^{d_1} \supset A \rightarrow \mathbb{R}^{d_2}$  は  $c \in A$  で微分可能であるとする.
- (D2)  $\forall n \in \mathbb{N} P[X_n \in A] = 1$  を満たす  $d_1$  次元確率変数列  $\{X_n\} \subset \mathbb{R}^{d_1}$  に対して,  $\infty$  に発散する係数列  $\{b_n\} \subset \mathbb{R}$  が定める線型変換は  $b_n(X_n - c) \xrightarrow{d} Z \in \mathbb{R}^{d_1}$  を満たすとする.

このとき,

$$b_n[f(X_n) - f(c)] - \partial_x f(c)b_n(X_n - c) \xrightarrow{P} 0 \quad (n \rightarrow \infty).$$

$\partial_x f(c)$  は  $d_2 \times d_1$ -行列であることに注意.

**要諦 3.6.19.** したがって Slutsky の定理の系から特に,  $b_n[f(X_n) - f(c)] \xrightarrow{d} \partial_x f(c)Z$  ( $n \rightarrow \infty$ ) である. 列  $(b_n[f(X_n) - f(c)])$  は  $(b_n \partial_x f(c)(X_n - c))$  に漸近同等であることがわかった (差が  $o_p(1)$ ). これは1次の Taylor 展開に拠ると思えることが名前の由来である. これを Banach 空間上で行う技法を関数デルタ法という.

## 3.7 安定収束

コンパクト性の射影を用いた特徴付けのように, 確率収束を定める位相に対する振る舞いに関して考える. 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.

### 3.7.1 定義と例

**定義 3.7.1 (stable convergence).** 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 3.7.2.** 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. □

## 3.7.2 安定収束極限定理

独立同分布列  $\{X_n\} \subset L^1(\Omega)$  が  $X_1 \notin L^2(\Omega)$  であるとき,  $\{b_n\}, \{a_n\}$  をうまく取れば  $\frac{1}{a_n} \left( S_n - \sum_{j=1}^n b_j \right)$  が収束するが, これは正規分布ではない. これを安定分布という.

## 第 4 章

# 独立性

独立確率変数の和の演算は、特性関数の積の演算に写されるため、分布が不明でも特性関数の性質から多くの摩訶不思議な性質を持ち還ってくる事が出来る。

- (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。一方で Markov, Bernstein による現代的な研究は、完全な独立性よりも条件式の数減らして得られる Markov 性の条件を課して考察している。
- (2) 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 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。

#### 4.1.1 動機

**議論 4.1.1** (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、 $E[X, B] = E[1_B X]$  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 4.1.2** (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  $\sigma$ -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  $\sigma$ -代数を考える。

$(B_i)$  を事象による  $\Omega$  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  $\sigma$ -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 4.1.3** (測度論的議論). 有限生成とは限らない一般の  $\sigma$ -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。



## 4.1.2 定義

$E[-|\mathcal{G}] : \text{Meas}_{\mathcal{F}}(\Omega, \mathbb{R}) \rightarrow L^1(\Omega, \mathcal{G})$  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 4.1.4 (conditional expectation).** 次の条件を満足する確率変数  $Y : \Omega \rightarrow \mathbb{R}$  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 4.1.5** (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 4.1.6 (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].**  $Q(B) := E[X, B] = E[1_B X]$  ( $B \in \mathcal{G}$ ) 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  $Y : \Omega \rightarrow \mathbb{R}$  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 4.1.7** (条件付き確率).  $(\Omega_j)_{j \in \mathbb{N}}$  を  $\Omega$  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。

□

## 4.1.3 性質

$E[-|\mathcal{G}] : \mathcal{L}^1(\Omega, \mathcal{F}) \rightarrow L^1(\Omega, \mathcal{G})$  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 4.1.8.**  $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F})$  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..<sup>†1</sup>
- (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)  $X \in \text{Meas}_{\mathcal{G}}(\Omega, \mathbb{R})$ ,  $XY \in \mathcal{L}^1(\Omega, \mathcal{F})$  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  $\sigma$ -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、 $f : \mathbb{R} \rightarrow \mathbb{R}$  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..

<sup>†1</sup> 除外集合  $N_{a,b}$  は任意の  $a, b \in \mathbb{R}$  について一様には取れない。

**[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  $Z \geq 0$  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??. 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 4.2.1.

■

**命題 4.1.9 (Jensen の不等式).**  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  を凸関数とする.  $X, \psi(X) \in \mathcal{L}^1(\Omega)$  のとき,

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

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

**命題 4.1.10 (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 4.1.11 (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,

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

#### 4.1.4 可測写像を与えたもとの条件付き期待値

部分  $\sigma$ -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 4.1.12.**  $(\Omega, \mathcal{F}, P)$  上の可積分確率変数  $X \in \mathcal{L}^1(\Omega)$  と, 可測空間  $(\mathcal{T}, \mathcal{G})$  への可測写像  $T: \Omega \rightarrow \mathcal{T}$  を考える.

**定義 4.1.13.** 次の 2 条件を満たす関数  $g: \mathcal{T} \rightarrow \mathbb{R}$  を,  $T = t$  の下での  $X$  の条件付き期待値という:

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

$g$  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,

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

**補題 4.1.14 (well-definedness).**

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

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

**定義 4.1.15 (条件付き確率).**

- (1) 部分  $\sigma$ -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 4.1.16.** 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.

**4.1.5 正則条件付き確率**

**定義 4.1.17 (regular conditional probability).** 族  $(p(\omega, A))_{\omega \in \Omega, A \in \mathcal{F}}$  が次の3条件を満たすとき, 部分  $\sigma$ -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)  $\forall \omega \in \Omega \ p(\omega, -) : \mathcal{F} \rightarrow [0, 1]$  は確率測度を定める.
- (2)  $\forall A \in \mathcal{F} \ p(-, A) : \Omega \rightarrow [0, 1]$  は  $\mathcal{G}$ -可測.
- (3)  $\forall A \in \mathcal{F} \ \forall B \in \mathcal{G} \ P(A \cap B) = \int_B p(\omega, A) P(d\omega).$

**定義 4.1.18.** 可測空間  $(\Omega, \mathcal{F})$  が条件 (S) を満たすとは,

- (1) [?] 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  $\sigma$ -加法族となること.
- (2) [?] 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 4.1.19 (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  $\sigma$ -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 4.1.20.**  $T : (\Omega, \mathcal{F}, P) \rightarrow (\mathcal{T}, \mathcal{B}), X : (\Omega, \mathcal{F}, P) \rightarrow (\mathcal{X}, \mathcal{A})$  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)  $\forall t \in \mathcal{T} \ p(t, -) : \mathcal{A} \rightarrow [0, 1]$  は  $(\mathcal{X}, \mathcal{A})$  上の確率測度である.
- (2)  $\forall A \in \mathcal{A} \ p(-, A) : \mathcal{T} \rightarrow [0, 1]$  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 4.1.21 (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 4.1.22 (Fubini の類似).**  $f : \mathcal{T} \times \mathcal{X} \rightarrow \mathbb{R}$  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,

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

**命題 4.1.23 (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,  $E[X_i] = \mu_i, \text{Cov}[X_i, X_j] = \Sigma_{ij}$  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 4.1.24.**  $\Sigma_{11}$  が退化しているときも,  $\Sigma_{11}^{-1}$  を一般化逆行列とすれば, 同様の結果が成り立つ.

## 4.2 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（ $\sigma$ -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。

**命題 4.2.1 (可測関数は独立性を保つ).** 可測空間  $(\mathcal{Y}_i, \mathcal{G}_i)$  への射  $f_i: \mathcal{X}_i \rightarrow \mathcal{Y}_i$  について、確率変数  $X_1, \dots, X_n$  が独立ならば、 $f_1(X_1), \dots, f_n(X_n)$  も独立である。

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

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

■

**要諦 4.2.2.**  $X_1, X_2$  の独立性を定義する方程式のうち、 $f_1(X_1), f_2(X_2)$  は定める同値類がより粗いため、これが独立であるためにはそのうち何本かだけが成り立てば十分。

**命題 4.2.3.** 確率変数  $A_1, \dots, A_n: \Omega \rightarrow \mathcal{G}$  が独立であるとする。任意の  $I := [n] \supset J$  について、 $\{A_j^c, A_k \mid j \in J, k \in I \setminus J\}$  は独立である。

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

■

**命題 4.2.4 (概収束極限も独立).**  $n$  を固定すると  $(X_\lambda^n)_{\lambda \in \Lambda}$  は独立で、 $\lambda$  を固定すると  $X_\lambda^n \rightarrow X_\lambda$  a.s. とする。このとき、 $(X_\lambda)$  も独立である。

**系 4.2.5 (分散が和を保つ条件).**  $(X_i)_{i \in [n]}$  を二乗可積分で  $E[|X_i|^2] < \infty$  ( $i = 1, \dots, n$ )、対独立な確率変数の列とする。この時、次が成り立つ： $\text{Var} \left[ \sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i]$ 。

**[証明].** 2変数の場合、

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

であるから、互いに独立な2変数の共分散は0であることを示せば良いが、

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

■

**要諦 4.2.6.** 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。

## 第 5 章

# 独立確率変数列

### 5.1 独立確率変数列の和の振る舞い

記法 5.1.1 (extension).

- (1) 条件  $\alpha \subset \Omega$  について,  $\alpha$  を成立させるような元からなる集合  $\{\alpha\} := \{\omega \in \Omega \mid \alpha(\omega)\}$  を  $\alpha$  の外延という. すると,  $\{X(\omega) \leq \alpha\} = X^{-1}((-\infty, \alpha])$  などと表せる.
- (2)  $(\Omega, P)$  を可分完全確率測度空間,  $\{\mu_n\} \subset P(\mathbb{R})$  を確率分布の列,  $\{X_n\} \subset \mathcal{L}(\mathbb{R}^\infty)$  をこれらを分布に持つ独立な実確率変数の列とする.
- (3)  $S_n := X_1 + \cdots + X_n$  を部分和の列とする.
- (4) 列  $(S_n)$  が収束するという事象を

$$C := \bigcap_{m \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{j > k > n} \left\{ |S_j(\omega) - S_k(\omega)| < \frac{1}{m} \right\}$$

とすると, これは  $P$ -可測集合である.

**定理 5.1.2** (独立確率変数の和は概収束するか概発散するかのいずれかである).  $P[C] \in \{0, 1\}$  である.

**[証明]**. 任意の  $n \in \mathbb{N}$  に対して,  $S_1, S_2, \dots$  が収束することと  $S_{n+1}, S_{n+2}, \dots$  が収束することは同値であるから, 任意の  $n \in \mathbb{N}$  に対して

$$C := \bigcap_{m \in \mathbb{N}} \bigcup_{r > n \in \mathbb{N}} \bigcap_{j > k > r} \left\{ |S_j - S_k| < \frac{1}{m} \right\}$$

とも表せることに注意して, これが独立な  $\sigma$ -代数の列  $(\sigma[X_k])$  の末尾事象に属することを示せば, Kolmogorov の 0-1 法則より,  $P[C] \in \{0, 1\}$  である.

- (1) いま,  $\mathcal{G}_k := \sigma[X_{k+1}, X_{k+2}, \dots] = \bigvee_{k > n} \sigma[X_k]$  とすると, 任意の  $j > k$  について,  $S_j - S_k = X_{k+1} + \cdots + X_j \in \mathcal{L}_{\mathcal{G}_k}$ . よって,  $\left\{ |S_j - S_k| < \frac{1}{m} \right\} \in \mathcal{G}_k$ . よって,  $\bigcap_{j > k > n} \left\{ |S_j - S_k| < \frac{1}{m} \right\} \in \mathcal{G}_n$ .
- (2) よって任意の  $n \in \mathbb{N}$  に対して  $\bigcup_{r > n} \bigcap_{j > k > r} \left\{ |S_j - S_k| < \frac{1}{m} \right\} \in \mathcal{G}_n$  より,  $C \in \mathcal{G}_n$ . したがって,  $C \in \bigwedge_{n \in \mathbb{N}} \mathcal{G}_n = \bigwedge_{n \in \mathbb{N}} \bigvee_{k > n} \sigma[X_k]$ .

■

### 5.2 概収束するための条件

独立確率変数列  $\{X_n\}$  については,  $\sum_{j=1}^n E[X_j] = E[S_n], \sum_{j=1}^n \text{Var}[X_j] = \text{Var}[S_n]$  に注意.

**定理 5.2.1 (Kolmogorov)**. 独立変数列  $\{X_n\} \subset L^2(\Omega)$  について, 平均と分散の級数  $\sum_{n \in \mathbb{N}} E[X_n], \sum_{n \in \mathbb{N}} \text{Var}[X_n]$  がいずれも収束するならば,  $S_n$  は概収束する.

**[証明].** 0 に収束する実数列  $(E[X_n])$  の分だけ平行移動しても概収束の如何は同値であるから,  $E[X_n] = 0$  仮定して示せば十分.  $(S_n)$  が殆ど確実に Cauchy 列を定めることを示す.

(1) 任意の  $n, m \in \mathbb{N}$  について,  $X_{n+1}, \dots, X_{n+m}$  に対する Kolmogorov の不等式より,

$$\forall \epsilon > 0 \quad P\left[\max_{1 \leq l \leq m} |S_{n+k} - S_n| > \epsilon\right] \leq \frac{1}{\epsilon^2} \sum_{k=1}^m \text{Var}[X_{n+k}].$$

(2) 三角不等式より  $|S_{n+k} - S_{n+l}| \leq |S_{n+k} - S_n| + |S_{n+l} - S_n|$  より,  $\max_{1 \leq k, l \leq m} |S_{n+k} - S_{n+l}| \leq 2 \max_{1 \leq k \leq m} |S_{n+k} - S_n|$  だから,

$$P\left[\max_{1 \leq k, l \leq m} |S_{n+k} - S_{n+l}| > 2\epsilon\right] \leq P\left[\max_{1 \leq k \leq m} |S_{n+k} - S_n| > \epsilon\right] \leq \frac{1}{\epsilon^2} \sum_{k=1}^m \text{Var}[X_{n+k}] \leq \sum_{k=1}^{\infty} \text{Var}[X_{n+k}].$$

(3) 上式の  $P$  の中身は  $m$  に関する増大列で,  $P$  はこれについて連続であるから,  $m \rightarrow \infty$  の極限を考えて

$$P\left[\sup_{k, l \in \mathbb{N}} |S_{n+k} - S_{n+l}| > 2\epsilon\right] \leq \frac{1}{\epsilon^2} \sum_{k=1}^{\infty} \text{Var}[X_{n+k}].$$

次に  $P$  の中身は  $n$  に関する減少列であるから, 再び  $P$  の連続性について

$$P\left[\lim_{n \rightarrow \infty} \sup_{k, l \in \mathbb{N}} |S_{n+k} - S_{n+l}| > 2\epsilon\right] \leq \frac{1}{\epsilon^2} \sum_{k=1}^{\infty} \text{Var}[X_{n+k}]$$

この右辺は,  $\sum_{n \in \mathbb{N}} \text{Var}[X_n]$  が収束することより 0 に収束する. よって,  $\epsilon > 0$  は任意としたから,  $\lim_{n \rightarrow \infty} \sup_{k, l \in \mathbb{N}} |S_{n+k} - S_{n+l}| = 0$  a.s. ■

**要諦 5.2.2.** このとき実は  $L^2$ -収束もしている.

**系 5.2.3.** 独立な正規確率変数列  $\{X_n\} \subset \mathcal{L}(\mathbb{R}^\infty)$  について, 次の 2 条件は同値.

- (1)  $S_n$  は概収束する.
- (2) 期待値と分散  $\sum_{n \in \mathbb{N}} E[X_n], \sum_{n \in \mathbb{N}} \text{Var}[X_n]$  がいずれも収束する.

**要諦 5.2.4.** 期待値と分散の代わりに, 中心値  $\gamma(S_n)$  と散布度  $\sigma(S_n)$  とを用いると, 一般の独立確率変数列についてこの系のような結果が成り立つ. いずれにしろ, 収束のモードは次の 3 種類に分類出来る:

- (1)  $\{\gamma(S_n)\}, \{\sigma(S_n)\}$  がいずれも収束する. これは  $\{S_n\}$  が概収束することに同値.
- (2)  $\{\gamma(S_n)\}$  が発散する. これは  $\{S_n\}$  は概発散だが, 中心を調整し続ける過程  $\{S_n - \gamma(S_n)\}$  とすると概収束する (擬発散型).
- (3)  $\{\sigma(S_n)\}$  が発散するときは, 任意の実数列  $\{a_n\}$  に対して  $\{S_n - a_n\}$  は概発散する (真発散型).

**定理 5.2.5 (Hincin の 3 級数定理).** 独立確率変数列  $(X_n)$  に対して,  $[-1, 1]$  以外の値を取る部分をカットオフした確率変数を  $X'_n := X_n 1_{[-1, 1]}(X_n)$  とする. このとき, 次の 2 条件は同値:

- (1)  $S_n$  は概収束する.
- (2) 期待値と分散  $\sum_{n \in \mathbb{N}} E[X'_n], \sum_{n \in \mathbb{N}} \text{Var}[X'_n]$ , そして  $\sum_{n \in \mathbb{N}} P[X_n \neq X'_n]$  がいずれも収束する.

**[証明].**

(2)  $\Rightarrow$  (1)  $(X_n)$  が独立のとき,  $(X'_n)$  も独立であるから, Kolmogorov の定理より  $(X'_n)$  の級数列は概収束する.  $\sum P[X_n \neq X'_n]$  が収束することから, Borel-Cantelli の補題より,  $P[X_n = X'_n \text{ f.e.}] = 1$ .

(1)  $\Rightarrow$  (2) (a)  $\sum_{n \in \mathbb{N}} P[X_n \neq X'_n]$  が発散するならば, Borel-Cantelli の補題より,  $P[X_n \neq X'_n \text{ i.o.}] = 1$ . すなわち,  $P[|X_n| > 1] = 1$  より, 概発散する.

(b) よって, Borel-Cantelli の補題より  $P[X_n = X'_n \text{ f.e.}] = 1$  より,  $\sum_{n \in \mathbb{N}} X'_n$  は概収束する. この概収束極限を 1 つとり,  $S: \Omega \rightarrow \mathbb{R}$  とする.

- (c)  $X'_n, S_n := \sum_{k=1}^n X'_k$  の特性関数を  $\varphi_n, \psi_n, \psi$  とし,  $X'_n \sim \mu_n$  とすると,  $\varphi_n = \varphi_1 \cdots \varphi_n$  であるから, ノルムの連続性から,

$$|\psi(z)|^2 = \lim_{n \rightarrow \infty} |\psi_n(z)|^2 = \lim_{n \rightarrow \infty} \prod_{k=1}^n |\varphi_k(z)|^2 = \prod_{k=1}^{\infty} |\varphi_k(z)|^2$$

が成り立つ.  $\psi$  は特性関数だから  $\psi(0) = 1$  で, さらにある近傍  $U \subset \mathbb{R}$  上で消えないから,  $U$  上では右辺の無限積は確かに収束しており, 特に  $\forall z \in U \lim_{k \rightarrow \infty} \varphi_k(z) = 1$  から  $\forall z \in U \sum_{n \in \mathbb{N}} (1 - |\varphi_n(z)|^2) < \infty$  を得る (Cauchy 列を定めるため).

- (d) 今  $|\varphi_n|^2 = \varphi_n \cdot \overline{\varphi_n}$  は,  $\overline{\varphi_n}$  が  $\mu_n$  の対称変換 ( $\mu_n(\cdot)$ ) の Fourier 変換であるから,  $\nu := \mu_n * \check{\mu}_n$  の特性関数であり,  $\nu_n$  は  $X'_n$  の独立な 2 つのコピーの差の分布なので,  $[-2, 2]$  に台を持ち, 0 を平均として対称な分布になる. よって先の不等式は,  $\sin$  が奇関数であることに注意して,

$$\forall z \in U \sum_{n \in \mathbb{N}} \left( 1 - \int_{\mathbb{R}} (\cos(z\xi) + i \sin(z\xi)) \nu_n(d\xi) \right) = \sum_{n \in \mathbb{N}} \int_{-2}^2 (1 - \cos(z\xi)) \nu_n(d\xi) < \infty$$

と換言できる.

- (e) さらに十分小さい  $\theta$  について  $1 - \cos \theta > \theta^2/3$  が成り立つため, 十分小さい近傍  $0 \in V \subset U$  について,

$$\forall z \in V \sum_{n \in \mathbb{N}} \int_{-2}^2 z^2 \xi^2 \nu_n(d\xi) < \sum_{n \in \mathbb{N}} 3 \int_{-2}^2 (1 - \cos(z\xi)) \nu_n(d\xi) < \infty.$$

左辺は  $\sum_{n \in \mathbb{N}} \int_{-2}^2 z^2 \xi^2 \nu_n(d\xi) = z^2 \sum_{n \in \mathbb{N}} \int_{-2}^2 \xi^2 \nu_n(d\xi)$  より,  $\sum_{n \in \mathbb{N}} \int_{\mathbb{R}} \xi^2 \nu_n(d\xi) < \infty$  を含意する.

- (f) いま, 分布  $\nu_n$  の平均は 0 で分散は  $\mu_2(\nu_n) = \mu_2(\mu_n) + \mu_2(\check{\mu}_n) = 2\mu_2(\mu_n)$  であるから, これは  $\sum_{n \in \mathbb{N}} \text{Var}[X'_n] < \infty$  を含意する.

- (g) よって Kolmogorov の定理から  $\sum_{n \in \mathbb{N}} (X'_n - E[X'_n])$  は概収束する. いま  $\sum_{n \in \mathbb{N}} X'_n$  の概収束はわかっているから,  $\sum_{n \in \mathbb{N}} E[X'_n]$  の概収束も得た.

■

**要諦 5.2.6.** (2) の 3 番目のアイデアは Borel-Cantelli の翻訳を噛ませると,  $P[X_n = X'_n \text{ f.e.}] = 1$  である.  $(X_n)$  が  $[-1, 1]$  以外の点にある程度の確率で収束したら, 級数  $\sum_{n \in \mathbb{N}} X_n$  は概収束しなくなってしまうため, 非常に自然な仮定である.

### 5.3 独立確率変数の和の収束同等性

**定理 5.3.1 (Levy の 3 収束同等定理).** 独立な実確率変数列  $\{X_n\} \subset \mathcal{L}(\mathbb{R}^\infty)$  について, 次の 3 条件は同値:

- (1)  $S_n$  の確率法則が弱収束する.
- (2)  $S_n$  が確率収束する.
- (3)  $S_n$  が概収束する.

**[証明].**  $X_n \sim \mu_n, \varphi_n := \mathcal{F}\mu_n$  とする. そして

$$S_{mn} := \sum_{k=m+1}^n X_k, \quad S_{mn} \sim \mu_{mn}, \quad \varphi_{mn} \sim \mathcal{F}\mu_{mn} \quad (m < n)$$

の記号も用意する.

- (1)  $\Rightarrow$  (2) 仮定より  $\mu_{1n} \rightarrow \mu$  だから,  $\varphi_{1n} \rightarrow \varphi := \mathcal{F}\mu$ ,

- (a)  $|\varphi|$  は  $z = 0$  のある近傍  $[-a, a]$  である正数  $b > 0$  より大きいから,  $\forall z \in [-a, a] \exists N \in \mathbb{N} \forall n \geq N |\varphi_{1n}(z)| \geq \frac{b}{2}$ .

- (b)  $\varphi_{1m} = \varphi_{1n} \cdot \varphi_{nm}$  より,  $\varphi_{1n} - \varphi_{1m} = \varphi_{1n}(1 - \varphi_{nm})$  かつ  $|\varphi_{1n}| \geq \frac{b}{2}$  on  $[-a, a]$  より,

$$\forall z \in [-a, a] \forall \epsilon > 0 \exists N \in \mathbb{N} \forall m > n > M \quad |1 - \varphi_{nm}(z)| = \left| \int (1 - e^{iz\xi}) \mu_{nm}(d\xi) \right| = |1 - \varphi_{nm}(z)| < \epsilon.$$



(c) よって, 三角不等式より,

$$\left| \frac{1}{2a} \int_{-a}^a dz \int (1 - e^{iz\xi}) \mu_{nm}(d\xi) \right| \leq \frac{1}{2a} \int_{-a}^a dz \left| \int (1 - e^{iz\xi}) \mu_{nm}(d\xi) \right| < \epsilon.$$

(d) Fubini の定理より, この積分を計算すると,

$$\begin{aligned} \frac{1}{2a} \int_{-a}^a dz \int (1 - e^{iz\xi}) \mu_{nm}(d\xi) &= \frac{1}{2a} \int \left[ z + \frac{ie^{iz\xi}}{\xi} \right]_{-a}^a \mu_{nm}(d\xi) \\ &= \frac{1}{2a} \int \left( 2a - \frac{2}{\xi} \frac{e^{ia\xi} - e^{-ia\xi}}{2i} \right) \mu_{nm}(d\xi) \\ &= \int \left( 1 - \frac{\sin(a\xi)}{a\xi} \right) \mu_{nm}(d\xi) \end{aligned}$$

で, 絶対値は要らないことが分かる.

(e) ここで,  $\exists \beta > 0 \quad \forall \xi \in \mathbb{R} \quad \beta \frac{a^2 \xi^2}{1 + a^2 \xi^2} \leq 1 - \frac{\sin a\xi}{a\xi}$  である. これは, Taylor の定理より  $\theta \in (0, 1)$  が存在して,  $\forall x \in \mathbb{R} \quad \frac{\theta^3}{3!} \frac{x^2}{1 + x^2} < \frac{\theta^3 x^2}{3!} = 1 - \frac{\sin x}{x}$  が成り立つことから,  $\beta < \frac{\theta^3}{3!}$  とすれば良い. これを用いて

$$\int_{\mathbb{R}} \frac{a^2 \xi^2}{1 + a^2 \xi^2} \mu_{nm}(d\xi) \leq \frac{1}{\beta} \int \left( 1 - \frac{\sin a\xi}{a\xi} \right) \mu_{nm}(d\xi) < \frac{\epsilon}{\beta}.$$

(f) 任意の  $\eta > 0$  に対して,

$$\mu_{nm}([- \eta, \eta]^c) = \int 1_{[- \eta, \eta]^c} \mu_{nm}(d\xi) \leq \left( \frac{a^2 \eta^2}{1 + a^2 \eta^2} \right)^{-1} \int_{\mathbb{R}} \frac{a^2 \xi^2}{1 + a^2 \xi^2} \mu_{nm}(d\xi) < \frac{1 + a^2 \eta^2}{a^2 \eta^2} \frac{\epsilon}{\beta}$$

$\mu_{nm}$  とは  $S_{nm} = S_m - S_n$  の確率法則で,  $\epsilon > 0$  は任意であったから, 任意の  $\eta > 0$  に対して,  $P[|S_m - S_n| > \eta]$  を任意に小さく出来ることを意味する.

(2)  $\Rightarrow$  (3)  $(S_n)$  が確率収束するあから,  $\exists N \in \mathbb{N} \quad \forall n, m > N \quad P[|S_m - S_n| > \epsilon] < \frac{1}{2}$ . よって, Ottaviani の不等式より, 任意の  $m > n > N$  に対して,

$$P[\max_{k \in [m]} |S_{n+k} - S_n| > 2\epsilon] \leq 2P[|S_{n+m} - S_n| > \epsilon].$$

したがって, 変数を  $k, l$  の 2 つに増やすと

$$P[\max_{k, l \in [m]} |S_{n+k} - S_{n+l}| > 4\epsilon] \leq 4P[|S_{n+m} - S_n| > \epsilon] \leq 4 \sup_{m \in \mathbb{N}} P[|S_{n+m} - S_n| > \epsilon].$$

したがって Kolmogorov の定理の証明と同様,  $n \rightarrow \infty$  のとき, 右辺は 0 に収束する.

■

**命題 5.3.2.** 独立な実確率変数列  $\{X_n\} \subset L^p(\Omega)$  は中心化されているとする.  $S_n$  が  $L^p$ -収束するならば, 概収束する.

**[証明].**  $L^p$  に関する Kolmogorov の不等式 1.9.6 による.

■

## 5.4 概発散の描像

任意の  $\{a_n\} \subset \mathbb{R}$  に対して  $\limsup_{n \rightarrow \infty} |S_n - a_n| = \infty$  a.s. となる. が, うまく  $\{b_n\} \subset \mathbb{R}_+$  を分母につけて, 0 に抑えることが考えられる. これが大量の法則である.

**定理 5.4.1.** 独立確率変数列  $(X_n)$  は, 任意の実数列  $\{a_n\} \subset \mathbb{R}$  に対して  $\{S_n - a_n\}$  が概発散するとする. このとき, 任意の実数列  $\{a_n\} \subset \mathbb{R}$  に対して,  $\limsup_{n \rightarrow \infty} |S_n - a_n| = \infty$  a.s.

## 5.5 大数の法則

### 平均過程周りの揺らぎは線型よりも遅い

一般に  $(S_n)$  は概発散するか概収束するかの2通りで、もし真発散型ならば  $\limsup_{n \rightarrow \infty} |S_n - a_n| = \infty$  a.s. になる。すなわち、任意の追い方  $\{a_n\} \subset \mathbb{R}$  に対して、 $S_n$  は殆ど確実に無限回逃避に成功する。しかし、その逃避行は線型よりも遅く、 $\limsup_{n \rightarrow \infty} \frac{|S_n - a_n|}{b_n} = \lim_{n \rightarrow \infty} \frac{S_n - a_n}{b_n} = 0$  a.s. と捕まえられる。実数和  $\sum_{n \in \mathbb{N}} x$  ならば線型に増大するが、「確率の取り分」が線型よりも、そして  $x^{1/2+\epsilon}$  ( $\epsilon > 0$ ) よりも遅くする。

### 5.5.1 本質的要件

$\frac{1}{b_n} \sum_{k=1}^n X_k$  という形の確率変数の収束の議論は極めて扱いにくい。Cauchy 列の議論をしようにも係数  $1/b_n$  が邪魔で、 $(X_n)$  が独立でなかったらおてあげであるが、Cauchy の積級数の収束判定条件のように、 $\sum_{n \in \mathbb{N}} \frac{1}{b_n} X_n$  の有界性に目を向けるアイデアが Kronecker の補題である。実際、Cauchy の収束判定法は部分和公式、Kronecker の補題は部分積分公式を用いる。

**補題 5.5.1 (Kronecker).**  $\{x_n\} \subset \mathbb{R}, \{b_n\} \subset \mathbb{R}_+$  は発散する単調増加列とする。このとき、

$$\sum_{n \in \mathbb{N}} \frac{x_n}{b_n} \Rightarrow \frac{1}{b_n} \sum_{k=1}^n x_k \xrightarrow{n \rightarrow \infty} 0.$$

**[証明].** 部分和を  $s_0 = 0, s_n = \sum_{k=1}^n \frac{x_k}{b_k}$  とすると、 $s_n \rightarrow s \in \mathbb{R}$  で、 $x_n = b_n(s_n - s_{n-1})$  が成り立つ。

(1) Abel の部分積分の公式より、

$$\begin{aligned} \frac{1}{b_n} \sum_{k=1}^n x_k &= \frac{1}{b_n} \sum_{k=1}^n b_k (s_k - s_{k-1}) \\ &= \frac{1}{b_n} \sum_{k=1}^n b_k s_k - \frac{1}{b_n} \sum_{k=1}^n b_k s_{k-1} \\ &= \frac{1}{b_n} \left( b_n s_n + \sum_{k=1}^n b_{k-1} s_{k-1} \right) - \frac{1}{b_n} \sum_{k=1}^n b_k s_{k-1} \\ &= s_n - \frac{1}{b_n} \sum_{k=1}^n (b_k - b_{k-1}) s_{k-1}. \end{aligned}$$

(2)  $\frac{1}{b_n} \sum_{k=1}^n (b_k - b_{k-1}) s_{k-1} \rightarrow s$  を示す。まず、 $s_{k-1} \rightarrow s$  より、 $\forall \epsilon > 0 \exists N \in \mathbb{N} \forall n > N |s_n - s| < \epsilon$ 。よって、

$$\begin{aligned} \forall \epsilon > 0 \exists N \in \mathbb{N} \forall n > N \frac{1}{b_n} \sum_{k=1}^N (b_k - b_{k-1}) s_{k-1} + \frac{1}{b_n} \sum_{k=N+1}^n (b_k - b_{k-1}) s + \frac{1}{b_n} \sum_{k=N+1}^n (b_k - b_{k-1}) (s_{k-1} - s) \\ < \frac{1}{b_n} \sum_{k=1}^N (b_k - b_{k-1}) s_{k-1} + \frac{b_n - b_N}{b_n} s + \epsilon \xrightarrow{n \rightarrow \infty} s + \epsilon. \end{aligned}$$

■

**要諦 5.5.2 (概収束への議論の還元).** この補題により、 $\frac{1}{r_n} \sum_{k=1}^n X_k \rightarrow 0$  を満たす係数列  $\{r_n\}$  を見つける問題は、 $\sum_{n \in \mathbb{N}} \frac{X_n}{r_n}$  が概収束するものを見つめる、という古典的な問題に変換された。よって  $\frac{S_n - E[S_n]}{n} \rightarrow 0$  a.s. も、 $\frac{1}{n} \sum_{k=1}^n X_k$  に議論が落ち着く。これが  $E[X_n]$  に収束するそのモードは、3 収束いずれも等価であることに注意。

**定理 5.5.3 (大数の法則).**  $\{X_n\} \subset L^2(\mathbb{R}^\infty)$  が独立,  $\{b_n\} \subset \mathbb{R}_+$  は発散する単調増加列とする.

(1)

$$\sum_{n \in \mathbb{N}} \frac{\text{Var}[X_n]}{b_n^2} < \infty \Rightarrow \frac{S_n - E[S_n]}{b_n} \xrightarrow{n \rightarrow \infty} 0 \text{ a.s.}$$

(2) 分散の和  $\sum_{n \in \mathbb{N}} \text{Var}[X_n]$  が収束するとき, 任意の発散する正数列  $\{b_n\}$  に対して,

$$\frac{S_n - E[S_n]}{b_n} \rightarrow 0 \text{ a.s.}$$

(3) 分散の和  $\sum_{n \in \mathbb{N}} \text{Var}[X_n]$  が発散するとき, 任意の  $\epsilon > 0$  に対して,

$$\frac{S_n - E[S_n]}{V[S_n]^{1/2+\epsilon}} \xrightarrow{n \rightarrow \infty} 0 \text{ a.s.}$$

**[証明].**

(1)  $Y_n := \frac{X_n - E[X_n]}{b_n}$  なる独立確率変数の級数  $\sum_{n \in \mathbb{N}} Y_n$  の概収束を示せば, Kronecker の定理より従う.  $E[Y_n] = 0$  で,

$$\sum_{n \in \mathbb{N}} \text{Var}[Y_n] = \sum_{n \in \mathbb{N}} \frac{\text{Var}[X_n]}{b_n^2} < \infty \text{ より, Kolmogorov の定理から } Y_n \text{ は概収束する.}$$

(2)  $S_n - E[S_n]$  の  $L^2$ -ノルムの極限が有限であるとき,  $L^1$ -ノルムの極限も有限.

(3)  $\text{Var}[S_1] > 0$  と仮定して一般性を失わない.  $\text{Var}[S_n]$  は単調増加であることに加えて,  $\text{Var}[S_n] \rightarrow \infty$  であるから,  $b_n := (\text{Var}[S_n])^{1/2+\epsilon}$  とおけば, (1) より  $\sum_{n \in \mathbb{N}} \frac{\text{Var}[X_n]}{b_n^2}$  が収束することを言えばよいが, これは関数  $f(x) = x^{-1-2\epsilon}$  が単調減少であることと,  $\text{Var}[X_n] = \text{Var}[S_n] - \text{Var}[S_{n-1}]$  が区間の幅とみなせることから,

$$\sum_{n=2}^{\infty} \frac{\text{Var}[X_n]}{b_n^2} = \sum_{n=2}^{\infty} \frac{1}{(\text{Var}[S_n])^{1+2\epsilon}} (\text{Var}[S_n] - \text{Var}[S_{n-1}]) \leq \sum_{n=2}^{\infty} \int_{\text{Var}[S_{n-1}]}^{\text{Var}[S_n]} \frac{dx}{x^{1+2\epsilon}} = \int_{\text{Var}[S_1]}^{\infty} \frac{dx}{x^{1+2\epsilon}} < \infty.$$

■

## 5.5.2 種々の十分条件

**系 5.5.4 (Kolmogorov).**  $\{X_n\} \subset L^2(\mathbb{R}^\infty)$  が独立で,  $\{V[X_n]\} \subset \mathbb{R}$  が有界ならば,

$$\frac{S_n - E[S_n]}{n} \xrightarrow{n \rightarrow \infty} 0 \text{ a.s.}$$

**[証明].**  $\sum \text{Var}[X_n]$  が収束するならば Kolmogorov の定理より中心化された確率変数の列  $S_n - E[S_n]$  が収束する (期待値の和は自明に収束するので). よって,  $\sum \text{Var}[X_n]$  が発散する場合を示せば良いが, 定理で  $\epsilon = 1/2$  とおくと

$$\frac{S_n - E[S_n]}{\text{Var}[S_n]} \rightarrow 0 \text{ a.s.}$$

いま  $\text{Var}[X_n]$  は有界であるから,  $\alpha := \sup_{n \in \mathbb{N}} \text{Var}[X_n] < \infty$  とおくと,

$$\text{Var}[S_n] = \sum_{k=1}^n \text{Var}[X_k] \leq n\alpha.$$

よって, 収束する優級数が見つかったので  $\frac{S_n - E[S_n]}{n}$  も 0 に概収束する. ■

**系 5.5.5 (独立分布の場合の必要十分条件).**  $\{X_n\} \subset L(\mathbb{R}^\infty)$  は独立同分布列とする. このとき, 次の 2 条件は同値:

(1)  $\{X_n\} \subset L^1(\mathbb{R}^\infty)$ .

(2)  $\frac{S_n}{n} \xrightarrow{n \rightarrow \infty} \alpha_1 \text{ a.s.}$

**[証明]**  $X'_n : X_n 1_{\{|X_n| \leq n\}}$  とし,  $S'_n := \sum_{k=1}^n X'_k$ ,  $|X_n| \sim \nu$  とする.

- (1)  $P[X_n = X'_n \text{ f.e.}] = 1$  である. 実際,  $y \mapsto \nu((y, \infty))$  は単調減少であることと,  $1_{[y, \infty)}(x) = 1_{[0, x]}(y)$  に注意すれば, Fubini の定理から,

$$\begin{aligned} \sum_{n=1}^{\infty} P[X_n \neq X'_n] &= \sum_{n=1}^{\infty} P[|X_n| > n] = \sum_{n=1}^{\infty} \nu((n, \infty)) \\ &\leq \int_0^{\infty} \nu(y, \infty) dy \\ &= \int_0^{\infty} \left( \int_y^{\infty} \nu(dx) \right) dy \\ &= \int_0^{\infty} \int_0^{\infty} 1_{[0, x]}(y) dy \nu(dx) \\ &= \int_0^{\infty} x \nu(dx) = E[|X_1|] < \infty. \end{aligned}$$

よって Borel-Cantelli の補題より,  $P[X_n \neq X'_n \text{ i.o.}] = 0$  かつ  $P[X_n = X'_n \text{ f.e.}] = 1$ . したがって,  $\frac{S'_n}{n} \rightarrow \alpha_1$  を言えば良い.

- (2) 過程  $\left(\frac{S'_n}{n}\right)_{n \in \mathbb{N}}$  の平均の過程は,  $\left(\sum_{k=1}^{\infty} \frac{E[X'_k]}{n}\right)$  で, 単調収束定理より  $E[X'_k] = E[X_k 1_{|X_k| \leq k}] \xrightarrow{k \rightarrow \infty} \alpha_1$ . よって,  $\forall \epsilon > 0 \exists N \in \mathbb{N} \forall n > N |E[X'_n] - \alpha_1| < \epsilon$  だから,

$$\begin{aligned} \sum_{k=1}^n \frac{E[X'_k]}{n} &= \sum_{k=1}^N \frac{E[X'_k]}{n} + \sum_{k=N+1}^n \frac{E[X'_k] - \alpha_1}{n} + \sum_{k=N+1}^n \frac{\alpha_1}{n} \\ &< \frac{E[S'_N]}{n} + \epsilon \left(1 - \frac{N+1}{n}\right) + \alpha_1 \left(1 - \frac{N+1}{n}\right) \xrightarrow{n \rightarrow \infty} \epsilon + \alpha_1 \end{aligned}$$

より,  $\frac{E[S'_n]}{n} \rightarrow \alpha_1$  が分かる. いま  $\{X'_n\} \subset L^2(\Omega)$  に注意すれば, 大数の法則から, あとは  $\sum_{n \in \mathbb{N}} \frac{\text{Var}[X'_n]}{n^2} < \infty$  を示せば良い.

- (3)  $1_{[0, y]}(x) = 1_{(x, \infty)}(y)$  a.e. と,  $\forall n \geq 1 \frac{1}{n^2} \leq \frac{4}{(n+1)^2}$  に注意して, Fubini の定理より,

$$\begin{aligned} \sum_{n \in \mathbb{N}} \frac{\text{Var}[X'_n]}{n^2} &\leq \frac{E[X_n'^2]}{n^2} \leq 4 \sum_{n \in \mathbb{N}} \frac{E[X_n'^2]}{(n+1)^2} \\ &\leq 4 \sum_{n \in \mathbb{N}} \frac{1}{(1+n)^2} \int_{[0, n]} x^2 \nu(dx) \\ &\leq 4 \int_0^{\infty} \frac{dy}{y^2} \int_{[0, y]} x^2 \nu(dx) \\ &= 4 \int_0^{\infty} \left( \int_x^{\infty} \frac{dy}{y^2} \right) x^2 \nu(dx) \\ &= 4 \int_0^{\infty} x \nu(dx) = E[|X_1|] < \infty. \end{aligned}$$

■

**要諦 5.5.6.** 独立同分布であるときは, 積分 (Fubini の定理) による議論が出来るので, この途を通れば  $L^1$  の仮定のみから大数の法則が成り立つ. またこの系は, 独立同分布とせずとも, 対独立な同分布で成り立つ.

### 5.5.3 弱法則の十分条件

**定理 5.5.7.**  $\{X_n\} \subset L^1(\Omega)$  は組ごとに独立で,  $\{\text{Var}[X_n]\} \subset \mathbb{R}$  は有界であるとする. このとき,  $\frac{S_n - E[S_n]}{n} \xrightarrow{p} 0$ .

## 5.5.4 例

例 5.5.8.  $\{X_i\} \subset L^\infty(\Omega)$  は独立同分布

$$P[X_i = 0] = 1 - p, \quad P[X_i = 1] = p, \quad p \in (0, 1).$$

について,

$$\sum_{n=1}^{\infty} \frac{X_n}{2^n}$$

は概収束し,

- (1)  $p = 1/2$  のとき, 分布  $U([0, 1])$  を持つ.
- (2)  $p \neq 1/2$  のとき, その分布は原子を持たないが Lebesgue 測度に関して特異連続になる.

□

## 5.6 中心極限定理

$(S_n)$  の平均過程周りの揺らぎを分散の過程でスケールし続けて追うと, 最後に残る不確定性は標準正規分布になる. トレンド  $E[S_n]$  を除去し, 広がっていく拡散性  $\sqrt{\text{Var}[S_n]}$  もピンチアウトして追い続けたときに残る不確定性は標準正規分布になるという不思議.

**要諦 5.6.1** (議論の中核). 特性関数を独立性を用いて積に分解し, Taylor 展開して  $E$  を各項積分してから Taylor 展開を戻すと, 正規分布の特性関数を得る, という物語が骨格である.  $T_n := \frac{S_n - E[S_n]}{(V[S_n])^{1/2}}$  とすると,

$$\begin{aligned} \mathcal{F}P^{T_n}(u) &= \prod_{k=1}^n E \left[ \exp \left( \frac{iu(X_k - E[X_k])}{\sqrt{V[S_k]}} \right) \right] \\ &= \prod_{k=1}^n E \left[ 1 + \frac{iu(X_k - E[X_k])}{\sqrt{V[S_k]}} - \frac{u^2(X_k - E[X_k])^2}{2V[S_k]} \right] \\ &= \prod_{k=1}^n E \left[ 1 - \frac{V[X_k]}{2V[S_k]} u^2 + \dots \right] \\ &\sim \prod_{k=1}^n \exp \left( -\frac{V[X_k]}{2V[S_k]} u^2 \right) = \exp \left( -\frac{z^2}{2} \right). \end{aligned}$$

## 5.6.1 指数関数の Taylor 展開に関する補題

$\lim_{n \rightarrow \infty} \left(1 + \frac{\alpha}{n}\right)^n = e^\alpha$  は,  $1 + \frac{\alpha}{n}$  の  $n$  乗でなくとも, 和が  $\alpha$  になる  $(1 + \alpha_{nk})$  の列を  $k = 1$  から  $k = N(n) \rightarrow \infty$  までかけ合わせれば成り立つ. また, 複素関数の Taylor の定理に対応する手頃な結果を用意する (複素積分を用いた表示などの一般形は使いにくい).

**補題 5.6.2.** 任意の  $n \in \mathbb{N}$  に複素数の組  $(\alpha_{n1}, \dots, \alpha_{nN(n)}) \in \mathbb{C}^{N(n)}$  ( $N(n) \in \mathbb{N}$ ) が対応しているとする. 次の条件が成り立つとき,

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

- (1)  $\alpha_n := \sum_{k=1}^{N(n)} \alpha_{nk} \rightarrow \alpha.$
- (2)  $\beta_n := \max_{k \in [N(n)]} |\alpha_{nk}| \rightarrow 0.$
- (3) 有界:  $\exists b \in \mathbb{R} \forall n \in \mathbb{N} \gamma_n := \sum_{k \in [N(n)]} |\alpha_{nk}| < b.$

**[証明].** 条件 (2) から, 任意の  $n, k \in \mathbb{N}$  に対して  $|\alpha_{nk}| < \frac{1}{2}$  として一般性を失わない.  $\log(1+z)$  の  $z=0$  で 0 となる分枝を取れば,  $|z| < 1$  上で一価で, Taylor の定理より, ある正則関数  $f \in \mathcal{O}(\Delta)$  が存在して, 任意の  $z \in \Delta$  に対して

$$\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots = z + f(z)z^2$$

が成り立つ. いま,  $|z| < \frac{1}{2}$  の範囲に話を限れば, 上式の実部と虚部を考えることにより,  $\operatorname{Re}(f(z)), \operatorname{Im}(f(z)) \in (-1/2, 1/2)$  を満たすから, 特に  $f(\Delta/2) \subset \Delta$ .

$$\begin{aligned} \prod_{k=1}^{N(n)} (1 + \alpha_{nk}) &= \prod_{k \in [N(n)]} \exp(\log(1 + \alpha_{nk})) \\ &= \prod_{k \in [N(n)]} \exp(\alpha_{nk} + f(\alpha_{nk})\alpha_{nk}^2) \\ &= \exp\left(\sum_{k \in [N(n)]} \alpha_{nk} + \sum_{k \in [N(n)]} f(\alpha_{nk})\alpha_{nk}^2\right) \\ &= \exp\left(\alpha_n + \beta_n \theta_n \sum_{k \in [N(n)]} |\alpha_{nk}| \right) \\ &= \exp(\alpha_n + \beta_n \theta'_n b) \\ &\xrightarrow{n \rightarrow \infty} e^\alpha. \end{aligned} \quad \begin{aligned} \theta_n &:= \frac{\sum_{k \in [N(n)]} |\alpha_{nk}|}{\sum_{k=1}^{N(n)} \frac{f(\alpha_{nk})}{\beta_n} |\alpha_{nk}|^2} \in \Delta \\ \theta'_n &:= \theta_n \frac{\sum_{k \in [N(n)]} |\alpha_{nk}|}{b} \in \Delta \end{aligned}$$

**補題 5.6.3.** 任意の実数  $z \in \mathbb{R}$  に対して, ある実数  $\theta \in [-1, 1]$  が存在して,

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

**[証明].**

$$f(z) := e^{iz} - \sum_{k=0}^n \frac{(iz)^k}{k!}$$

とおくと,  $f^{(n+1)}(z) = i^{n+1} e^{iz}$  で, そこまでの微分は消えている. よって,  $f(z)$  は  $f^{(n+1)}$  の  $n+1$  重積分で表せる.  $|f(z)|$  は 1 の  $n+1$  重積分よりも小さい. よって,  $|f(z)| \leq \frac{|z|^{n+1}}{(n+1)!}$ . ■

## 5.6.2 Lindeberg の条件

$T_n = T_n^{1/2} := \frac{S_n - E[S_n]}{\sqrt{\operatorname{Var}[S_n]}}$  が  $N(0, 1)$  に弱収束するための十分条件を考える. また, 周辺の  $T_n^\alpha := \frac{S_n - E[S_n]}{\operatorname{Var}[S_n]^\alpha}$  の振る舞いも考える.

**定理 5.6.4 (Lindeberg).**  $\{X_n\} \subset L^2(\mathbb{R}^\infty)$  は独立で, 次を満たすならば,  $T_n$  に関する中心極限定理が成り立つ.

$$\forall \epsilon > 0 \quad \frac{1}{\operatorname{Var}[S_n]} \sum_{k=1}^n E[(X_k - E[X_k])^2, |X_k - E[X_k]| \geq \epsilon \sqrt{\operatorname{Var}[S_n]}] \rightarrow 0. \quad (\operatorname{Var}[S_n] > 0 \text{ f.e.})$$

**[証明].**

$$Y_{nk} := \frac{X_k - E[X_k]}{\sqrt{\operatorname{Var}[S_n]}}$$

とおくと, これは中心化された独立確率変数列で,  $T_n = \sum_{k=1}^n Y_{nk}$  が成り立つ.  $\operatorname{Var}[T_n] = \sum_{k=1}^n \operatorname{Var}[Y_n] = \sum_{k=1}^n E[Y_{nk}^2]$  に注意すれば, Lindeberg の条件とは,

$$\sum_{k=1}^n E\left[\left(\frac{X_k - E[X_k]}{\sqrt{\operatorname{Var}[S_n]}}\right)^2, \left|\frac{X_k - E[X_k]}{\sqrt{\operatorname{Var}[S_n]}}\right| \geq \epsilon\right] = \sum_{k=1}^n E[Y_{nk}^2, |Y_{nk}| \geq \epsilon] \rightarrow 0$$

というある種の絶対連続性の仮定他ならない。

方針  $\mu_n := P^{T_n}$  とおくと,  $\{Y_{nk}\}_{k \in [n]}$  の独立性より,

$$\mathcal{F}\mu_n = E[e^{izT_n}] = \prod_{k=1}^n E[e^{izY_{nk}}] =: \prod_{k \in [n]} (1 + \alpha_{nk})$$

とみる。

補題の要件 (1), (3)

$$\begin{aligned} \alpha_{nk} &= E[e^{izY_{nk}}] - 1 = E[e^{izY_{nk}}, |Y_{nk}| < \epsilon] + E[e^{izY_{nk}}, |Y_{nk}| \geq \epsilon] - 1 \\ &= E\left[1 + izY_{nk} - \frac{1}{2}z^2Y_{nk}^2 + \frac{\theta_3}{6}|zY_{nk}|^3, |Y_{nk}| < \epsilon\right] \\ &\quad + E\left[1 + izY_{nk} + \frac{\theta_2}{2}z^2Y_{nk}^2, |Y_{nk}| \geq \epsilon\right] - 1 \\ &= izE[Y_{nk}] - \frac{1}{2}z^2E[Y_{nk}^2] + \frac{1}{2}z^2E[Y_{nk}^2, |Y_{nk}| \geq \epsilon] \\ &\quad + \frac{1}{6}|z|^3E[\theta_3|Y_{nk}|^3, |Y_{nk}| < \epsilon] + \frac{1}{2}z^2E[\theta_2Y_{nk}^2, |Y_{nk}| \geq \epsilon] \\ &= -\frac{1}{2}z^2E[Y_{nk}^2] + \frac{1}{2}z^2(1 + \theta'_2)E[Y_{nk}^2, |Y_{nk}| \geq \epsilon] + \frac{1}{6}|z|^3\theta'_3E[Y_{nk}^2] \end{aligned} \quad \begin{aligned} &\theta_2, \theta \in [-1, 1] \\ &\theta'_2, \theta'_3 \in [-1, 1] \end{aligned}$$

第3項は十分小さな  $\epsilon$  を取り, 第1項は  $\max_{k \in [n]} E[Y_{nk}^2] \leq \epsilon^2 + \sum_{k \in [n]} E[Y_{nk}^2, |Y_{nk}| \geq \epsilon] \rightarrow \epsilon^2$ , 第2項も同様に評価出来るから, 総じて

$$\limsup_{n \rightarrow \infty} \max_{k \in [n]} |\alpha_{nk}| \leq \frac{1}{2}z^2\epsilon^2 + \frac{1}{6}|z|^3\epsilon^3 \rightarrow 0.$$

したがって特に,  $\max_{k \in [n]} |\alpha_{nk}| \rightarrow 0$ . また,  $\sum_{k \in [n]} |\alpha_{nk}|$  は  $n \in \mathbb{N}$  について有界で,

$$\sum_{k \in [n]} |\alpha_{nk}| \leq \frac{1}{2}z^2 + z^2 + \frac{1}{6}|z|^3\epsilon.$$

補題の要件 (2)

$$\sum_{k \in [n]} \alpha_{nk} = -\frac{1}{2}z^2 + \frac{1}{2}z^2(1 + \theta''_2) \sum_{k \in [n]} E[Y_{nk}^2, |Y_{nk}| \geq \epsilon] + \frac{1}{6}|z|^3\theta''_3\epsilon.$$

全く同様にして,  $\sum_{k \in [n]} \alpha_{nk} \rightarrow -\frac{1}{2}z^2$  を得る.

結論 補題より,  $\mathcal{F}\mu_n(z) \rightarrow e^{-z^2/2}$  に各点収束する. Glivenko の定理より,  $\mu_n \rightarrow N(0, 1)$ .

■

系 5.6.5. Lindeberg 条件の下で

- (1) たしかに一様可積分性の条件になっている: Lindeberg 条件は  $v = \lim_{n \rightarrow \infty} V[S_n] = \infty$  を含意している.
- (2)  $T_n^\alpha$  ( $\alpha > 1/2$ ) は  $\delta_0$  に法則収束する.
- (3)  $T_n$  は概収束はしない (無限幅の振動をする).
- (4)  $T_n$  は確率収束もしない.

[証明].

- (1) 背理法による.  $\text{Var}[S_n] \rightarrow \infty$  でないのなら,  $\text{Var}[S_n] > 0$  f.e. よりある  $Y_{nk}$  は  $n \rightarrow \infty$  のとき 0 に収束しないが, これについて  $E[Y_{nk}^2, |Y_{nk}| \geq \epsilon]$  は十分小さい  $\epsilon$  について正であり続け,  $\sum_{k=1}^n E[Y_{nk}^2, |Y_{nk}| \geq \epsilon]$  は 0 に収束しえない.
- (2) よって大数の法則より,  $T_n^\alpha = \frac{S_n - E[S_n]}{(\text{Var}[S_n])^\alpha}$  ( $\alpha > 1/2$ ) は 0 に概収束する.
- (3)



**要諦 5.6.6.** 実は Lindeberg の条件は必要条件でもある。

**定理 5.6.7 (Lindeberg-Feller).**  $\{X_{nj}\}_{j \in [k_n]} \subset L^2(\Omega)$  は独立であるとし,  $S_n := \sum_{j \in [k_n]} X_{nj}$  とおき,  $\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} \text{Var}[S_n - E[S_n]] = v \in \mathbb{R}_+$  とする. このとき, 次の2条件は同値:

- (1)  $S_n - E[S_n] \Rightarrow N(0, v)$  かつ  $\max_{1 \leq j \leq k_n} \text{Var}[X_{nj} - E[X_{nj}]] \xrightarrow{n \rightarrow \infty} 0$ .
- (2) Lindeberg の条件が成り立つ:  $\forall \epsilon > 0 \sum_{j=1}^{k_n} E[X_{nj}^2, |X_{nj}| \geq \epsilon] \xrightarrow{n \rightarrow \infty} 0$ .

### 5.6.3 種々の十分条件

独立同分布のとき, Lindeberg 条件は単に定積分が絶対連続であることに帰着する。

**命題 5.6.8.**

- (1) ( $L^\infty$ -有界のとき)  $M := \sup_{n \in \mathbb{N}} \|X_n\|_\infty < \infty$  かつ  $\sum_{n \in \mathbb{N}} \text{Var}[X_n] = \infty$  を満たす独立な  $\{X_n\} \subset L^\infty$  は中心極限定理が成り立つ.
- (2) (同分布のとき, Levy)  $\{X_n\} \subset L^2$  が独立で, 正の分散を持つ同分布に従うとき, 中心極限定理が成り立つ.
- (3) (Ljapunov)  $\{X_n\} \subset L^2$  が独立で, 次の条件を満たすとき, 中心極限定理が成り立つ:

$$\exists \delta > 0 \frac{1}{\text{Var}[S_n]^{1+\delta/2}} \sum_{k=1}^n E[|X_k - E[X_k]|^{2+\delta}] \rightarrow 0.$$

**[証明].**

- (1)  $\{X_n\}$  が  $L^\infty$ -有界のとき,  $\{X_k - E[X_k]\}$  も  $L^2$ -有界である. よって, Lindeberg 条件が成り立つ:

$$\forall \epsilon > 0 \frac{1}{\text{Var}[S_n]} \sum_{k=1}^n E[(X_k - E[X_k])^2, |X_k - E[X_k]| \geq \epsilon \sqrt{\text{Var}[S_n]}] \rightarrow 0. \quad (\text{Var}[S_n] > 0 \text{ f.e.})$$

- (2)  $X_n$  が同分布に従うとき, Lindeberg 条件は  $\sum$  記号が消えて

$$\forall \epsilon > 0 \frac{1}{\text{Var}[X_1]} E[(X_1 - E[X_1])^2 1_{\{|X_1 - E[X_1]| \geq \epsilon \sqrt{n \text{Var}[X_1]}\}}] \rightarrow 0$$

となるが, これは定積分の絶対連続性から明らか. Lindeberg 条件は  $\text{Var}[S_n] > 0$  を含意しているので, この論法を行うには  $\text{Var}[X_n] > 0$  が必要であることに注意.

- (3) (1) の一般化にあたる. Markov の不等式から直接示せるが,  $\{X_k - E[X_k]\}$  が  $L^{2+\delta}$ -有界ならば  $L^2$ -有界でもあり, あとは定積分の絶対連続性による.

## 5.7 重複対数の法則

### 5.7.1 ルートより少し速い収束

$\{X_n\} \subset L^2$  を独立とする. Lindeberg 条件の下で  $\frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}}$  は法則収束するが, 概発散する. これは  $V[S_n]$  よりも対数の積だけ速い消息を捕まえそこねていることを意味する.  $S_n - E[S_n]$  は  $\sqrt{V[S_n]}$  より無限回大きく,  $-\sqrt{V[S_n]}$  より無限回小さくなる. 特に,  $X_i$  が Rademacher 変数の場合,  $S_n$  は無限回 0 を通ることがわかる.

**定理 5.7.1.**  $\{X_n\} \subset L^2$  を独立とする. Lindeberg 条件の下で次が成り立つ:

- (1)  $\limsup_{n \rightarrow \infty} \frac{S_n - E[S_n]}{V[S_n]^{1/2}} = \infty$  a.s.  
 (2)  $\liminf_{n \rightarrow \infty} \frac{S_n - E[S_n]}{V[S_n]^{1/2}} = -\infty$  a.s.

### 5.7.2 大数の法則の精緻化

**定理 5.7.2.**  $\{X_n\} \subset L^2(\Omega)$  は独立,  $\sum_{n \in \mathbb{N}} \text{Var}[X_n]$  は発散するとする.

- (1)  $\varphi: (a, \infty) \rightarrow \mathbb{R}_+$  ( $a \in \mathbb{R}_+$ ) は単調増加関数で,  $\infty$  に発散し,  $\int_a^\infty \frac{dt}{\varphi(t)^2} < \infty$  を満たすとする. このとき,  $\frac{S_n - E[S_n]}{\varphi(\text{Var}[S_n])}$  は 0 に概収束する.  
 (2)  $\varphi(t) = \sqrt{t(\log t)^{1+\epsilon}}, \sqrt{t \log t \log^2 t \cdots \log^{k-1} t (\log^k t)^{1+\epsilon}}$  ( $\epsilon > 0$ ) においても 0 に概収束する.  
 (3)  $\{\text{Var}[X_n]\}$  が有界ならば,

$$\frac{S_n - E[S_n]}{\sqrt{n \log n \log^2 n \cdots \log^{k-1} n (\log^k n)^{1+\epsilon}}} \rightarrow 0 \text{ a.s.}$$

### 5.7.3 重複対数の法則

Hincin により種々の仮定の下で示され, 次第に精緻化された.

**定理 5.7.3.** 次のいずれかの条件を満たす独立確率変数列  $\{X_n\} \subset L^2$  は, 次の 3 式を満たす:

$$\limsup_{n \rightarrow \infty} \frac{S_n - E[S_n]}{(2V[S_n] \log^2 V[S_n])^{1/2}} = 1 \text{ a.s.} \quad \liminf_{n \rightarrow \infty} \frac{S_n - E[S_n]}{(2V[S_n] \log^2 V[S_n])^{1/2}} = -1 \text{ a.s.} \quad \limsup_{n \rightarrow \infty} \frac{|S_n - E[S_n]|}{(2V[S_n] \log^2 V[S_n])^{1/2}} = 1 \text{ a.s.}$$

- (1)  $\{X_n\}$  は独立な Gauss 変数列で,  $V[S_n] \rightarrow \infty$  かつ  $V[X_n]/V[S_n] \rightarrow 0$  を満たす.

## 5.8 大数の法則的現象

位相とは空間の観測であり, 数論とは統計の実験である. 数学的現象と物理的現象とを峻別しない形式であるところが, 確率論の好きなところの 1 つである. 大数の法則とは,  $S/n^\epsilon$  は  $\epsilon > 1/2$  について 0 に収束するという理論である.

### 5.8.1 積分

**命題 5.8.1.**

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 \cdots \int_0^1 \frac{x_1^q + x_2^q + \cdots + x_n^q}{x_1^p + x_2^p + \cdots + x_n^p} dx_1 dx_2 \cdots dx_n = \frac{p+1}{q+1} \quad (0 < p < q < \infty).$$

**[証明].**  $X_1, X_2, \dots$  を  $U((0, 1))$  に従う独立同分布列とすると,  $E \left[ \frac{S_{q,n}}{S_{p,q}} \right] S_{r,n} := \sum_{k=1}^n X_k^r$  を求めれば良い. 大数の法則より,  $n^{-1} S_{r,n} \rightarrow E[X_1^r] = (1+r)^{-1}$  a.s.. よって,  $S_{q,n}/S_{p,n} \rightarrow \frac{p+1}{q+1}$  a.s. 仮定  $0 < p < q < 1$  より,  $|S_{q,n}/S_{p,n}| \leq 1$  だから, 有界収束定理より平均も同じ値に収束する. ■

### 5.8.2 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,  $k/n \rightarrow x$  に概収束するから,  $f(x)$  は  $P_n(x)$  で近似できる. すると, 全ての関数は確率変数の退化 (特殊化) なのかもしれない. となると, 物理学理論が確率論化したのは自然で, いずれ全ての理論がそうなるであろうという新たな自然法則に向

き合いつつあるのかもしれない。2項展開の各項を 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 5.8.2 (Bernstein polynomial).**  $b_{k,n}(x) := \binom{n}{k} x^k (1-x)^{n-k}$  ( $k \in n+1$ ) の形で表される多項式を、 $n$  次の **Bernstein**

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

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

**定理 5.8.3 (Weierstrass の多項式近似).** 任意の連続関数  $f \in C([0,1])$  について、多項式の列  $(P_n)_{n \in \mathbb{N}}$  ( $\deg P_n = n$ ) が存在して、 $\lim_{n \rightarrow \infty} \max_{x \in [0,1]} |P_n(x) - f(x)| = 0$ .

**[証明].**

**構成** 任意の  $x \in [0,1]$  について、これを成功確率とする Bernoulli 試行  $B(n, x)$  に従う Bernoulli 列  $(X_i^x)_{i \in \mathbb{N}}$  を取る (独立同分布に従う確率変数の列の存在定理??). これが定める確率変数を  $S_n^x := \sum_{i=1}^n X_i^x$  とすると、この Bernoulli 試行

$B(n, x)$  の期待値の  $f$  による押し出しの期待値を  $P_n(x) := E\left[f\left(\frac{S_n^x}{n}\right)\right]$  とすると、 $k \in n+1$  回成功する確率はそれぞれ

$$P(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k} \text{ と表せるため,}$$

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

とも表せる。

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

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

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

■

**要諦 5.8.4.**  $f \in C([0,1])$  を、確率空間  $([0,1], \mathcal{B}([0,1]), P)$  上の実確率変数だと思えば、少し難しすぎる。そこで、離散的な確率

空間へと引き戻して考え、これらの離散空間からの実確率変数の列  $\lim_{n \rightarrow \infty} (n^{-1})^* f$  の極限だと考える：

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

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

### 5.8.3 Borel の正規数定理

正規列とは、有限列のパターン  $w \in \Sigma^{<\omega}$  の出現確率が一様分布に従う文字列をいう。

**定義 5.8.5 (normal number (Borel 1909)).**  $\Sigma$  を  $r$ -文字からなるアルファベットとする。  $N_S : \Sigma^{<\omega} \times \mathbb{N} \rightarrow \mathbb{N}$  を最初の  $n$  ブロックにパターン  $w$  が出現する回数  $N_S(w, n) := \#\{i \in [n] \mid S_i \cdots S_{i+|w|-1} = w\}$  とする。

(1)  $S \in \Sigma^\omega$  が正規であるとは、  $\lim_{n \rightarrow \infty} \frac{N_S(w, n)}{n} = \frac{1}{|\Sigma|^{|w|}}$  が成り立つことをいう。

(2)  $x \in [0, 1]$  が  $r$ -進正規数であるとは、  $r$ -進無限小数表示が正規であることをいう。

**定理 5.8.6 (Borel (1909)).**  $[0, 1]$  上の正規数全体の集合の Lebesgue 測度は 1 である。

**注 5.8.7.** 正規数の例は Sierpinski によって 1917 に与えられた。

### 5.8.4 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 5.8.8.** 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、  $(x_1, \dots, x_k)$  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 5.8.9.** 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 5.8.10.**  $\forall_{k \in \mathbb{N}} \sigma_n^k \Rightarrow \mu^k (n \rightarrow \infty)$ 。

### 5.8.5 熱力学的極限

**記法 5.8.11.**  $\Lambda_L := [-L, L]^d \subset \mathbb{R}^d$  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、  $L, N \rightarrow \infty$  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  $\lambda$  の Poisson 点過程という。

$S := \Lambda_L, p(A) := \frac{m(A)}{(2L)^d} (A \in \mathcal{G}(\mathbb{R}^d) \cap \mathcal{P}(\Lambda_L))$  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 5.8.12.** 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  $n(D, -) : \Omega \rightarrow \mathbb{Z}$  を

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

と定めると,

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

## 5.9 中心極限定理的現象

中心化過程  $S_n - E[S_n]$  が Lindeberg の十分条件を満たすとき, 1 次の項は  $\frac{1}{n}N(0, 1)$  となるという 1 次の確率的 Taylor 展開理論である.

### 5.9.1 極限計算

**命題 5.9.1.**

$$(1) \lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}.$$

### 5.9.2 Gauss の誤差論

「偶然誤差  $X$  が Gauss 分布に従うことの説明としては, それが無数の独立な微小偶然誤差  $X_1, \dots, X_n$  の和と考えて  $X$  の確率法則が Gauss 分布に近いことを証明する.

**定理 5.9.2.** 任意の  $n \in \mathbb{N}$  に対して,  $X_{n1}, \dots, X_{nN}$  ( $N(n) \in \mathbb{N}$ ) は独立とし,  $S_n := \sum_{k=1}^{N(n)} X_{nk}$  とする. 次の条件が満たされたとき,  $S_n$  の法則  $\mu_n$  は  $N(m, v)$  に弱収束する.

- (1)  $\epsilon_n := \sup_{k \in [N(n)], \omega \in \Omega} |X_{nk}(\omega)| \rightarrow 0.$
- (2)  $E[S_n] := \sum_{k \in N(n)} E[X_{nk}] \rightarrow m.$
- (3)  $\text{Var}[S_n] := \sum_{k \in N(n)} \text{Var}[X_{nk}] \rightarrow v.$

**[証明].** 中心極限定理の証明と全く平行に進む. ■

**要諦 5.9.3.** (1) 系統誤差を消去したとすれば, 測定値は真の値  $a$  と偶然誤差の和となるから,  $X \sim N(a, v)$  となるはずである.  
 (2) 測定回数を増やして平均を取った確率変数  $\bar{X} := \frac{1}{n}(X_1 + \dots + X_n)$  は  $N(a, v/n)$  に従う.  
 (3) しかし,  $X$  が中心値  $a$  の Cauchy 分布に従うとき,  $\bar{X}$  も  $X$  も測定の精度は変わらない.

### 5.9.3 Poisson の少数の法則

独立な事象列  $\alpha_1, \alpha_2, \dots, \alpha_n$  に対して,

- (1) 各確率  $p_k$  は偶然誤差のように極めて小さく,
- (2) 起こる確率の総和  $\sum_{k=1}^n p_k$  はほぼ  $\lambda$  で一定である

とき, 起こる個数  $N \in n+1$  はパラメータ  $\lambda$  の Poisson 分布に等しい.

**定理 5.9.4.** 任意の  $n \in \mathbb{N}$  に対して,  $X_{n1}, \dots, X_{nm}$  ( $m(n) \rightarrow \infty$ ) は独立で,  $P[X_{nk} = 1] = p_{nk}$ ,  $P[X_{nk} = 0] = 1 - p_{nk}$  とする.

$$\bar{p}_n := \max_{k \in [m(n)]} p_{nk} \rightarrow 0, \quad p_n := \sum_{k \in [m(n)]} p_{nk} \rightarrow \lambda$$

を満たすならば,  $N_n := \sum_{k=1}^{m(n)} X_{nk} \xrightarrow{n \rightarrow \infty} \text{Pois}(\lambda)$ .

**[証明].**  $X_n = (X_{n1}, \dots, X_{nm})$  の独立性より,

$$\begin{aligned} E[e^{izN_n}] &= \prod_{k \in [m]} E[e^{izX_{nk}}] = \prod_{k \in [m]} (p_{nk}e^{iz} + (1 - p_{nk})) \\ &= \prod_{k \in [m]} (1 + p_{nk}(e^{iz} - 1)) =: \prod_{k \in [m]} (1 + \alpha_{nk}). \end{aligned}$$

とおくと,

$$\begin{aligned} (1) \quad & \sum_{k \in [m]} \alpha_{nk} = \sum_{k \in [m]} p_{nk}(e^{iz} - 1) \rightarrow \lambda(e^{iz} - 1). \\ (2) \quad & \max_{k \in [m]} |\alpha_{nk}| \leq 2 \max_{k \in [m]} p_{nk} \rightarrow 0. \\ (3) \quad & \sum_{k \in [m]} |\alpha_{nk}| \leq 2 \sum_{k \in [m]} p_{nk} \rightarrow 2\lambda. \end{aligned}$$

だから,  $E[e^{izN_n}] \rightarrow \exp(\lambda(e^{iz} - 1))$ . これは Poisson 分布の特性関数である. ■

**命題 5.9.5.**  $X_1, \dots, X_k$  は独立で, それぞれ  $\text{Pois}(\lambda_i)$  に従うとする.  $X_1 + X_2 + \dots + X_n = n$  の下での  $(X_1, \dots, X_k)$  の条件付き確率分布は多項分布に等しい.

### 5.9.4 Weyl の一様分布定理

**定理 5.9.6.**  $\alpha \in \mathbb{R}^+ \setminus \mathbb{Q}$  について,  $a_n := \{n\alpha\} := n\alpha - [n\alpha]$  とすると, この数列は  $[0, 1]$  上に一様分布する. すなわち, 次の同値な条件が成り立つ:

$$\begin{aligned} (1) \quad & \frac{1}{N} \sum_{n=1}^N \delta_{a_n} \Rightarrow \mu_{U([0,1])}. \\ (2) \quad & \text{任意の } f \in C([0,1]) \text{ について, } \lim_{N \rightarrow \infty} \frac{1}{N} \int_{n=1}^N f(a_n) = \int_0^1 f(x) dx. \\ (3) \quad & \text{任意の } (a, b) \subset [0,1] \text{ について,} \\ & \lim_{N \rightarrow \infty} \frac{\#\{n \in \mathbb{N} \mid a_n \in [a, b]\}}{N} = b - a. \end{aligned}$$

## 5.10 経験分布論

### 5.10.1 大数の法則の精緻化

**定理 5.10.1 (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:  $P[\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0] = 1$ .

### 5.10.2 中心極限定理の精緻化

## 5.11 大偏差原理

$\{X_n\} \subset L^1(\Omega)$  が独立同分布のとき,  $\frac{S_n}{n}$  は定数  $\alpha_1$  に概収束するのであった. このとき分布は  $\delta_0$  に弱収束し,  $m \notin \bar{A}$  を満たす  $A \in \mathcal{G}(\mathbb{R})$  について  $P[S_n/n \in A] \rightarrow 0$  でもある. (積率母関数が 0 の近傍で存在するならば) この収束の速さは,  $n$  の指数関数の速さであり, 係数  $\gamma(a) := \lim_{n \rightarrow \infty} \frac{1}{n} \log P[S_n/n \geq a]$  も特定出来る.

## 歴史 5.11.1.

- (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.

## 5.11.1 定義

## 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  $P(X_n = a) \xrightarrow{n \rightarrow \infty} 0$ , 中心極限定理により指数減衰  $P(X_n \geq a) = \int_{z \geq a} p_n^{X_n}(z) e^{-nI(z)} dz = e^{-I(a)}$  (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.

**記法 5.11.2.**  $X$  を可分完備距離空間とし, 同時にある  $\sigma$ -代数  $\mathcal{F}$  により可測空間でもあるとする.

**定義 5.11.3 (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  $I: X \rightarrow \mathbb{R}_+$  が存在して, 任意の可測集合  $\Gamma \in \mathcal{F}$  に対して

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

が成り立つことをいう. このとき,  $I$  を **rate 関数** という.

**定義 5.11.4 (good rate function).**  $I$  が良いレート関数であるとは, 任意の  $l \geq 0$  に対して, 等位集合  $I^{-1}((-\infty, l]) = \{x \in X \mid I(x) \leq l\}$  がコンパクトであることをいう.

## 5.11.2 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 5.11.5.** 有界閉区間  $[a, b]$  上の連続関数  $f, g > 0$  について,

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

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

**注 5.11.6.** 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 5.11.7.** (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  $n! = \int_0^\infty x^n e^{-x} dx \approx n^n e^{-n} \ (n \rightarrow \infty)$  を Laplace の原理から説明する. 積分変換によって

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

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



**定理 5.11.8 (Laplace's method (1774)).** 関数  $f: [a, b] \rightarrow \mathbb{R}$  は点  $x_0 \in (a, b)$  でただ一つの最小点を取るとする:  $f''(x_0) > 0$ .

このとき,  $n \rightarrow \infty$  において,

$$\int_a^b e^{-nf(n)} dx \sim \sqrt{\frac{2\pi}{nf''(x_0)}} e^{-nf(x_0)}.$$

**要諦 5.11.9 (n-CatLab<sup>†1</sup>).**  $e^{-nf(x)}$  の積分は,  $f$  を極小点  $x_0$  にて 2 次近似したもの

$$g(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$$

で置き換えた  $e^{-ng(x)}$  の  $\mathbb{R}$  上の積分

$$\int_{-\infty}^{\infty} g(x) dx = \int_{-\infty}^{\infty} \left( \frac{2}{nf''(x_0)} (x \text{ のモニックな 2 次式}) \right) = \sqrt{\frac{2\pi}{nf''(x_0)}}$$

を最大値  $e^{-nf(x_0)}$  でスケールした分だけの違いを除いて, 漸近的に等しい, という結果である. なお, 積分区間については, 正規分布の減衰は極めて速いためと理解出来る. したがって結果的に Gauss 積分と等しくなる. 正規分布が 3 次以上のキュムラントが消えるのも, 偶然の一致ではないだろう.

### 5.11.3 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,  $\lim_{n \rightarrow \infty} P(Y_n \in A) = 0$  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 5.11.10.** 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  $\text{Dom}(M_v) := \{t \in \mathbb{R} \mid M(t) := E[e^{tX_1}] < \infty\}$  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)  $\forall a > E[X_1] \lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq an) = -I(a)$ . 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,  $\lim_{z \rightarrow \pm\infty} I(z) = \infty, \forall z \in \mathbb{R} \ I(z) \geq 0 = I(E[X_1])$  を満たす.

**要諦 5.11.11.**  $E[X_1] < a$  について  $A := [a, \infty)$  とおくと,

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

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,  $n \rightarrow \infty$  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.

### 5.11.4 Schilder の理論

#### 確率過程に関する大偏差原理

$I(\phi)$  を, 連続関数  $\phi$  の持つエネルギーだとして, 「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.

<sup>†1</sup> [https://golem.ph.utexas.edu/category/2021/10/stirlings\\_formula.html](https://golem.ph.utexas.edu/category/2021/10/stirlings_formula.html)

## 5.11.5 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 5.11.12.**  $(X, \mathcal{F})$  上の確率測度の列  $(\mu_n)$  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、

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

## 5.12 集中不等式

集中不等式は、確率分布の偏差  $|X - \mu|$  を捉える。極限分布が Gauss であるとき、これを用いて近似することがまず考えられるが、近似誤差が大きい 5.12.1。そこで、極限定理に依らずに、「裾が軽い」ことを示す方法論が必要となる。その一つが集中不等式である。（極限定理では特性関数が活躍したが、）ここでは、積率母関数に注目することとなる。

## 5.12.1 Berry-Esseen の不等式

## 中心極限定理の誤差の評価

極限分布として正規分布を持つ確率変数を、その極限分布によって近似したときの近似誤差は、 $1/\sqrt{N}$  のオーダーを持つ。これはあまりにも遅い、収束が線型よりも遅い。そのために、別の方法で目標の分布が「裾が軽い」ことを示す必要が応用上出てくる、これが集中不等式である。

**定理 5.12.1 (Berry-Esseen central limit theorem).** 独立同分布を持つ確率変数列  $\{X_n\} \subset \mathcal{L}^1(\Omega, \mathcal{F}, P)$  は  $E[X_n] = \mu \in \mathbb{R}, \text{Var}[X_n] = \sigma^2 \in \mathbb{R}_+$  を満たすとする。  $S_N := X_1 + \cdots + X_N$  について、

$$Z_N := \frac{S_N - E[S_N]}{\sqrt{\text{Var}[S_N]}} = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^N (X_i - \mu) \xrightarrow{d} N(0, 1)$$

であるが、 $\rho := \frac{E[|X_1 - \mu|^3]}{\sigma^3}$  と  $g \sim N(0, 1)$  について、

$$\forall N \in \mathbb{N} \quad \forall t \in \mathbb{R} \quad |P[Z_N \geq t] - P[g \geq t]| \leq \frac{\rho}{\sqrt{N}}$$

## 5.12.2 Chebyshev の不等式

最も即時的で古典的な上界を与え、多くの場合線型に過ぎない。

**定理 5.12.2.**  $X \sim (\mu, \sigma^2)$  を実確率変数とする。

- (1) (Markov)  $X \geq 0$  ならば、 $\forall t > 0 \quad P[X \geq t] \leq \frac{E[X]}{t}$ .
- (2) (Chebyshev)  $\forall t > 0 \quad P[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}$ .

## 5.12.3 Hoeffding の不等式

Chebyshev の不等式 1.7.13 より,

$$P[|S - E[S]| \geq a\sqrt{\text{Var}[S]}] \leq a^{-2}$$

であるが,  $X$  が一様に有界であるとき, 右辺は  $e^{-a^2/2}$  に置き換えられる [Kotz et al., 2005] (PROBABILITY INEQUALITIES FOR SUMS OF BOUNDED RANDOM VARIABLES). これを一般化して, Gauss 分布と全く同様な収束レート  $e^{-t^2/2}$  による評価を与えるのが Hoeffding の不等式である. これは理想的な形で, 中心極限定理の代替となっている. 証明はモーメント母関数により, これは一般の有界な確率変数に一般化出来るが, 鋭さは落ちる.

**定理 5.12.3.**  $X_1, \dots, X_N$  を独立な Rademacher 確率変数とし,  $a = (a_1, \dots, a_N) \in \mathbb{R}^N$  とする. このとき,

$$\forall t \in \mathbb{R}_+ \quad P\left[\sum_{i=1}^N a_i X_i \geq t\right] \leq \exp\left(-\frac{t^2}{2\|a\|_2^2}\right).$$

**定理 5.12.4 (Bernstein).**  $\text{Im}(X_i) \subset [0, 1]$  を独立な可積分関数とし,

$$S := X_1 + \dots + X_n, \quad \mu := \frac{E[S]}{n}$$

を考える. 任意の  $t \in (0, 1 - \mu)$  について,

$$P[S - E[S] \geq nt] \leq e^{-2nt^2}.$$

この一般化が Hoeffding の不等式である.

**定理 5.12.5 (Hoeffding's inequality for general bounded random variable [Hoeffding, 1963]).**  $X_1, \dots, X_N$  を独立な有界確率変数とする:  $\forall i \in [N] \exists m_i, M_i \in \mathbb{R} \text{ Im } X_i \subset [m_i, M_i]$ . このとき,

$$\forall t > 0 \quad P\left[\sum_{i=1}^N (X_i - E[X_i]) \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^N (M_i - m_i)^2}\right)$$

**系 5.12.6 (正規化した場合).**  $(Z_i)_{i \in [n]}$  を  $[0, 1]$  上に台を持つ独立同分布に従うとする. このとき,

$$\forall \epsilon > 0 \quad P\left[\left|\frac{1}{n} \sum_{i \in [n]} Z_i - E[Z_1]\right| \geq \epsilon\right] \leq 2e^{-2n\epsilon^2}.$$

**応用例 5.12.7 (平均の頑健推定).** 独立な観測  $X_1, \dots, X_N \sim (\mu, \sigma^2)$  から平均  $\mu$  を, 誤差  $\epsilon$  以内で推定したい.

- (1)  $N = O(\sigma^2/\epsilon^2)$  個の標本が存在すれば, 確率  $3/4$  以上で  $(\mu - \epsilon, \mu + \epsilon)$  に入る推定量が構成できる.
- (2) 任意の  $\delta \in (0, 1)$  について,  $N = O(\log(\delta^{-1})\sigma^2/\epsilon^2)$  個の標本が存在すれば, 確率  $1 - \delta$  以上で  $(\mu - \epsilon, \mu + \epsilon)$  に入る推定量が構成できる.

## 5.12.4 Chernoff の不等式

Bernoulli 変数について, 母数  $p_i$  が小さすぎるとき (Poisson の少数の法則的な現象のとき), Hoeffding の不等式は見当違いの結果を与える.

**定理 5.12.8.**  $X_i \sim B(p_i)$  を Bernoulli 確率変数とする. 和  $S_N := \sum_{i=1}^N X_i$  の平均を  $\mu := E[S_N]$  で表すと,

$$\forall t > \mu \quad P[S_N \geq t] \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t.$$

**要諦 5.12.9.** これは Poisson 分布的な結果である. 実際  $X \sim \text{Pois}(\lambda)$  ならば,

$$\forall t > \lambda \quad P[X \geq t] \leq e^{-\lambda} \left(\frac{e\lambda}{t}\right)^t.$$

これを、各第  $i$  回での成功確率  $p_i$  がバラバラな場合でも行っている。

**注 5.12.10** (Poisson 尾部の観察).  $k!$  を Stirling の公式で近似すると

$$P[X = k] \sim \frac{1}{\sqrt{2\pi k}} e^{-\lambda} \left(\frac{e\lambda}{k}\right)^k$$

を得るから、Poisson 分布の尾部と言っても、一番小さい部分が殆どすべてを占めてしまう。大偏差原理的な現象である。

**定理 5.12.11** (小偏差の場合).  $X_i \sim B(p_i)$  を Bernoulli 確率変数とする。和  $S_N := \sum_{i=1}^N X_i$  の平均を  $\mu := E[S_N]$  で表すと、

$$\exists c > 0 \quad \forall \delta \in (0, 1] \quad P[|S_N - \mu| \geq \delta \mu] \leq 2e^{-c\mu\delta^2}$$

**要諦 5.12.12.** これは Poisson 分布の

$$\forall t \in (0, \lambda] \quad P[|X - \lambda| \geq t] \leq 2 \exp\left(-\frac{ct^2}{\lambda}\right)$$

に対応する。Pois( $\lambda$ ) は平均  $\lambda$  近くでは  $N(\lambda, \lambda)$  にすごく似ているが、大偏差部分では裾は重い： $(\lambda/t)^t$ 。

### 5.12.5 確率グラフ

ネットワークの確率モデルとなる。 $p$  が十分大きく、平均して  $O(\log n)$  の次数を持つとき、どの頂点の次数も  $[0.9d, 1.1d]$  の間にあって、他と結ばれ過ぎていたり、孤立していたりすることは殆どない。

**定義 5.12.13.** Erdos-Renyi model  $G(n, p)$  ( $n \in \mathbb{N}, p \in [0, 1]$ ) とは、 $n$  個の頂点と、その任意の 2 頂点の間を確率  $p$  で辺が存在するグラフである。

**補題 5.12.14.** 任意の頂点の次数の期待値は  $d := (n-1)p$  となる。

**定理 5.12.15 (dense graphs are almost regular).** ある  $C > 0$  が存在して、任意の  $d \geq C \log n$  を満たすランダムグラフ  $G \sim G(n, p)$  は、十分大きな確率 (0.9 以上) で、任意の頂点の次数が  $[0.9d, 1.1d]$  の間に入る。

### 5.12.6 劣 Gauss 分布

Bernoulli 確率変数  $X_i$  について集中不等式を示したが、どこまで一般の確率変数に適用できるのか？

**命題 5.12.16.** 確率変数  $X$  について、次の 4 条件は同値で、 $E[X] = 0$  ならば (5) も同値。また、ある定数  $C > 0$  が存在して、パラメータ  $K_i > 0$  はそれぞれ互いの  $C$  倍を超えない。

- (1)  $X$  の尾部は  $\forall t \in \mathbb{R}_+ \quad P[|X| \geq t] \leq 2 \exp\left(-\frac{t^2}{K_1^2}\right)$  を満たす。
- (2)  $X$  のモーメントは  $\forall p \geq 1 \quad \|X\|_{L^p} \leq K_2 \sqrt{p}$  を満たす。
- (3)  $X^2$  の積率母関数は  $\forall |\lambda| \leq 1/K_3 \quad E[e^{\lambda^2 X^2}] \leq \exp(K_3^2 \lambda^2)$  を満たす。
- (4)  $X^2$  の積率母関数は  $E[e^{X^2/K_4^2}] \leq 2$ 。
- (5)  $X$  の積率母関数は  $\forall \lambda \in \mathbb{R} \quad E[e^{\lambda X}] \leq e^{K_5^2 \lambda^2}$ 。

**定義 5.12.17 (sub-gaussian norm).** 命題の条件を満たす確率変数  $X$  を劣ガウスであるといい、劣ガウスノルムを (4) を満たす最小の定数  $K_4$ 、すなわち

$$\|X\|_{\psi_2} = \inf \left\{ t > 0 \mid E[e^{X^2/t^2}] \leq 2 \right\}$$

で定める。

**例 5.12.18.**

- (1)  $X \sim N(0, \sigma^2)$  ならば、 $\|X\|_{\psi_2} \leq C\sigma$ 。

(2)  $X$  が Rademacher ならば,  $\|X\|_{\psi_2} \leq \frac{1}{\sqrt{\log 2}} =: C$ .

(3)  $X$  が有界ならば,  $\|X\|_{\psi_2} \leq C\|X\|_{\infty}$ .

一方で, Poisson 分布, 指数分布, Pareto 分布, もちろん Cauchy 分布は裾が重く, 劣 Gauss ではない. □

### 5.12.7 Orlicz 空間

劣 Gauss 分布をさらに一般化すると, Orlicz 空間にいきつく.

### 5.12.8 一般化 Hoeffding 不等式

劣 Gauss ノルムの言葉を用いて, Hoeffding 不等式は一般の劣 Gauss 確率変数に一般化出来る.

### 5.12.9 劣指数分布

**補題 5.12.19.**  $X$  を確率変数とする. 次の 2 条件は同値:

- (1)  $X$  は劣 Gauss である.
- (2)  $X^2$  は劣指数である.

このとき,  $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$ .

### 5.12.10 Bernstein の不等式

劣指数分布

**定理 5.12.20.**  $X_1, \dots, X_N$  を平均 0 の劣指数確率変数とする. このとき,

$$\exists c > 0 \quad \forall t \in \mathbb{R}_+ \quad P \left[ \left| \sum_{i=1}^N X_i \right| \geq t \right] \leq 2 \exp \left( -c \min \left( \frac{t^2}{\sum_{i=1}^N \|X_i\|_{\psi_1}^2}, \frac{t}{\max_{i \in [N]} \|X_i\|_{\psi_1}} \right) \right).$$

## 第 6 章

# 離散確率過程に対する極限定理

独立でない確率変数の列への扱いを考える．代数的な特性で特徴付ける関数解析的な方法と，独立性を緩める確率論的な方法の 2 つが考えられる．後者では Kolmogorov の不等式の類似物が期待される．

### 6.1 非独立過程の大数の法則

独立でない場合でも大数の法則が成り立ち，結局は直交性（あるいはさらにそれを弱めたもの）が大数の法則を誘発しているのであり，独立性はその一つの表現でしかない．

#### 6.1.1 $L^2$ -有界な直交系の列

確率変数列  $\{X_n\} \subset L^2(\Omega)$  が有界な直交系ならば，大数の法則が成り立つ．

**定理 6.1.1 (Rademacher-Men'shov).**  $(X, \mathcal{B}, \mu)$  を測度空間， $\{f_n\} \subset L^2(X; \mathbb{C})$  を  $L^2$ -有界な直交系とする．このとき， $\sum_{n \in \mathbb{N}} |c_n|^2 (\log n)^2 < \infty$  を満たす任意の複素数列  $\{c_n\} \subset \mathbb{C}$  について， $\sum_{n=1}^{\infty} c_n f_n$  は概収束する．特に，Kronecker の定理より

$$\forall \epsilon > 0 \quad \frac{1}{\sqrt{N}(\log N)^{3/2+\epsilon}} \sum_{n=1}^N f_n \xrightarrow{N \rightarrow \infty} 0 \text{ a.e.}$$

**要諦 6.1.2.** これは Hilbert 空間  $L^2(X)$  上の Fourier 級数論も含んだ消息である． $f_n(x) = e^{2ni\pi x} \in L^2([0, 1])$  とすると正規直交基底だから明らかに有界な直交系で，Fourier 係数の級数が条件  $\sum_{n \in \mathbb{N}} |\hat{f}(n)|^2 (\log n)^2 < \infty$  を満たすとき，Fourier 級数  $\sum_{n \in \mathbb{N}} \hat{f}(n) e^{2ni\pi x}$  は概収束する．収束先は  $f$  である．しかしこの消息は非本質的で，何の仮定を置かずとも， $f \in L^2([0, 1])$  の Fourier 級数は  $f$  に概収束する (Carleson-Hunt 1968)

**命題 6.1.3 (直交系の最大不等式).**  $\{\psi_i\}_{i \in \mathbb{N}+1} \subset L^2(X; \mathbb{C})$  を直交系とする．

$$\forall a_0, \dots, a_N \in \mathbb{C} \quad \int_X \left( \max_{0 \leq n \leq N} \left| \sum_{k=0}^n a_k \psi_k \right| \right)^2 d\mu \leq \left( \max_{0 \leq k \leq N} \int_X |\psi_k|^2 d\mu \right) \left( \sum_{k=0}^N (A_k^{-1/2})^2 \right)^2 \left( \sum_{k=0}^N |a_k|^2 \right).$$

#### 6.1.2 乗法系

一般の  $\{X_n\} \subset L^1(\Omega)$  の列についても，直交性と  $L^2$ -有界性に似た条件が成り立つならば，大数の法則が成り立つ．

**定義 6.1.4 (multiplicative system).**  $\{X_n\} \subset L^1(\Omega)$  が乗法系であるとは，

$$\forall s \in \mathbb{N}+ \quad \forall 1 \leq n_1 < \dots < n_s \in \mathbb{N}+ \quad X_{n_1} \cdots X_{n_s} \in L^1(\Omega) \wedge E[X_{n_1} \cdots X_{n_s}] = 0$$

を満たすことをいう．特に，乗法系は中心化されている．

**定理 6.1.5.**  $\{X_n\} \subset L^1(\Omega)$  は次の2条件を満たすとする.

- (1) 一様有界である:  $\exists M \in \mathbb{R} \quad \forall n \in \mathbb{N}^+ \quad \forall \omega \in \Omega \quad |X_n(\omega)| \leq M$ .
- (2) 乗法系である.

このとき, 任意の  $\alpha \in l^2(\mathbb{N})$  について,  $\sum_{j \in \mathbb{N}} \alpha_j X_j$  は概収束する.

**系 6.1.6.** 係数列  $\{p_k\} \subset \mathbb{R}$  と発散列  $\{P_k\} \subset l_0(\mathbb{N}; \mathbb{R}^+)$  は  $\sum_{k \in \mathbb{N}} \frac{p_k^2}{P_k^2} < \infty$  を満たすとする. このとき,

$$\frac{1}{P_n} \sum_{k=1}^n p_k X_k \xrightarrow{n \rightarrow \infty} 0 \quad \text{a.s.}$$

なお,  $p_k = 1, P_k = k$  の場合は, 大数の強法則を含意する.

**補題 6.1.7 (乗法系の最大不等式).** 定理の条件 (1), (2) の下で, 任意の実数列  $\{b_k\} \subset \mathbb{R}$  について, 次の評価が成り立つ:

$$\left( E \left[ \max_{1 \leq i \leq n} \sum_{k=1}^i b_k X_k \right] \right)^2 \leq 2K^2 \sum_{p=1}^n b_p^2.$$

**要諦 6.1.8 (最大不等式).** 議論の方針の一番の違いは Kolmogorov の不等式の類似物が見当たらないことである. しかし, この議論は大数の法則が独立性によるものというよりも, 乗法性という代数的条件に起因することをえぐり出すことになる.

### 6.1.3 エルゴード定理

大数の法則はエルゴード理論の言葉で一般化される.

**定義 6.1.9.** 確率過程  $\xi: T \rightarrow L^2(\Omega)$  が

- (1)  $E[\xi(t)] = m, K_\xi(t, s) = k(t-s)$  を満たすとき **広義定常** という.
- (2) 任意の有限次元分布が平行移動  $\tau > 0$  について不変であるとき, **狭義定常** という.
- (3) 広義定常過程  $\xi$  についてエルゴード性が成り立つとは,  $\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(\tau) d\tau = m$  をいう.

**定理 6.1.10.** 広義定常な  $L^2$ -過程  $\xi$  の共分散が

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T k(\tau) d\tau = 0$$

を満たすならば, エルゴード性が成り立つ.

**定理 6.1.11.** 狭義定常な可測過程  $\xi: T \rightarrow L^1(\Omega)$  は, 個別エルゴード性を満たす:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \int_0^N \xi(t, \omega) dt \quad \text{a.s.}$$

一般にはこの極限は  $E\xi(t, \omega)$  とは一致しないが, 条件付き期待値  $E[\xi|J]$  に a.s. で等しい.

**定義 6.1.12.** 任意の  $\tau \in \mathbb{R}$  について, 時刻の  $\tau$  だけのずらしは 1-パラメータ変換群をなす. これに関する不変集合の全体を  $J$  で表すと, これは零集合と充満集合の全体を含む  $\sigma$ -代数である. これが 2 に一致するとき, 過程を **エルゴード的** であるという.

## 6.2 McLeish の中心極限定理

マルチンゲール中心極限定理を導く, Lindeberg の中心極限定理を含意する主張である.

**定義 6.2.1 (triangular array).** 発散列  $\{k_n\} \subset \mathbb{R}$  について, 二重数列  $(X_{nj})_{n \in \mathbb{N}, j \in [k_n]}$  を **三角整列** という. これについて中心極限定理が成り立つとは,

$$\sum_{j=1}^{k_n} X_{nj} \xrightarrow{w} N(0, v)$$



を満たすことをいう。

**定理 6.2.2 (McLeish).**  $\{X_{nj}\} \subset L(\Omega)$  を次の 3 条件を満たす二重数列とする：

$$(M1) \quad \forall t \in \mathbb{R} \quad \left\{ Z_n(t) := \prod_{j=1}^{k_n} (1 + itX_{nj}) \right\}_{n \in \mathbb{N}^+} \subset L^1(\Omega) \text{ は一様可積分である.}$$

$$(M2) \quad \sum_{j \in [k_n]} X_{nj}^2 \xrightarrow{P} \nu \in \mathbb{R}_+.$$

$$(M3) \quad \max_{j \in [k_n]} |X_{nj}| \xrightarrow{P} 0.$$

このとき、 $(X_{nj})$  について中心極限定理が成り立つことは、 $\forall t \in \mathbb{R} \quad E[Z_n(t)] \rightarrow 1$  に同値。

**系 6.2.3 (大乘法系に関する中心極限定理).** 乗法系  $\{X_n\} \subset L_0^1(\Omega)$  が次の 2 条件を満たすならば、中心極限定理  $\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \xrightarrow{w} N(0, \nu)$  が成り立つ。

(1)  $(X_n)$  は一様有界である： $\exists M \in \mathbb{R} \quad \forall n \in \mathbb{N}^+ \quad \forall \omega \in \Omega \quad |X_n(\omega)| \leq K$ .

(2)  $(X_n^2 - \nu)$  ( $\nu \in \mathbb{R}_+$ ) は乗法系である。

**補題 6.2.4 (augmented multiplicative system の特徴付け).** 上の系の 2 条件が成り立つことは、次の条件に同値：

$$\forall k \in \mathbb{N} \quad \forall 1 \leq n_1 < \dots < n_k \quad E[X_{n_1}^2 \cdots X_{n_k}^2] = \nu^k, \quad E[X_{n_1} \cdots X_{n_k}] = 0.$$

このような乗法系を**大乘法系**とも言える。

**例 6.2.5 (Lindeberg の中心極限定理と議論する  $\nu$  が違い得る).**  $X_1$  を Rademacher 確率変数とする。

$$P[X_j = \pm j] = \frac{1}{2j^2}, \quad P[X_j = \pm 1] = \frac{1}{2} \left( 1 - \frac{1}{j^2} \right) \quad (j = 2, 3, \dots)$$

を満たす独立列  $\{X_j\}_{j \in \mathbb{N}^+}$  を考えると、極限では再び  $\lim_{j \rightarrow \infty} P[X_j = \pm 1] = 1/2$  となる。

(1)  $\{X_j\}$  は中心化されており、 $\nu = \lim_{n \rightarrow \infty} \text{Var}[S_n] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n E[X_j^2] = 2$ .

(2) 任意の  $\epsilon > 0$  について、 $\frac{1}{n} \sum_{j=1}^n E[X_j^2, |X_j| \geq \epsilon \sqrt{n}] \xrightarrow{n \rightarrow \infty} 1$  より、Lindeberg-Feller の定理より、 $N(0, 2)$  には弱収束しない。

(3)  $\prod_{j=1}^n (1 + itX_j)$  は任意の  $t \in \mathbb{R}$  について  $L^2$ -有界であり、特に一様可積分である。また、 $\sum_{j=1}^n \frac{X_j^2}{n} \xrightarrow{P} 1$ 。  $P[\max_{1 \leq j \leq n} |X_j| > \epsilon \sqrt{n}] = 1/j^2 \rightarrow 0$  であるから  $\max_{1 \leq j \leq n} \frac{|X_j|}{\sqrt{n}} \xrightarrow{P} 0$  でもある。よって、McLeish の中心極限定理によれば、

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n X_j \rightarrow N(0, 1).$$

□

## 6.3 第 1 逆正弦法則

Andersen (1953,[?]) により発見された独立確率変数の和に見られる確率変動の性質を、Bernolli 試行を例に取ってみる。この確率過程は空間内での対称な単純ランダムウォークであり、その滞在時間は、極限としての Brown 運動の滞在時間の確率分布が  $\arcsin$  を用いて表せるために、 $\arcsin$  の消息が観測される。

### 6.3.1 投票でのリード

**補題 6.3.1 (鏡像原理).** 点  $A$  から  $B$  に行く道のうち、 $x$  軸に接するか交わる道は、 $A$  の  $x$  に関する対称点  $A'$  から  $B$  に行く道の数に等しい。



**系 6.3.2 (投票問題: Bertrand 1887).** 候補者  $P$  が  $p$  票を得て、候補者  $Q$  が  $q < p$  票を得たとする。このとき、票読の間、常に  $P$  が  $Q$  よりも得票数が多かった確率は  $\frac{p-q}{p+q}$  である。

**系 6.3.3.** 銅貨投げゲームにおいて、一度もスコアの逆転現象が起こらない確率が一番高い。

**例 6.3.4** (リードの逆転は思ったよりも起こらない). 処置群と対照群で、降順のデータ  $a_1 > \cdots > a_n, b_1 > \cdots > b_n$  が得られたとする。これらを混ぜて、長さ  $2n$  の有限な単調減少列を定める。完全に無効な処置である場合、

$$|\{i \in [n] \mid a_i > b_i\}| \geq k$$

となる確率は  $\frac{n-k+1}{n+1}$  である。これを用いて Galton は 1876 年に検定を行なっていて、そのデータを Charles Darwin が引用している。このとき、 $n = 15, k = 13$  であったから、このような現象は確率  $3/16$  で起こるので全く棄却できないが、Galton は有効としてしまった。□

**例 6.3.5** (タイも驚くほど起こらない). 硬貨投げを 10000 回行う。タイが 140 回以上の確率は 0.157 で、14 回以下の確率は 0.115 である。□

### 6.3.2 第 1 逆正弦法則

**定理 6.3.6 (第 1 逆正弦法則).** ある  $\alpha \in (0, 1)$  に対して、 $n \rightarrow \infty$  のとき、正の側で費やされる時間の割合  $k/n$  が  $k/n < \alpha$  となる確率は次に等しい：

$$P[k/n < \alpha] = \pi^{-1} \int_0^\alpha \frac{dx}{\sqrt{x(1-x)}} = \frac{2}{\pi} \arcsin \sqrt{\alpha}.$$

**系 6.3.7.**

- (1) 粒子が同じ側で約 97.6% の時間を過ごす確率は 20% である。
- (2) 10 回中に 1 回は、粒子は 99.4% の時間を同じ側で過ごす。

## 6.4 Donsker の定理

## 第 7 章

## 参考文献

## 参考文献

- [Billingsley, 1999] Billingsley, P. (1999). Convergence of probability measures. Wiley series in probability and mathematical statistics. New York: Wiley, 2 edition.
- [Bogachev, 2018] Bogachev, V. I. (2018). Weak Convergence of Measures, volume 234 of Mathematical surveys and monographs. Providence, R.I. : American Mathematical Society.
- [Hoeffding, 1963] Hoeffding (1963). Probability inequalities for sums of bounded random variables. Journals of the American Statistical Association, 58(301):13–30.
- [Kolmogorov, 1963] Kolmogorov, A. N. (1963). On tables of random numbers. Sankhyā: The Indian Journal of Statistics, Series A (1961-2002), 25(4):369–376.
- [Kotz et al., 2005] Kotz, S., Read, C. B., Balakrishnan, N., and Vidakovic, B. (2005). Encyclopedia of Statistical Science. Wiley.
- [Landkov, 1972] Landkov, N. S. (1972). 確率論入門. 森北出版. Markov 過程に詳しく, ここに多くの読み残しあり.
- [Mumford, 2000] Mumford, D. (2000). The Dawning of the Age of Stochasticity, pages 197–218. American Mathematical Society.
- [Rota, 2001] Rota, G.-C. (2001). Twelve problems in probability no one likes to bring up. In Algebraic Combinatorics and Computer Sciences, pages 57–93. Springer, Milano.
- [Simpson, 2017] Simpson, A. (2017). Probability sheaves and the giry monad. In Bonchi, F. and König, B., editors, 7th Conference on Algebra and Coalgebra in Computer Science, pages 1:1–1:6.
- [von Mises, 1919] von Mises, R. (1919). Fundamentalsätze der wahrscheinlichkeitsrechnung. Mathematische Zeitschrift, 4:1–97.
- [von Mises and Geiringer, 1957] von Mises, R. and Geiringer, H. (1957). Probability, Statistics and Truth. Allen and Unwin, second revised english edition edition.
- [伊藤清, 1991] 伊藤清 (1991). 確率論. 岩波基礎数学選書. 岩波書店, 2 edition.
- [高信敏, 2015] 高信敏 (2015). 確率論, volume 4 of 共立講座数学の魅力. 共立出版.

# 参考文献

- [1] Billingsley76 Patrick Billingsley. (1976). *Probability and Measures*.
- [2] Billingsley68 Patrick Billingsley. (1968). *Convergence of Probability Measures*.
- [3] Billingsley71 Patrick Billingsley. (1971). *Weak Convergence of Measures*.
- [4] Bogachev Bogachev, V. I. (2018) *Weak Convergence of Measures*.
- [5] Rudin Rudin, W. *Functional Analysis*.
- [6] Dunford-Schwartz Dunford and Schwartz. *Linear Operators*.
- [7] HarmonicAnalysisOnSemigroups Berg, and Christensen, and Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*.
- [8] 伊藤清伊藤清. 『確率論』
- [9] 柴田義貞柴田義貞 (1981) 『正規分布』 (UP 応用数学選書, 東京大学出版会).
- [10] ハンドブック 『確率論ハンドブック』
- [11] 西尾西尾真喜子. 『確率論』
- [12] 高信高信敏 『確率論』 (共立出版, 数学の魅力 4).
- [13] Alexandorff Alexandorff, A. D. (1943). Additive set-functions in abstract spaces.
- [14] 清水良一清水良一. (1976). 中心極限定理.
- [15] 盛田盛田健彦 (2004) 『実解析と測度論の基礎』 (数学レクチャーノート, 培風館).
- [16] Yoshida Yoshida, Kousaku. *Functional Analysis*.
- [17] Butzer Butzer, P. L., and Oberdorster, W. (1975). Linear Functionals Defined on Various Spaces of Continuous Functions on  $\mathbb{R}$ . *Journal of Approximation Theory*. 13: 451-469.
- [18] Folland Folland, Gerald. *Real Analysis: Modern Techniques and Their Applications*.
- [19] Voevodsky Vladimir Voevodsky "Notes on categorical probability"
- [20] Kolmogorov31 Kolmogorov, A. N. (1933). Analytical methods in probability theory. 「私は Kolmogorov のこの論文 (「解析的方法」) の序文にあるアイデアからヒントを得て, マルコフ過程の軌道を表す確率微分方程式を導入したが, これが私のその後の研究の方向を決めることになった。」
- [21] Kolmogorov33 Kolmogorov, A. N. (1933). Grundbeegriffe der Wahrscheinlichkeitsrechnung. (確率論の基礎概念).
- [22] Levy37 Lévy, P. (1937). Théorie de l'addition des variables aléatoires. (独立確率変数の和の理論).
- [23] Doob37 Doob, J. L. (1937). Stochastic Processes Depending on a Continuous Parameter. *Transactions of the American Mathematical Society*. 42. 「正則化」の概念の初出.
- [24] ItoPhDthesis Ito, K. (1942). Differentiation of Stochastic Processes (Infinitely Divisible Laws of Probability). *Japanese Journal of Mathematics*. 18: 261-301. Levy-Ito の定理が示されている.
- [25] Ito42 Ito, K. (1942). Differential Equations Determining Markov Processes. 全国紙上数学談話会誌 (1077): 1352 – 1400.
- [26] Ito51 Ito, K. (1951). On Stochastic Differential Equations. *Memoir of American Mathematical Society*. 4: 1-51. 戦後間もなかったため, Doob の計らいでメモワールシリーズの 1 つとして米国で発行された. Levy による確率過程の見方と, Kolmogorov による Markov 過程への接近方法とを統一することにより, 確率微分方程式とそれに関連する確率解析の理論を創出した. 「Levy 過程を Markov 過程の接線として捉える」
- [27] Ito53 Ito, K. (1953). Stationary Random Distributions. *Memoirs of the College of Science, Kyoto Imperial University, Series. A*. 28: 209-223.
- [28] BeurlingDeny Beurling, A., and Deny, J. (1959). Dirichlet spaces. *Proceedings of the National Academy of Sciences*

- of the United States of America*. 45 (2): 208 – 215. Dirichlet 形式が初めて定義された.
- [29] KunitaWatanabe Kunita, H., and Watanabe, S. (1967). On Square Integrable Martingales. *Nagoya Mathematics Journal*. 30: 209-245.
- [30] Meyer Meyer, P. A. (1967). Intégrales Stochastiques. *Séminaire de Probabilités I*. Lecture Notes in Math., 39: 72-162. 劣 martingale の Doob-Meyer 分解を用いて, 確率積分が一般の半マルチンゲールについて定義された. こうして確率解析の復権が起こった.
- [31] Ito70 Ito, K. (1970). Poisson Point Processes Attached to Markov Processes. *Berkeley Symposium on Mathematical Statistics and Probability*. 3: 225-239.
- [32] Ito-Selected Kiyosi Ito Selected Papers, edited by Stroock, D. W., and Varadhan S. R. S. (1986). Springer-Verlag.
- [33] Strook03 Stroock, D. (2003). *Markov Processes from K. Ito's Perspective*. Princeton University Press.
- [34] Adams Adams, W. J. (1974). *The Life and Times of the Central Limit Theorem*.
- [35] Kolmogorov Gnedenko, B. V., and Kolmogorov A. N. (1954). *Limit Distribution for Sums of Independent Random Variables*
- [36] Encyclopaedia3 Hazewinkel, M. (1995). *Encyclopaedia of Mathematics Volume 3*. Springer U.S.
- [37] Feller William Feller (1950). *An Introduction to Probability Theory and its Applications, Volume I*.
- [38] Andersen Erik Sparre Andersen. (1953). On the fluctuations of sums of random variables. *Mathematica Scandinavica*. 1:263-285, 2:195-223.
- [39] Frechet Maurice Fréchet. (1940). Les probabilités associées à un système d'événements compatibles et dépendants. *Actualités scientifiques et industrielles*.