確率の基礎

滝根 哲哉*

目 次

1　確率空間

1.1　標本空間と事象

確率モデル（あるいは確率的実験）は以下の３つの要素から構成されていると考えることが出来る．

- **標本空間**（sample space）: 起こり得る結果，すなわち**標本**（sample）の集合であり，実験が行われるとこの中の一つだけが実際に起こる
- **事象**（event）の集合 : それぞれの事象は標本空間の部分集合
- それぞれの事象へ確率を割り当てる規則 : 確率は０以上１以下の実数

注意 1.1 標本と事象の違いには特に注意を払う必要がある．標本とは実験の結果そのものであり，一つの実験に対して単一の標本が対応する．一方，事象は，一般には，実験の結果を部分的に特徴付けるものであり，実験の結果が（その事象に対応する）標本空間の部分集合に含まれていることを示す．□

実験の最も単純な例は標本の数が有限か，あるいは可算無限個（個々の標本に対して $\omega_1, \omega_2,$ というふうに番号をつけることができる）の場合である．この場合は，それぞれの標本に対応する事象を考えることで，標本そ

*大阪大学大学院工学研究科電気電子情報工学専攻（〒 565-0871 吹田市山田丘 2-1）
電話 : (06)6879-7740　FAX : (06)6875-5901　電子メール : takine@comm.eng.osaka-u.ac.jp
URL : http://www2b.comm.eng.osaka-u.ac.jp/~takine/

のものに対して確率を割り当てることができる．また，このとき，任意の事象の確率は，その事象を構成する標本の確率を全て足し合わせたもので与えられる．さらに全ての標本に割り当てられた確率の総和は 1 である．

例 1.1 コインを 2 回投げる実験を考える．ここで標本空間は四つの標本 { (表，表), (表，裏), (裏，表), (裏，裏) } から成るとし，それぞれの標本点に 1/4 の確率を割り当てる．標本空間の部分集合はそれぞれ事象である．例えば「最初に表が出る」という事象は { (表，表), (表，裏) } であり，確率 1/4+1/4=1/2 をもつ． □

注意 1.2 標本空間を上記のように選ぶということは，コインが立つ，あるいはコインが 2 回目に投げられる前になくなってしまうといった現象を無視したことを意味する．また，確率を上記のように選ぶということはコインに歪みがあるという可能性を排除したことになる．このように，実際の物理現象と確率モデルの違いに注意しなければならない． □

例 1.1 をみると事象という概念を持ち出す必要はなく，個々の標本に確率を割り当てれば十分と思えるかも知れない．しかし，次の例では，個々の標本に確率を割り当てるということは意味がなく，なぜ，事象という概念を導入しなければならないかが明らかになる．

例 1.2 $[0, 1]$ の実数からなる標本空間を考え，実験から得られる結果は区間 $[0, 1]$ 内で一様に分布しているとする[1]．もし，ある標本に正の確率 $p > 0$ が割り当てられているならば，一様に分布しているという仮定より，他の標本も同じ確率 p を持つことになるが，一方，標本の数は無限にあるので，確率の総和は $p \times \infty = \infty$ となってしまい矛盾する．よって各標本に割り当てられる確率は 0 でなければならない．しかし，確率 0 をもつ標本を幾つ足し合わせても確率の総和は 0 にしかならず，個々の事象に対する確率（例えば標本が区間 $[0, 1/2]$ 内に入る確率）を得る合理的な方法はない． □

例 1.2 の場合，任意の区間に対して，区間長に等しい確率を割り当て，互いに交わらない区間の和に対応する確率は，それぞれの区間の確率の和で表すことが合理的であると思われる．例えば標本が区間 $[0, 1/2]$ 内に入る確率は 1/2 であり，区間 $[0, 1/2]$ あるいは区間 $[3/4, 1]$ のいずれかに含まれる確率は 3/4 である．

この例から分かるように，確率は，一般には，標本ではなく事象に対して割り当てなければならない．さらに，割り当てられた確率は非負の値であり，標本空間全体に対しては 1 であり，かつ，任意の互いに交わらない事象の列 $E_1, E_2, \ldots$ に対してこれらの事象の和集合の確率はそれぞれの事象の確率の和で与えられるようなものでなくては我々が通常の生活で使っている「確率」という概念と一致しない．

標本空間が可算の場合，上の議論は事象に対する確率を事象に含まれる個々の標本に対する確率の和として定義できることを示している．しかし，より一般的な例 1.2 のような場合においては，全ての事象に対して確率が決定できるようにするためには，どのような事象に確率を割り当てておけば十分かという問題に直面することになる．これを解決するため，事象は以下のようにして定義される．

定義 1.1 (事象と事象の集合) 標本空間 Ω の部分集合の集まり $\mathcal{F}$ が，条件

1. $\Omega \in \mathcal{F}$（標本空間 Ω は事象である）
2. $A \in \mathcal{F} \Rightarrow A^c = \Omega - A \in \mathcal{F}$ （A が事象ならば A の余事象 A^c も事象である）
3. $A_i \in \mathcal{F}\ (i = 1, 2, \ldots) \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$ （事象の無限列に対してそれらの和事象も事象となる[2]）

を満たすとき，$\mathcal{F}$ を Ω に対する事象の集合といい，$\mathcal{F}$ の要素を事象と呼ぶ． □

直観的には，標本空間 Ω を分割して得られる集合が事象であり，それらを集めたものが事象の集合 $\mathcal{F}$ である．定義より，以下のことが成り立つ．

補題 1.1 (事象の性質)

1. $\phi \in \mathcal{F}$（空集合 ϕ も事象である）

[1] 一様に分布しているとは取りうるどの値も同様に確からしいということである．

[2] $\cup_{i=1}^{\infty} A_i = \{\omega \mid$ ある自然数 i が存在して $\omega \in A_i\}$.

2. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$, $A - B \in \mathcal{F}$（事象 A, B の和事象 $A \cup B$，積事象 $A \cap B$，差事象 $A - B$ も事象である）
3. $A_i \in \mathcal{F}$ $(i = 1, 2, \ldots) \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$（事象の無限列に対してそれらの積事象も事象となる[3]）

□

1. $\phi = \Omega^c \in \mathcal{F}$. 2. $A_1 = A, A_2 = B, A_i = \phi$ $(i \geq 3)$ で 3. を適用．$A \cap B = (A^c \cup B^c)^c$. $A - B = A \cap B^c$. 3. $\cap_{i=1}^{\infty} A_i = (\cup_{i=1}^{\infty} A_i^c)^c$.

事象 A_i $(i = 1, \ldots, n)$ が全ての i, j $(i \neq j)$ に対して $A_i \cap A_j = \phi$ であるとき，事象 A_i $(i = 1, \ldots, n)$ は**排反**（exclusive, disjoint）と言われる．

標本空間 Ω に対するの事象の集合は一意ではない．例えば，$\mathcal{F}_1 = \{\phi, \Omega\}$ や，ある $A \in \Omega$ に対して $\mathcal{F}_2 = \{\phi, A, A^c, \Omega\}$，あるいは $\mathcal{F}_3 = \{\Omega$ の全ての部分集合$\}$ は全て標本空間 Ω に対する事象の集合である．以下では

$$2^{\Omega} = \{\Omega \text{ の全ての部分集合}\}$$

という記号を用いる．全ての部分集合には空集合 ϕ と全体集合 Ω も含まれていることに注意する．

標本空間の中で応用上特に重要なものに，有限あるいは可算無限個の標本からなる Ω と，実数全体（$= \mathcal{R}$）からなる Ω がある．有限あるいは可算無限個の標本からなる Ω に対しては，通常，$\mathcal{F} = 2^{\Omega}$ とする．一方，実数全体 $\mathcal{R}$ からなる Ω に対しては

$$\mathcal{F} = \mathcal{B}(\mathcal{R}) = \text{全ての半開区間 } (a, b] \text{ を要素にもち，かつ，定義 1.1 の性質をみたす集合}$$

とする．このように事象の集合を選ぶことで，確率を矛盾なく定義することができることが知られている．この事象の集合 $\mathcal{B}(\mathcal{R})$ には 1 点集合 $\{a\}$, 開区間 (a, b), 任意の閉区間 $[a, b]$ や (a, ∞), $(-\infty, b)$, $[a, \infty)$, $(-\infty, b]$ が全て要素として含まれる．

$$\{a\} = \cap_{n=1}^{\infty}(a - 1/n, a] \in \mathcal{B}(\mathcal{R}),\ (a, b) = \cup_{n=1}^{\infty}(a, b - 1/n] \in \mathcal{B}(\mathcal{R}),\ [a, b] = \cap_{n=1}^{\infty}(a - 1/n, b] \in \mathcal{B}(\mathcal{R}) \text{ etc}$$

1.2 確率分布

次に標本空間と事象の集合に組 $(\Omega, \mathcal{F})$ が与えられたとき，各事象に対して確率分布を定義する．以前に述べたように確率は以下の三つの公理を満足する必要がある．

定義 1.2 (確率の公理) 以下の性質をもつ，$\mathcal{F}$ 上で定義された実数値関数 P は**確率分布**（probability distribution）と呼ばれる[4]．

1. 全ての $A \in \mathcal{F}$ に対して $\mathrm{P}(A) \geq 0$ （確率は非負の値を取る）
2. $\mathrm{P}(\Omega) = 1$ （標本空間全体に対する確率は１）
3. $A_i \in \mathcal{F}$ $(i = 1, 2, \ldots)$ かつ $A_i \cap A_j = \phi$ $(i \neq j)$ ならば

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)$$

が成立する. (排反な事象の無限列の和事象の確率はそれぞれの事象の確率の和に等しい)

□

$\mathrm{P}(A)$ は事象 A の**確率**（probability）と呼ばれる．特に $\mathrm{P}(A) = 1$ のとき，事象 A は**確率１で成り立つ**（with probability 1, w.p.1）という．$\mathrm{P}(A) = 1$ は $A = \Omega$ を意味しないことに注意する．

[3] $\cap_{i=1}^{\infty} A_i = \{\omega \mid \text{全ての自然数 } i \text{ に対して} \omega \in A_i\}$.
[4] P は Pr と書かれることもある．

例 1.3 例 1.2 において A を標本が開区間 (0,1) 内にある事象とすれば $\mathrm{P}(A)=1$ であるし，B を標本が閉区間 [0,1] から高々可算個の点 $x_1, x_2, \ldots$ を取り除いた残りの区間に含まれる確率とすれば $\mathrm{P}(B)=1$ である． □

言い換えると，ある事象 C に対して $\mathrm{P}(C)=0$ であるからといって，その事象に含まれる標本 $\omega \in C$ が決して起こり得ないかというと，必ずしもそうではない．たとえば，例 1.2 において，実験を一回行えば，必ずある標本 $\omega \in [0,1]$ が得られるが，$\mathrm{P}(\{\omega\})=0$ である．このように標本空間が連続な場合，我々が生活で用いている「確率」という概念はそのまま適用することが出来ない．

標本空間 Ω，事象の集合 $\mathcal{F}$，確率分布（事象に確率を割り当てる規則）P の三つ組 $(\Omega, \mathcal{F}, P)$ を**確率空間**（probability space）という．確率空間は，標本空間が有限または可算な場合（離散型確率空間）とそうでない場合に大別することができる．離散型確率空間の場合は，$\Omega=\{\omega_1, \omega_2, \ldots\}$，$\mathcal{F}=2^{\Omega}$ に対して，p_i $(i=1,2,\ldots)$ を総和が 1 である非負の実数としたとき，

$$\mathrm{P}(A)=\sum_{\omega_i \in A} p_i$$

とすれば，$(\Omega, \mathcal{F}, P)$ は確率空間となる．このような p_i は**確率関数**（probability function, または probability mass function）と呼ばれる．

一方，標本空間 Ω が実数全体の場合（$\Omega=\mathcal{R}$），明らかに，各標本と確率を 1 対 1 に対応させる方法では確率空間を作り出すことは出来ない．このような場合，確率空間は実数値関数を用いて構築される．実数値関数 $F(x)$ $(x \in \mathcal{R})$ が以下の性質をもつと仮定する．

定義 1.3 (分布関数の性質)

1. $\lim_{x \to -\infty} F(x)=0$ [5]
2. $F(x)$ は非減少（$x<y \Rightarrow F(x) \leq F(y)$）かつ右連続（$\lim_{\epsilon \to 0+} F(x+\epsilon)=F(x)$ [6]）
3. $\lim_{x \to \infty} F(x)=1$ [7]

このような性質をもつ関数 $F(x)$ は**分布関数**（distribution function）あるいは単に**分布**（distribution）と呼ばれる． □

ここで，$\Omega=\mathcal{R}$，$\mathcal{F}=\mathcal{B}(\mathcal{R})$ とする．ある分布関数 $F(x)$ が与えられたとき，

$$\mathrm{P}((-\infty, x])=F(x) \tag{1}$$

を満たすような確率分布 P が一意に定まることが知られている．よって，標本空間が実数全体の場合，確率空間を定めることと分布関数を定めることは等価であり，式 (1) を満たすように確率分布 P を定めると，$(\Omega, \mathcal{F}, P)=(\mathcal{R}, \mathcal{B}(\mathcal{R}), P)$ は確率空間となる．式 (1) ならびに定義 1.2 で与えた確率の公理 3. より，任意の a, b $(a<b)$ に対して

$$F(b)=\mathrm{P}((-\infty, b])=\mathrm{P}((-\infty, a] \cup (a, b])=\mathrm{P}((-\infty, a])+\mathrm{P}((a, b])=F(a)+\mathrm{P}((a, b])$$

が成立する．すなわち

$$\mathrm{P}((a, b])=F(b)-F(a), \qquad a<b \tag{2}$$

となる．

ある分布関数 $F(x)$ が与えられたとき

$$\int_{-\infty}^{x} f(y) dy=F(x) \tag{3}$$

が全ての実数 x に対して成り立つような関数 $f(x)$ が存在すれば，$f(x)$ は $F(x)$ の**密度関数**（density function）と呼ばれる．定義より，分布関数 $F(x)$ が微分可能ならば密度関数 $f(x)$ が存在し，$\epsilon>0$ に対して

$$f(x)=\frac{d}{dx} F(x)=\lim_{\epsilon \to 0} \frac{F(x+\epsilon)-F(x)}{\epsilon}=\lim_{\epsilon \to 0} \frac{F(x)-F(x-\epsilon)}{\epsilon}$$

[5] 任意の正数 ϵ に対して正数 δ が存在し，$x<-\delta$ ならば $|f(x)-a|<\epsilon$ となるとき $\lim_{x \to -\infty} f(x)=a$ と書く．

[6] $\lim_{\epsilon \to a+}$ は ϵ を正の側から a へ近付けることを意味する．よって，$\lim_{\epsilon \to 0+} f(x+\epsilon)=f(x)$ は，任意の正数 ϵ に対して正数 δ が存在し，$y \in [x, x+\delta)$ ならば $|f(y)-f(x)|<\epsilon$ となるということである．

[7] 任意の正数 ϵ に対して正数 δ が存在し，$x>\delta$ ならば $|f(x)-a|<\epsilon$ となるとき $\lim_{x \to \infty} f(x)=a$ と書く．

が成立する．式 (2) より，

$$F(x+\epsilon) - F(x) = \mathrm{P}((x, x+\epsilon])$$

であり，$F(x+\epsilon)$ を x の周りでテーラー展開すると

$$F(x+\epsilon) = F(x) + \frac{d}{dx}F(x)\epsilon + \frac{d^2}{dx^2}F(x)\frac{\epsilon^2}{2} + \frac{d^3}{dx^3}F(x)\frac{\epsilon^3}{6} + \cdots$$

となる．よって $f(x)$ は

$$\mathrm{P}((x, x+\epsilon]) = f(x)\epsilon + o(\epsilon)$$

を満たす[8]．同様に，

$$F(x) - F(x-\epsilon) = \mathrm{P}((x-\epsilon, x])$$

であり，$F(x-\epsilon)$ を x の周りでテーラー展開すると

$$F(x-\epsilon) = F(x) + \frac{d}{dx}F(x)\cdot(-\epsilon) + \frac{d^2}{dx^2}F(x)\frac{(-\epsilon)^2}{2} + \cdots$$

となる．よって $f(x)$ は

$$\mathrm{P}((x-\epsilon, x]) = f(x)\epsilon + o(\epsilon)$$

も満たす.

以上より，十分小さな ϵ に対しては

$$f(x)\epsilon = f(x)\frac{\epsilon}{2} + f(x)\frac{\epsilon}{2} = \mathrm{P}\left((x-\epsilon/2, x]\right) + \mathrm{P}\left((x, x+\epsilon/2]\right) + o(\epsilon)$$

となるため

$$f(x)\epsilon = \mathrm{P}\left((x-\epsilon/2, x+\epsilon/2]\right) + o(\epsilon) \tag{4}$$

と解釈できる．また $F(\infty) = 1$ なので，非負の値を取る関数 $f(x)$ が

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

を満たすならば，式 (3) で定義された関数 $F(x)$ は $f(x)$ を密度関数にもつ分布関数となる．

問 1.1　例 1.2 の実験に対応する分布関数を求めよ．また，密度関数はどのようになるか.

標本空間が連続な場合と同様に，標本空間 Ω が $\Omega \subset \mathcal{R}$ であるような離散型確率空間 $(\Omega, 2^\Omega, P)$ に対しても分布関数は定義できる．すなわち，可算個の $x_1, x_2, \ldots \in \Omega$, に対して確率関数 $p_i = \mathrm{P}(x_i)$ が与えられたとき，

$$F(x) = \sum_{x_i \le x} p_i$$

とすれば，$F(x)$ は分布関数となる．

問 1.2　$x_1 < x_2 < x_3$ に対して，$p_1 = p_2 = p_3 = 1/3$ のとき，分布関数 $F(x)$ を求めよ．

上記では標本空間が 1 次元の分布関数を考えたが，これを多次元に拡張することができる．標本空間が n 次元実ベクトル空間 $\mathcal{R}^n$ であるとする．このとき，$\mathcal{R}^n$ 上の全ての長方形領域 $\cap_{i=1}^n (a_i, b_i]$ を要素のもつ集合 $\mathcal{B}(\mathcal{R}^n)$ を $\mathcal{R}^n$ 上の事象の集合と呼ぶ．1 次元の場合と同様に以下のような性質をもつ関数 $F(x_1, x_2, \ldots, x_n)$ を考える．

1. $\lim_{x_1\to-\infty}\lim_{x_2\to-\infty}\cdots\lim_{x_n\to-\infty} F(x_1, x_2, \ldots, x_n) = 0$
2. $F(x_1, x_2, \ldots, x_n)$ は各引数 x_i に対して非減少かつ右連続
3. $\lim_{x_1\to\infty}\lim_{x_2\to\infty}\cdots\lim_{x_n\to\infty} F(x_1, x_2, \ldots, x_n) = 1$

[8] $o(\epsilon)$ は $\lim_{\epsilon\to0} o(\epsilon)/\epsilon = 0$ となる項を表す．この授業では，ϵ で展開したとき，ϵ に関する 2 次以上の項をまとめたものに対応する．

このような性質をもつ関数 $F(x_1, x_2, \ldots, x_n)$ は n 次元分布関数と呼ばれる．n 次元分布関数 $F(x_1, x_2, \ldots, x_n)$ が与えられると，標本空間と事象の集合の組 $(\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$ に対して

$$\mathrm{P}\left(\cap_{i=1}^{n}(-\infty, x_i]\right) = F(x_1, x_2, \ldots, x_n), \qquad -\infty < x_i < \infty$$

によって確率分布 P を決定することができる．n 次元分布関数によって決定される確率分布は n 次元確率分布，または，n 次元分布と呼ばれる．

また，
$$\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n} f(y_1, y_2, \ldots, y_n)dy_1dy_2\cdots dy_n = F(x_1, x_2, \ldots, x_n)$$
を満たす関数 $f(x_1, x_2, \ldots, x_n)$ を $F(x_1, x_2, \ldots, x_n)$ の密度関数といい，

$$\sum_{i_1 \le x_1}\sum_{i_2 \le x_2}\cdots\sum_{i_n \le x_n} p_{i_1, i_2, \ldots, i_n} = F(x_1, x_2, \ldots, x_n)$$

を満たす関数 $p_{i_1, i_2, \ldots, i_n}$ を $F(x_1, \ldots, x_n)$ の確率関数という．

さて，確率の定義より以下が成立する．

補題 1.2 (確率の性質) 事象 $A, B \in \mathcal{F}$ に対して

1. $\mathrm{P}(\phi) = 0$
2. $A \cap B = \phi$ ならば $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$
3. $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$
4. $A \subset B$ ならば $\mathrm{P}(A) \le \mathrm{P}(B)$

□

性質 1. は $A_i = \Omega$, $A_i = \phi$ $(i = 2, 3, \ldots)$ とすれば公理の 3. より $\mathrm{P}(\Omega) = \mathrm{P}(\Omega) + \sum_{i=2}^{\infty}\mathrm{P}(\phi)$, すなわち $\sum_{i=2}^{\infty}\mathrm{P}(\phi) = 0$. 公理の 1. より $\mathrm{P}(\phi) \ge 0$ なので $\mathrm{P}(\phi) = 0$. 性質 2. は $A_1 = A$, $A_2 = B$, $A_i = \phi$ $(i = 3, 4, \ldots)$ として公理の 3. と性質 1. を用いる．性質 3. は $A_1 = A$, $A_2 = A^c$, $A_i = \phi$ $(i = 3, 4, \ldots)$ として公理の 2., 3. と性質 1. を用いる．性質 4. は $B = A \cup (B - A)$ かつ $A \cap (B - A) = \phi$ に留意して性質 2. を適用し，公理 1. を用いる．

一般に，n 個の事象 $A_1, A_2, \ldots, A_n$ が次の二つの条件を満たすとき，事象の組 $\{A_1, A_2, \ldots, A_n\}$ を標本空間 Ω の**分割**（partition）という．

$$A_i \cap A_j = \phi, \quad i \ne j \quad \text{（各事象は互いに排反）}$$
$$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega \quad \text{（各事象の和集合は標本空間となる）}$$

例えば，事象 A が与えられたとき，$\{A, A^c\}$ は標本空間 Ω の分割である．

問 1.3 [[3]] 今，ある標本空間 Ω における二つの事象 A, B がある．以下の問いに答えよ．

(a) 事象 A, B によって定まる四つの事象からなる集合を用いて標本空間 Ω を分割することができる．このような分割を求めよ（ヒント：ベン図を書いてみよ）．
(b) $\mathrm{P}(A) = a$, $\mathrm{P}(B) = b$, $\mathrm{P}(A \cap B) = c$ であるとする．前問 (a) の分割を与える各事象の確率を a, b, c を用いて表せ．
(c) a, b, c がこのような確率として正しく定義できるための条件を示せ（ヒント：確率は非負であり，かつ総和は 1）．

問 1.4 [[3]] 事象 A, B に対して $\mathrm{P}(A) = 3/4$, $\mathrm{P}(B) = 1/3$ であるとする．前問の結果を利用して以下の問いに答えよ．

(a) $\mathrm{P}(A \cup B)$ が取り得る可能性のある最大ならびに最小の値を求めよ．

(b) $\mathrm{P}(A\cup B)$ が最大であるとき，$\mathrm{P}(A\cap B)$ の値を求めよ．また，$\mathrm{P}(A\cup B)$ が最小であるとき，$\mathrm{P}(A\cap B)$ の値を求めよ．

次に条件付き確率を定義する．

定義 1.4 (条件付き確率) 事象 $A, B \in \mathcal{F}$ に対して，$\mathrm{P}(B) > 0$ のとき，

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

を事象 B が起こったという条件の下での事象 A の**条件付き確率**（conditional probability）という． □

定義より，積事象に対する確率 $\mathrm{P}(A\cap B)$ は条件付き確率 $\mathrm{P}(A \mid B)$ を用いて

$$\mathrm{P}(A\cap B) = \mathrm{P}(A \mid B)\mathrm{P}(B) \tag{5}$$

と表すことができる．この関係を繰り返し用いると

$$\mathrm{P}(A\cap B\cap C) = \mathrm{P}(A\cap(B\cap C)) = \mathrm{P}(A \mid B\cap C)\mathrm{P}(B\cap C) = \mathrm{P}(A \mid B\cap C)\mathrm{P}(B \mid C)\mathrm{P}(C) \tag{6}$$

という関係も得られる．さらに，条件付き確率の定義から

$$\mathrm{P}(A\cap B \mid C) = \frac{\mathrm{P}((A\cap B)\cap C)}{\mathrm{P}(C)} = \frac{\mathrm{P}(A\cap B\cap C)}{\mathrm{P}(C)} \tag{7}$$

なので，式 (6), (7) から

$$\mathrm{P}(A\cap B \mid C) = \mathrm{P}(A \mid B\cap C)\mathrm{P}(B \mid C) \tag{8}$$

を得る．

今，事象 B_i $(i = 1, \ldots, n)$ が排反（全ての i, j $(i \neq j)$ に対して $B_i \cap B_j = \phi$）でかつ $\Omega = \cup_{i=1}^n B_i$ であるとする．このとき，任意の事象 A に対して

$$\mathrm{P}(A) = \mathrm{P}(A\cap\Omega) = \mathrm{P}\left(A \bigcap \left(\bigcup_{i=1}^n B_i\right)\right) = \mathrm{P}\left(\bigcup_{i=1}^n (A\cap B_i)\right) = \sum_{i=1}^n \mathrm{P}(A\cap B_i)$$

となるので，式 (5) を用いると

$$\mathrm{P}(A) = \sum_{i=1}^n \mathrm{P}(A \mid B_i)\mathrm{P}(B_i) \tag{9}$$

を得る．ここで n は無限大でも良い．これは**全確率の法則**（law of total probabilities）と呼ばれる．

また，ある事象 B に対して，事象 B_i $(i = 1, \ldots, n)$ が排反でかつ $B = \cup_{i=1}^n B_i$ であるとする．このとき，任意の事象 A に対して

$$\begin{aligned}
\mathrm{P}(A \mid B) &= \mathrm{P}(A\cap B \mid B) = \mathrm{P}\left(A \bigcap \left(\bigcup_{i=1}^n B_i\right) \middle| B\right) = \mathrm{P}\left(\bigcup_{i=1}^n (A\cap B_i) \middle| B\right) \\
&= \frac{\mathrm{P}\left(\left(\bigcup_{i=1}^n (A\cap B_i)\right)\cap B\right)}{\mathrm{P}(B)} = \frac{\mathrm{P}\left(\bigcup_{i=1}^n (A\cap B_i\cap B)\right)}{\mathrm{P}(B)} = \frac{\mathrm{P}\left(\bigcup_{i=1}^n (A\cap B_i)\right)}{\mathrm{P}(B)} \\
&= \frac{\sum_{i=1}^n \mathrm{P}\left(A\cap B_i\right)}{\mathrm{P}(B)} = \sum_{i=1}^n \mathrm{P}(A\cap B_i \mid B)
\end{aligned}$$

となるので，式 (8) を用いると

$$\mathrm{P}(A \mid B) = \sum_{i=1}^n \mathrm{P}(A \mid B_i\cap B)\mathrm{P}(B_i \mid B) = \sum_{i=1}^n \mathrm{P}(A \mid B_i)\mathrm{P}(B_i \mid B) \tag{10}$$

を得る．

一方，$\mathrm{P}(B_j \mid A) = \mathrm{P}(B_j \cap A)/\mathrm{P}(A)$ であるので，この式の分子，分母をそれぞれ式 (5)，式 (9) を用いて書き換えると

$$\mathrm{P}(B_j \mid A) = \frac{\mathrm{P}(A \mid B_j)\mathrm{P}(B_j)}{\displaystyle\sum_{i=1}^{n} \mathrm{P}(A \mid B_i)\mathrm{P}(B_i)}$$

となる．これは**ベイズの公式**（Bayes' formula）と呼ばれる．

問 1.5 [ポリヤの壷]r 個の赤玉と b 個の黒玉の入った壷がある．ここで，この壷から玉を一つ取りだし，取り出した玉と同じ色の玉を c 個加えて壷へ戻すという操作を繰り返す．k 回目に赤玉を取り出す事象を R_k，黒玉を取り出す事象を B_k としたとき，$\mathrm{P}(R_1)$，$\mathrm{P}(R_2)$，$\mathrm{P}(R_1 \mid R_2)$ を求めよ．

定義 1.5 (独立) 事象 $A, B \in \mathcal{F}$ に対して，$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$ が成り立つとき，事象 A と事象 B は**独立**（independent）であるといわれる．また，一般に n 個の事象 A_i $(i = 1, \ldots, n)$ が与えられたとき，

$$\mathrm{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathrm{P}(A_i)$$

が成り立つならば，事象 A_i $(i = 1, \ldots, n)$ は**互いに独立**（mutually independent）であるといわれる． □

注意 1.3 もし，事象 A_i $(i = 1, \ldots, n)$ が互いに独立ならば，任意の m $(m = 2, \ldots, n)$ と任意の $i_j \in \{1, \ldots, n\}$ に対して

$$\mathrm{P}\left(\bigcap_{j=1}^{m} A_{i_j}\right) = \prod_{j=1}^{m} \mathrm{P}(A_{i_j})$$

が成立する．しかし，$m = n$ を除く全ての m $(m = 2, \ldots, n-1)$ と任意の $i_j \in \{1, \ldots, n\}$ に対して上式が成立したとしても，事象 A_i $(i = 1, \ldots, n)$ は互いに独立とは限らない． □

問 1.6 例 1.1 において，A_1 を１回目が表，A_2 を２回目が裏，A_3 を１回目，２回目ともに表あるいは裏という事象とする．A_i と A_j $(i \neq j)$ は互いに独立となるか．また，A_1，A_2，A_3 は互いに独立となるか．

問 1.7 [[3]] 大相撲の巴戦を考える[9]．３人の実力は全く同等であるが，各対戦ではその前の対戦で控えていた力士の方が有利で，勝つ確率は p であったとする．このとき，最初に対戦した力士と，控えていた力士では，どちらが優勝する確率が高いか．ただし，最初の対戦だけは両者互いに五分であるとする．

1.3 確率変数とその分布

最初に確率変数を定義する．

定義 1.6 (確率変数) 確率空間 $(\Omega, \mathcal{F}, P)$ において，Ω から実数の集合 $\mathcal{R}$ への関数 $X(\omega)$ が任意の実数 x に対して

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$$

であるならば，$X = X(\omega)$ を**確率変数**（random variable）という． □

すなわち，確率変数とは，各標本に対して実数が対応しており，さらに，その実数が取り得る範囲を限定したとき，対応関係を用いて事象を特定することが可能であり，かつ，そのような事象に対する確率が定まるものである．

[9]巴戦とは，千秋楽に３人の力士 A, B, C が同じ勝ち星で並んだとき，その中から一人の優勝者を決める方法である．まず，A と B が対戦し，その勝者が C と対戦する．ここで連勝すればその力士が優勝，C が勝てば，前の対戦で負けて控えていた力士と C が対戦し，C が連勝すれば C の優勝，もし負ければ，勝った力士と控えていた力士が再度対戦し，.... このようにして２連勝する力士が現れるまで対戦を続け，最初に２連勝した力士が優勝を勝ち取る．

確率変数の定義において注意すべき点は，各標本に対して実数が定義されていなければならないことである．

例 1.4 例 1.1 において，N を最初に表が出るまでにコインを投げた回数とする．このとき（表，表）なら $N = 1$,（表，裏）なら $N = 1$,（裏，表）なら $N = 2$ であるが,（裏，裏）の場合は値が定義されていない．よって N は確率変数ではない[10]. □

以下では，誤解を生じない限り確率空間を特に指定せず，単に「X は確率変数である」と書くことにする．また $\{\omega \in \Omega \mid X(\omega) \leq x\}$ も単に $\{X \leq x\}$ と書き，さらに $\mathrm{P}(\{X \leq x\})$ は単に $\mathrm{P}(X \leq x)$ と書くことにする．

さて，確率変数 X に対して $F(x) = \mathrm{P}(X \leq x)$ とおけば，$F(x)$ は分布関数となる．このような $F(x)$ を確率変数 X の分布関数または分布という．さらに，確率変数 X は分布関数（あるいは分布）$F(x)$ をもつ，または，確率変数 X は分布関数（あるいは分布）$F(x)$ に従うという．

問 1.8 以下の問いに答えよ．

(a) 互いに独立な $n+1$ 個の連続確率変数 X_i $(i = 0, 1, \ldots, n)$ がある．X_i の分布関数と密度関数をそれぞれ $F_i(x)$, $f_i(x)$ とする．$\mathrm{P}(X_0 = \min(X_0, X_1, \ldots, X_n))$ は分布関数，密度関数を用いてどのように表現できるか示せ．

(b) X_i $(i = 0, 1, \ldots,)$ を分布関数 $F(x)$，密度関数 $f(x)$ をもつ，**独立で同一な分布に従う**（independent and identically distributed, 略して i.i.d.）連続確率変数の列とする．問 (a) の結果を利用して，$\mathrm{P}(X_0 = \min(X_0, X_1, \ldots, X_n))$ と $\mathrm{P}(X_0 = \min(X_0, X_1, \ldots, X_{n-1}))$ の間に成り立つ関係を示せ．さらに，$\mathrm{P}(X_0 = \min(X_0)) = 1$ に注意して，$\mathrm{P}(X_0 = \min(X_0, X_1, \ldots, X_n))$ を求めよ．

1.4 確率ベクトルと結合分布，周辺分布

n 個の確率変数 $X_1, X_2, \ldots, X_n$ に対して，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を**確率ベクトル**（probability vector）という．また，$\boldsymbol{B} \in \mathcal{B}(\mathcal{R}^n)$ に対して，$\boldsymbol{X} \in \boldsymbol{B}$ は事象となり，確率空間 $(\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n), P)$ を作ることが出来る．$\mathrm{P}(\boldsymbol{X} \in \boldsymbol{B})$ は $\boldsymbol{X}$ の分布，あるいは 確率変数 $X_1, X_2, \ldots, X_n$ の**結合分布**，あるいは**同時分布**（joint distribution）と呼ばれる．確率変数 $X_1, X_2, \ldots, X_n$ の結合分布は n 次元分布関数 $F(x_1, x_2, \ldots, x_n)$ を用いて一意に定めることが出来る．

$$F(x_1, x_2, \ldots, x_n) = \mathrm{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots X_n \leq x_n)$$

このとき，$F(x_1, x_2, \ldots, x_n)$ は $X_1, X_2, \ldots, X_n$ の結合分布関数と呼ばれる．

n 次元分布関数 F に対して $\mathcal{R}$ 上の関数 $F_1(x)$ を

$$F_1(x) = \lim_{x_2 \to \infty} \lim_{x_3 \to \infty} \cdots \lim_{x_n \to \infty} F(x, x_2, x_3, \ldots, x_n)$$

とすると，$F_1(x)$ は（１次元の）分布関数である．このように多次元の分布関数 F において幾つかの引数に対して無限大の極限をとることで得られる分布関数で特徴づけられる分布を F の**周辺分布**（marginal distribution）という．

定義 1.7 (確率変数の独立性) 確率変数 $X_1, X_2, \ldots, X_n$ が全ての $B_i \in \mathcal{B}(\mathcal{R})$ $(i = 1, \ldots, n)$ に対して

$$\mathrm{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \mathrm{P}(X_1 \in B_1)\mathrm{P}(X_2 \in B_2) \cdots \mathrm{P}(X_n \in B_n)$$

を満たすならば，確率変数 $X_1, X_2, \ldots, X_n$ は互いに独立であると呼ばれる． □

確率変数 $X_1, X_2, \ldots, X_n$ がそれぞれ周辺分布 $F_i(x)$ $(i = 1, \ldots, n)$ をもっているとき，これらが互いに独立であるということは結合分布が周辺分布の積に等しいことと等価である．

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n)$$

[10] このような場合,（裏，裏）に対して ∞ を割り当て，確率変数 N の取り得る値に無限大を含めて考えることが多い．このように本来の確率変数の定義を拡張した結果，$\mathrm{P}(-\infty < X < \infty) < 1$ となるような確率変数を不完全な（defective）確率変数という．

また，このとき，結合密度関数と周辺密度関数，あるいは結合確率関数と周辺確率関数に対しても同様の結果を得ることが出来る．

$$f(x_1,\dots,x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n), \qquad p(x_1,\dots,x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n)$$

問 1.9 [[3]] 二つのサイコロを振ったとき，大きい方の目を X，小さい方の目を Y とする（同じ目なら $X=Y$）．X と Y の結合確率関数，および，X，Y それぞれの周辺確率関数を求めよ．

1.5 確率変数を引数とする関数と確率変数の畳み込み

密度関数 $f(x)$ をもつ確率変数 X と，実数から実数への関数 $h(x)$ が与えられたとき，$Y=h(X)$ も確率変数となる．このとき，確率変数 Y の分布関数 $G(x)$ は

$$G(x) = \mathrm{P}(h(X)\le x) = \int_{h(u)\le x} f(u)du$$

で与えられる．もし，確率変数 X が確率関数 $p(x)$ をもつならば

$$G(x) = \sum_{h(u)\le x} p(u)$$

である．

問 1.10 確率変数 X の分布関数を $F(x)$ とする．このとき $Y=X^2$ の分布関数 $G(x)$ を $F(x)$ を用いて表せ．

次に確率変数の和が従う分布を考える．今，確率変数 X_1, X_2 がそれぞれ分布関数 $F_1(x)$, $F_2(x)$ ならびに密度関数 $f_1(x)$, $f_2(x)$ をもっており，互いに独立であるとする．このとき，新しい確率変数 $Y=X_1+X_2$ の分布関数 $G(x)=\mathrm{P}(Y\le x)$ は

$$\begin{aligned} G(x) = \mathrm{P}(X_1+X_2\le x) &= \int\int_{u_1+u_2\le x} f_1(u_1)f_2(u_2)du_1du_2 \\ &= \int_{u_2=-\infty}^{\infty}\int_{u_1=-\infty}^{x-u_2} f_1(u_1)f_2(u_2)du_1du_2 = \int_{-\infty}^{\infty} F_1(x-u_2)f_2(u_2)du_2 \end{aligned}$$

となる．左辺の最後の積分は分布関数 $F_1(x)$ と $F_2(x)$ の**畳み込み**（convolution）と呼ばれ，$F_1*F_2(x)$ と書かれることが多い．

$$F_1*F_2(x) = \int_{-\infty}^{\infty} F_1(x-u)f_2(u)du$$

もし，上記において X_1 と X_2 の役割を交換して計算すると

$$\begin{aligned} G(x) = \mathrm{P}(X_1+X_2\le x) &= \int\int_{u_1+u_2\le x} f_2(u_2)f_1(u_1)du_2du_1 \\ &= \int_{u_1=-\infty}^{\infty}\int_{u_2=-\infty}^{x-u_1} f_2(u_2)f_1(u_1)du_2du_1 = \int_{-\infty}^{\infty} F_2(x-u_1)f_1(u_1)du_1 \end{aligned}$$

となる．すなわち，分布関数の畳み込みは計算の順序と無関係である．

$$G(x) = F_1*F_2(x) = F_2*F_1(x)$$

特に独立な非負の確率変数 X_1 と X_2 に対しては

$$G(x) = \int_0^x F_1(x-u_2)f_2(u_2)du_2 = \int_0^x F_2(x-u_1)f_1(u_1)du_1$$

となる．

さらに

$$\frac{d}{dx}G(x) = \frac{d}{dx}\int_{-\infty}^{\infty} F_1(x-u)f_2(u)du = \int_{-\infty}^{\infty} f_1(x-u)f_2(u)du$$

となるので，分布関数 $G(x)$ の密度関数 $g(x)$ は

$$g(x) = \int_{-\infty}^{\infty} f_1(x-u)f_2(u)du$$

で与えられる．左辺の積分は密度関数 $f_1(x)$ と $f_2(x)$ の畳み込みと呼ばれ，$f_1 * f_2(x)$ と書かれる．上式で $v = x-u$ と変数変換すれば

$$g(x) = \int_{-\infty}^{\infty} f_1(v)f_2(x-v)dv = f_2 * f_1(x)$$

となるので，密度関数の畳み込みに関しても計算の順序とは無関係である．特に独立な非負の確率変数 X_1 と X_2 に対しては

$$g(x) = \int_0^x f_1(x-u_2)f_2(u_2)du_2 = \int_0^x f_1(u_1)f_2(x-u_1)du_1$$

となる．

例 1.5 (正規分布の再現性) パラメタ m, σ（ただし $\sigma > 0$）をもつ密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

に従う分布を**正規分布**（normal distribution）といい，$N(m, \sigma^2)$ と書く．畳み込みの例として，独立な正規分布 $N(m_1, \sigma_1^2)$，$N(m_2, \sigma_2^2)$ の密度関数の畳み込み $g(x)$ を考える．定義より

$$\begin{aligned}
g(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{1}{2}\frac{(x-u-m_1)^2}{\sigma_1^2}\right)\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{1}{2}\frac{(u-m_2)^2}{\sigma_2^2}\right)du \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2\sigma_1^2\sigma_2^2}\left[(\sigma_1^2+\sigma_2^2)\left(u-\frac{(x-m_1)\sigma_2^2+m_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2+\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}(x-m_1-m_2)^2\right]\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(u-\frac{(x-m_1)\sigma_2^2+m_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2\right)du \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\int_{-\infty}^{\infty}e^{-y^2/2}dy
\end{aligned}$$

ここで $\int_{-\infty}^{\infty} e^{-y^2/a^2}dy = \sqrt{a\pi}$ を用いると

$$g(x) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\exp\left(-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)$$

を得る．すなわち，独立な正規分布 $N(m_1, \sigma_1^2)$，$N(m_2, \sigma_2^2)$ の密度関数の畳み込み $g(x)$ は正規分布 $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ である．言い換えれば，正規分布 $N(m_1, \sigma_1^2)$，$N(m_2, \sigma_2^2)$ に従う独立な二つの確率変数 X_1, X_2 の和 $Y = X_1 + X_2$ は正規分布 $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ に従う．これは正規分布の**再現性**と呼ばれている． □

問 1.11　独立な確率変数 X, Y がともに $(0, 1]$ 上の一様分布に従うとする．このとき，$Z = X + Y$ の密度関数を求めよ．

次に整数値をとる離散型分布の畳み込みを考える．独立な二つの確率変数 X_1, X_2 がそれぞれ分布 $\{p_1(k)\}$, $\{p_2(k)\}$ をもつとする．このとき確率変数 $Y = X_1 + X_2$ は整数値をとり，その確率関数 $q(k) = \mathrm{P}(Y = k)$ は

$$q(k) = \sum_{i=-\infty}^{\infty} p_1(k-i)p_2(i) = \sum_{i=-\infty}^{\infty} p_1(i)p_2(k-i)$$

で与えられる．特に独立な非負の整数値を取る確率変数 X_1 と X_2 に対しては，確率変数 $Y = X_1 + X_2$ は非負の整数値をとり，

$$q(k) = \sum_{i=0}^{k} p_1(k-i)p_2(i) = \sum_{i=0}^{k} p_1(i)p_2(k-i)$$

となる.

問 1.12 [ポワソン分布の再現性] 正のパラメタ λ をもつ確率関数

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

に従う分布を**ポワソン分布**（Poisson distribution）という．独立なパラメタ λ_1, λ_2 をもつポワソン分布に従う確率変数 X_1, X_2 の和 Y に対する確率関数 $q(k)$ $(k = 0, 1, \ldots)$ はパラメタ $\lambda_1 + \lambda_2$ をもつポワソン分布になることを示せ.

2　期待値

2.1　期待値の定義

最初に非負の確率変数 X の期待値を定義する.

定義 2.1 (非負の確率変数の期待値) 密度関数 $f(x)$ をもつ非負の確率変数 X の**期待値**（expectation）$\mathrm{E}(X)$ は

$$\mathrm{E}(X) = \int_0^\infty xf(x)dx \tag{11}$$

で与えられる．また，確率関数 $p(x)$ をもつ非負の確率変数 X の期待値 $\mathrm{E}(X)$ は

$$\mathrm{E}(X) = \sum_{i=1}^\infty x_i p(x_i) \tag{12}$$

で与えられる．$\mathrm{E}(X)$ は無限大となることもある． □

式 (12) で与えられる確率関数が存在する場合の期待値は，日常でも普通に用いられている．例えば，サイコロを 1 度投げたときに出る目 X の平均は

$$\sum_{k=1}^6 k\mathrm{P}(X = k) = \sum_{k=1}^6 k \cdot \frac{1}{6} = 3.5$$

と求められる．一方，式 (11) で与えられる密度関数をもつ非負の確率変数に対する期待値は積分を用いて与えられているが，これは本質的に確率関数をもつ場合と同じであることを以下で説明する.

まず，定義から

$$\mathrm{E}(X) = \lim_{b\to\infty}\int_0^b xf(x)dx$$

に注意する．ここで右辺の積分は以下のように解釈される．まず，閉区間 $[0, b]$ を n 等分することにより，n 個の小区間 $[x_{i-1}, x_i]$ $(i = 1, 2, \ldots, n)$ を作る.

$$x_0 = 0, \qquad x_i = \frac{ib}{n}\ (i = 1, 2, \ldots, n)$$

このとき，小区間の長さは全て等しく $\Delta x_i = b/n$ であり，それらの最大値 $|\Delta| = |\Delta(n)|$ も b/n で与えられる．各小区間 $[x_{i-1}, x_i]$ の左端の点を $\xi_i = x_{i-1}$ とすると，Riemann 積分[11]の定義から

$$\int_0^b xf(x)dx = \lim_{|\Delta(n)|\to 0}\sum_{i=1}^n \xi_i f(\xi_i)|\Delta(n)| = \lim_{n\to\infty}\sum_{i=1}^n x_{i-1}f(x_{i-1})|\Delta(n)| = \lim_{n\to\infty}\sum_{i=0}^{n-1} x_i f(x_i)|\Delta(n)|$$

を得る．ここで，式 (4) より

$$f(x_i)|\Delta(n)| = \mathrm{P}\left(x_i - \frac{|\Delta(n)|}{2} < X \le x_i + \frac{|\Delta(n)|}{2}\right) + o(|\Delta(n)|)$$

[11]区分求積法により定義される定積分．付録を参照.

と解釈されることに注意する．すなわち，密度関数 $f(x)$ をもつ非負の確率変数 X の期待値 $\mathrm{E}(X)$ は

$$\lim_{n\to\infty} o(|\Delta(n)|) = \lim_{n\to\infty} o(b/n) = 0$$

に注意すると、式 (11) は

$$\mathrm{E}(X) = \int_0^\infty xf(x)dx = \lim_{b\to\infty}\int_0^b xf(x)dx = \lim_{b\to\infty}\left[\lim_{n\to\infty}\sum_{i=0}^{n-1} x_i \mathrm{P}\left(x_i - \frac{b}{2n} < X \le x_i + \frac{b}{2n}\right)\right] \tag{13}$$

と書き換えることができる．

式 (12) と式 (13) の類似性に注意する．式 (13) となる連続な確率変数の場合，標本空間を微少な小区間に分割し，確率変数の取る値が特定の小区間にある確率とその区間の中央値を掛け，全ての場合について和を取ったものの極限になっており，これは式 (12) の確率関数が存在する離散な確率変数の期待値と同形である．

次に任意の確率変数 X の期待値を考える．まず，$X^+ = \max(0, X)$，$X^- = -\min(0, X)$ によって新しい二つの確率変数を導入する．定義より $X^+ \ge 0$，$X^- \ge 0$ であり，$X = X^+ - X^-$ である．

定義 2.2 (確率変数の期待値) もし，$\mathrm{E}(X^+)$ と $\mathrm{E}(X^-)$ の少なくとも一方が有限であるならば，X の期待値は

$$\mathrm{E}(X) = \mathrm{E}(X^+) - \mathrm{E}(X^-)$$

で与えられる．$\mathrm{E}(X^+) = \mathrm{E}(X^-) = \infty$ の場合は期待値は存在しない． □

もし，期待値 $\mathrm{E}(X)$ が存在するならば，定義 2.2 より，密度関数 $f(x)$ をもつ場合は

$$\mathrm{E}(X) = \int_{-\infty}^\infty xf(x)dx$$

となり，離散的ならば

$$\mathrm{E}(X) = \sum_{i=1}^\infty x_i p(x_i)$$

となる．これらの右辺は分布 $F(x)$ の**平均**（mean）と呼ばれる．

問 2.1　$\exp(\lambda) = \sum_{k=0}^\infty \lambda^k/k!$ に注意して，問 1.5 で与えたポワソン分布の平均を求めよ．

期待値は以下の性質をもつ．

定理 2.1 (期待値の性質) 確率変数 X と Y の期待値 $\mathrm{E}(X)$，$\mathrm{E}(Y)$ が共に有限であるとする．このとき

1. $X \le Y$ ならば $\mathrm{E}(X) \le \mathrm{E}(Y)$
2. a を定数としたとき $\mathrm{E}(aX) = a\mathrm{E}(X)$
3. $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$
4. X と Y が互いに独立であるならば $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$

が成立する． □

一般に，確率変数の関数に対しては次の定理が成り立つ．

定理 2.2 密度関数 $f(x)$ をもつ確率変数 X と関数 $u(x)$ に対して期待値 $\mathrm{E}(u(X))$ が存在するならば，それは

$$\mathrm{E}(u(X)) = \int_{-\infty}^\infty u(x)f(x)dx$$

で与えられる．また，確率関数 $p(x_i)$ $(i = 1, 2, \ldots)$ をもつ場合は

$$\mathrm{E}(u(X)) = \sum_{i=1}^\infty u(x_i)p(x_i)$$

で与えられる． □

問 2.2 正確な硬貨を初めて表が出るまで投げ続ける実験を行う．硬貨を投げた回数を N とする．実験の結果が $N = k$ $(k = 1, 2, \ldots)$ であるとき，2^k 円の報酬を受け取ることが出来るとすると，報酬の期待値はいくらになるか．

例 2.1 非負の確率変数 X の期待値を別の方法で求める．

$$\sum_{n=0}^{\infty} \mathrm{P}(X > n) = \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} \mathrm{P}(X = k) = \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} \mathrm{P}(X = k) = \sum_{k=1}^{\infty} k\mathrm{P}(X = k)$$

となるため，

$$\mathrm{E}(X) = \sum_{n=0}^{\infty} \mathrm{P}(X > n)$$

で与えられる． □

なお，例 2.1 と同様の結果が，非負の確率変数 X が連続確率変数の場合にも以下のように成立する．

$$\mathrm{E}(X) = \int_0^{\infty} (1 - F(x))dx$$

問 2.3 [不幸のパラドックス（ill-luck paradox）[1]] 大勢の人がおみくじを引く場面を想定する．おみくじには，連続な分布からの標本値が一つ書かれており，大きい数字ほどより吉兆であるとする．この設定において，最初におみくじを引いた人よりも，悪いおみくじを引く人が出てくるまで，平均，何人の人がさらにおみくじを引くことになるかを考える．$\{X_n;\ n = 0, 1, \ldots\}$ を独立同一な分布に従う連続確率変数の列であるとする．このとき，$N = \min(n;\ X_n < X_0)$ と定義すると，この問いの答えは $\mathrm{E}(N)$ である．確率 $\mathrm{P}(N > n)$ が $n + 1$ 個の独立同一な分布に従う確率変数 X_i $(i = 0, 1, \ldots, n)$ のうち，X_0 が最小の値を取る確率（問 1.3 参照）と等価であることに注意して $\mathrm{E}(N)$ を求めよ．

最後に分布と期待値の間に成立する不等式を導く．密度関数 $f(x)$ をもつ非負の確率変数 X に対して

$$\mathrm{P}(X \geq x) = \int_{x-}^{\infty} f(y)dy \leq \int_{x-}^{\infty} \frac{y}{x} f(y)dy \leq \frac{1}{x} \int_0^{\infty} y f(y)dy = \frac{E[X]}{x}$$

となるので，次の定理を得る．

定理 2.3 (マルコフ不等式) 非負の確率変数 X に対して

$$\mathrm{P}(X \geq x) \leq \frac{\mathrm{E}(X)}{x}, \qquad x > 0$$

が成立する．これは**マルコフ不等式**（Markov's Inequality）と呼ばれている． □

もし，$x \leq \mathrm{E}(X)$ ならば $\mathrm{E}(X)/x \geq 1$ となるので，マルコフ不等式は $x > \mathrm{E}(X)$ の場合にのみ意味がある．

2.2 積率，分散，共分散

X を分布関数 $F(x)$ をもつ確率変数とする．このとき，$\mathcal{R}$ から $\mathcal{R}$ への可測関数 $g(x)$ に対する期待値 $\mathrm{E}(g(X))$ は，分布関数 $F(x)$ の関数と見なすことが出来る[12]．それゆえ，期待値 $\mathrm{E}(g(X))$ を調べることで分布関数 $F(x)$ の特徴をある程度把握することが出来る．このような目的で良く用いられるものには $g(x) = x^n$ や $g(x) = (x - \mathrm{E}(X))^n$ などがある．

定義 2.3 分布関数 $F(x)$ をもつ確率変数 X に対して，$\mathrm{E}(X^n)$ を確率変数 X（あるいは分布 $F(x)$）の ***n* 次積率**（nth moment）という．$n = 1$ のとき，これは確率変数 X の期待値，あるいは分布 $F(x)$ の平均に等しい．また，$\mathrm{E}(\{X - \mathrm{E}(X)\}^n)$ を確率変数 X（あるいは分布 $F(x)$）の ***n* 次中心積率**（nth central moment）という．特に，$n = 2$ の場合，$\mathrm{E}(\{X - \mathrm{E}(X)\}^2)$ を**分散**（variance）と呼び，$\mathrm{V}(X)$ と書く． □

[12] このような，関数の関数を汎関数（functional）という．

分散 $\mathrm{V}(X) = \mathrm{E}(\{X - \mathrm{E}(X)\}^2)$ は確率変数 X の値が期待値からどの程度離れた値を取るかという，バラツキを表す指標として用いられる．分散には次のような性質がある．

定理 2.4 $\mathrm{E}(X^2) < \infty$ ならば

1. $\mathrm{V}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$
2. 定数 a, b に対して $\mathrm{V}(aX + b) = a^2\mathrm{V}(X)$
3. $\mathrm{V}(X) = 0$ ならば $\mathrm{P}(X = \mathrm{E}(X)) = 1$

が成立する． □

さて，分散 $\mathrm{V}(X)$ をもつ確率変数 X に対して $Y = (X - \mathrm{E}(X))^2$ により，新しい確率変数 Y を定義する．Y に定理 2.3 のマルコフ不等式を適用すると

$$\mathrm{P}(Y \geq y) \leq \frac{\mathrm{E}(Y)}{y}$$

を得る．ここで $y > 0$ に対して $y = x^2$ $(x > 0)$ とおくと

$$\mathrm{P}(Y \geq x^2) = \mathrm{P}((X - \mathrm{E}(X))^2 \geq x^2) = \mathrm{P}(|X - \mathrm{E}(X)| \geq x)$$

となり，$\mathrm{E}(Y) = \mathrm{V}(X)$ に注意すると

$$\mathrm{P}(|X - \mathrm{E}(X)| \geq x) \leq \frac{\mathrm{V}(X)}{x^2} \tag{14}$$

を得る．これは**チェビシェフの不等式**（Chebyshev's inequality）と呼ばれている．マルコフ不等式が分布の上側の裾野の上界値のみを与え，また，非負の確率変数のみに適用できるのに対し，チェビシェフの不等式は分布の両側の裾野の上界値を与えていることに注意する．また，マルコフ不等式の上界値は 0 からの距離の逆数に比例する形で 0 へ向かうのに対し，チェビシェフの不等式の上界値は平均からの距離の二乗の逆数に比例する形で 0 へ向かう．

問 2.4 [**標準化**（standardization）] 確率変数 X に対して $0 < \mathrm{V}(X) < \infty$ のとき

$$Y = \frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{V}(X)}} \tag{15}$$

によって新しい確率変数 Y を定義する．Y の期待値と分散を求めよ．

式 (15) の変換は標準化と呼ばれており，分母に現れる $\sqrt{\mathrm{V}(X)}$ は**標準偏差**（standard deviation）と呼ばれる．また，非負の確率変数 X に対しては，バラツキを表す無次元の量として**変動係数**（coefficient of variation）$\mathrm{C}(X)$ がある．

$$\mathrm{C}(X) = \frac{\sqrt{\mathrm{V}(X)}}{\mathrm{E}(X)}$$

問 2.5 ある正数 $\mu > 0$ に対して，分布関数 $F(x)$ $(x \geq 0)$ が

$$F(x) = 1 - e^{-\mu x}$$

で与えられる分布をパラメタ μ をもつ**指数分布**という．パラメタ μ をもつ指数分布の平均，分散，変動係数を求めよ．

次に二つの確率変数の間の依存性を表す指標を導入する．

定義 2.4 (共分散) 平均をもつ二つの確率変数 X と Y に対して定義された

$$\mathrm{Cov}(X, Y) = \mathrm{E}(\{X - \mathrm{E}(X)\}\{Y - \mathrm{E}(Y)\}) \tag{16}$$

を確率変数 X と Y の**共分散**（covariance）という． □

問 2.6　共分散 $\mathrm{Cov}(X,Y)$ は期待値の性質を用いると

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

と書くことができることを示せ．

もし $X = Y$ ならば共分散は分散に等しい．また，X と Y が独立ならば $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ となるので，共分散は 0 である．確率変数 X と Y の共分散が 0 であるとき，X と Y は**無相関**（uncorrelated）であるという．しかし，X と Y が無相関であっても独立であるとは限らない．

確率変数 X，Y の結合確率が

$$\mathrm{P}(X=1, Y=1) = \mathrm{P}(X=2, Y=0) = \mathrm{P}(X=3, Y=1) = 1/3$$

で与えられているとする．このとき以下の問いに答えよ．

(a) X, Y の周辺確率関数を求め，X と Y が独立であるか否か，論じよ．
(b) $\mathrm{E}(X), \mathrm{E}(Y), \mathrm{E}(XY)$ を求め，X と Y は無相関であるか否か，論じよ．

一般に n 個の確率変数 $X_1, X_2, \ldots, X_n$ の和の分散は共分散を用いて

$$\mathrm{V}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{V}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

で与えられる．よって，n 個の確率変数 $X_1, X_2, \ldots, X_n$ が互いに独立ならば，それらの和の分散はそれぞれの分散の和で与えられる．

補題 2.1 (コーシー - シュワルツの不等式) $\mathrm{E}(X^2)$，$\mathrm{E}(Y^2)$ が有限で正の値を取るならば

$$\{\mathrm{E}(XY)\}^2 \le \mathrm{E}(X^2)\mathrm{E}(Y^2)$$

が成立する．等号成立はある定数 a に対して $Y = aX$ となるときのみである．これは**コーシー - シュワルツの不等式**（Cauchy-Schwarz inequality）と呼ばれている．□

任意の実数 t に対して $\mathrm{E}((X-tY)^2) \ge 0$ であるので，$\mathrm{E}((X-tY)^2) = \mathrm{E}(X^2) - 2\mathrm{E}(XY)t + \mathrm{E}(Y^2)t^2 \ge 0$ が成立する．これを t の 2 次関数とみれば判別式は非正でなければならない．よって $\{\mathrm{E}(XY)\}^2 - \mathrm{E}(X^2)\mathrm{E}(Y^2) \le 0$ を得る．また，等号が成り立つのは $X - tY = 0$ の場合に限る．

共分散の $\mathrm{Cov}(X,Y)$ の取る値は，例えば X を a 倍すると a 倍となる．このように共分散は単位の取り方に依存する．そこでこのような単位の取り方に依存しないように正規化した指標が相関係数である．

定義 2.5 (相関係数) 確率変数 X, Y が有限でかつ 0 でない分散を持つとする．このとき

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}} \tag{17}$$

で定義される $\rho_{X,Y}$ を確率変数 X と Y の**相関係数**（correlation coefficient）という．□

補題 2.1 より $-1 \le \rho_{X,Y} \le 1$ である[13]．特にある正数 a に対して $Y = aX$ ならば $\rho_{X,Y} = 1$ となり，$Y = -aX$ ならば $\rho_{X,Y} = -1$ となる．

3　確率変数と分布の収束

この章では確率変数ならびに分布の様々な収束の概念について解説する．これらの収束概念に対して正確な感覚を身につけることは，確率・統計における最も重要な定理である大数の法則や中心極限定理の意味するところを正しく理解するために必須である．

[13] 補題 2.1 において，X の代わりに $X - \mathrm{E}(X)$，Y の代わりに $Y - \mathrm{E}(Y)$ とすると $\mathrm{Cov}(X,Y)^2 \le \mathrm{V}(X)\mathrm{V}(Y)$ を得る．

3.1　確率変数列の概収束と確率収束

1.3 節で述べたように，確率変数は標本空間 Ω から実数 $\mathcal{R}$ への関数である．よって確率変数の収束は，数列の収束とは異なり，様々な定義がある．以下では $X_1(\omega), X_2(\omega), \ldots$ を確率空間 $(\Omega, \mathcal{F}, P)$ で定義された確率変数として，代表的な確率変数の収束の定義を与える．

定義 3.1 (概収束) 事象 Ω_1 が $\mathrm{P}(\Omega_1) = 1$ であるとする．このとき，任意の $\omega \in \Omega_1$ に対して

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \tag{18}$$

となるような確率変数 $X(\omega)$ が存在するとき，確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は確率変数 X へ**概収束**（almost sure convergence) する，あるいは**確率 1 で収束**（convergence with probability 1）するといい，$X_n \xrightarrow{a.s.} X$ または $\lim_{n\to\infty} X_n = X$ w.p.1 と書く． □

概収束するということは，全ての $\omega \in \Omega_1$ に対して，各標本 ω 毎に収束するということであり，上で与えた定義は $\mathrm{P}(\lim_{n\to\infty} X_n = X) = 1$ と等価である．すなわち，式 (18) を満たす ω 全体が形成する事象 $\{\omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ の確率は 1 である（このような事象を Ω_1 とした）．言い換えれば，もし $\lim_{n\to\infty} X_n(\omega) \neq X(\omega)$ であるような ω が存在するならば，そのような ω は $\omega \in \Omega_0 = \Omega - \Omega_1$（すなわち ω は $\mathrm{P}(\Omega_0) = 0$ であるような事象 Ω_0 の要素）である．

系 3.1 $X_n \xrightarrow{a.s.} X$ であるということは，全ての $\epsilon > 0$ に対して

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{m\geq n} |X_m - X| \geq \epsilon\right) = 0$$

となることと等価である． □

$A_n^{(\epsilon)} = \{\omega \mid |X_n(\omega) - X(\omega)| \geq \epsilon\}$, $A^{(\epsilon)} = \limsup_{n\to\infty} A_n^{(\epsilon)} = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k^{(\epsilon)}$ とする．このとき，$A = \{\omega \mid \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}$ とすると $A = \cup_{\epsilon>0} A^{(\epsilon)} = \cup_{\epsilon>0}[\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k^{(\epsilon)}] = \cup_{m=1}^{\infty}[\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k^{(1/m)}] = \cup_{m=1}^{\infty} A^{(1/m)}$ である．よって $X_n \xrightarrow{a.s.} X \Leftrightarrow \mathrm{P}(A) = \mathrm{P}(\cup_{m=1}^{\infty} A^{(1/m)}) = 0 \Leftrightarrow \forall m \geq 1, \mathrm{P}(A^{(1/m)}) = 0 \Leftrightarrow \forall \epsilon > 0, \mathrm{P}(A^{(\epsilon)}) = 0$.

定義 3.2 (確率収束) 任意の正数 ϵ に対して

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| > \epsilon) = 0$$

が成立するとき，確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は確率変数 X へ**確率収束**（convergence in probability) するといい，$X_n \xrightarrow{P} X$ と書く． □

定義より $X_n \xrightarrow{P} X$ は $\lim_{n\to\infty} \mathrm{P}(|X_n - X| \leq \epsilon) = 1$ と等価である．確率収束は，概収束のような関数の収束を意味しているのではなく，実数列 $\mathrm{P}(|X_n - X| > \epsilon)$ $(n = 1, 2, \ldots)$ の 0 への収束を意味している．

例 3.1 確率空間 $(R, \mathcal{B}(\mathcal{R}), P)$ において，P は区間 $[0, 1]$ 上の一様分布とする．この確率空間上で確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$，$\{Y_n;\ n = 1, 2, \ldots\}$ を以下のように定義する．

$$X_n(\omega) = \begin{cases} 1, & \omega \in [0, \frac{1}{n}] \\ 0, & \text{その他} \end{cases}$$

$$Y_n(\omega) = \begin{cases} 1, & \omega \in \left[\dfrac{n \bmod 2^i}{2^i}, \dfrac{n \bmod 2^i + 1}{2^i}\right] \\ 0, & \text{その他} \end{cases} \qquad (\text{ただし } 2^i \leq n < 2^{i+1})$$

ここで $n \bmod m$ は n を m で割ったときの余りを表す．確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は任意の選ばれたある $\omega \in [0, 1]$ に対して

$$\lim_{n\to\infty} X_n(\omega) = \begin{cases} 1, & \omega = 0, \\ 0, & \text{その他}, \end{cases} \qquad \mathrm{P}(|X_n| \geq \epsilon) = \frac{1}{n} \ \ (0 < \epsilon < 1)$$

である．$\mathrm{P}(\{0\})=0$ であるので確率変数列 $\{X_n;\ n=1,2,\ldots\}$ は

$$X(\omega)=0, \qquad \omega\in\mathcal{R}$$

へ概収束する．また $\lim_{n\to\infty}\mathrm{P}(|X_n|\geq\epsilon)=0$ なので $\lim_{n\to\infty}\mathrm{P}(|X_n-X|>\epsilon)=0$ となり，確率変数列 $\{X_n;\ n=1,2,\ldots\}$ は X へ確率収束する．

一方，任意の選ばれたある $\omega\in[0,1]$ に対して確率変数列 $\{Y_n;\ n=1,2,\ldots\}$ は

$$\limsup_{n\to\infty}Y_n(\omega)=1, \qquad \liminf_{n\to\infty}Y_n(\omega)=0$$

となるので極限は存在せず，概収束しない．しかし，任意の $0<\epsilon<1$ に対して，$2^i\leq n<2^{i+1}$ のとき

$$\mathrm{P}(|Y_n-X|>\epsilon)=\mathrm{P}(Y_n\neq 0)=\frac{1}{2^i}$$

となるので，$n\to\infty$ のとき $2^i\to\infty$ となることに注意すれば，$\lim_{n\to\infty}\mathrm{P}(|Y_n-X|>\epsilon)=0$ となり，確率変数列 $\{Y_n;\ n=1,2,\ldots\}$ は X へ確率収束する． □

この例から分かるように概収束は確率収束よりも強い収束である．実際，次の定理が成立する．

定理 3.1 (概収束 ⇒ 確率収束) $X_n\xrightarrow{a.s.}X$ ならば $X_n\xrightarrow{P}X$ である． □

概収束，確率収束以外の重要な確率変数収束概念として次の法則収束がある．

定義 3.3 (法則収束) 確率変数列 $\{X_n;\ n=1,2,\ldots\}$ と確率変数 X があるとする．X_n, X の分布関数をそれぞれ $F_n(x), F(x)$ としたとき，$F(x)$ が連続である全ての x で

$$\lim_{n\to\infty}F_n(x)=F(x)$$

が成立するならば，確率変数列 $\{X_n;\ n=1,2,\ldots\}$ は確率変数 X へ**法則収束**（convergence in law, convergence in distribution）するといい，$X_n\xrightarrow{D}X$ と書く．また $X_n\xrightarrow{D}X$ であるとき，X の分布 $F(x)$ を X_n の**漸近分布**（asymptotic distribution）または**極限分布**（limiting distribution）という． □

例 3.2 (例 3.1 の続き) 例 3.1 で与えた非負確率変数 $X_n(\omega)$ の分布関数 $F_n(x)$ は

$$F_n(x)=\begin{cases}0, & x<0\\ \dfrac{n-1}{n}, & 0\leq x<1\\ 1, & x\geq 1\end{cases}$$

であるので，極限分布

$$F(x)=\begin{cases}0, & x<0\\ 1, & x\geq 0\end{cases}$$

をもつ．すなわち $X_n\xrightarrow{D}X$ である．また，例 3.1 で与えた非負確率変数 $Y_n(\omega)$ の分布関数 $G_n(x)$ は $2^i\leq n<2^{i+1}$ なる n に対して

$$G_n(x)=\begin{cases}0, & x<0\\ 1-\dfrac{1}{2^i}, & 0\leq x<1\\ 1, & x\geq 1\end{cases}$$

なので，極限分布

$$F(x)=\begin{cases}0, & x<0\\ 1, & x\geq 0\end{cases}$$

をもつ．すなわち $Y_n\xrightarrow{D}X$ である． □

一方，法則収束するが確率収束しない例は簡単に作ることが出来る．

例 3.3 確率空間 $(R, \mathcal{B}(\mathcal{R}), P)$ において，P は区間 $[0,1]$ 上の一様分布とする．この確率空間上で確率変数列 $\{Z_n;\ n=1,2,\ldots\}$ を以下のように定義する．

$$n\text{ が奇数のとき}: Z_n(\omega) = \begin{cases} 1 & \omega \in [0,1/2] \\ 0, & \text{その他} \end{cases} \qquad n\text{ が偶数のとき}: Z_n(\omega) = \begin{cases} 1 & \omega \in (1/2,1] \\ 0, & \text{その他} \end{cases}$$

定義より $\mathrm{P}(Z_n=0)=\mathrm{P}(Z_n=1)=1/2$ であるので，全ての n に対して Z_n は同じ分布をもつ．よって $Z_n \xrightarrow{D} Z_1$ である．しかし n が偶数のとき $\mathrm{P}(|Z_n - Z_1| > \epsilon) = 1$ となり，確率収束の意味では収束しない． □

このように法則収束は確率収束よりも弱い収束であり，実際，確率収束するならば法則収束することが知られている．

定理 3.2 (確率収束 $\Rightarrow$ 法則収束) $X_n \xrightarrow{P} X$ ならば $X_n \xrightarrow{D} X$ である． □

x を $F(x)$ の連続点とする．このとき任意の正数 ϵ に対して $F_n(x)-F(x) = \mathrm{P}(X_n \le x) - \mathrm{P}(X \le x) = \mathrm{P}(X_n \le x, X > x) + \mathrm{P}(X_n \le x, X \le x) - \mathrm{P}(X \le x) \le \mathrm{P}(X_n \le x, X > x) \le \mathrm{P}(X_n \le x, X > x+\epsilon) + \mathrm{P}(x < X \le x+\epsilon) \le \mathrm{P}(|X_n - X| > \epsilon) + F(x+\epsilon) - F(x)$ となり，$\epsilon \to 0$ とすると $\lim_{n\to\infty} F_n(x) - F(x) \le 0$ を得る．同様に $F(x) - F_n(x) = \mathrm{P}(X \le x) - \mathrm{P}(X_n \le x) = \mathrm{P}(X \le x, X_n > x) + \mathrm{P}(X \le x, X_n \le x) - \mathrm{P}(X_n \le x) \le \mathrm{P}(X \le x, X_n > x) \le \mathrm{P}(X \le x-\epsilon, X_n > x) + \mathrm{P}(x-\epsilon < X \le x) \le \mathrm{P}(|X_n - X| > \epsilon) + F(x) - F(x-\epsilon)$ となり，$\epsilon \to 0$ とすると $\lim_{n\to\infty} F(x) - F_n(x) \le 0$ を得る．

補題 3.1 ある定数 c に対して

$$X_n \xrightarrow{D} c \ \Leftrightarrow\ X_n \xrightarrow{P} c$$

である． □

定理 3.2 より $X_n \xrightarrow{P} c$ ならば $X_n \xrightarrow{D} c$ である．そこで $X_n \xrightarrow{D} c \Rightarrow X_n \xrightarrow{P} c$ を示す．X_n の分布関数を $F_n(x)$, c の分布関数を $F(x)$ とする．すなわち $F(x) = 0\ (x < c)$, $F(x) = 1\ (x \ge c)$ である．もし $X_n \xrightarrow{D} c$ ならば，全ての正数 ϵ に対して $\lim_{n\to\infty} F_n(c-\epsilon) = F(c-\epsilon) = 0$ であり，$\lim_{n\to\infty} F_n(c+\epsilon) = F(c+\epsilon) = 1$ である．よって $\lim_{n\to\infty} \mathrm{P}(|X_n - c| \ge \epsilon) = \lim_{n\to\infty} \mathrm{P}(X_n \ge c+\epsilon) + \lim_{n\to\infty} \mathrm{P}(X_n \le c-\epsilon) = 1 - \lim_{n\to\infty} \mathrm{P}(X_n < c+\epsilon) + \lim_{n\to\infty} F(c-\epsilon) = 0.$

3.2 大数の法則

多くの回数，同じ条件で確率的実験を繰り返したときに得られる数値の平均は，ある値に近付いていくことが知られている．一般にこのような性質は**大数の法則**（law of large numbers）と呼ばれている．大数の法則が成り立つ条件は色々と知られているが，ここでは確率変数列 $\{X_n;\ n=1,2,\}$ が互いに独立な場合の結果について紹介する．以下では確率変数列 $\{X_n;\ n=1,2,\}$ に対して最初の n 個の確率変数の和を

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

で表す．

定理 3.3 (大数の弱法則 (1)) $\{X_n;\ n=1,2,\}$ を独立な確率変数列とし，$\mathrm{V}(X_n) = \sigma_n^2 < \infty$ であるとする．もし，$\lim_{n\to\infty} \sum_{k=1}^{n} \sigma_k^2/n^2 = 0$ ならば，

$$\frac{S_n - \mathrm{E}(S_n)}{n} \xrightarrow{P} 0$$

が成立する．これをチェビシェフの**大数の弱法則**（weak law of large numbers）という． □

式 (14) で与えたチェビシェフの不等式を $(S_n - \mathrm{E}(S_n))/n$ へ適用すると $\mathrm{P}(|S_n - \mathrm{E}(S_n)| \ge n\epsilon) \le \frac{1}{n^2\epsilon^2} \sum_{k=1}^{n} \sigma_k^2$ を得る．よって仮定より，$\lim_{n\to\infty} \mathrm{P}(|S_n - \mathrm{E}(S_n)|/n \ge \epsilon) = 0.$

分散に関する条件 $\lim_{n\to\infty} \sum_{k=1}^{n} \sigma_k^2/n^2 = 0$ が成立する例として，全ての k に対して $\sigma_k^2 \le K$ となるような定数 K が存在する場合などがある．

大数の弱法則 (1) は分散が有限であるという制約があった．しかし，この制約は外すことができる．

定理 3.4 (大数の弱法則 (2)) $\{X_n;\ n = 1, 2, \}$ を独立で同一な分布に従う確率変数列とする．このとき，$\mathrm{E}(X_1) < \infty$ ならば

$$\frac{S_n}{n} \xrightarrow{P} \mathrm{E}(X_1)$$

が成立する． □

この定理の証明は**切断**（truncation）と呼ばれる技法を用いて，無限の分散をもつ確率変数を有限な分散をもつ確率変数へ変換することで行われる．詳細は付録 B.1 を参照すること．

上記の二つの大数の弱法則はいずれも確率収束を与えていた．一方，この収束が概収束の意味で成り立つとき，**大数の強法則**（strong law of large numbers）という．

定理 3.5 (大数の強法則) $\{X_n;\ n = 1, 2, \}$ を独立で同一な分布に従う確率変数列とする．もし，$\mathrm{E}(X_1) < \infty$ ならば

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathrm{E}(X_1)$$

が成立する． □

証明するには多くの準備を必要とするため省略する（例えば [2] を参照）．

大数の弱法則と大数の強法則の違い，すなわち確率収束と概収束の違いをしっかりと理解する必要がある．系 3.1 より大数の強法則は

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{m\geq n}\left|\frac{S_m}{m} - \mathrm{E}(X_1)\right| \geq \epsilon\right) = 0$$

と書くことが出来る．ここで事象 $\{\sup_{m\geq n}|S_m/m - \mathrm{E}(X_1)| > \epsilon\}$ は $|S_m/m - \mathrm{E}(X_1)|$ の値が ϵ より大きくなるような m （$m \geq n$）が少なくとも一つあるということである．よって大数の強法則が示していることは，n が十分に大きければ，n 以上の m で $|S_m/m - \mathrm{E}(X_1)| > \epsilon\}$ となってしまうことはほとんどないということである．言い換えれば，このことは $S_1/1, S_2/2, \ldots, S_n/n, \ldots$ という列が n が増加するにつれて，徐々に，それより大きな $m > n$ のいずれの項も平均との差が ϵ より大きくならないようになるということである．一方，大数の弱法則がいうことは n が増加するにつれて，徐々に，S_n/n と平均との差が ϵ より大きくならないようになるということである．このように大数の強法則は列 $\{S_n/n;\ n = 1, 2, \ldots\}$ の最初の $n-1$ 個を除いた残り全てに関する性質に言及しているのに対し，大数の弱法則は項 S_n/n それ自身に対する性質を述べている．

例 3.4 例 1.1 を拡張して，コインを無限に投げ続ける実験を考える．X_n $(n = 1, 2, \ldots)$ を n 回目に表が出れば 1，裏が出れば 0 の値を取る確率変数とする．このとき S_n/n は確率 1 で 1/2 へ収束する．一方，S_n/n が 1/2 でないような標本は簡単に作ることが出来る．例えば全て表が出る，全て裏が出る，全体の 1/3 だけ表が出る，などである．大数の強法則は，このような S_n/n が 1/2 に収束しないような標本全体からなる事象は確率 0 をもつということを主張している．一方，大数の弱法則は n を大きくすれば事象 $\{|S_n/n - 1/2| \geq \epsilon\}$ が起こる確率をいくらでも小さくできるということを主張している． □

3.3 分布の弱収束と中心極限定理

確率変数列の法則収束は，対応する分布関数列の収束と考えることが出来る．このような意味で定義された分布関数の収束を**弱収束**（weak convergence）といい，$F_n(x) \xrightarrow{D} F(x)$ と書く．法則収束の定義から明らかなように，弱収束では $F(x)$ が連続であるような点でのみ収束が要求されており，$F(x)$ の不連続点での収束は要求されていない．

例 3.5 以下で定義される分布関数 $F_n(x)$，$F(x)$ を考える．

$$F_n(x) = \begin{cases} 0, & x < 0 \\ nx, & x \in [0, 1/n] \\ 1, & x > 1/n \end{cases} \qquad F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

明らかに $F_n(x)$ は $x=0$ を除いて $F(x)$ へ収束するが，$x=0$ では収束していない．しかし，$x=0$ は $F(x)$ の不連続点であるので，定義より $F_n(x) \xrightarrow{D} F(x)$ である． □

分布関数が弱収束することと等価な条件は幾つか知られており，例えば次のようなものがある．

定理 3.6 $F_n(x) \xrightarrow{D} F(x)$ は，任意の有界で連続な関数 $g(x)$ に対して

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g(x) dF_n(x) = \int_{-\infty}^{\infty} g(x) dF(x)$$

が成立することと等価である． □

弱収束を用いて表現された非常に重要な定理に中心極限定理がある．

定理 3.7 (中心極限定理) $\{X_n;\ n = 1, 2, \}$ を独立で同一な分布に従う確率変数列とする．もし，$\mu = \mathrm{E}(X_1) < \infty$ かつ $\sigma^2 = \mathrm{V}(X_1)$ が $0 < \sigma^2 < \infty$ を満たすならば，$S_n = \sum_{k=1}^n X_k$ としたとき

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \tag{19}$$

が成立する．これを**中心極限定理**（central limit theorem）という． □

定理 3.7 の証明は割愛する．

中心極限定理 (19) の左辺は平均 0，分散 1 をもつ標準正規分布 $N(0,1)$ の分布関数である．よって

$$\lim_{n\to\infty} \mathrm{P}\left(a \leq \frac{1}{\sqrt{n}\sigma} \sum_{k=1}^{n} (X_k - \mu) \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

と等価である．ガウス（Gauss）は実験から得られたデータの期待値からのバラツキに一定の法則があることを見い出した．それゆえ正規分布は**ガウス分布**（Gaussian distribution）とも呼ばれる．

この節を終えるにあたり，中心極限定理の意味するところについて述べる．まず，S_n が平均 $n\mu$, 分散 $n\sigma^2$ をもつことに注意する．中心極限定理 (19) の左辺に現れる

$$Z_n = (S_n - n\mu)/(\sqrt{n}\sigma)$$

は式 (15) に従って S_n を標準化したものであり，全ての自然数 n に対して平均 0，分散 1 をもつ確率変数である．中心極限定理はこの確率変数 Z_n の従う分布が，確率変数 S_n の分布，すなわち，確率変数 X_1 の分布が如何なるものであろうとも，有限の平均と分散をもつならば，n を増加させると標準正規分布 $N(0,1)$ に弱収束するということを主張している．すなわち，n を増加させると $(S_n - n\mu))/\sqrt{n}$ は平均 0，分散 σ^2 をもつ正規分布に弱収束するということである．

中心極限定理は確率変数 S_n が必ずしも（標準正規分布 $N(0,1)$ の密度関数で近似できるような）密度関数をもつことを主張しているわけではない．実際，X_1 が離散的確率変数ならば全ての n に対して確率変数 S_n は密度関数をもたない．このような場合，大きな n に対して，その分布関数は非常に細かな階段状の変化をする形状をもち，その形状は積分する（すなわち分布関数を考える）ことで平滑化されることになる．

絶対値が非常に大きい負の値 x に対しては，式 (19) の右辺は非常に 0 に近い値となる．よって n がそれほど大きくなければ，右辺と（左辺に現れる）確率 $\mathrm{P}(Z_n \leq x)$ の差の絶対値は小さいが，その比は 1 から懸け離れたものになっていると思われる．特に非負の確率変数の和を考えると，全ての標本に対して $S_n \geq 0$ であるが，正規

分布の定義域は $(-\infty, \infty)$ であるので，有限の n に対して $x < -n\mu/(\sqrt{n}\sigma) = -\sqrt{n}\mu/\sigma$ では明らかに近似的評価として採用できない．また同様に，非常に大きい正数 x に対しては右辺と $\mathrm{P}(Z_n \leq x)$ の両方とも 1 に非常に近く，この場合も右辺は左辺に現れる確率の良い近似とはなっていないと思われる．すなわち左辺に現れる確率は適当な大きさの n に対しては，x が 0 に比較的近い場合にのみ良い近似と見なせると考えられ，この定理が極限定理ではなく，中心極限定理と呼ばれている理由もそこにある．

前節で紹介した大数の法則や中心極限定理では独立な確率変数列を仮定していた．実は，大数の法則は独立でない確率変数列に対しても，一定の条件下で成立する．この意味で，大数の法則の適用範囲は（中心極限定理に比べて）相当広い．一方，独立でないような確率変数列に対する中心極限定理はほとんど知られていない．すなわち，中心極限定理においては独立性が非常に重要な仮定となっており，中心極限定理を応用で用いる際には，この点に十分注意を払う必要がある．

A　Riemann 積分と Riemann-Stieltjes 積分について

A.1　Riemann 積分

ここでは [4] に沿って **Riemann 積分**を説明する．これは，いわゆる区分求積法と呼ばれているものであり，定積分が面積に対応することを明確に表している．

まず，閉区間 $[a, b]$ に対して

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

となるような有限個の点 x_i（$i = 0, 1, \ldots, n$）を取ることをこの区間の**分割**といい，点 x_i を**分点**，閉区間 $[x_i, x_{i-1}]$（$i = 1, \ldots, n$）を**小区間**とよぶ．さらにこの分割を Δ で表し，小区間の長さを Δx_i と記す.

$$\Delta x_i = x_i - x_{i-1}, \qquad i = 1, 2, \ldots, n$$

明らかに

$$\sum_{i=1}^{n} \Delta x_i = b - a$$

である．閉区間 $[a, b]$ の分割 Δ に対して，さらにいくつかの分点を追加して得られる分割を Δ の**細分**という．閉区間 $[a, b]$ の任意の二通りの分割 Δ，Δ' に対して，常にこれらの共通の細分 Δ'' を作ることができる．実際，Δ, Δ' の分点を合わせたものを Δ'' の分点とすればよい.

閉区間 $[a, b]$ で定義された関数 $u(x)$ が与えられたとき，閉区間 $[a, b]$ の分割 Δ および各小区間 $[x_{i-1}, x_i]$ 上の点 ξ_i $(i = 1, 2, \ldots, n)$ をとり，和

$$S(\Delta) = \sum_{i=1}^{n} u(\xi_i) \Delta x_i$$

を作る．これを **Riemann 和**という．小区間の長さ Δx_i $(i = 1, 2, \ldots, n)$ のうち最大のものを $|\Delta|$ としたとき，分割 Δ および点 ξ_i の取り方に関係なく，極限値

$$\lim_{|\Delta| \to 0} S(\Delta) = A$$

が定まるとき，関数 $u(x)$ は $[a, b]$ において**積分可能**であるといい，この値を $f(x)$ の $[a, b]$ における**定積分**，または a から b までの**定積分**という．そして，これを

$$A = \int_a^b u(x) dx$$

で表す.

関数 $u(x)$ が区間 $[a,b]$ において積分可能であるか否かは，以下のようにして検証できる．m_i と l_i $(i=1,\dots,n)$ をそれぞれ小区間 $[x_i, x_{i-1}]$ における関数 $u(x)$ の上界値，下界値を与える点とし，これらの点を分点とした Riemann 和

$$S(\mathit{\Delta}) = \sum_{i=1}^{n} u(\xi_i)\mathit{\Delta} x_i, \qquad M(\mathit{\Delta}) = \sum_{i=1}^{n} u(m_i)\mathit{\Delta} x_i, \qquad L(\mathit{\Delta}) = \sum_{i=1}^{n} u(l_i)\mathit{\Delta} x_i$$

を作れば，$M(\mathit{\Delta})$，$L(\mathit{\Delta})$ の値は分割 $\mathit{\Delta}$ によって定まり

$$L(\mathit{\Delta}) \le S(\mathit{\Delta}) \le M(\mathit{\Delta})$$

となる．さらに，分割 $\mathit{\Delta}$ の細分 $\mathit{\Delta}''$ をとれば，$M(\mathit{\Delta})$，$L(\mathit{\Delta})$ の作り方から

$$L(\mathit{\Delta}) \le L(\mathit{\Delta}''). \qquad M(\mathit{\Delta}'') \le M(\mathit{\Delta})$$

となる．よって

$$\lim_{|\mathit{\Delta}|\to 0} |M(\mathit{\Delta}) - L(\mathit{\Delta})| = 0$$

となることを示せばよい．

そこで，関数 $u(x)$ は閉区間 $[a,b]$ で**連続**[14]であると仮定する．有界閉集合において連続な関数は，その集合内で必ず最大値と最小値をもち，かつ，**一様連続**[15]であることが知られている．よって次式が成立する．

$$\min_{x\in[a,b]} u(x)\cdot(b-a) \le L(\mathit{\Delta}), \qquad M(\mathit{\Delta}) \le \max_{x\in[a,b]} u(x)\cdot(b-a)$$

閉区間 $[a,b]$ の任意の二通りの分割 $\mathit{\Delta}$，$\mathit{\Delta}'$ に対して，これらの共通の細分 $\mathit{\Delta}''$ をとれば

$$L(\mathit{\Delta}) \le L(\mathit{\Delta}'') \le M(\mathit{\Delta}'') \le M(\mathit{\Delta}')$$

である．すなわち，閉区間 $[a,b]$ のあらゆる分割 $\mathit{\Delta}$ に対して，$L(\mathit{\Delta})$ は上に有界であり，その上限 L が存在する．同様に $M(\mathit{\Delta})$ は下に有界であり，その下限 M が存在する．

$$L(\mathit{\Delta}) \le L \le M \le M(\mathit{\Delta})$$

$u(x)$ は閉区間 $[a,b]$ において一様連続であるので，任意の正数 ϵ に対して正数 δ が存在し

$$x, x' \in [a,b], |x-x'| < \delta \quad \Rightarrow \quad |u(x) - u(x')| < \epsilon$$

となる．よって，$|\mathit{\Delta}| < \delta$ となる任意の分割 $\mathit{\Delta}$ をとれば，$\mathit{\Delta}$ の小区間 $[x_{i-1}, x_i]$ において，$u(x)$ を最大，最小とする点 m_i, l_i に対しても

$$0 \le u(m_i) - u(l_i) < \epsilon$$

となる．この不等式の各辺に小区間の長さ $\mathit{\Delta} x_i$ を掛けて，全ての i について加えれば

$$0 \le M(\mathit{\Delta}) - L(\mathit{\Delta}) < (b-a)\epsilon$$

となる．よって $\epsilon \to 0$ の極限を取れば

$$\lim_{|\mathit{\Delta}|\to 0} |M(\mathit{\Delta}) - L(\mathit{\Delta})| = 0, \qquad L = M = \lim_{|\mathit{\Delta}|\to 0} S(\mathit{\Delta})$$

となり，閉区間 $[a,b]$ において連続な関数 $u(x)$ の定積分は存在する．

上記の証明から以下のことが分かる．関数 $u(x)$ が開区間 (a,b) で一様連続であれば，$u(x)$ は $[a,b]$ において積分可能である．この場合，区間の両端 a，b における $u(x)$ の値は，この定積分とは全く無関係である．一般に，関数 $u(x)$ は，閉区間 $[a,b]$ において，有限個の不連続点をもつものとし，隣り合う二つの不連続点の間では一様連続となっていれば，$u(x)$ は $[a,b]$ において積分可能である．

[14]点集合 D を定義域にもつ関数 $u(x)$ が与えられたとし，D 上の 1 点 c を考える．任意の正数 ϵ に対して，正数 $\delta = \delta(c,\epsilon)$ が存在し，$|x-c| < \delta$ ならば $|u(x) - u(c)| < \epsilon$ となるとき，関数 $u(x)$ は点 c において連続であるという．さらに，関数 $u(x)$ が定義域 D の全ての点において連続であるとき，$u(x)$ は D において連続であるという．

[15]連続の定義において，δ を ϵ のみの関数として定めることができる場合，一様連続という．正確には下記の通りである．点集合 D を定義域にもつ関数 $u(x)$ が与えられたとき，任意の正数 ϵ に対して，正数 $\delta = \delta(\epsilon)$ が存在し，$x, x' \in D$ かつ $|x-x'| < \delta$ ならば $|u(x) - u(x')| < \epsilon$ となるとき，関数 $u(x)$ は D において一様連続であるという．

A.2　Riemann-Stieltjes 積分

次に，[5] に沿って，簡単に **Riemann-Stieltjes 積分**の概略を紹介する．

閉区間 $[a,b]$ で定義された関数 $u(x)$ と区間 $[a,b]$ で定義された右連続単調非減少関数 $F(x)$ を考える．閉区間 $[a,b]$ の分割は Reimann 積分の場合と同様に定義されているものとする．さらに，$\mathit{\Delta} F_i = F(x_i) - F(x_{i-1})$ とする．このとき，上界，下界の重み付き和

$$L(\mathit{\Delta}) = \sum_{i=1}^{n} u(l_i)\mathit{\Delta} F_i, \qquad M(\mathit{\Delta}) = \sum_{i=1}^{n} u(m_i)\mathit{\Delta} F_i$$

が，$|\mathit{\Delta}| \to 0$ の極限において一致する時，$u(x)$ の $F(x)$ に関する Riemann-Stieltjes 積分が以下のように定義される．

$$\int_a^b u(x)dF(x) = \lim_{|\mathit{\Delta}|\to 0} L(\mathit{\Delta}) = \lim_{|\mathit{\Delta}|\to 0} M(\mathit{\Delta})$$

以下のことに注意する．ある点 $\tau \in [a,b]$ で $u(x)$ は連続であるが $F(x)$ が不連続な場合，$n \to \infty$ の極限において点 τ における上界あるいは下界の積和への貢献は $u(\tau)(F(\tau) - F(\tau-))$ となる．ただし，$F(\tau-)$ は $F(x)$ の $x = \tau$ における左極限（x を左から τ に近付けた時の極限）である．よって，$F(x)$ が x_i で不連続となる階段関数ならば，$x = x_i$ において定義可能な $u(x)$ に対して

$$\int_a^b u(x)dF(x) = \sum_{i=1}^{n} u(x_i)\mathit{\Delta} F_i$$

が成立する．

Riemann 積分は $F(x) = x$ の場合に対応する．もし，区間 $[a,b]$ で $F(x)$ が微分可能であれば，Riemann-Stieltjes 積分 は Riemann 積分で表現できる．

$$\int_a^b u(x)dF(x) = \int_a^b u(x)\frac{dF(x)}{dx}dx$$

なお，異常積分は Riemann 積分と同様に極限として定義される．

$$\int_a^\infty u(x)dF(x) = \lim_{b\to\infty}\int_a^b u(x)dF(x)$$

B　定理の証明

B.1　大数の弱法則の証明

この節では定理 3.4 の証明を行う．互いに独立で同じ分布に従い，有限の平均 μ_X をもつ確率変数列 $\{X_n;\ n = 1, 2, \}$ と任意の正数 $\epsilon > 0$ が与えられたとき，ある正数 $b > 0$ を用いて新しい確率変数列 $\{Y_n;\ n = 1, 2, \ldots\}$ を以下で定義する．

$$Y_n = \begin{cases} X_n, & |X_n - \mu_X| \le b \\ \mu + b, & X_n - \mu_X > b \\ \mu - b, & X_n - \mu_X < -b \end{cases}$$

確率変数列 $\{Y_n;\ n = 1, 2, \ldots\}$ は互いに独立で同一の分布に従っており，その平均を μ_Y とする．Y_1 の分散 σ_Y^2 は

$$\mathrm{E}((Y_1 - \mu_X)^2) = \mathrm{E}(\{(Y_1 - \mu_Y) + (\mu_Y - \mu_X)\}^2) = \sigma_Y^2 + (\mu_Y - \mu_X)^2$$

であるので，$\sigma_Y^2 \le \mathrm{E}((Y_1 - \mu_X)^2)$ である．さらに

$$\sigma_Y^2 \le \mathrm{E}((Y_1 - \mu_X)^2) = \int_{-\infty}^{\infty} (x - \mu_X)^2 dF_Y(x) \le b\int_{-\infty}^{\infty} |x - \mu_X| dF_Y(x)$$

を得る．ただし $F_Y(x)$ は Y_1 の分布関数である．最後の不等号は $|Y-\mu_X|\le b$ を用いた．さらに $\mu_X-b\le x<\mu_X+b$ では $F_Y(x)$ は X_1 の分布関数に等しいことに注意すると，

$$\sigma_Y^2\le b\int_{-\infty}^{\infty}|x-\mu_X|dF_Y(x)\le b\int_{-\infty}^{\infty}|x-\mu_X|dF_X(x)$$

である．

さて，$T_n=Y_1+\cdots+Y_n$ とし，式 (14) のチェビシェフの不等式を ϵ を $\epsilon/2$ に変更して適用すると

$$\mathrm{P}\left(\left|\frac{T_n}{n}-\mu_Y\right|\ge\frac{\epsilon}{2}\right)\le\frac{4\sigma_Y^2}{n\epsilon^2}\le\frac{4b\alpha}{n\epsilon^2}$$

となる．ただし $\alpha=\int_{-\infty}^{\infty}|x-\mu_X|dF_X(x)<\infty$ とした．$|\mu_Y-\mu_X|$ は $b\to\infty$ としたとき 0 へ近付くので，任意の $\epsilon>0$ に対して，十分大きな b が存在し $|\mu_Y-\mu_X|<\epsilon/2$ とすることができる．よって，そのような b に対して

$$\mathrm{P}\left(\left|\frac{T_n}{n}-\mu_X\right|\ge\epsilon\right)\le\frac{4b\alpha}{n\epsilon^2}\tag{20}$$

を得る[16]．

S_n と T_n は全ての k $(k=1,\ldots,n)$ で $|X_k-\mu_X|\le b$ となるような標本に対しては同じ値を取る．よって A_k で事象 $|X_k-\mu_X|>b$ を表すと

$$\mathrm{P}(T_n\ne S_n)\le\mathrm{P}\left(\cup_{k=1}^{n}A_k\right)=n\mathrm{P}(A_1)=n\mathrm{P}(|X_1-\mu_X|>b)$$

を得る．さらに $G(x)=\mathrm{P}(|X_1-\mu_X|\le x)$ とすると

$$\mathrm{P}(|X_1-\mu_X|>b)=\int_b^{\infty}dG(x)\le\frac{1}{b}\int_b^{\infty}xdG(x)$$

である．α が有限であるので最後の積分は $b\to\infty$ の極限で 0 へ収束する．よって，任意の $\delta>0$ に対して十分大きな b が存在し $\mathrm{P}(|X_1-\mu_X|>b)\le\delta^2/b$ とすることが出来る．すなわち，十分大きな b に対して

$$\mathrm{P}(T_n\ne S_n)\le n\mathrm{P}(|X_1-\mu_X|>b)\le\frac{n\delta^2}{b}\tag{21}$$

事象 $|S_n/n-\mu_X|\ge\epsilon$ は事象 $|T_n/n-\mu_X|\ge\epsilon$ あるいは $S_n\ne T_n$ のいずれかが起こったときのみ起こり得るので，式 (20) と式 (21) を組合せて，$\delta=b/n$ とすると

$$\mathrm{P}\left(\left|\frac{S_n}{n}-\mu_X\right|\ge\epsilon\right)\le\frac{4\delta\alpha}{\epsilon^2}+\delta\tag{22}$$

を得る．式 (20) と式 (21) は任意の n と十分大きな b について成立するので，式 (22) は任意の正数 $\delta>0$ と十分大きな n について成立する．よって，任意の正数 $\epsilon>0$ に対して，適当な δ を選ぶことで式 (22) の右辺第 1 項，第 2 項共に，いくらでも小さくすることが出来る．よって定理が成立する．

C　練習問題の略解

問 1.1.

$$F(x)=\begin{cases}0, & x<0\\ x, & 0\le x<1,\\ 1, & x\ge 1\end{cases},\qquad f(x)=\begin{cases}0, & x<0\quad(\text{あるいは } x\le 0),\\ 1, & 0\le x<1\quad(\text{あるいは } 0<x<0,\ 0\le x<0),\\ 0, & x\ge 1\quad(\text{あるいは } x>1),\end{cases}$$

[16] $|\mu_Y-\mu_X|<\epsilon/2$ のとき，$\{|T_n/n-\mu_X|\ge\epsilon\}=\{|T_n/n-\mu_Y+\mu_X-\mu_Y|\ge\epsilon\}\Rightarrow\{|T_n/n-\mu_Y|+|\mu_X-\mu_Y|\ge\epsilon\}\Rightarrow\{|T_n/n-\mu_Y|\ge\epsilon/2\}$ であるので，$A\Rightarrow B$ ならば $\mathrm{P}(A)\le\mathrm{P}(B)$ に注意すると式 (20) を得る．

（密度関数に関しては，区間の境界をいずれに含めるかは任意，全区間で積分して 1 になればよい）

問 1.2.

$$F(x) = \begin{cases} 0, & x < x_1 \\ 1/3, & x_1 \le x < x_2, \\ 2/3, & x_2 \le x < x_3, \\ 1, & x \ge x_3 \end{cases},$$

（$x = x_1, x_2, x_3$ での取り扱いに注意）

問 1.3. (a) $A\cap B$, $A^c\cap B$, $A\cap B^c$, $A^c\cap B^c$ (b) $\mathrm{P}(A^c\cap B) = b-c$, $\mathrm{P}(A\cap B^c) = a-c$, $\mathrm{P}(A^c\cap B^c) = 1-a-b+c$ (c) $c \ge 0$, $a-c \ge 0$, $b-c \ge 0$, $1-a-b+c \ge 0$

問 1.4. (a) $3/4 \le \mathrm{P}(A\cup B) \le 1$ (b) $1/12$, $1/3$

問 1.5. $\mathrm{P}(R_1) = r/(r+b)$，$\mathrm{P}(R_2) = r/(r+b)$，$\mathrm{P}(R_1 \mid R_2) = (r+c)/(r+b+c)$

問 1.6. $\mathrm{P}(A_1) = \mathrm{P}(A_2) = \mathrm{P}(A_3) = 1/2$, $\mathrm{P}(A_1\cap A_2) = 1/4$, $\mathrm{P}(A_2\cap A_3) = 1/4$, $\mathrm{P}(A_1\cap A_3) = 1/4$ より A_i と A_j $(i \ne j)$ は互いに独立．$\mathrm{P}(A_1\cap A_2\cap A_3) = 0 \ne 1/8 = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3)$ より A_1，A_2，A_3 は互いに独立ではない．

問 1.7. 各対戦の勝者を並べたものを標本空間 Ω とする．すなわち，例えば $ACBACBACC$ を $A(CBA)^2CC$ と書くとすると $\Omega = \{A(CBA)^nA, A(CBA)^nCC, A(CBA)^nCBB, B(CAB)^nB, B(CAB)^nCC, B(CAB)^nCAA, n = 0,1,\ldots\}$ である．A が優勝する場合は $A(CBA)^nA$ と $B(CAB)^nCAA$ なので，その確率は $\sum_{n=0}^{\infty}(1/2)(p^3)^n(1-p) + \sum_{n=0}^{\infty}(1/2)(p^3)^n pp(1-p) = (1+p^2)/[2(1+p+p^2)]$, B が優勝する確率は A と同じ．C が優勝する確率は $p/(1+p+p^2)$．$p > 1/2$ ならば C が不利となる．

問 1.8. (a) $\mathrm{P}(X_0 = \min(X_0, X_1, \ldots, X_n)) = \int_{-\infty}^{\infty}[\prod_{n=1}^{n}(1-F_i(x))]f_0(x)dx$
(b) $p(n) = \mathrm{P}(X_0 = \min(X_0, X_1, \ldots, X_n))$ とおく．

$$\begin{aligned} p(n) &= \int_{-\infty}^{\infty}(1-F(x))^n f(x)dx = [(1-F(x))^n F(x)]_{-\infty}^{\infty} + n\int_{-\infty}^{\infty}(1-F(x))^{n-1}F(x)f(x)dx \\ &= n\int_{-\infty}^{\infty}(1-F(x))^{n-1}f(x)dx - n\int_{-\infty}^{\infty}(1-F(x))^n f(x)dx = np(n-1) - np(n) \end{aligned}$$

より $(n+1)p(n) = np(n-1)$．よって $(k+1)p(k) = kp(k-1)$ の両辺をそれぞれ全ての $k = 1, 2, \ldots, n$ について足し合わせると $(n+1)p(n) = p(0)$ より，$p(n) = 1/(n+1)$．

問 1.9. $p_{i,j} = \mathrm{P}(X = i, Y = j)$ $(i \ge j)$ とする．$1 \le j < i \le 6$（全 15 通り）に対して $p_{i,j} = 1/18$，$1 \le j = i \le 6$ に対して $p_{i,i} = 1/36$. $\mathrm{P}(X = i) = (i-1)/18 + 1/36 = (2i-1)/36$ $(i = 1, 2, \ldots, 6)$, $\mathrm{P}(Y = j) = (6-j)/18 + 1/36 = (13-2i)/36$.

問 1.10. $x \ge 0$ に対して $G(x) = \mathrm{P}(X^2 \le x) = \mathrm{P}(-\sqrt{x} \le X \le \sqrt{x}) = \mathrm{P}(X \le \sqrt{x}) - \mathrm{P}(X < -\sqrt{x}) = F(\sqrt{x}) - \lim_{u\to x-} F(-\sqrt{u})$.

問 1.11. X の密度関数は $x \in (0, 1]$ のとき 1，それ以外は 0 となる関数．Z の密度関数を $f(x)$ とすると $x \le 0$ ならびに $x > 2$ では $f(x) = 0$, $0 < x \le 1$ では $f(x) = \int_0^x 1\cdot 1dx = x$，$1 < x \le 2$ では $\int_{x-1}^{1} 1\cdot 1dx = -x+2$,

問 1.12. $p_i(k) = \exp(-\lambda_i)\lambda_i^k/k!$ としたとき，$q(k) = \sum_{n=0}^{k} p_1(n)p_2(k-n)$ である．

$$\begin{aligned} q(k) &= \sum_{n=0}^{k} e^{-\lambda_1}\frac{\lambda_1^n}{n!}e^{-\lambda_2}\frac{\lambda_2^{k-n}}{(k-n)!} = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}\sum_{n=0}^{k}\frac{k!}{n!(k-n)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^n\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{k-n} \\ &= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2} + \frac{\lambda_2}{\lambda_1+\lambda_2}\right)^k = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!} \end{aligned}$$

問 2.1. $\sum_{k=0}^{\infty} kp(k) = \sum_{k=1}^{\infty} kp(k) = \sum_{k=1}^{\infty} \exp(-\lambda)\lambda^k/(k-1)! = \lambda \exp(-\lambda) \sum_{k=1}^{\infty} \lambda^{k-1}/(k-1)!$
$= \lambda \exp(-\lambda) \sum_{k=0}^{\infty} \lambda^k/k! = \lambda$.

問 2.2. $\mathrm{P}(N = k) = (1/2)^k$ なので $\sum_{k=1}^{\infty} 2^k \mathrm{P}(N = k) = \sum_{k=1}^{\infty} 1 = \infty$ となり，報酬は無限大.

問 2.3. 例 2.1 ならびに問 1.3 の結果を用いると $\mathrm{E}(N) = \sum_{n=0}^{\infty} \mathrm{P}(N > n) = \sum_{n=0}^{\infty}(n+1)^{-1} > \int_0^{\infty}(x+1)^{-1}dx = \infty$ より $\mathrm{E}(N)$ は無限大.

問 2.4. $\mathrm{E}(Y) = \mathrm{E}((X - \mathrm{E}(X))/\sqrt{\mathrm{V}(X)}) = \mathrm{E}((X - \mathrm{E}(X)))/\sqrt{\mathrm{V}(X)} = [\mathrm{E}(X) - \mathrm{E}(X)]/\sqrt{\mathrm{V}(X)} = 0$. $\mathrm{V}(Y) = \mathrm{V}(X - \mathrm{E}(X))/\mathrm{V}(X) = \mathrm{V}(X)/\mathrm{V}(X) = 1$.

問 2.5. 平均 $1/\mu$，分散 $1/\mu$，変動係数 1.

問 2.6. $\mathrm{Cov}(X, Y) = \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) = \mathrm{E}(XY - \mathrm{E}(X)Y - X\mathrm{E}(Y) + \mathrm{E}(X)\mathrm{E}(Y)) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(X)\mathrm{E}(Y) + \mathrm{E}(X)\mathrm{E}(Y)$ から導かれる.

問 2.6. (a) $\mathrm{P}(X = i) = 1/3$ $(i = 1, 2, 3)$，$\mathrm{P}(Y = 0) = 1/3$，$\mathrm{P}(Y = 1) = 2/3$ なので，例えば $\mathrm{P}(X = 1, Y = 1) = 1/3 \neq 2/9 = \mathrm{P}(X = 1)\mathrm{P}(Y = 1)$ より，X と Y は独立でない． (b) $\mathrm{E}(X) = 2$，$\mathrm{E}(Y) = 2/3$，$\mathrm{E}(XY) = 1 \times 1/3 + 0 \times 1/3 + 3 \times 1/3 = 4/3$ なので $\mathrm{E}(X)\mathrm{E}(Y) = \mathrm{E}(XY)$ が成立．よって無相関.

参考文献

[1] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications, Vol.II.* John Wiley & Sons, New York.

[2] Gallager, R. G. (1996) *Discrete Stochastic Processes.* Kluwer, Boston.

[3] 高橋幸雄 (2008) 確率論．朝倉書店.

[4] 瀧澤精二, 微分積分学 (上), 廣川書店, 1976.

[5] R. W. Wolff, *Stochastic Modeling and the Theory of Queues*, Prentice Hall, Englewood Cliffs, NJ, 1989.