確率数理要論 メモ

2009 年 担当 竹村

注意事項

1. このメモは，一昨年の駒木先生の講義メモに昨年竹村が若干加筆したもので，講義はこのノートに基づいておこなう．
2. 教科書は講義要綱に参考文献は Lamperti (1996) “Probability : a survey of the mathematical theory”, 2nd edition である．ただし講義と完全に一致しているわけではない．
3. やや進んだ話題を扱った節や問題，証明等には $*$ 印をつける．これらの部分は，最初は飛ばしても差し支えない．

1　確率空間

1.1　確率空間の定義

Ω を (空集合ではない) ある集合とする．Ω は直観的にいえば，想定できる結果をすべてあつめた集合であり，要素 ω を与えると，想定される事象の成否がすべて決まるようなものである．Ω を**標本空間**， $\omega \in \Omega$ を **標本** あるいは**根元事象** と呼ぶ．

例． (サイコロ) $\Omega = \{1, 2, \ldots, 6\}$, ω: サイコロの目
例． (無限回のコイン投げ) $\Omega = \{0, 1\}^\infty$, $\omega = \omega_1\omega_2\ldots$, $\omega_i = 0$ or 1

定義． Ω の部分集合を要素とする集合 (集合族) $\mathcal{F}$ が

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F}$ $(i = 1, 2, 3, \ldots) \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

を満たすとき，$\mathcal{F}$ は σ**-加法族** であるという．Ω と σ-加法族 $\mathcal{F}$ との組 $(\Omega, \mathcal{F})$ を**可測空間**と呼ぶ． $\square$

[**補足**] Ω が可算集合ならば，べき集合 2^Ω を $\mathcal{F}$ とすればよい．すなわちすべての部分集合が可測としてよい．しかし実軸のように Ω が非可算集合の時に，どのような部分集合の「長さが測れるか」は自明ではない．従って上で可算加法性を要求している．長さの測れる集合の可算和は面積が測れてほしいという要請である．

定義． $\mathcal{F}$ 上の非負関数 P が

1. $P(\Omega) = 1$
2. (可算加法性) $A_i \in \mathcal{F}$ $(i = 1, 2, 3, \ldots)$ かつ $A_i \cap A_j = \emptyset$ (任意の $i \neq j$) ならば,

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

を満たすとき, P を $(\Omega, \mathcal{F})$ 上の**確率測度**, 3つ組 $(\Omega, \mathcal{F}, P)$ を **確率空間** と呼ぶ. また, 確率空間を考えるとき, $\mathcal{F}$ の要素を**事象**と呼ぶ. □

問題 1. $A_1, A_2, \ldots \in \mathcal{F}$ で, かつ $A_1 \subset A_2 \subset \cdots$ のとき, $\lim_{i\to\infty} A_i = \cup_{i=1}^{\infty} A_i$ と定義すると,

$$\lim_{i\to\infty} P(A_i) = P(\lim_{i\to\infty} A_i)$$

が成立することを示せ. 同様に, $A_1, A_2, \ldots \in \mathcal{F}$ で, かつ $A_1 \supset A_2 \supset \cdots$ のとき, $\lim_{i\to\infty} A_i = \cap_{i=1}^{\infty} A_i$ と定義すると,

$$\lim_{i\to\infty} P(A_i) = P(\lim_{i\to\infty} A_i)$$

が成立することを示せ.

1.2 実数軸上の確率測度

Ω として実数の集合 $\mathbb{R}$ をとり, σ-加法族 $\mathcal{F}$ として, Borel 集合族 (開集合の全体を含む最小の σ-加法族) を考える. P を $(\Omega, \mathcal{F})$ 上の確率測度とする. この例は, 応用上大変重要である. (なお, この講義では特に断らない場合には, $\mathbb{R}$ 上の σ-加法族として, Borel 集合族を考えることとし, いちいち明記しないことにする.)

注. ルベーグ積分について学習したことが無い人は, これから現れるいろいろな $\mathbb{R}$ の部分集合が, Borel 可測では無いのではないか, と心配になるかもしれない. しかし, 普通に応用上あらわれるおよそありとあらゆる集合は Borel 可測だと思って実は差し支えない. 可測でない集合の例は, 例えば伊藤 (1963) p.49 以下にある (ルベーグ非可測集合の例が書いてあるが, ルベーグ非可測集合はボレル非可測集合) がかなり苦労しないと作れない. Emile Borel 自身はボレル可測集合についてかなり具体的に詳しく考えている (田中尚夫「選択公理と数学」増補版 3 章 15 節).

定義. (**分布関数**)

$$F(t) := P(\{u : -\infty < u \leq t\})$$

□

分布関数の性質.

1. F は単調非減少
2. F は右連続
3. $\lim_{t\to-\infty} F(t)=0, \quad \lim_{t\to\infty} F(t)=1$

逆に上の 1, 2, 3 の性質をもつ $\mathbb{R}$ 上の関数 F が与えられると，対応する $\mathbb{R}$ 上の測度が決まることが知られている．

問題 2．上記の分布関数の性質のうち，2, 3 を証明せよ．

問題 3*．分布関数 F に対し

$$\lim_{t\uparrow a} F(t) < F(a)$$

が成立するとき，点 a でジャンプがあるということにする．任意の分布関数 F が与えられたとき，ジャンプがある点は高々可算無限個であることを示せ．

2 確率変数

2.1 確率変数の定義

$(\Omega,\mathcal{F})$ をある可測空間，X を Ω 上の実数値関数とする．

定義． 任意の Borel 集合 $C\subset\mathbb{R}$ に対して，$X^{-1}(C)\in\mathcal{F}$ のとき，X は**確率変数**であるという．

この定義は，ひとことで言えば，<u>確率変数とは確率空間上の可測関数である</u> ということであり，極めて重要である．

例． (1 回のコイン投げ) $\Omega=\{$“表”, “裏”$\}$, $X(\text{表})=1$, $X(\text{裏}=0)$. この例は 1 対 1 なのでやや形式的.
例． (無限回のコイン投げ) $\Omega=\{0,1\}^\infty$, $\omega=\omega_1\omega_2\ldots$, $\omega_i=0$ or 1. $X_n(\omega)=\omega_n$. つまり n 回目を取り出す写像.

注． $C\subset\mathbb{R}$ を任意の Borel 集合とするとき，$X^{-1}(C)\in\mathcal{F}$ ならば，X は $\mathcal{F}$ に関して Borel 可測とテキスト Lamperti(1996) では呼んでいる．この定義には 2 つの σ-加法族が関係している．$\mathbb{R}$ の部分集合族である Borel 集合族と Ω の部分集合族である $\mathcal{F}$ である．同じ場合に $\mathcal{F}$-可測ということがむしろ多いので，どちらの意味か確認が必要である．誤解がなければ，両方省略して単に**可測関数**ということが多い．なお，$X^{-1}(C)$ は C の原像，すなわち，$X(\omega)\in C$ を満たすような値 $\omega\in\Omega$ の集合である (念のため).

定義．積分
$$\int_{\Omega} X(\omega)\mathrm{d}P$$
が存在するとき，X の**期待値**と呼び，$\mathrm{E}(X)$ と表す．

注．

1. $\int X(\omega)\mathrm{d}P$ は関数 X の測度 P でのルベーグ積分の意味である．測度の定義されている空間を引数ではっきりさせるために，$\int X(\omega)\mathrm{d}P(\omega)$ のように表すこともある．また，$\int X(\omega)P(\mathrm{d}\omega)$ のような記法が使われることもあるが，すべて同じ意味である．
2. 積分はルベーグ積分の意味であるから，$\int |X|\mathrm{d}P < \infty$ のときのみ，X は可積分である．例えば $(1/x)\sin x$ の広義積分はルベーグ積分の意味では存在しない．

[**補足**] ルベーグ積分について．通常のリーマン積分が「縦切り」なのに対して，ルベーグ積分は「横切り」である．気分的には，等間隔で横切りにする．そしてその間隔を狭めていくことによる極限操作で積分を定義する．1 次元の場合は，長方形の和で近似するわけだが，高さを先に決めてやり，底辺の長さを確率測度 P で測る形となる．横切りの利点は，例えば有理数の定義関数の積分などが定義できる点にある．

確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数 X が与えられたとき，任意の Borel 集合 $C \subset \mathbb{R}$ に対し，
$$P_X(C) := P(X^{-1}(C)) \tag{1}$$
と定義すると，自然に $\mathbb{R}$ 上の測度 P_X が決まる (測度の「誘導」push-forward)．

問題 4．(1) は，$\mathbb{R}$ 上の確率測度となっていることを確認せよ．

このとき，
$$\mathrm{E}(X) := \int_{\Omega} X(\omega)\mathrm{d}P(\omega) = \int_{\mathbb{R}} t\mathrm{d}P_X(t) \tag{2}$$
が成立する．

例． (サイコロ) $\Omega = \{1, 2, \ldots, 6\}$．$X(\omega) = 2I_{\{\omega=6\}} + I_{\{\omega=5\}}$．6 が出たら 2 円もらえ，5 が出たら 1 円もらえるクジ．左辺は $\omega = 1, \ldots, 6$ と順次動かして $0 \times 1/6 + 0 \times 1/6 + 0 \times 1/6 + 0 \times 1/6 + 1 \times 1/6 + 2 \times 1/6$ と計算．右辺は X の確率分布を先に計算しておき $0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2)$ と計算．

このことを，少し一般的な設定で示すことにする．$(\Omega, \mathcal{F}, P)$ を確率空間，$(\Omega', \mathcal{F}')$ を可測空間とする．可測写像
$$\Phi : \Omega \to \Omega'$$
(すなわち，$\forall A' \in \mathcal{F}', \Phi^{-1}(A') \in \mathcal{F}$) が与えられたとき，$A' \in \mathcal{F}'$ に対し，
$$P_{\Phi}(A') := P(\Phi^{-1}(A))$$

と定義すると，$(\Omega', \mathcal{F}')$ 上の確率測度 P_Φ が決まる．なお，Ω' を $\mathbb{R}$, $\mathcal{F}'$ を Borel 集合族ととれば，(1) で考えた状況になる．

　X' が $(\Omega', \mathcal{F}')$ 上の確率変数のとき，$X = X' \circ \Phi$ を

$$X(\omega) := X'(\Phi(\omega)) \tag{3}$$

と定義すれば，X は $(\Omega, \mathcal{F})$ 上の確率変数になっている．x

問題 5. (3) で定義される $(\Omega, \mathcal{F})$ 上の関数 X は，$(\Omega, \mathcal{F})$ 上の確率変数であることを示せ．

　このとき，次が成り立つ．

定理．(積分の変数変換)

$$\Big(\mathrm{E}(X) :=\Big) \int_\Omega X \mathrm{d}P = \int_{\Omega'} X' \mathrm{d}P_\Phi$$

□

　この定理は２つの積分のうち片方が定積分の値をもてば，もう片方も定積分の値をもち，両者の値が一致することを意味する，

　定理の意味は，直観的には理解しやすいと思われる．証明をとばしてもよいのだが，この定理の証明で使われる論法，すなわち，まず定義関数の場合に証明し，次に非負の単関数の場合に証明し，次に非負の可測関数で証明し，次に一般の可測関数の場合に証明するという手続きは，可測関数に関する定理を証明するときの常套手段であるので，慣れるためにここで証明も見ておくことにする．

証明*．定義より，X は $(\Omega, \mathcal{F})$ 上可測関数である．以下いくつかの段階に分けて証明する．

(第１段階：X' が定義関数の場合)

　ある $A' \in \mathcal{F}'$ に対し，

$$X'(\omega') = \mathbf{1}_{A'}(\omega') := \begin{cases} 1, & \omega' \in A' \\ 0, & \omega' \notin A' \end{cases}$$

の場合を考える．まず，

$$\int_{\Omega'} X'(\omega') \mathrm{d}P_\Phi = \int_{\Omega'} \mathbf{1}_{A'}(\omega') \mathrm{d}P_\Phi = P_\Phi(A').$$

また，

$$\int_\Omega X(\omega) \mathrm{d}P = \int_\Omega \mathbf{1}_{A'}(\Phi(\omega)) \mathrm{d}P = \int_\Omega \mathbf{1}_{\Phi^{-1}(A')}(\omega) \mathrm{d}P = P(\Phi^{-1}(A')).$$

P_Φ の定義より，

$$P_\Phi(A') = P(\Phi^{-1}(A'))$$

だから，

$$\int_{\Omega'} X'(\omega')\mathrm{d}P_\Phi = \int_\Omega X(\omega)\mathrm{d}P.$$

(第 2 段階：X' が非負の可測単関数の場合)

集合 Ω' を $\Omega' = A'_1 + A'_2 + \cdots + A'_n$ と**有限個**の集合の直和にあらわして，

$$X'(\omega') = \sum_{i=1}^n c_i \mathbf{1}_{A'_i}(\omega) \quad (c_i \in \mathbb{R}, c_i \neq c_j\ (i \neq j))$$

の形に表せる関数を単関数という．非負の可測単関数であれば，A'_i $(i = 1, \ldots, n)$ は可測で，$c_i \geq 0$ $(i = 1, \ldots, n)$.　　この場合は定義関数の重みつき $\underline{\text{有限和}}$ を考えれば十分なので，収束に関する面倒なことを考える必要がなく，第 1 段階の結果から，

$$\int_{\Omega'} X'(\omega')\mathrm{d}P_\Phi = \int_\Omega X(\omega)\mathrm{d}P$$

がいえる．(有限和の表し方に依存しないことも細分をとることにより容易にチェックできる．)

(第 3 段階：X' が非負の可測関数の場合)

X' が非負の可測関数とする．このとき，

$$X'_n(\omega') := \begin{cases} \dfrac{k-1}{2^n}, & \dfrac{k-1}{2^n} \leq X'(\omega') < \dfrac{k}{2^n} \quad (k = 1, 2, \ldots, 2^n n) \\ \quad n, & X'(\omega') \geq n \end{cases}$$

と定義すると，$\{X'_n(\omega')\}$ は非負の可測な単関数の単調非減少列で，Ω' の各点 ω' で

$$\lim_{n\to\infty} X'_n(\omega') = X'(\omega')$$

となっている．すると単調収束定理（下の注参照）より，

$$\lim_{n\to\infty} \int_{\Omega'} X'_n(\omega')\mathrm{d}P_\Phi = \int_{\Omega'} X'(\omega')\mathrm{d}P_\Phi.$$

また，

$$X_n(\omega) := X'_n(\Phi(\omega))$$

とおくと，$X_n(\omega)$ は $(\Omega, \mathcal{F})$ 上の可測な非負単関数．$\{X_n(\omega)\}$ は単調非減少列で，Ω の各点 ω で

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

となっている．よって単調収束定理より，

$$\lim_{n\to\infty} \int_\Omega X_n(\omega)\mathrm{d}P = \int_\Omega X(\omega)\mathrm{d}P.$$

また，第2段階の結果より任意の n で，

$$\int_{\Omega'} X_n'(\omega')\mathrm{d}P_\Phi = \int_\Omega X_n(\omega)\mathrm{d}P$$

だから，

$$\lim_{n\to\infty}\int_{\Omega'} X_n'(\omega')\mathrm{d}P_\Phi = \lim_{n\to\infty}\int_\Omega X_n(\omega)\mathrm{d}P$$

となるので，

$$\int_{\Omega'} X'(\omega')\mathrm{d}P_\Phi = \int_\Omega X(\omega)\mathrm{d}P$$

が成立する．

(第4段階：X' が一般の可測関数の場合)

$X' = X'^+ - X'^-$, $X'^+ \geq 0$, $X'^- \geq 0$ と表せる．ここで，X'^+, X'^- はともに可測である．X'^+, X'^- について第3段階と同様に考えればよい． □

注． (ルベーグの単調収束定理) f_n $(n = 1, 2, \ldots)$ が $(\Omega, \mathcal{F})$ 上の非負可測関数列で，各 n で，$f_n(\omega) \leq f_{n+1}(\omega)$, $\forall\omega$, が成り立っているとき，

$$\lim_{n\to\infty}\int_\Omega f_n\mathrm{d}P = \int_\Omega \lim_{n\to\infty} f_n\mathrm{d}P \leq \infty$$

が成立する．(右辺は ∞ でもよい．) □

参考文献

Lamperti, J. W. (1996). Probability: A Survey of the Mathematical Theory, 2nd ed., Wiley, New York.
伊藤清三 (1963). ルベーグ積分入門, 裳華房.

2.2 確率変数列の収束

確率変数列 X_n $(n = 1, 2, 3, \ldots)$ の収束にはいくつかの異なる種類がある．よく使われるものに，概収束，平均収束，確率収束，法則収束がある．それぞれの意味と違いをよく理解しておく必要がある．法則収束は，確率変数というより，本質的には確率分布に関する概念なので後回しにする．

定義． (概収束) ある確率変数 X が存在して，

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

となるとき，確率変数列 X_n は確率変数 X に**概収束** (almost sure convergence) するといい，

$$\lim_{n\to\infty} X_n = X \quad \text{a.s.}$$

のように表す．a.s. は almost sure の略である．全く同じ意味で a.e. (almost everywhere の略) が使われることもある．日本語では**ほとんどいたるところ**という．また，**確率 1 で収束する**ということもある． □

定義からわかるように，概収束は，関数列の (測度 0 の集合を無視した) 各点収束のことである．

各 X_n および X が同一の確率空間上で定義されていると設定するのが測度論的確率論の特徴である．つまり ω がいったん抽出されると，確率変数の値はすべて定まり，ω を所与とする時に $\{X_n(\omega)\}$ は通常の数列と考えることができる．

定義．(確率収束) ある確率変数 X が存在して，任意の $\varepsilon > 0$ に対し，

$$\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0$$

となるとき，確率変数列 X_n は確率変数 X に**確率収束** (convergence in probability) するという． □

定義．(α 次平均収束) ある $\alpha > 0$ に対し，確率変数 X が存在して，

$$\lim_{n\to\infty} \mathrm{E}(|X_n(\omega) - X(\omega)|^{\alpha}) = 0$$

となるとき，確率変数列 X_n は確率変数 X に α 次**平均収束** (convergence in the mean of order α) するという．とくに，$\alpha = 2$ の場合が一番よく現れる．このとき，$\alpha(=2)$ を略して，単に平均収束ということが多い． □

３つの異なる種類の収束には以下の関係がある．

定理．

1) X_n が X に概収束すれば，X_n は X に確率収束する．

2) X_n が X に $\alpha(>0)$ 次平均収束すれば，X_n は X に確率収束する．

□

なお，概収束するからといって平均収束するとは限らず，平均収束するからといって概収束するとは限らない．

証明．

2) (α 次平均収束 $\Rightarrow$ 確率収束)

$P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\})$ を $P(|X_n - X| > \varepsilon)$ と略記することにする．$\varepsilon > 0$ に対し，

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^\alpha}\mathrm{E}(|X_n - X|^\alpha).$$

だから，

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|^\alpha) = 0$$

なら，

$$\forall\varepsilon > 0, \quad \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

1) (概収束 $\Rightarrow$ 確率収束)

$$A_m^\varepsilon := \{\omega : |X_m(\omega) - X(\omega)| \leq \varepsilon\}$$

とおく．すると，$\omega \in \Omega$ が集合　　(後で「下極限」としてあらためて定義する)

$$\bigcup_{n=1}^\infty \bigcap_{m=n}^\infty A_m^\varepsilon$$

に属するということは，$\omega \in \Omega$ に対し，ある正の整数 N が存在してすべての $m \geq N$ に対して

$$|X_m(\omega) - X(\omega)| \leq \varepsilon$$

となることと同じことである．したがって，X_m が X に概収束すれば，任意の $\varepsilon > 0$ に対し，

$$P(\bigcup_{n=1}^\infty \bigcap_{m=n}^\infty A_m^\varepsilon) = 1$$

となる．

$$B_n^\varepsilon := \bigcap_{m=n}^\infty A_m^\varepsilon$$

とおくと，B_n^ε は n に関して (集合列として) 単調増加 ($B_1^\varepsilon \subset B_2^\varepsilon \subset \cdots$) なので，

$$\lim_{n\to\infty} P(B_n^\varepsilon) = P(\lim_{n\to\infty} B_n^\varepsilon) = P(\bigcup_{n=1}^\infty B_n^\varepsilon) = P(\bigcup_{n=1}^\infty \bigcap_{m=n}^\infty A_m^\varepsilon) = 1$$

となる (p. 2 の問題 1 参照)．$B_n^\varepsilon \subset A_n^\varepsilon$ より，

$$\lim_{n\to\infty} P(A_n^\varepsilon) = \lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\}) = 1.$$

□

問題 6．確率収束するが概収束はしない例をあげよ．

慣れないうちは確率収束と概収束の区別はあまり簡単では無いかもしれない．しかし，上の問題を考えた人は，ある種の確率システムの安定性を示したいときに，確率変数 X_n がある定数 c に確率収束することを示したのでは意味がなく，概収束することを示さなければならないことなどが理解できるはずである．

3 独立性

「独立性」が確率論に独自の意味をあたえる概念であることが，Kolmogorov をはじめ，多くの研究者により強調されてきた．

定義． (事象の独立性)
事象 $A_1, A_2, \ldots, A_n$ が独立
$\overset{\mathrm{d}}{\Leftrightarrow}$
任意の空でない $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$ に対して，

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

□

問題 7. (復習)

1) 事象 A_1, A_2, A_3 について，

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

が成立しても，A_1, A_2, A_3 が独立にならない例を与えよ．

2) 事象 A_1, A_2, A_3 について，A_1 と A_2 が独立，A_2 と A_3 が独立，A_3 と A_1 が独立であっても，A_1, A_2, A_3 が独立にならない例を与えよ．

[**補足**] 上の定義で部分集合族をとらなければならないのは，補集合を考えていないからである．次に述べる確率変数の独立性の定義にならい，I_{A_i} を集合 A_i の定義関数とする．確率変数の族として $I_{A_1}, \ldots, I_{A_n}$ がであることと，事象 $A_1, A_2, \ldots, A_n$ が独立であることが同値となる．

定義． (確率変数の独立性)
確率変数 $X_1, X_2, \ldots, X_n$ が独立
$\overset{\mathrm{d}}{\Leftrightarrow}$
$\mathbb{R}$ の任意の Borel 集合 $S_1, S_2, \ldots, S_n$ に対して，n 個の事象

$$A_k := \{\omega : X_k(\omega) \in S_k\} \qquad (k = 1, 2, \ldots, n)$$

が独立．

なお，無限個の確率変数の集合 X_λ, $\lambda \in \Lambda$ (Λ は適当な添字集合) に対しては，任意の有限部分集合 $\{i_1, i_2, \dots, i_k\} \subset \Lambda$ に対して $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ が独立であるとき，独立であるという．無限個の事象 A_λ, X_λ, $\lambda \in \Lambda$ の独立性については，それらの定義関数 I_{A_λ} の独立性として定義すればよいが，これは任意の有限個の独立性による定義と同じこととなる． □

確率空間 $(\Omega, \mathcal{F}, P)$ 上の n 個の確率変数 $X_1, X_2, \dots, X_n$ を考える．このとき，

$$Z(\omega) := (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$$

とおいて，$\mathbb{R}^n$ の Borel 集合 C に対し，

$$P_Z(C) := P(Z^{-1}(C))$$

と定義すると，P_Z は，$\mathbb{R}^n$ 上の確率測度となる (メモ p.4 参照)．P_Z を $X_1, X_2, \dots, X_n$ の確率分布という．多次元であることを強調するときは，**同時確率分布** という．

定義． (多次元の分布関数)

$$F_{X_1,X_2,\dots,X_n}(t_1, t_2, \dots, t_n) := P(\{\omega : X_1(\omega) \le t_1, X_2(\omega) \le t_2, \dots, \text{and } X_n(\omega) \le t_n\})$$

を $X_1, \dots, X_n$ の **分布関数** または **確率分布関数** と呼ぶ．多次元であることを強調するときは **同時確率分布関数** と呼ぶ． □

X_i の確率分布の分布関数を $F_{X_i}(t_i)$ とすると，

$$F_{X_i}(t_i) = \lim_{s_1 \to \infty} \cdots \lim_{s_{i-1} \to \infty} \lim_{s_{i+1} \to \infty} \cdots \lim_{s_n \to \infty} F_{X_1,X_2,\dots,X_n}(s_1, \dots, s_{i-1}, t_i, s_{i+1}, \dots, s_n)$$

が成り立つことはすぐにわかる．このとき，$F_{X_i}(t_i)$ を**周辺分布関数** という．

定理．
確率変数 $X_1, X_2 \dots, X_n$ が独立
$\Leftrightarrow$

$$\forall (t_1, t_2, \dots, t_n) \in \mathbb{R}^n, \quad F_{X_1,X_2,\dots,X_n}(t_1, t_2, \dots, t_n) = \prod_{i=1}^{n} F_{X_i}(t_i)$$

□

(証明略)

定義 (密度関数)
確率変数 X の分布関数 $F_X(t)$ に対し，

$$\forall t \in \mathbb{R}, \quad F_X(t) = \int_{-\infty}^{t} f_X(t) \mathrm{d}t$$

となる $\mathbb{R}$ 上の関数 $f_X(t)$ が存在するとき，f_X を X の確率分布の**密度関数** または **確率密度関数** とよぶ. $\square$

よく知られているように，確率分布の分布関数は必ず存在するが，密度関数は存在するとは限らない．密度関数が存在するような分布は，ルベーグ測度に関して絶対連続な分布である (以下の Radon-Nikodym の定理を参照).

定義 (多次元の密度関数)

確率変数 $X_1, X_2, \ldots, X_n$ の同時分布関数 $F_{X_1,X_2,\ldots,X_n}(t)$ に対し，

$$\forall(t_1, t_2, \ldots, t_n) \in \mathbb{R}^n,$$
$$F_{X_1,X_2,\ldots,X_n}(t_1, t_2, \ldots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f_{X_1,X_2,\ldots,X_n}(t_1, t_2, \ldots, t_n)\mathrm{d}t_1\mathrm{d}t_2 \cdots \mathrm{d}t_n$$

となる $\mathbb{R}$ 上の関数 $f_{X_1,X_2,\ldots,X_n}$ が存在するとき，$f_{X_1,X_2,\ldots,X_n}$ を $X_1, X_2, \ldots, X_n$ の確率分布の**密度関数** または **確率密度関数** あるいはとよぶ．多次元であることを強調するときは **同時確率密度関数** と呼ぶこともある. $\square$

定理．

確率変数 X_i $(i = 1, \ldots, n)$ のしたがう確率分布が，それぞれ確率密度関数 $f_{X_i}(t)$ $(i = 1, \ldots, n)$ をもつとする．このとき，

$\prod_{i=1}^{n} f_{X_i}(t_i)$ が $X_1, \ldots, X_n$ の同時密度関数 $\Leftrightarrow$ $X_1, \ldots, X_n$ は独立. $\square$

(証明略)

問題 8. (復習)

1) 確率変数 X と Y の相関係数が 0 でも独立とは限らないことを例をあげて説明せよ（相関係数の定義については省略する）.

2) 確率変数 X と Y それぞれの分散が存在するとき，X と Y が独立であれば，X と Y とは無相関であることを示せ.

定理．

X と Y はそれぞれ確率分布関数 F, G の確率分布に独立にしたがう確率変数とする．このとき，$X + Y$　の確率分布関数は

$$P(X + Y \leq x) = \int_{-\infty}^{\infty} F(x - t)\mathrm{d}G(t) = \int_{-\infty}^{\infty} G(x - t)\mathrm{d}F(t)$$

となる. $\square$

証明．x に対し，関数 $g_x(u,v)$ を

$$g_x(u,v) := \begin{cases} 1, & \text{if} \quad u+v \le x \\ 0, & \text{otherwise} \end{cases}$$

と定義する．F と G に対応する確率測度をそれぞれ P_X, P_Y とおく．また, P_X と P_Y との直積測度 (product measure) を $P_{X,Y}$ とおくと，Fubini の定理より，

$$\begin{aligned} \int_{\mathbb{R}^2} g_x \mathrm{d}P_{X,Y} &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g_x(u,v) \mathrm{d}P_X(u) \right) \mathrm{d}P_Y(v) \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g_x(u,v) \mathrm{d}P_Y(v) \right) \mathrm{d}P_X(u) \end{aligned}$$

となる．よって示せた． □

系．X と Y はそれぞれ確率分布関数 F, G の確率分布に独立にしたがう確率変数とする．X の確率分布が確率密度関数 $f(u)$ をもつとき，$X+Y$ のしたがう確率分布は確率密度関数

$$\int_{-\infty}^{\infty} f(t-v) \mathrm{d}G(v)$$

をもつ． □

証明．Fubini の定理より，

$$\begin{aligned} \int_{-\infty}^{t} \left(\int_{-\infty}^{\infty} f(u-v) \mathrm{d}G(v) \right) \mathrm{d}u &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{t} f(u-v) \mathrm{d}G(v) \right) \mathrm{d}u \\ &= \int_{-\infty}^{\infty} F(t-v) \mathrm{d}G(v) = P(X+Y \le t) \end{aligned}$$

よりいえる． □

[**補足**] Fubini の定理は Lamperti の Appendix 参照．より詳しい説明は Billingsley “Probability and Measure”, 3rd ed., Wiley, の Section 18 を見るとよい．2 変数関数 $f(x,y)$ が (同時) 可測ならば，すべての y について y を固定した時に x について可測であり，x で積分した後は y について可測である．Lamperti の 1.4 章の条件つき期待値は後回し．1.5 章のコルモゴロフの拡張定理は略.

4 大数の法則

大数の法則には，弱法則と呼ばれるものと，強法則と呼ばれるものがあり，弱法則が確率収束に，強法則が概収束に対応する．

4.1　大数の弱法則

定理． (復習) (Chebyshev の不等式)．X は分散 $\mathrm{var}(X)$ をもつ確率変数とする．任意の $a > 0$ に対し，

$$P(|X - \mathrm{E}(X)| \geq a) \leq \frac{\mathrm{var}(X)}{a^2}$$

□

(証明略)

大数の弱法則 (の基本形) は Chebyshev の不等式を使って容易に証明できる．これは 2 次平均収束が確率収束を含意することにもとづく．あとで見る大数の強法則の証明と比較せよ．

定理． (大数の弱法則の基本形,「2 次収束 ⇒ 確率収束」を用いる).

$X_1, X_2, \ldots$ が，期待値 μ, 分散 σ^2 の同一の分布に独立にしたがう確率変数列であるとき，任意の $\varepsilon > 0$ に対し，

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

が成立する．□

これは，ひとことで言えば，サンプル平均が期待値に確率収束するということである．

証明．$S_n := X_1 + \cdots X_n$ とおくと，S_n/n は期待値 μ, 分散 σ^2/n の確率変数．Chebyshev の不等式により，

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

□

大数の弱法則にはさまざまな一般化がある．一般化された定理も，確率収束を扱っている限り，大数の弱法則と呼ばれるのが普通である．Chebyshev の不等式は分散だけを使って上界を与えるため，確率変数列の独立性ではなく無相関性を仮定するだけで十分であることを利用した一般化が次の定理である．

定理．

X_i $(i = 1, 2, \ldots)$ はそれぞれ期待値 μ_i，分散 σ_i をもつ互いに相関の無い確率変数列とする．

このとき，

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = 0$$

であれば，任意の $\varepsilon > 0$ に対して，

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\cdots+X_n}{n} - \frac{\mu_1+\cdots+\mu_n}{n}\right| \geq \varepsilon\right) = 0$$

が成立する． □

問題 9. 上の定理を証明せよ．

4.2 大数の強法則

事象の系列 $A_1, A_2, \dots$ に関して考察するとき，集合

$$\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n, \quad \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n$$

を扱う必要がしばしば現れる．これらは，それぞれ (集合列の) 上極限，下極限と呼ばれ，

$$\limsup_{n\to\infty} A_n := \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n, \quad \liminf_{n\to\infty} A_n := \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n$$

のように表される．

$\omega \in \limsup_{n\to\infty} A_n$ は，ω を含む A_n が無限個あることを意味する．また，$\omega \in \liminf_{n\to\infty} A_n$ は，ω を含まない A_n がたかだか有限個であることを意味する．自分で考えて納得してほしい．

また $\limsup_{n\to\infty} A_n$ を "A_n *i.o.*" (*i.o.* は infinitely often) と表すことも多い．補集合をとる操作との関係では，定義より

$$(\limsup_{n\to\infty} A_n)^C = \liminf_{n\to\infty} A_n^C$$

となることが容易にわかる．

問題 10. 事象 A の定義関数を $\mathbf{1}_A(\omega)$ で表す．すると，

$$\limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega), \quad \liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega) \tag{4}$$

(この $\limsup$, $\liminf$ は通常の (実数列の) 上極限，下極限の意味) はそれぞれ，

$$\limsup_{n\to\infty} A_n, \quad \liminf_{n\to\infty} A_n$$

の定義関数になっていること，すなわち，

$$\limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega) = \mathbf{1}_{\limsup_{n\to\infty} A_n}(\omega), \quad \liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega) = \mathbf{1}_{\liminf_{n\to\infty} A_n}(\omega)$$

が成立することを示せ．

問題 11．$P(\limsup_{n\to\infty} A_n) \geq \limsup_{n\to\infty} P(A_n)$, $P(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} P(A_n)$ を示せ．

定理 (Borel-Cantelli lemma).
　$A_1, A_2, \ldots$ を事象の系列とし，

$$B := \limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

とおく．このとき，

1) $\sum_{n=1}^{\infty} P(A_n) < \infty$ ならば $P(B) = 0$.

2) 事象 $A_1, A_2, \ldots$ が独立で，かつ $\sum_{n=1}^{\infty} P(A_n) = \infty$ であれば $P(B) = 1$.

□

証明．1) 任意の k に対し，

$$P(B) = P(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n) \leq P(\bigcup_{n=k}^{\infty} A_n) \leq \sum_{n=k}^{\infty} P(A_n).$$

ここで，$\sum_{n=1}^{\infty} P(A_n) < \infty$ であるから，$\lim_{k\to\infty} \sum_{n=k}^{\infty} P(A_n) = 0$．したがって，$P(B) = 0$ がいえる．

2) 事象系列 C_i $(i = 1, 2, \ldots)$ で，任意の i について $P(C_i) = 1$ であれば，$P(\cap_{i=1}^{\infty} C_i) = 1$ がいえる．したがって，任意の k について

$$P(\bigcup_{n=k}^{\infty} A_n) = 1$$

が成立する事を示せば十分．しかるに，任意の $K > k$ に対し，

$$1 - P(\bigcup_{n=k}^{\infty} A_n) \leq 1 - P(\bigcup_{n=k}^{K} A_n) = P((\bigcup_{n=k}^{K} A_n)^{\mathrm{c}}) = P(\bigcap_{n=k}^{K} {A_n}^{\mathrm{c}}) = \prod_{n=k}^{K} (1 - P(A_n)).$$

最後の等号は A_n $(n = k, \ldots, K)$ が独立であれば ${A_n}^{\mathrm{c}}$ $(n = k, \ldots, K)$ も独立であることによる．ここで，

$$\log\{\prod_{n=k}^{K} (1 - P(A_n))\} = \sum_{n=k}^{K} \log\{1 - P(A_n)\} \leq -\sum_{n=k}^{K} P(A_n).$$

$\lim_{K\to\infty} \sum_{n=k}^{K} P(A_n) = \infty$ より $\lim_{K\to\infty} \log\{\prod_{n=k}^{K} (1 - P(A_n))\} = -\infty$ だから，
$\lim_{K\to\infty} \prod_{n=k}^{K} (1 - P(A_n)) = 0$．よって $1 - P(\bigcup_{n=k}^{\infty} A_n) = 0$ がいえる． □

問題 12. Borel-Cantelli lemma の 2) で，$A_1, A_2, \ldots$ の独立性の仮定がないと，$\sum_{n=1}^{\infty} P(A_n) = \infty$ であっても $P(B) = 1$ となるとは限らないことを例をあげて示せ．

簡単のため，やや強い仮定 (４次モーメントの存在) をおいた上で，大数の強法則を証明する．

定理 (４次モーメントの存在を仮定した大数の強法則).
$X_1, X_2, \ldots$ を，独立に同一の分布にしたがう確率変数列とする．X_i の分布の４次モーメントは有限であると仮定し，期待値と分散をそれぞれ μ, σ^2 とおく．このとき，

$$P\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1.$$

□

証明.

$$\mathrm{E}\left(\{\sum_{k=1}^{n}(X_k - \mu)\}^4\right) = n\mathrm{E}\left((X_k - \mu)^4\right) + \binom{n}{2}\binom{4}{2}\sigma^4$$

より，ある $C > 0$ が存在して，

$$\mathrm{E}\left(\{\sum_{k=1}^{n}(X_k - \mu)\}^4\right) \le Cn^2.$$

ここで，４次モーメントに基づく (一般化した) Chebyshev の不等式

$$P(|Y| \ge b) \le \frac{\mathrm{E}(|Y|^k)}{b^k} \qquad (b > 0, \quad k > 0)$$

を用いると，

$$P\left(\left|\sum_{k=1}^{n}(X_k - \mu)\right| \ge \varepsilon n\right) \le \frac{\mathrm{E}\left(\left|\sum_{k=1}^{n}(X_k - \mu)\right|^4\right)}{(\varepsilon n)^4} \le \frac{Cn^2}{(\varepsilon n)^4} = \frac{C}{\varepsilon^4 n^2}.$$

したがって，

$$\sum_{n=1}^{\infty} P\left(\left|\sum_{k=1}^{n}(X_k - \mu)\right| \ge \varepsilon n\right) < \infty.$$

Borel-Cantelli lemma より，ω を与えたとき $\left|\sum_{k=1}^{n}(X_k - \mu)\right| \ge \varepsilon n$ が成り立つ n が有限個しかないことが確率 1 でいえる ($A_n := \{\omega : \left|\sum_{k=1}^{n}(X_k(\omega) - \mu)\right| \ge \varepsilon n\}$ とおけばよい).
よって，任意の $\varepsilon > 0$ に対し，

$$P\left(\limsup_{n\to\infty} \frac{\left|\sum_{k=1}^{n}(X_k - \mu)\right|}{n} \le \varepsilon\right) = 1. \tag{5}$$

$\epsilon_i = 1/i$ $(i = 1, 2, 3, \ldots)$ とおくと，各 $\epsilon = \epsilon_i$ について，(5) が成立する．可算個の確率 1 の事象の intersection の確率は 1 であることは前に (Borel-Cantelli lemma の 2 の証明中で) 見たので，

$$P\left(\limsup_{n\to\infty} \frac{\left|\sum_{k=1}^{n}(X_k - \mu)\right|}{n} = 0\right) = P\left(\bigcap_{i=1}^{\infty}\left\{\omega : \limsup_{n\to\infty} \frac{\left|\sum_{k=1}^{n}(X_k(\omega) - \mu)\right|}{n} \le \frac{1}{i}\right\}\right) = 1.$$

□

問題 13. $X_1, X_2, \ldots$ を独立に $[0,1]$ 上の一様分布にしたがう確率変数列とする．このとき，

$$\lim_{n\to\infty} (X_1 X_2 \cdots X_n)^{1/n} = \mathrm{e}^{-1} \quad \text{a. s.}$$

を示せ．

４次モーメントの存在を仮定しない，大数の強法則を証明抜きで示しておく．(Lamperti, Section 9, Theorem 4.)

定理 (大数の強法則).
$X_1, X_2, \ldots$ を，期待値 μ をもつ同一の分布に独立にしたがう確率変数列とする．このとき，

$$P\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1.$$

□

(証明略)

[**補足**] 独立同一分布の設定で $E|X_1| = \infty$ ならば $\limsup_n |\bar{X}_n| = \infty$ a.s. となることが割合容易に示されるので，この意味で期待値 μ の存在は必要条件でもある．これについても Lamperti, Section 9, Theorem 4 を参照．Lamperti の 2 章にはこの他にも重複対数の法則 (Law of the iterated logarithm) も述べられているが，ノートでは省略する．

5 確率分布の弱収束

確率論では，確率分布の列がある分布 (たとえば正規分布や Poisson 分布) に収束することが重要であることが多い．有名な中心極限定理は，その典型的な例である．確率分布全体の空間は無限次元なので，収束の定義の仕方はいろいろある．ここでは，いろいろな確率分布の「収束」のなかでも，非常に良く使われて，いろいろな重要な定理を理解するうえで不可欠の弱収束と呼ばれる概念を扱う．

定義. P_n $(n=1,2,\ldots)$, P をそれぞれ $\mathbb{R}^k$ 上の確率測度とする．$\mathbb{R}^k$ 上の任意の有界連続関数 f に対し，

$$\lim_{n\to\infty}\int_{\mathbb{R}^k} f\mathrm{d}P_n = \int_{\mathbb{R}^k} f\mathrm{d}P$$

が成立するとき，P_n は P に**弱収束** (weak convergence, P_n converges weakly to P) するという．また，弱収束と同じ意味で**分布収束** (convergence in distribution) または**法則収束** (convergence in law) ということもある． □

注． Euclid 空間とは限らない一般の距離空間上の確率測度の弱収束も，同様に定義される．

P_n が P に弱収束するとき，

$$P_n \rightsquigarrow P$$

と表す．全く同じ意味で $P_n \Rightarrow P$ という表記も良く使われるが以下では採用しない．

定理． P_n $(n=1,2,\ldots)$, P を $\mathbb{R}^k$ 上の確率測度，F_n $(n=1,2,\ldots)$, F をそれぞれに対応する分布関数とする．F の任意の連続点 x で，

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

が成立することは，P_n が P に弱収束するための必要十分条件である． □

(証明は次回)

分布関数 F_n に対応する確率測度 P_n が，分布関数 F に対応する確率測度 P に弱収束するとき，F_n は F に弱収束，分布収束，あるいは法則収束するいい，

$$F_n \rightsquigarrow F$$

と表す．また，確率変数 X_n の確率分布 P_n が確率変数 X の確率分布 P に弱収束するとき，X_n は X に弱収束，分布収束，あるいは法則収束するといい，

$$X_n \rightsquigarrow X$$

のように表す．また，

$$X_n \rightsquigarrow P, \quad X_n \rightsquigarrow F$$

のような表記も用いる．

例．$\mathbb{R}$ 上の分布関数を考える．

$$F_n(x) := \begin{cases} 1, & x \geq 1/n \\ 0, & x < 1/n \end{cases}, \quad F(x) := \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

のとき，$x = 0$ は F の連続点ではない．このとき，

$$\lim_{n\to\infty} F_n(0) = 0 \neq F(0) = 1.$$

だが，$F_n \rightsquigarrow F$ は成立している． □

問題 14．区間 $[0,1]$ 上の確率測度を考える．P_n を $(n+1)$ 個の点 $0, 1/n, 2/n, \ldots, n/n = 1$ にそれぞれ確率 $1/(n+1)$ を与える（離散型）確率測度とする．$n \to \infty$ で，P_n は $[0,1]$ 上の Lebesgue 測度に弱収束することを示せ．

参考: レポート課題 (2007年の駒木先生の時のもの)

問題 1. $A_1, A_2, \ldots \in \mathcal{F}$ で，かつ $A_1 \subset A_2 \subset \cdots$ のとき，$\lim_{i\to\infty} A_i = \cup_{i=1}^{\infty} A_i$ と定義すると，

$$\lim_{i\to\infty} P(A_i) = P(\lim_{i\to\infty} A_i)$$

が成立することを示せ．同様に，$A_1, A_2, \ldots \in \mathcal{F}$ で，かつ $A_1 \supset A_2 \supset \cdots$ のとき，$\lim_{i\to\infty} A_i = \cap_{i=1}^{\infty} A_i$ と定義すると，

$$\lim_{i\to\infty} P(A_i) = P(\lim_{i\to\infty} A_i)$$

が成立することを示せ.

問題 6. 確率収束するが概収束はしない例をあげよ.

問題 10. 事象 A の定義関数を $\mathbf{1}_A(\omega)$ で表す．すると，

$$\limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega), \quad \liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega) \tag{6}$$

(この $\limsup$, $\liminf$ は通常の (実数列の) 上極限，下極限の意味) はそれぞれ，

$$\limsup_{n\to\infty} A_n, \quad \liminf_{n\to\infty} A_n$$

の定義関数になっていること，すなわち，

$$\limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega) = \mathbf{1}_{\limsup_{n\to\infty} A_n}(\omega), \quad \liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega) = \mathbf{1}_{\liminf_{n\to\infty} A_n}(\omega)$$

が成立することを示せ.

問題 11. $P(\limsup_{n\to\infty} A_n) \geq \limsup_{n\to\infty} P(A_n)$, $P(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} P(A_n)$ を示せ.

以上

参考：レポート課題 (2008 年竹村)

問題 1. 任意個 (非可算個でもよい) の σ-field の積 (共通集合) が σ-field となることを示せ．これにより Ω の任意の部分集合族 $\mathcal{A}$ に対して $\mathcal{A}$ を含む最小の σ-field が存在することを示せ.

問題 2. Lamperti の Chapter 1, Section 1, Problem 1. (Ω が可算の場合の問題)

問題 3. Lamperti の Chapter 1, Section 2, Problem 4. ($\lim X_n$ の可測性．ただし "if X_n converges almost surery" の条件は不要です．)

問題 4. X と Y が独立であれば，X の任意の可測関数 $f(X)$ と Y の任意の可測関数 $g(Y)$ は独立であることを示せ.

問題 5. $A_m^\varepsilon := \{\omega : |X_m(\omega) - X(\omega)| \le \varepsilon\}$ とおく．X_n が X に概収束することと，すべての $\epsilon > 0$ に対して

$$P(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^\varepsilon) = 1$$

となることの同値性を証明せよ.

p. 20 の定理の証明．簡単のため 1 次元 Euclid 空間 $\mathbb{R}$ の場合に限って証明を与える．
(十分性) 定理の条件が成立すると仮定する．任意の $\varepsilon > 0$ をとり，

$$g_\varepsilon(t) := \begin{cases} 0 & (t \ge x + \varepsilon), \\ \dfrac{-t + \varepsilon + x}{\varepsilon} & (x < t < x + \varepsilon), \\ 1 & (t \le x) \end{cases}$$

とおくと，g_ε は有界連続関数であるから，

$$\lim_{n\to\infty} \int_{\mathbb{R}} g_\varepsilon \mathrm{d}P_n = \int_{\mathbb{R}} g_\varepsilon \mathrm{d}P.$$

また，不等式

$$F_n(x) = \int_{-\infty}^{x} g_\varepsilon \mathrm{d}P_n \le \int_{-\infty}^{\infty} g_\varepsilon \mathrm{d}P_n$$

と

$$\int_{-\infty}^{\infty} g_\varepsilon \mathrm{d}P \le F(x + \varepsilon)$$

は明らかに成立する．これらより，

$$\limsup_{n\to\infty} F_n(x) \le \lim_{n\to\infty} \int_{\mathbb{R}} g_\varepsilon \mathrm{d}P_n = \int_{\mathbb{R}} g_\varepsilon \mathrm{d}P \le F(x + \varepsilon).$$

$\varepsilon \downarrow 0$ の極限をとると，$F(x)$ は右連続なので，

$$\limsup_{n\to\infty} F_n(x) \leq F(x).$$

この不等式は x が連続点でなくても成立することに注意する．

　また，

$$f_\varepsilon(t) := \begin{cases} 0 & (t \geq x), \\ \dfrac{-t+x}{\varepsilon} & (x-\varepsilon < t < x), \\ 1 & (t \leq x-\varepsilon) \end{cases}$$

とおくと，f_ε は有界連続関数であるから，

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_\varepsilon \mathrm{d}P_n = \int_{\mathbb{R}} f_\varepsilon \mathrm{d}P.$$

また，不等式

$$F_n(x) \geq \int_{-\infty}^{\infty} f_\varepsilon \mathrm{d}P_n$$

と

$$\int_{-\infty}^{\infty} f_\varepsilon \mathrm{d}P \geq F(x-\varepsilon)$$

は明らかに成立する．これらより，

$$\liminf_{n\to\infty} F_n(x) \geq \lim_{n\to\infty} \int_{\mathbb{R}} f_\varepsilon \mathrm{d}P_n = \int_{\mathbb{R}} f_\varepsilon \mathrm{d}P \geq F(x-\varepsilon).$$

$\varepsilon \downarrow 0$ の極限をとると，

$$\liminf_{n\to\infty} F_n(x) \geq F(x-).$$

x が連続点であれば，$F(x-) = F(x)$ であるから，

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

(必要性)

　$\varepsilon > 0$ に対し，a, b $(a < b)$ を

$$F(a) < \varepsilon, \quad F(b) > 1-\varepsilon$$

を満たす F の連続点とする．$F(x-) \neq F(x)$ となる不連続点は高々可算無限個 (p.3 問題 2 参照) なので，必ずこのような a, b をとることができる．

　a, b は連続点なので，n を十分大きくとると，

$$F_n(a) < 2\varepsilon, \quad F_n(b) > 1-2\varepsilon$$

とできる．

f を $\mathbb{R}$ 上の有界連続関数とする．f は $[a,b]$ 上で一様連続なので，a, b 間を適当な個数に分割して，$a = a_0 < a_1 < a_2 < \cdots < a_k = b$ ととり，

$$\sup_{x\in[a_i,a_{i+1}]} |f(x) - f(a_i)| < \varepsilon \quad (i = 0, 1, \ldots, k-1)$$

かつ a_i $(i = 0, 1, \ldots, k)$ はすべて F の連続点となるようにできる．

すると，

$$\begin{aligned}
&\left| \int_{\mathbb{R}} f \mathrm{d}P_n - \int_{\mathbb{R}} f \mathrm{d}P \right| \\
&\leq \left| \int_{(-\infty,a]} f \mathrm{d}P_n - \int_{(-\infty,a]} f \mathrm{d}P \right| + \sum_{i=0}^{k-1} \left| \int_{(a_i,a_{i+1}]} f \mathrm{d}P_n - \int_{(a_i,a_{i+1}]} f \mathrm{d}P \right| \\
&\quad + \left| \int_{(b,\infty)} f \mathrm{d}P_n - \int_{(b,\infty)} f \mathrm{d}P \right|. \qquad (7)
\end{aligned}$$

$\sup_{x\in\mathbb{R}} |f(x)| < C$ とすると，第１項，第３項は，それぞれ $3C\varepsilon$ 以下である．第２項は，

$$\begin{aligned}
&\sum_{i=0}^{k-1} \left| \int_{(a_i,a_{i+1}]} f \mathrm{d}P_n - \int_{(a_i,a_{i+1}]} f \mathrm{d}P \right| \\
&= \sum_{i=0}^{k-1} \Bigg| \int_{(a_i,a_{i+1}]} f \mathrm{d}P_n - f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P_n + f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P_n \\
&\qquad - \int_{(a_i,a_{i+1}]} f \mathrm{d}P + f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P - f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P \Bigg| \\
&\leq \sum_{i=0}^{k-1} \left| \int_{(a_i,a_{i+1}]} f \mathrm{d}P_n - f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P_n \right| + \sum_{i=0}^{k-1} \left| f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P_n - f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P \right| \\
&\quad + \sum_{i=0}^{k-1} \left| \int_{(a_i,a_{i+1}]} f \mathrm{d}P - f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P \right| \\
&\leq \sum_{i=0}^{k-1} \left| \int_{(a_i,a_{i+1}]} f \mathrm{d}P_n - f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P_n \right| \\
&\quad + \sum_{i=0}^{k-1} \left\{ |f(a_i)| \Big| \{F_n(a_{i+1}) - F_n(a_i)\} - \{F(a_{i+1}) - F(a_i)\} \Big| \right\} \\
&\quad + \sum_{i=0}^{k-1} \left| \int_{(a_i,a_{i+1}]} f \mathrm{d}P - f(a_i) \int_{(a_i,a_{i+1}]} \mathrm{d}P \right| \qquad (8)
\end{aligned}$$

ここで，(8) の第１項，第３項はそれぞれ ε で抑えられる．また，k は固定しているので，n を十分大きくとれば，(8) の第２項も ε で抑えられる．まとめると，(7) は，$(6C+3)\varepsilon$ で抑えられることになる． □

定理. ある確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数列 X_n $(n = 1, 2, 3, \ldots)$ が確率変数 X に確率収束するとする．P_n を X_n の確率分布，P を X の確率分布とするとき，P_n $(n = 1, 2, 3, \ldots)$ は P に弱収束する．

証明．P_n の分布関数を F_n, P の分布関数を F とする．任意の $\varepsilon > 0$ に対し

$$\begin{aligned}\{\omega : X_n(\omega) \leq x\} \subset &\Big(\{\omega : X(\omega) \leq x + \varepsilon\} \cap \{X_n(\omega) \leq x\}\Big) \\ &\cup \Big(\{\omega : X(\omega) > x + \varepsilon\} \cap \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}\Big) \\ &\subset \{\omega : X(\omega) \leq x + \varepsilon\} \cup \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}\end{aligned}$$

だから

$$F_n(x) = P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| \geq \varepsilon)$$

ある N が存在して，任意の $n \geq N$ に対し，

$$P(|X_n - X| \geq \varepsilon) \leq \varepsilon$$

とできるから，任意の $n \geq N$ に対し，

$$F_n(x) \leq F(x + \varepsilon) + \varepsilon.$$

また，

$$\{\omega : X_n(\omega) \leq x\} \supset \{\omega : X(\omega) \leq x - \varepsilon\} \cap \{\omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\}$$

より，

$$\begin{aligned}F_n(x) =& P(X_n \leq x) \geq P\Big(\{\omega : X(\omega) \leq x - \varepsilon\} \cap \{\omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\}\Big) \\ \geq& P(\{\omega : X(\omega) \leq x - \varepsilon\}) - P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\})\end{aligned}$$

だから，任意の $n \geq N$ に対し，

$$F_n(x) \geq F(x - \varepsilon) - \varepsilon.$$

よって

$$F(x - \varepsilon) - \varepsilon \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x + \varepsilon) + \varepsilon$$

$\varepsilon > 0$ は任意に小さくとれるので x が F の連続点であれば,

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

□

Lévy 距離と呼ばれる距離を定義することにより，$\mathbb{R}$ 上確率測度全体の空間は，距離空間となる．

定義 (Lévy 距離). P, Q を $\mathbb{R}$ 上の確率測度, F, G を対応する確率分布関数とする. このとき,

$$L(F,G) := \inf\{h > 0 : F(x-h) - h \le G(x) \le F(x+h) + h \text{ for all } x\}$$

を F と G 間の **Lévy 距離** (Lévy distance) と呼ぶ. なお, P と Q の Lévy 距離ともいう. □

問題 15*. L は距離の公理を満たすことが知られている. 距離の公理のうち 3 角不等式 $L(P,Q) \le L(P,R) + L(R,Q)$ が成立することを示せ.

Lévy 距離は弱収束の概念と結びついている. それを示すのが次の定理である.

定理. P_n $(n = 1, 2, \ldots)$, P はそれぞれ $\mathbb{R}$ 上確率測度であるとする. このとき,

$$P_n \rightsquigarrow P \quad \text{if and only if} \quad L(P_n, P) \to 0.$$

(証明略)

上の定理から, 確率分布の弱収束について議論するときには, 距離空間に対する一般論がそのまま適用できることになる. Lévy 距離以外にも確率分布間の距離にはいろいろなものがある. ここでは, 全変動距離 (total variation distance) と呼ばれるものについて見ておくことにする. この他にも, Helliger 距離と呼ばれる量や, 距離ではないが, Kullback-Leibler ダイバージェンスと呼ばれる量に基づいた収束を考えることも多い.

定義 (全変動距離 total variation distance). 可測空間 $(\Omega, \mathcal{F})$ 上の確率測度 P, Q に対し,

$$d(P,Q) := \sup_{A \in \mathcal{F}} |P(A) - Q(A)|$$

を P, Q 間の**全変動距離** (total variation distance) と呼ぶ. 確率分布関数 F, G に対して, $d(F,G)$ は, F, G にそれぞれ対応する $\mathbb{R}$ 上の確率測度間の全変動距離を表す. □

問題 16. d は距離の公理を満たすことが知られている. 距離の公理のうち 3 角不等式 $d(P,Q) \le d(P,R) + d(R,Q)$ が成立することを示せ.

確率分布間のいろいろな距離は, それぞれ意味が違うので注意が必要である.

問題 17. 確率分布関数列 F_n $(n = 1, 2, \ldots)$ が確率分布関数 F に Lévy 距離の意味では収束するのに, 全変動距離の意味では収束しない例を示せ. (ヒント: Lévy 距離の意味で収束するということは, 弱収束するということなので, Lévy 距離そのものを直接扱う必要はない. 簡単な例が作れる).

逆に，全変動距離の意味で収束すれば Lévy 距離の意味でも収束することが知られている．

以下の 2 つの定理は確率測度の弱収束に関する基本的で重要な結果である．

定理． (Helly-Bray の定理) $\{F_n\}(n=1,2,3,\ldots)$ を任意の確率分布関数列とする．適当な $\{n\}$ の部分列 $\{n_i\}$ $(i=1,2,3,\ldots)$ と $\mathbb{R}$ 上の単調非減少で右連続な関数 F $(0 \le F \le 1)$ をとることにより，F の任意の連続点 x で,

$$\lim_{i\to\infty} F_{n_i}(x) = F(x)$$

を満たすようにできる． □

証明．証明は対角線論法による．

すべての有理数を一列に並べ，$r_1, r_2, r_3, \ldots$ とおく．$\{F_n(r_1)\}$ $(n=1,2,3,\ldots)$ は有界な数列なので，$\{n\}$ のある部分列 $\{n(1,j)\}$ をとることにより，数列 $\{F_{n(1,j)}(r_1)\}$ $(j=1,2,3,\ldots)$ が収束するようにできる．ここで,

$$L(r_1) = \lim_{j\to\infty} F_{n(1,j)}(r_1)$$

とおく．

次に，数列 $\{F_{n(1,j)}(r_2)\}$ $(j=1,2,3,\ldots)$ を考える．$\{F_{n(1,j)}(r_2)\}$ $(j=1,2,3,\ldots)$ は有界な数列なので，$\{n(1,j)\}$ のある部分列 $\{n(2,j)\}$ をとることにより，数列 $\{F_{n(2,j)}(r_2)\}$ $(j=1,2,3,\ldots)$ が収束するようにできる．この収束先を

$$L(r_2) = \lim_{j\to\infty} F_{n(2,j)}(r_2)$$

とおく．

次に，数列 $\{F_{n(2,j)}(r_3)\}$ $(j=1,2,3,\ldots)$ を考え，$\{n(2,j)\}$ の部分列 $\{n(3,j)\}$ を，数列 $\{F_{n(3,j)}(r_3)\}$ $(j=1,2,3,\ldots)$ が収束するようにとり，$L(r_3) = \lim_{j\to\infty} F_{n(3,j)}(r_3)$ とおく．

以下同様に手続きを繰り返して，部分列を次々と構成していく．このとき，$\{n\}$ の部分列 $\{n(j,j)\}$ $(j=1,2,3,\ldots)$ を考えると，任意の r_i に対し,

$$\lim_{j\to\infty} F_{n(j,j)}(r_i) = L(r_i)$$

となることがわかる．L は，有理数上の単調非減少関数になる．

ここで,

$$F(x) := \inf_{r>x} L(x)$$

とおくと，F は実数上の単調非減少関数となる．明らかに $0 \le F \le 1$ である．

まず，F が右連続関数であることを示す．F は単調非減少関数であるから，任意の x に対し,

$$\lim_{\varepsilon\downarrow 0} F(x+\varepsilon) = \inf_{y>x} F(y).$$

また，F の定義より，$x < r \leq y < r'$ を満たす任意の実数 y, 有理数 r, r' に対し，$F(x) \leq L(r) \leq F(y) \leq L(r')$ が成立する．したがって，

$$F(x) = \inf_{r>x} L(r) = \inf_{y>x} F(y) = \inf_{\varepsilon \downarrow 0} F(x+\varepsilon)$$

となり，F は右連続関数であることが示せた．

次に，x が F の任意の連続点であるとき，数列 $\{F_{n(j,j)}(x)\}$ $j = 1, 2, 3, \ldots$ が $F(x)$ に収束することを示す．

r を $r > x$ を満たす有理数とすると，

$$\limsup_{j\to\infty} F_{n(j,j)}(x) \leq \lim_{j\to\infty} F_{n(j,j)}(r) = L(r) \leq F(r).$$

また，r' を $r' \leq x$ を満たす有理数，$\varepsilon > 0$ を任意の正の数とすると，

$$\liminf_{j\to\infty} F_{n(j,j)}(x) \geq \lim_{j\to\infty} F_{n(j,j)}(r') = L(r') \geq F(r'-\varepsilon).$$

よって，

$$F(r'-\varepsilon) \leq \liminf_{j\to} F_{n(j,j)}(x) \leq \limsup_{j\to\infty} F_{n(j,j)}(x) \leq F(r).$$

有理数 r, $r'-x$ は x にいくらでも近くとることができることと，x が F の連続点であることから，

$$\lim_{j\to\infty} F_{n(j,j)}(x) = F(x).$$

□

Helly-Bray の定理で収束先の $F(x)$ は右連続な単調非減少関数ではあるが，確率分布関数とは限らない．2 ページで見たように，$F(x)$ が，$\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$ を満たせば確率分布関数となる．収束先が確率分布関数になるのを保証するのが次の性質である．

定義． 確率分布関数の列 $\{F_n\}$ $(n = 1, 2, 3, \ldots)$ を考える．P_n を F_n に対応する確率測度とする．任意の与えられた $\varepsilon > 0$ に対し，ある $C > 0$ が存在して，任意の n に対し，

$$P_n([-C, C]) = F_n(C) - F_n(-C-) > 1 - \varepsilon$$

となるとき，確率分布関数の列 $\{F_n\}$ $(n = 1, 2, 3, \ldots)$ は**タイト** (tight) であるという．また，確率測度の列 $\{P_n\}$ $(n = 1, 2, 3, \ldots)$ は**タイト** (tight) であるともいう．なお，$F_n(-C-) := \lim_{x\uparrow -C} F_n(x)$ である．

□

(タイトの条件の中で,「任意の n」は「有限個を除いてすべての n」としてもいいです．つまり $\forall\varepsilon > 0\ \exists C\ \exists n_0\ \forall n \geq n_0$ としてもよい．)

確率測度の列がタイトであるということは，直観的には確率が無限大あるいは無限小に逃げていかないということである．

定理. (Prohorov の定理の特別な場合) 確率分布関数の列 $\{F_n\}$ $(n = 1, 2, 3, \ldots)$ がタイトであるとき，ある確率分布関数 F に弱収束する $\{F_n\}$ の部分列が存在する． □

問題 18. 上の定理を (Helly-Bray の定理を用いて) 証明せよ．

極値分布 (弱収束の例)

確率的な現象のリスクを考えるとき等には，確率変数列の平均値周辺の挙動よりも最大値の分布が問題になることがある．

確率変数 $X_1, X_2, \ldots, X_n$ が独立に分布関数 $G(x)$ をもつ確率分布にしたがうとき，最大値 $M_n := \max\{X_1, X_2, \ldots, X_n\}$ の確率分布の分布関数 F_n は,

$$F_n(x) := P(M_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = G^n(x)$$

となる．最大値の分布を**極値分布** (extreme value distribution) と呼ぶ．

例 1. 期待値 $1/\alpha$ $(\alpha > 0)$ の指数分布の分布関数は,

$$G(x) = \begin{cases} 1 - \mathrm{e}^{-\alpha x} & (x \geq 0), \\ 0 & (x < 0). \end{cases}$$

確率変数 $X_1, X_2, \ldots, X_n$ が独立に期待値 $1/\alpha$ の指数分布にしたがう時，最大値 M_n の確率分布の分布関数 F_n は,

$$F_n(x) = \begin{cases} (1 - \mathrm{e}^{-\alpha x})^n & (x \geq 0), \\ 0 & (x < 0). \end{cases}$$

ここで $n \to \infty$ の極限をとると，任意の x に対し，$\lim_{n\to\infty} F_n(x) = 0$ となり，F_n は確率分布関数には弱収束しないことがわかる．このことは，$n \to \infty$ の極限では，M_n が無限に大きな値をとるようになることに対応している．

しかし，確率変数 $M_n - \alpha^{-1} \log n$ を考えると，その分布関数は,

$$\begin{aligned} P(M_n - \alpha^{-1} \log n \leq x) &= P(M_n \leq x + \alpha^{-1} \log n) \\ &= \begin{cases} \{1 - \mathrm{e}^{-(\alpha x + \log n)}\}^n & (x \geq -\alpha^{-1} \log n), \\ 0 & (x < -\alpha^{-1} \log n). \end{cases} \end{aligned}$$

ここで，$\{1 - \mathrm{e}^{-(\alpha x + \log n)}\}^n = \{1 - (1/n)\mathrm{e}^{-\alpha x}\}^n \to \mathrm{e}^{-\mathrm{e}^{-\alpha x}}$ $(n \to \infty)$ となり,

$$F(x) := \mathrm{e}^{-\mathrm{e}^{-\alpha x}}, \qquad \forall x \in \mathbb{R}$$

は確率分布関数なので，$M_n - \alpha^{-1} \log n$ の分布の分布関数は，$F(x)$ に弱収束することがいえる．すなわち,

$$M_n - \alpha^{-1} \log n \rightsquigarrow F(x)$$

である．このことから，n が大きいとき，確率変数 $M_n - \alpha^{-1}\log n$ の分布は $F(x)$ で近似できることがわかる．ここで　$-\alpha^{-1}\log n \downarrow -\infty$ であり，$F(x)$ のサポートが $\mathbb{R}$ 全域であることに注意する．$F(x)$ は**ガンベル分布** (Gumbel distribution) あるいはガンベル型の分布と呼ばれる分布の分布関数である． □

例 1 で見たように，確率論では確率変数列 Z_n で $n \to \infty$ の極限での分布を考えるとき，n が無限に大きくなると，Z_n も無限に大きい値をとるようになったり，あるいは 1 点に退化したりするとき，適当な実数列 a_n, b_n をとり，Z_n の代わりに，確率変数列 $(Z_n - b_n)/a_n$ の分布の極限を考えることがよく行われる．上の例は，$a_n = 1,\ b_n = \alpha^{-1}\log n$ の場合である．

例 2．確率分布関数

$$G(x) = \begin{cases} 1 - x^{-\alpha} & (x \geq 1), \\ 0 & (x < 1) \end{cases}$$

$(\alpha > 0)$ により定まる確率分布にしたがう n 個の独立な確率変数 $X_1, X_2, \ldots, X_n$ の最大値を M_n とする．M_n の確率分布の分布関数 F_n は，

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

ここで $n \to \infty$ の極限をとると，任意の x に対し，$\lim_{n\to\infty} F_n(x) = 0$ となり，F_n はやはり確率分布関数には弱収束しないことがわかる．

ここで，$n^{-1/\alpha}M_n$ の確率分布の分布関数は，

$$P(n^{-1/\alpha}M_n \leq x) = P(M_n \leq n^{1/\alpha}x) = \begin{cases} (1 - n^{-1}x^{-\alpha})^n & (x \geq n^{-1/\alpha}), \\ 0 & (x < n^{-1/\alpha}) \end{cases}$$

であり，$(1 - n^{-1}x^{-\alpha})^n \to \mathrm{e}^{-x^{-\alpha}}\ (n \to \infty)$ だから，$n^{-1/\alpha} \to 0$ に注意すると，$n^{-1/\alpha}M_n$ の確率分布は，分布関数

$$F(x) = \begin{cases} \mathrm{e}^{-x^{-\alpha}} & (x \geq 0), \\ 0 & (x < 0). \end{cases}$$

をもつ確率分布に弱収束することがわかる．$F(x)$ は**フレシェ分布** (Frechet distribution) あるいはフレシェ型の分布と呼ばれる分布の分布関数である．この場合は，$a_n = n^{1/\alpha}$, $b_n = 0$ である． □

例 3．例 1，例 2 と同様の問題を，確率分布関数

$$G(x) = \begin{cases} 1 & (x \geq 1) \\ 1 - (1-x)^{\alpha} & (0 \leq x < 1), \\ 0 & (x < 0). \end{cases}$$

$(\alpha > 0)$ について考える．最大値 M_n の確率分布の分布関数 F_n は,

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

なので，$n \to \infty$ のとき，M_n の確率分布は確率 1 で値 1 をとる 1 点分布に弱収束する．$n^{1/\alpha}(M_n - 1)$ の分布関数は,

$$P(n^{1/\alpha}(M_n - 1) \leq x) = P(M_n \leq n^{-1/\alpha}x + 1) = \begin{cases} 1 & (x \geq 0), \\ \{1 - n^{-1}(-x)^\alpha\}^n & (-n^{1/\alpha} \leq x < 0), \\ 0 & (x < -n^{1/\alpha}). \end{cases}$$

$n \to \infty$ のとき，これは，分布関数

$$F(x) = \begin{cases} 1 & (x \geq 0), \\ \mathrm{e}^{-(-x)^\alpha} & (x < 0) \end{cases}$$

に弱収束する．これは**ワイブル分布** (Weibull distribution) あるいはワイブル型の分布と呼ばれる分布になっている.（確率変数 Z の確率分布関数が上で与えられた $F(x)$ のとき，Z は負の値しかとらない．実際には，$-Z$ の分布をワイブル分布と呼ぶことが多い.）□

以上の 3 例では特殊な分布に基づく極値分布の極限分布を導出した．実は，極値分布の極限分布としては，本質的にこの 3 種類だけしか現れないことが知られている．この驚くような結果は，極値分布がもとの分布の右側の裾の落ち方の挙動だけで決まることによっている．

なお，$X_1, X_2, \ldots, X_n$ が独立に標準正規分布にしたがうとき,

$$a_n = \frac{1}{\sqrt{2\log n}}, \quad b_n = \sqrt{2\log n} - \frac{\log(4\pi \log n)}{2\sqrt{2\log n}}$$

ととれば,

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \mathrm{e}^{-\mathrm{e}^{-x}} \quad \text{for all } x$$

となることが知られている．このことは標準正規分布の右側の裾確率 $1-\Phi(x) = \int_x^\infty \phi(u)\mathrm{d}u$ に関する次の事実から導かれる．

$$\lim_{x\to\infty} \frac{1-\Phi(x)}{(1/x)\phi(x)} = 1.$$

(Lamperti, Section 24, Problem 3.)

問題 19. $X_1, X_2, \ldots, X_n$ は独立に同一の分布にしたがう確率変数列，$M_n := \max(X_1, X_2, \ldots, X_n)$ とする．$G(x)$ を X_i の確率分布の分布関数として,

$$\lim_{x\to\infty} \mathrm{e}^x\{1 - G(x)\} = c, \quad c > 0$$

が成立するとき,

$$\lim_{n\to\infty} P(M_n - \log(cn) \le x) = \mathrm{e}^{-\mathrm{e}^{-x}}$$

となることを示せ.

参考文献

Galambos, J. (1987). The Asymptotic Theory of Extreme Order Statistics, 2nd ed., Krieger, Melbourne.

6　特性関数

確率分布の弱収束を考える際に,強力な道具となるのが特性関数である.中心極限定理は,まず,2項分布などの特別な場合に直接分布を扱う方法による証明が与えられた.しかし,1900年ごろロシアの数学者 Lyapunov により始められた特性関数を用いたアプローチにより,統一的な証明を見通しよく与えることが可能になった.

定義． 確率変数 X に対し定まる λ の関数

$$\phi(\lambda) := \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda X})$$

を X の**特性関数** (characteristic function) と呼ぶ.

また,$\mathbb{R}$ 上の確率測度 P に対して定まる λ の関数

$$\phi(\lambda) := \int \mathrm{e}^{\mathrm{i}\lambda x} \mathrm{d}P(x)$$

を P の特性関数と呼ぶ.

$\phi(\lambda)$ は (少なくとも) 任意の実数 λ に対して定義される.(場合によっては,λ が複素数の値をとることを考えることがある.) □

注． 複素数値関数の積分に関する注意.$f(x)$ が複素数値関数であるときには,$\mathrm{Re} f(x)$, $\mathrm{Im} f(x)$ がともに可積分であるときに限って $f(x)$ は可積分であるといい,

$$\int f(x)\mathrm{d}P = \int \mathrm{Re} f(x)\mathrm{d}P + \mathrm{i}\int \mathrm{Im} f(x)\mathrm{d}P$$

により f の積分を定義する.Lebesgue の収束定理等が複素数値関数についても実数値関数の場合と同様に成立することはこの積分の定義から容易に確認できる. □

注． Lebesgue の収束定理 (dominated convergence theorem と呼ばれることも多い)

「可測関数の列 $f_n(x)$ $n = 1, 2, 3, \dots$ があって,測度 P に対しほとんど至るところ $\lim f_n(x) = f(x)$ が存在するものとする.また,測度 P に対し可積分な関数 $g(x)$ があっ

て，すべての n に対しほとんど至るところ $|f_n(x)| \leq g(x)$ が成立するものとする．このとき，$f(x)$ も可積分となって，

$$\lim_{n\to\infty} \int f_n \mathrm{d}P = \int f \mathrm{d}P$$

が成立する.」というのが Lebesugue の収束定理である．Lebesgue 積分の定理の中でも最も良く使われるもののひとつなので，具体例を考えて意味をよく理解しておく必要がある． □

問題 20．特性関数 $\phi(\lambda)$ を定義する積分 $\phi(\lambda) := \int \mathrm{e}^{\mathrm{i}\lambda x}\mathrm{d}P(x)$ について考える．任意の $\lambda \in \mathbb{R}$ を固定したとき，x の関数 $\mathrm{e}^{\mathrm{i}\lambda x}$ は測度 P について可積分であることを示せ．(このことから $\phi(\lambda)$ は任意の実数 λ に対して定義されることがわかる．)

問題 21．$\phi(\lambda)$ はある確率変数の特性関数であるとするとき，$\phi(0) = 1$ であること，任意の実数 λ に対し $|\phi(\lambda)| \leq 1$ であること，ϕ は実数上の関数として一様連続であること，を示せ.

問題 22．正規分布 $N(0, \sigma^2)$ の特性関数は，

$$\phi(\lambda) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\lambda x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2\right)\mathrm{d}x = \mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}$$

となることを示せ.

問題 23．Cauchy 分布は，密度関数が，

$$f(x) := \frac{1}{\pi}\frac{c}{x^2 + c^2} \quad (c > 0)$$

で与えられる確率分布である．Cauchy 分布の特性関数が，

$$\phi(\lambda) = \mathrm{e}^{-c|\lambda|}$$

となることを示せ.

独立な確率変数の和の特性関数に関する次の事実は，容易に示すことができる.

補題． 独立な確率変数 $X_1, X_2, \ldots, X_n$ の特性関数が $\phi_1(\lambda), \phi_2(\lambda), \ldots, \phi_n(\lambda)$ であるとき，確率変数 $S_n := X_1 + X_2 + \cdots + X_n$ の特性関数 $\psi(\lambda)$ は

$$\psi(\lambda) := \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda S_n}) = \prod_{i=1}^{n} \phi_i(\lambda)$$

□

次の定理により，特性関数を与えれば確率分布が一意的に決まることが容易にいえる．このことから，確率分布と特性関数とは１対１に対応する ことがわかる．

定理． (Lévy の反転公式) F を $\mathbb{R}$ 上のある確率測度の分布関数，ϕ をその特性関数とする．α, β $(\alpha < \beta)$ を F の連続点とする．このとき，

$$F(\beta) - F(\alpha) = \lim_{\sigma \to 0+} \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(\lambda) \mathrm{e}^{-\frac{\sigma^2 \lambda^2}{2}} \frac{\mathrm{e}^{-\mathrm{i}\lambda\beta} - \mathrm{e}^{-\mathrm{i}\lambda\alpha}}{-\mathrm{i}\lambda} \mathrm{d}\lambda. \qquad \square$$

注． $\mathbb{R}$ 上の関数 f が L^1 に属し，その Fourier 変換

$$\phi(\lambda) := \int_{-\infty}^{\infty} f(x) \mathrm{e}^{\mathrm{i}\lambda x} \mathrm{d}x \tag{9}$$

も L^1 に属すとき，

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(\lambda) \mathrm{e}^{-\mathrm{i}\lambda x} \mathrm{d}\lambda \tag{10}$$

となることは良く知られており，(10) は Fourier 逆変換と呼ばれている．なお，L^1 は，$\mathbb{R}$ 上の関数 f で $\int_{-\infty}^{\infty} |f(x)| \mathrm{d}x < \infty$ となるもの全体の空間のことである (念のため)．

注． 反転公式は，$\alpha < \beta$ を F の連続点として，

$$F(\beta) - F(\alpha) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \phi(\lambda) \frac{\mathrm{e}^{-\mathrm{i}\lambda\beta} - \mathrm{e}^{-\mathrm{i}\lambda\alpha}}{-\mathrm{i}\lambda} \mathrm{d}\lambda. \tag{11}$$

の形で書かれることも多い．下端の $T_1 = -T$ と上端の $T_2 = T$ はそれぞれ別々に $-\infty$ および ∞ に発散してもよい．α, β に point mass (離散的な正の確率) $F(\{\alpha\}) > 0, F(\{\beta\}) > 0$ がある場合には，上の定理あるいは (11) のように対称的に反転してやると，反転公式には $F(\{\alpha\})/2$ 及び $F(\{\beta\})/2$ が加わる．この重みを変えてやることもできる．

確率分布が密度関数をもつとき，特性関数は Fourier 変換 (9) に他ならない．Fourier 逆変換が存在すれば，特性関数から元の密度関数を復元できることになる．

ただ，一般に確率分布は密度関数をもつとは限らず，密度関数が存在する場合であっても，特性関数 ϕ が L^1 に属するとは限らない．確率分布と特性関数が１対１に対応することを示すためにはこの問題を回避する必要がある．それをうまく解決するのが Lévy の反転公式である．

以下の証明では，小さい分散を持つ正規分布とのたたみ込みを用いることによって，上の注にある L^1 に関する事実を前提とせず，反転公式の証明が完結している．

証明 (Lévy の反転公式)．

分布関数 F は一般には滑らかとは限らない．そこで，平均 0, 分散 σ^2 の正規分布 N_σ とのたたみ込み

$$F_\sigma := F \star N_\sigma$$

を考える ($\star$ はたたみ込みを表す)．すると，F_σ は滑らかな $\mathbb{R}$ 上の関数となり，F_σ の微分 (密度関数) は,

$$F'_\sigma(x) = f_\sigma(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-u)^2}{2\sigma^2}} \mathrm{d}F(u) \tag{12}$$

と書ける．

(12) の特性関数 ϕ_σ は，N_σ の特性関数が $\mathrm{e}^{-\sigma^2\lambda^2/2}$ であることから，

$$\phi_\sigma(\lambda) = \mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\phi(\lambda) \tag{13}$$

となる．この特性関数のフーリエ逆変換をとってみよう．(13) は，$|\phi(\lambda)| \leq 1$ であるから，可積分 (言い換えると L^1 の要素) なので，Fourier 逆変換をとることができて，

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\lambda x}\phi_\sigma(\lambda)\mathrm{d}\lambda &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\lambda x}\mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\phi(\lambda)\mathrm{d}\lambda \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\lambda x}\mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\lambda u}\mathrm{d}F(u)\mathrm{d}\lambda \\
&= \int_{-\infty}^{\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\lambda(u-x)}\mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\mathrm{d}\lambda\mathrm{d}F(u) \qquad (14) \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-u)^2}{2\sigma^2}}\mathrm{d}F(u) = f_\sigma(x). \qquad (15)
\end{aligned}$$

(14) で Fubini の定理を使った．(15) は ϕ_σ を Fourier 逆変換して，密度関数 f_σ に戻す式になっている．このことから，正規分布とたたみ込みをとってなました分布については，普通の Fourier 変換，Fourier 逆変換の公式がそのまま使えることが確認できる．

(15) より，

$$\begin{aligned}
F_\sigma(\beta) - F_\sigma(\alpha) = \int_\alpha^\beta f_\sigma(x)\mathrm{d}x &= \int_\alpha^\beta \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\lambda x}\mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\phi(\lambda)\mathrm{d}\lambda\mathrm{d}x \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_\alpha^\beta \mathrm{e}^{-\mathrm{i}\lambda x}\mathrm{d}x\right)\mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\phi(\lambda)\mathrm{d}\lambda \qquad (16) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\phi(\lambda)\frac{\mathrm{e}^{-\mathrm{i}\lambda\beta} - \mathrm{e}^{-\mathrm{i}\lambda\alpha}}{-\mathrm{i}\lambda}\mathrm{d}\lambda. \qquad (17)
\end{aligned}$$

(16) で再び Fubini の定理を使った．

ここで，X を確率分布関数 F をもつ分布にしたがう確率変数，Z_σ を X とは独立に平均 0, 分散 1 の正規分布 $\mathrm{N}(0,1)$ にしたがう確率変数とすると，$X + \sigma Z$ の確率分布の分布関数は，F_σ となる．$\sigma \to 0+$ の極限では，$X + \sigma Z$ は X に確率収束することは容易に示せる．確率収束すれば弱収束するという関係から，

$$F_\sigma \rightsquigarrow F \quad (\sigma \to 0+)$$

がいえる．したがって，α, β は F の連続点であることより，

$$\lim_{\sigma\to 0+} F_\sigma(\alpha) = F(\alpha), \quad \lim_{\sigma\to 0+} F_\sigma(\beta) = F(\beta).$$

(17) より，

$$F(\beta)-F(\alpha)=\lim_{\sigma\to 0+}\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(\lambda)\mathrm{e}^{-\frac{\sigma^2\lambda^2}{2}}\frac{\mathrm{e}^{-\mathrm{i}\lambda\beta}-\mathrm{e}^{-\mathrm{i}\lambda\alpha}}{-\mathrm{i}\lambda}\mathrm{d}\lambda$$

がいえる． □

注． 大雑把にいえば，Fubini の定理は，$\mathbb{R}^2$ 上の関数 $f(x,y)$ が $\mathbb{R}^2$ 上の Lebesgue 測度に関して可積分であれば，$\mathbb{R}^2$ 上の Lebesgue 測度に対して直接定義される 2 重積分 $\int_{\mathbb{R}^2} f(x,y)\mathrm{d}x\mathrm{d}y$ と，まず y を固定して x について (1 重) 積分してから，y について (1 重) 積分して得られる逐次積分 $\int_{\mathbb{R}}\{\int_{\mathbb{R}} f(x,y)\mathrm{d}x\}\mathrm{d}y$ と等しくなることを保証する定理である．逐次積分の順序を交換して $\int_{\mathbb{R}}\{\int_{\mathbb{R}} f(x,y)\mathrm{d}y\}\mathrm{d}x$ としても結果は等しくなることも保証される．詳しくは Lebesgue 積分の本を参照.

系. $\mathbb{R}$ 上の確率測度 P と Q それぞれの特性関数が等しければ，確率測度 P と Q は等しい. □

問題 24. (正規分布の再生性) $X_1,X_2,\ldots,X_n$ が平均 0, 分散 σ^2 の正規分布 $\mathrm{N}(0,1)$ に独立にしたがうとき，$\sum_{i=1}^n X_i/\sqrt{n}$ のしたがう分布が N(0.1) であることを特性関数を用いて示せ.

問題 25. (Cauchy 分布の再生性) また，$X_1,X_2,\ldots,X_n$ がともに密度関数

$$f(x):=\frac{1}{\pi}\frac{c}{x^2+c^2}\quad(c>0)$$

の Cauchy 分布に独立にしたがうとき，$\sum_{i=1}^n X_i/n$ のしたがう分布が同じ Cauchy 分布であることを特性関数を用いて示せ.

F_n $(n=1,2,\ldots)$, F を $\mathbb{R}$ 上の確率測度の分布関数，ϕ_n $(n=1,2,\ldots)$, ϕ を対応する特性関数とするとき，$F_n\rightsquigarrow F$ ならば，任意の $\lambda\in\mathbb{R}$ に対し，$\phi_n(\lambda)\to\phi(\lambda)$ が成立することが，確率測度の弱収束の定義より直ちにいえる．

次の定理はこの逆を示すものである．このことから，特性関数の列 ϕ_n がある確率分布の特性関数 ϕ に各点で収束することと ϕ_n に対応する確率分布列が ϕ に対応する確率分布に弱収束することが同値であることがわかる．このことから，特性関数に基づく方法は分布の弱収束を示すのに大変便利になる．総変動距離などの分布間の他の距離に基づく分布の収束は，特性関数の各点での収束とは同値にはならないので，収束を示すには別の方法を使わなければならない.

定理． (連続性定理 (1)，continuity theorem) F_n $(n=1,2,\ldots)$, F を $\mathbb{R}$ 上の確率測度の分布関数，ϕ_n $(n=1,2,\ldots)$, ϕ を対応する特性関数とする．任意の $\lambda\in\mathbb{R}$ に対し，$\phi_n(\lambda)\to\phi(\lambda)$ のとき，$F_n\rightsquigarrow F$ が成立する．

証明．確率分布の特性関数 $\phi(\lambda)$ は連続関数であることが，Lebesgue の収束定理よりいえる (確認せよ)．したがって，下の連続性定理 (2) よりいえる． $\square$

上の定理は，$\phi_n(\lambda)$ の収束先 $\phi(\lambda)$ がある確率分布の特性関数であることが分かっている時に使えるものである．$\phi(\lambda)$ がある確率分布の特性関数であるかどうかが分かっていないときに有用なのが次の定理である．この定理も連続性定理 (continuity theorem) と呼ばれる．便宜的に連続性定理 (2) と呼ぶことにする．

定理． (連続性定理 (2)，continuity theorem) P_n $(n = 1, 2, \ldots)$ を $\mathbb{R}$ 上の確率測度，ϕ_n $(n = 1, 2, \ldots)$ を対応する特性関数とする．任意の $\lambda \in \mathbb{R}$ に対し，極限

$$\lim_{n\to\infty} \phi_n(\lambda) = \phi(\lambda)$$

が存在し，$\phi(\lambda)$ が $\lambda = 0$ で連続であるとき，ϕ は R 上のある確率測度 P の特性関数で，$P_n \rightsquigarrow P$ が成立する．

証明にはいる前に次の補題を示す．次の補題は，分布の裾の確率の挙動が特性関数の原点回りの挙動に反映されることを示すものである．

補題． G を $\mathbb{R}^1$ 上の確率分布とし $\psi(\lambda)$ をその特性関数とする．任意の $\delta > 0$ に対して

$$\frac{1}{\delta}\int_0^\delta \{2 - \psi(\lambda) - \psi(-\lambda)\}\mathrm{d}\lambda \geq 1 - G\left(\left[-\frac{2}{\delta}, \frac{2}{\delta}\right]\right).$$

証明.

$$\psi(\lambda) + \psi(-\lambda) = \int_{-\infty}^\infty 2\cos(\lambda x)\mathrm{d}G(x) \tag{18}$$

となり $\psi(\lambda) + \psi(-\lambda)$ は実数となる．

$$\begin{aligned}
\frac{1}{\delta}\int_0^\delta \{2 - \psi(\lambda) - \psi(-\lambda)\}\mathrm{d}\lambda &= \frac{1}{\delta}\int_0^\delta \int_{-\infty}^\infty \{2 - 2\cos(\lambda x)\}\mathrm{d}G(x)\mathrm{d}\lambda \\
&= \frac{2}{\delta}\int_{-\infty}^\infty \int_0^\delta \{1 - \cos(\lambda x)\}\mathrm{d}\lambda\mathrm{d}G(x) \qquad (19) \\
&= 2\int_{-\infty}^\infty \left\{1 - \frac{\sin(\delta x)}{\delta x}\right\}\mathrm{d}G(x) \\
&\geq 2\int_{|x|>\frac{2}{\delta}} \left(1 - \frac{\sin|\delta x|}{|\delta x|}\right)\mathrm{d}G(x) \qquad (|\sin t| \leq |t|, \forall t \in \mathbb{R} \text{ に注意}) \\
&\geq 2\int_{|x|>\frac{2}{\delta}} \left(1 - \frac{1}{|\delta x|}\right)\mathrm{d}G(x) \\
&\geq \int_{|x|>\frac{2}{\delta}} \mathrm{d}G(x) = 1 - G\left(\left[-\frac{2}{\delta}, \frac{2}{\delta}\right]\right). \qquad (20)
\end{aligned}$$

(19) で Fubini の定理を使った． $\square$

[**補足**] 補題の積分は

$$\frac{1}{\delta}\int_{-\delta}^{\delta}(1-\psi(\lambda))\mathrm{d}\lambda = \frac{1}{\delta}\int_{-\delta}^{\delta}(1-\mathrm{Re}\psi(\lambda))\mathrm{d}\lambda$$

とも書ける．

証明 (連続性定理 (2))．F_n を P_n の確率分布関数とする．まず，確率測度の列 $\{P_n\}$ はタイト (第 4 回のプリント p. 29) であることを示す．(18) は λ の実数値関数であるから，$\phi(\lambda)+\phi(-\lambda)$ も実数値関数になる．また $\phi(0)=1$ であり，仮定より，$\phi(\lambda)$ は $\lambda=0$ で連続であるから，任意の $\varepsilon>0$ に対して，ある $\delta>0$ をとり，$|\lambda|<\delta$ であれば，

$$|1-\phi(\lambda)|<\frac{\varepsilon}{4}$$

とできる．すると，

$$0<\frac{1}{\delta}\int_0^{\delta}\{2-\phi(\lambda)-\phi(-\lambda)\}\mathrm{d}\lambda<\frac{\varepsilon}{2}$$

となる．また，ルベーグの収束定理により，

$$\lim_{n\to\infty}\int_0^{\delta}\{2-\phi_n(\lambda)-\phi_n(-\lambda)\}\mathrm{d}\lambda = \int_0^{\delta}\{2-\phi(\lambda)-\phi(-\lambda)\}\mathrm{d}\lambda$$

だから，任意の $n\geq N$ に対して，

$$\frac{1}{\delta}\int_0^{\delta}\{2-\phi_n(\lambda)-\phi_n(-\lambda)\}\mathrm{d}\lambda<\varepsilon$$

となる正の整数 N が存在する．

このことと，(20) より，任意の $\varepsilon>0$ に対し，ある $\delta>0$ と正の整数 N が存在して，任意の $n\geq N$ に対し，

$$1-P_n\left(\left[-\frac{2}{\delta},\frac{2}{\delta}\right]\right)<\varepsilon$$

となることがいえる．よって，確率測度の列 $\{P_n\}$ はタイトである．

$\{P_n\}$ はタイトであるから Prohorov の定理 (第 4 回のプリント p. 29) により，$\{n\}$ のある部分列 $\{n(i)\}$ $(i=1,2,3,\ldots)$ をとり，$P_{n(i)}$ が $\mathbb{R}$ 上のある確率測度 P に弱収束するようにできる．$\phi(\lambda)$ を P の特性関数とすると，任意の $\lambda\in\mathbb{R}$ に対し，

$$\lim_{i\to\infty}\phi_{n(i)}(\lambda)=\phi(\lambda)$$

となる．

次に，P_n が P に弱収束しないと仮定して矛盾を導く．F を P の確率分布関数とすると，P_n が P に弱収束しないので，n が大きくなるとき $F_n(x)$ が $F(x)$ に収束しない F の連続点 x が存在する．このとき，ある $\mu>0$ とある $\{n\}$ の部分列 $\bar{n}(j)$ $(j=1,2,3,\ldots)$ をとり，任意の j に対し，

$$|F_{\bar{n}(j)}-F(x)|>\mu \tag{21}$$

とできる．ところが，$\{F_{\bar{n}(j)}\}$ もタイトなので，$\bar{n}(j)$ $(j=1,2,3,\ldots)$ の適当な部分列 $m(k)$ $(k=1,2,3,\ldots)$ をとることにより，$\{P_{m(k)}\}$ $(k=1,2,3)$ がある確率分布 $\bar{P}$ に弱収束するようにできる．このとき (21) より，$\bar{P}$ と P とは異なる確率測度になる．

すると，$\{P_{m(k)}\}$ $(k=1,2,3)$ の特性関数の列 $\phi_{m(k)}(\lambda)$ は，$\bar{P}$ の特性関数 $\bar{\phi}(\lambda)$ に各点で収束する．

ところが，定理の仮定より，任意の $\lambda \in \mathbb{R}$ で $\phi(\lambda)=\bar{\phi}(\lambda)$ でなければならない．特性関数と確率分布は 1 対 1 に対応するのでこれは矛盾である． □

今まで $\mathbb{R}$ 上の確率分布と特性関数の関係について見てきた．同様に，$\mathbb{R}^k$ $(k \geq 2)$ 上の確率分布の場合でも特性関数の方法は強力である．

定義． X を k 次元確率ベクトル (確率変数が k 個並んだ縦ベクトル)，P を X の分布 ($\mathbb{R}^k$ 上の確率測度) とする．また，$\lambda \in \mathbb{R}^k$ とする．このとき，

$$\phi(\lambda) := \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda^\top X}) = \int_{\mathbb{R}^k} \mathrm{e}^{\mathrm{i}\lambda^\top x} \mathrm{d}P(x)$$

を X の**特性関数** (characteristic function) と呼ぶ． □

定理． 特性関数 $\phi(\lambda)$ により，R^k 上の確率測度は一意に決まる． □

$\mathbb{R}^k$ $(k \geq 2)$ の場合に，連続性定理についても対応する結果が成立することが知られている．

7　中心極限定理と安定分布

特性関数を用いることにより中心極限定理が容易に証明できるようになる．また，安定分布とよばれる正規分布や Cauchy 分布を含むクラスの確率分布の性質を調べることができる．

補題. X を特性関数 ϕ をもつ確率変数とする．正の整数 k に対して，モーメント $\mathrm{E}(|X|^k)$ が存在するならば，ϕ は k 階連続微分可能で，

$$\phi^{(k)}(0) = \mathrm{i}^k \mathrm{E}(X^k)$$

が成立する．

証明. $k=1$ の場合について見る．特性関数の定義より，

$$\phi(\lambda) := \int \mathrm{e}^{\mathrm{i}\lambda x} \mathrm{d}P(x)$$

ϕ の微分は，

$$\lim_{h \to 0} \frac{\phi(\lambda+h)-\phi(\lambda)}{h}$$

で定義される．極限が存在するとき，$\phi(\lambda)$ は微分可能となる．ここで，

$$\begin{aligned}\frac{\phi(\lambda+h)-\phi(\lambda)}{h} &= \int \frac{1}{h}\left\{\mathrm{e}^{\mathrm{i}(\lambda+h)x}-\mathrm{e}^{\mathrm{i}\lambda x}\right\}\mathrm{d}P(x) \\ &= \int \frac{\mathrm{e}^{\mathrm{i}hx}-1}{h}\mathrm{e}^{\mathrm{i}\lambda x}\mathrm{d}P(x)\end{aligned}$$

被積分関数の $h \to 0$ の極限をとると，

$$\lim_{h\to 0}\frac{\mathrm{e}^{\mathrm{i}hx}-1}{h}\mathrm{e}^{\mathrm{i}\lambda x} = \mathrm{i}x\mathrm{e}^{\mathrm{i}\lambda x}.$$

また，任意の $u \in \mathbb{R}$ に対し，$|\mathrm{e}^{\mathrm{i}u}-1| \leq |u|$ であることから，

$$\left|\frac{\mathrm{e}^{\mathrm{i}hx}-1}{h}\mathrm{e}^{\mathrm{i}\lambda x}\right| \leq \left|\frac{hx}{h}\right| |\mathrm{e}^{\mathrm{i}\lambda x}| \leq |x|$$

となる．仮定より，期待値 $\mathrm{E}(|X|)$ は存在するので，Lebesgue の収束定理 (5 回目のプリント) により，

$$\begin{aligned}\lim_{h\to 0}\frac{\phi(\lambda+h)-\phi(\lambda)}{h} &= \lim_{h\to 0}\int \frac{\mathrm{e}^{\mathrm{i}hx}-1}{h}\mathrm{e}^{\mathrm{i}\lambda x}\mathrm{d}P(x) = \int \lim_{h\to 0}\frac{\mathrm{e}^{\mathrm{i}hx}-1}{h}\mathrm{e}^{\mathrm{i}\lambda x}\mathrm{d}P(x) \\ &= \int \mathrm{i}x\mathrm{e}^{\mathrm{i}\lambda x}\mathrm{d}P(x)\end{aligned}$$

となる．これは，λ の連続関数である．また，

$$\phi'(0) = i\mathrm{E}(X)$$

であることがわかる．$k \geq 2$ の場合も同様の議論を繰り返すことにより示せる． □

[**補足**] $k \geq 2$ については

$$\left|\mathrm{e}^{\mathrm{i}u} - \left(1+\mathrm{i}u+\frac{1}{2}(\mathrm{i}u)^2+\cdots+\frac{1}{(k-1)!}(\mathrm{i}u)^{k-1}\right)\right| \leq \frac{|u|^k}{k!}, \quad \forall u \in \mathbb{R}$$

を用いて証明することもできる．

[**補足**] k が偶数で $\phi^{(k)}(0)$ が存在すれば $E(X^k) < \infty$ となることも知られている．(Lamperti p.94 Remark.)

定理．(中心極限定理) $X_1, X_2, \ldots$ を平均 μ, 分散 σ^2 の同一の分布に独立にしたがう確率変数とする．$S_n := X_1 + \cdots + X_n$ とおくと，

$$F_n(x) := P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) \rightsquigarrow \mathrm{N}(0,1)$$

が成り立つ．

証明．$X_i - \mu$ をあらためて X_i と置き直すことにより，$\mu = 0$ と仮定できる．X_i の特性関数を ϕ とおくと，このとき，$S_n/\sqrt{n}\sigma$ の特性関数 ϕ_n は，

$$\phi_n(\lambda) = \mathrm{E}(\mathrm{e}^{\frac{\mathrm{i}\lambda S_n}{\sqrt{n}\sigma}}) = \prod_{k=1}^{n} \mathrm{E}(\mathrm{e}^{\frac{\mathrm{i}\lambda X_k}{\sqrt{n}\sigma}}) = \left\{\phi\left(\frac{\lambda}{\sqrt{n}\sigma}\right)\right\}^n$$

となる．

p. 41 の補題より，

$$\phi'(0) = \mathrm{i}\mu = 0, \quad \phi''(0) = -\sigma^2$$

で $\phi''(\lambda)$ は λ の連続関数だから，

$$\phi(\lambda) = 1 - \frac{\sigma^2}{2}\lambda^2 + \mathrm{o}(\lambda^2).$$

したがって，

$$\phi_n(\lambda) = \left\{\phi\left(\frac{\lambda}{\sqrt{n}\sigma}\right)\right\}^n = \left(1 - \frac{1}{2n}\lambda^2 + \mathrm{o}(n^{-1})\right)^n$$

より，

$$\lim_{n\to\infty} \phi_n(\lambda) = \mathrm{e}^{-\lambda^2/2}.$$

右辺は標準正規分布 $\mathrm{N}(0,1)$ の特性関数なので，continuity theorem より，定理の結果がいえる． □

[**補足**] 多次元の場合の中心極限定理も同様な形で証明される．Lamperti の 16 章には，Berry-Esseen 定理や Lindeberg 条件も紹介されている．また密度の収束に関する Local Central Limit theorem (p.100, Thm.5) は他の教科書にない形で述べられており，証明はすぐには見つからない．さらに，i.i.d. 確率変数の和を $\sqrt{n}$ が基準化した時に正規分布に収束するならば，分散が有限であるかということが気になるが，(竹村には) すぐにはわからない．

$X_1, X_2, \ldots, X_n$ が独立に正規分布 $\mathrm{N}(0,1)$ に独立にしたがうとき，和 $S_n := X_1 + X_2 + \cdots + X_n$ はやはり正規分布にしたがうことは良く知られている．このとき，S_n の分布と X_i の分布はスケール変換を除けば同じである．また，$X_1, X_2, \ldots, X_n$ がともに Cauchy 分布に独立にしたがうとき，和 $S_n := X_1 + X_2 + \cdots + X_n$ の分布はやはり Cauchy 分布になり，S_n の分布と X_i の分布はスケール変換を除けば同じになる．

このような正規分布や Cauchy 分布の性質と類似の性質を持つ分布は他にもあることが知られており，**安定分布** (stable distribution) と呼ばれている．安定分布の統一的な理論は，フランスの数学者 Paul Lévy (1886-1971) により構築された．このため，安定分布は Lévy 分布と呼ばれることもある．

Holtsmark 分布と呼ばれる安定分布は，天文学者 Holtsmark により Lévy の仕事より以前に発見されていた．Holtsmark が考察した，ランダムに配置された星の系のつくる重力の強さの問題に類似した問題をとおして，安定分布の具体例について見てみる．

区間 $[-n,n]$ 上に n 個の星 (点) が $[-n,n]$ 上の一様分布にしたがって独立に分布しているものとする．n 個の星の位置を $X_1, X_2, \ldots, X_n$ で表す．すべての星の質量は等しく m であるとする．簡単のため重力定数を 1 とすると，原点における質量 1 の質点に作用する力は，

$$Y_n = \sum_{k=1}^{n} \operatorname{sign}(X_k) \frac{m}{X_k^2}$$

となる．$n=1$ の時単一の要素の分布の裾を考えると，$P(|Y| \geq t) = O(t^{-1/2})$ となり裾が重いことがわかる．この裾の重さが安定分布の「特性指数」に関連している．

特性関数を用いて，Y_n の確率分布は $n \to \infty$ の極限で弱収束することが示せる．

X_k は $[-n,n]$ 上の一様分布にしたがうので，

$$\mathrm{E}\left(\mathrm{e}^{\mathrm{i}\lambda m \operatorname{sign}(X_k) \frac{m}{X_k^2}}\right) = \int_{-n}^{n} \mathrm{e}^{\mathrm{i}\lambda m \operatorname{sign}(x) \frac{m}{x^2}} \frac{1}{2n} \mathrm{d}x = \frac{1}{n} \int_{0}^{n} \cos\left(\frac{\lambda m}{x^2}\right) \mathrm{d}x.$$

Y_n の特性関数は，

$$\begin{aligned}
\mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda Y_n}) &= \left\{\frac{1}{n} \int_0^n \cos\left(\frac{\lambda m}{x^2}\right) \mathrm{d}x\right\}^n = \left[1 - \frac{1}{n} \int_0^n \left\{1 - \cos\left(\frac{\lambda m}{x^2}\right)\right\} \mathrm{d}x\right]^n \\
&= \left[1 - \frac{1}{n} \int_0^\infty \left\{1 - \cos\left(\frac{\lambda m}{x^2}\right)\right\} \mathrm{d}x + \mathrm{o}\left(\frac{1}{n}\right)\right]^n
\end{aligned}$$

より，

$$\lim_{n\to\infty} \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda Y_n}) = \exp\left[-\int_0^\infty \left\{1 - \cos\left(\frac{\lambda m}{x^2}\right)\right\} \mathrm{d}x\right]$$

となる．$y := |\lambda| m / x^2$ とおくと，

$$\lim_{n\to\infty} \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda Y_n}) = \exp\left\{-|\lambda|^{1/2} m^{1/2} \frac{1}{2} \int_0^\infty (1 - \cos y)\, y^{-3/2} \mathrm{d}y\right\}.$$

となり，c を正の定数として，

$$\lim_{n\to\infty} \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda Y_n}) = \mathrm{e}^{-c|\lambda|^{1/2}},$$

の形になることがわかる．これは，λ の連続関数なので，連続性定理より，ある確率分布の特性関数であることがわかる．特性関数 $\mathrm{e}^{-c|\lambda|^{1/2}}$ に対応する確率分布に Y_n の確率分布が弱収束していることになる．分布の特性関数は簡単な形をしているが，分布関数や密度関数は初等関数を用いた簡単な形には表せないことが知られている．

点 x に存在する星から原点の質点に作用する力が，$1/|x|^2$ ではなく $1/|x|^p$ $(p > 1/2)$ に比例するものとして，

$$Y_n = \sum_{k=1}^{n} \operatorname{sign}(X_k) \frac{m}{|X_k|^p}$$

の特性関数の $n \to \infty$ の極限を求めると

$$\lim_{n\to\infty} \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda Y_n}) = \mathrm{e}^{-c|\lambda|^{1/p}}, \qquad (c > 0) \tag{22}$$

となることが示せる．

特性関数 (22) に対応する確率分布は，正規分布やコーシー分布に似た性質を持っている．

X_i $(i = 1, 2, \ldots, k)$ を独立に特性関数 (22) に対応する確率分布にしたがう確率変数とし，$S_k := X_1 + X_2 + \cdots + X_k$ とおくと，S_k の特性関数は,

$$\mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda S_k}) = \{\mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda X_1})\}^k = (\mathrm{e}^{-c|\lambda|^{1/p}})^k = \mathrm{e}^{-c|k^p\lambda|^{1/p}}$$

より，S_k/k^p の分布が X_1 の分布と一致することがわかる．

安定分布についての定義を与えるために，分布の型 (type) という概念を準備する．

定義． 2つの確率分布関数 F, G に対し，ある定数 $a > 0, b$ が存在して，任意の $t \in \mathbb{R}$ に対し,

$$G(t) = F\left(\frac{t-b}{a}\right)$$

が成立するとき，F と G は同じ型 (type) であるといい，$F \sim G$ のように表す． □

確率変数 X の分布関数が $F(t)$ のとき $aX + b$ $(a > 0)$ の分布関数は $F((t-b)/a)$ になり，同じ型になる．

定義． (安定分布) $X_1, X_2, \ldots, X_k$ は同一の退化していない確率分布 P に独立にしたがう確率変数とする．任意の正の整数 k に対し，$S_k := X_1 + X_2 + \cdots + X_k$ の確率分布が P と同じ型のとき，P は**安定分布**であるという． □

注． 1点に確率がすべて集中している分布を退化しているという．

定理． P が $x = 0$ に対し対称な安定分布であるとき，P の特性関数は,

$$\phi(\lambda) = \mathrm{e}^{-c|\lambda|^{\alpha}}$$

で与えられる．ここで，$c > 0, \alpha \in (0, 2]$ は P に対応して決まる定数である． □

(証明略)

α は安定分布の**特性指数** (characteristic exponent) と呼ばれている．$\alpha = 2$ の場合は正規分布，$\alpha = 1$ の場合は Cauchy 分布になる．$\alpha = 3/2$ の場合が，天文学者 Holtsmark に

より発見された Holtsmark 分布である．ただし，Holtsmark 分布の密度関数を初等関数を用いて陽に表すことはできない．

また，対称でない安定分布も知られている．一般の安定分布の特性関数は

$$\phi(\lambda) = \begin{cases} \exp\left[\mathrm{i}d\lambda - c|\lambda|^{\alpha}\left\{1 + \mathrm{i}\theta\,\mathrm{sign}(\lambda)\tan\left(\frac{\pi\alpha}{2}\right)\right\}\right], & (\alpha \neq 1) \\ \exp\left\{\mathrm{i}d\lambda - c|\lambda|\left(1 + \mathrm{i}\theta\,\mathrm{sign}(\lambda)\log|\lambda|\right)\right\}, & (\alpha = 1) \end{cases} \tag{23}$$

の形に表せることが知られている．ここで，$0 < \alpha \leq 2$ は安定分布の特性指数，$-1 \leq \theta \leq 1$ は分布の歪み (非対称性) を表すパラメータである．また，$c > 0$, $-\infty < d < \infty$ はそれぞれ分布の尺度，位置に対応するパラメータにすぎない．したがって，安定分布の型は，α, θ の 2 つのパラメータで指定されることになる．なお，特性指数 $\alpha = 2$ のときは，$\tan((\pi\alpha)/2) = 0$ となるので，θ を含む項は消える．このことから，特性指数 2 の安定分布は正規分布しかないことがわかる．

対称でない安定分布として，密度関数が陽に表せる例について，ランダムウォークの再帰時間の問題をとおして見てみることにする．密度関数が初等関数を用いて陽に表せるのはこの例と正規分布，Cauchy 分布の 3 つしか知られていない．

整数の集合上の離散時間ランダムウォークを考える．時刻 0 で原点から出発し，時刻 t に整数 n 上にいる人が，時刻 $t+1$ に等確率 (1/2) で $n-1$ か $n+1$ に移動するものとする．ランダムウォークをしている人が最初に原点に帰ってくる時刻を T_1, k 回目に原点に返ってくる時刻を T_k とおく．このとき，以下の定理が示せる．

定理． 上でのべたランダムウォークの再帰時刻 T_n に関して，極限分布

$$\lim_{n\to\infty} P\left(\frac{T_n}{n^2} \leq x\right) = F(x), \qquad x \geq 0$$

が存在する．F に対応する特性関数は，

$$\phi(\lambda) = \int_0^{\infty} \mathrm{e}^{\mathrm{i}\lambda x}\mathrm{d}F(x) = \mathrm{e}^{-|\lambda|^{1/2}(1-\mathrm{i}\cdot\mathrm{sign}(\lambda))} \tag{24}$$

対応する密度関数は，

$$f(x) = F'(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2x}}x^{-\frac{3}{2}}, \qquad x \geq 0 \tag{25}$$

となる． □

(証明略)

特性関数 (24) は (23) で $\alpha = 1/2$, $\theta = 1$, $d = 0$, $c = 1$ とおいたものになっている．

問題 26．$X_1, X_2, \ldots, X_n$ が独立に密度関数 (25) をもつ確率分布にしたがうとき，$(X_1 + X_2 + \cdots + X_n)/n^2$ の分布は X_1 の分布と一致することを特性関数 (24) を用いて示せ．

この分布は逆ガウス分布と呼ばれる分布の特別の場合 (極限) になっている．逆ガウス分布は，正のドリフトをもつブラウン運動 (ブラウン運動について詳しくはあとで扱う) の初通過時間の研究の関連に関連して Schrödinger (量子力学の Schrödinger と同一人物) と Smoluchowski によって 1915 年に独立に発見された．逆ガウス分布の密度関数は,

$$f(x) = \left(\frac{\tau}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\tau(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0, \ \ \tau > 0, \ \ \mu > 0 \tag{26}$$

で与えられる．期待値は μ, 分散は μ^3/τ である．(26) で $\mu \to \infty$ とした極限が安定分布になる．ただし，$\mu < \infty$ のときは逆ガウス分布は安定分布にはならない．非負の値のみをとる連続型の確率分布としては，ガンマ分布，対数正規分布が有名だが，逆ガウス分布も美しい性質をもった分布である．たとえば，$X_1, X_2, \ldots, X_n$ が逆ガウス分布にしたがう独立な確率変数のとき，サンプル平均 $\bar{X} = X_1 + X_2 + \cdots + X_n$ も逆ガウス分布にしたがうことが知られている．また，逆ガウス分布は指数型分布族とよばれるクラスになっており，統計学の意味からも扱いやすい分布族である．比較的最近になってまた注目され，統計的モデリングにも利用されるようになっている．

問題 27.

1) X, Y がそれぞれ平均 μ_X, μ_Y の Poisson 分布に独立にしたがう確率変数のとき，$X+Y$ の分布は平均 $\mu_X + \mu_Y$ の Poisson 分布であることを示せ．

2) Poisson 分布は安定分布ではないことを示せ．

ランダムウォークの n 回目の再帰時間を T_n とすると，T_n/n^2 の分布は $n \to \infty$ の極限で安定分布 (25) に弱収束することをみた．$T_0 = 0$ とおいて，$D_i := T_i - T_{i-1}$ $(i = 1, 2, \ldots, n)$ とおくと，$D_1, D_2, \ldots, D_n$ は独立に同一の分布にしたがっている．独立に同一の分布にしたがう確率変数の和 $T_n = D_1 + \cdots + D_n$ を $1/n^2$ 倍して規格化すると $n \to \infty$ で安定分布に弱収束するわけである．これは中心極限定理における正規分布に似た役割を安定分布が果たすことを示唆するように思われる．実際に中心極限定理の一般化とみなせる次の定理が成立する．

定理． (Levy). $X_1, X_2, \ldots$ を独立に同一の確率分布にしたがう確率変数列とする．$S_n := X_1 + \cdots + X_n$ とおく．ある定数の列 $a_n > 0, b_n$ とある退化していない分布関数 $F(x)$ が存在して,

$$F_n(x) := P\left(\frac{S_n - b_n}{a_n} \leq x\right) \rightsquigarrow F(x)$$

となるとき，F は安定分布である． □

この定理の証明は本質的には難しいものではない．証明の前に必要となる補題を用意する．

補題．$F_n\ (n=1,2,\ldots)$, F, $G_n\ (n=1,2,\ldots)$, G をそれぞれ確率分布関数とする．$F_n \rightsquigarrow F$ かつ $G_n \rightsquigarrow G$ のとき，

$$F_n \star G_n \rightsquigarrow F \star G$$

が成立する．ただし，$\star$ はたたみ込みを表す． □

証明．$F_n\ (n=1,2,\ldots)$, F, $G_n\ (n=1,2,\ldots)$, G の特性関数をそれぞれ ϕ_n, ϕ, ψ_n, ψ とおくと，$F_n \star G_n$ の特性関数は $\phi_n\psi_n$ である．任意の λ で $\phi_n(\lambda) \to \phi(\lambda)$, $\psi_n(\lambda) \to \psi(\lambda)$ であるから，$\phi_n(\lambda)\psi_n(\lambda) \to \phi(\lambda)\psi(\lambda)$．$\phi\psi$ は $F \star G$ の特性関数であるから連続性定理より $F_n \star G_n \rightsquigarrow F \star G$． □

補題．$F_n\ (n=1,2,\ldots)$, G をそれぞれ確率分布関数とする．定数の列 $a_n > 0$, b_n が存在して，

$$F_n(a_n x + b_n) \rightsquigarrow G(x)$$

が成立するものとする．

このとき，$\alpha_n > 0$, β_n が

$$\lim_{n\to\infty} \frac{\alpha_n}{a_n} = 1, \quad \lim_{n\to\infty} \frac{\beta_n - b_n}{a_n} = 0$$

を満たす定数の列であれば，

$$F_n(\alpha_n x + \beta_n) \rightsquigarrow G(x)$$

が成立する． □

問題 28.　上の補題を証明せよ．(ヒント：$F_n(a_n x + b_n) \rightsquigarrow G(x)$ であるから，x を G の任意の連続点とするとき，$\lim_{n\to\infty} F_n(a_n x + b_n) = G(x)$ が成立する．このことを使って $\lim_{n\to\infty} F_n(\alpha_n x + \beta_n) = G(x)$ が成立することを示せばよい．)

補題．$G_n\ (n=1,2,\ldots)$ を確率分布関数の列，G, H を退化していない確率分布関数とする．ある定数の列 $a_n > 0$, b_n に対し，

$$G_n(x) \rightsquigarrow G(x), \quad G_n(a_n x + b_n) \rightsquigarrow H(x)$$

が成立するとき，G と H は同じ型 (type) になる． □

証明．まず，正の実数列 $\{a_n\}$ がある正の有限な値 a に収束する部分列をもつことを示す．

$\{a_n\}$ に，ある正の有限な値に収束する部分列が存在しないとき，

$$\liminf_{n\to\infty} a_n = 0$$

か

$$\limsup_{n\to\infty} a_n = \infty$$

のいずれかが成立する（Bolzano-Weierstrass の定理）．したがって，0 に収束するか，∞ に発散する $\{a_n\}$ の部分列 $\{a_{n'}\}$ がとれることになる．

最初に，すべての n に対し，$b_n = 0$ の場合を考え，あとで，$b_n = 0$ の仮定を除くことにする．0 に収束する $\{a_n\}$ の部分列 $\{a_{n'}\}$ がとれると仮定して矛盾を導く．

$b_n = 0$ なので，仮定より $G_n(a_n x) \rightsquigarrow H(x)$ だから，x を H の連続点とすると，

$$\lim_{n\to\infty} G_n(a_n x) = H(x) \tag{27}$$

である．

このとき，H の任意の正の連続点 $x > 0$ に対し，不等式

$$\lim_{n\to\infty} G_n(a_n x) \leq G(0+) \tag{28}$$

が，H の任意の負の連続点 $x < 0$ に対し，不等式

$$\lim_{n\to\infty} G_n(a_n x) \geq G(0-) \tag{29}$$

が成立することを以下で示す．

まず，(28) を示す．固定した任意の $x > 0, y > 0$ に対し，$n \geq N$ なら $a_n x \leq y$ となる，ある N が存在する．$y > 0$ が G の連続点なら，

$$\lim_{n\to\infty} G_n(a_n x) \leq \lim_{n\to\infty} G_n(y) = G(y)$$

となる．左辺の極限の存在は，(27) よりいえる．また，確率分布関数の不連続点は高々可算個しかないことにより，任意の $\varepsilon > 0$ に対し，

$$G(y) < G(0+) + \varepsilon$$

となる G の連続点 $y > 0$ が存在する．よって，不等式 (28) が示せた．不等式 (29) も同様に示せる．

(27), (28), (29) より，

$$H(x) \leq G(0+) \qquad (x > 0 \text{ が } H \text{ の連続点のとき}), \tag{30}$$

$$H(x) \geq G(0-) \qquad (x < 0 \text{ が } H \text{ の連続点のとき}) \tag{31}$$

となる．

連続点 $x > 0$ は任意に大きくとることができるので，(30) より，$1 \leq G(0+)$ となり，連続点 $x < 0$ は任意に小さくとることができるので，(31) より，$0 \geq G(0-)$ となる．よって G が 0 に退化した分布になるので，仮定に矛盾する．したがって，$b_n = 0$ $(n = 1, 2, \ldots)$ のとき，0 に収束する $\{a_n\}$ の部分列 $\{a_{n'}\}$ は存在しない．

すべての n に対し，$b_n = 0$ の場合に，∞ に発散する $\{a_n\}$ の部分列 $\{a_{n'}\}$ が存在することをを仮定して，矛盾を導く．後で $b_n = 0$ の仮定を除く．

まず，x が H の連続点のとき，

$$\lim_{n\to\infty} G_n(a_n x) = H(x)$$

が成り立つ．また，(28), (29) と同様に，H の任意の正の連続点 $x > 0$ に対し，不等式

$$\lim_{n\to\infty} G_n(a_n x) \geq 1 \tag{32}$$

が，H の任意の負の連続点 $x < 0$ に対し，不等式

$$\lim_{n\to\infty} G_n(a_n x) \leq 0 \tag{33}$$

が成立することが示せる．このとき，$H(x) = 1$ $(x > 0)$, $H(x) = 0$ $(x < 0)$ より，$H(x)$ は $x = 0$ に退化した分布となってしまうので，仮定に矛盾する．したがって，∞ に発散する $\{a_n\}$ の部分列 $\{a_{n'}\}$ は存在しない．

以上から，b_n $(n = 1, 2, \ldots)$ が 0 のとき，$\{a_n\}$ がある正の有限な値 a に収束する部分列をもつことが示せた．

次に $b_n = 0$ $(n = 1, 2, \ldots)$ という仮定を除いても，$\{a_n\}$ がある正の有限な値 a に収束する部分列をもつことを示す．

分布関数 F に対し，F^- を，F と原点に対し対称な分布の分布関数とする．つまり，確率変数 X の分布関数が F のとき，$-X$ の分布関数が F^- である．x が F の連続点のとき，$F^-(x) = 1 - F(-x)$ となる．

仮定より $G_n \rightsquigarrow G$, $G_n(a_n x + b_n) \rightsquigarrow H(x)$ だから，明らかに $G_n^- \rightsquigarrow G^-$, $G_n^-(a_n x + b_n) \rightsquigarrow H^-(x)$ が成り立つ．

よって，p. 45 上の補題より，

$$G_n \star G_n^- \rightsquigarrow G \star G^-$$

と

$$G_n(a_n x + b_n) \star G_n^-(a_n x + b_n) \rightsquigarrow H(x) \star H^-(x)$$

が成立する．

一方，

$$G_n(a_n x + b_n) \star G_n^-(a_n x - b_n) = G_n(a_n x) \star G_n^-(a_n x) \tag{34}$$

が成立することがすぐ見てとれる．なぜなら，確率変数 X の分布関数が $G_n(a_n x)$，Y の分布関数が $G_n^-(a_n x)$ で，X と Y とが独立のとき，$G_n(a_n x) \star G_n^-(a_n x)$ は，$X + Y$ の分布関数，$G_n(a_n x + b_n) \star G_n^-(a_n x - b_n)$ は $(X - b_n) + (Y + b_n)$ の分布関数であるから，(34) の両辺は等しくなる．

したがって，$\tilde{G}_n := G_n \star G_n^-$, $\tilde{G} := G \star G^-$, $\tilde{H} := H \star H^-$ とおけば，$\tilde{G}_n \rightsquigarrow \tilde{G}$, $\tilde{G}_n(a_n x) \rightsquigarrow H(x)$ が成り立ち，補題の仮定で $b_n = 0$ となる場合に帰着する．したがって，$b_n = 0$ が成立しない場合でも，$\{a_n\}$ がある正の有限な値 a に収束する部分列をもつことが示せた．

ここでの議論のように，分布関数 F の代わり対称化した分布 $F \star F^-$ を扱うことは，symmetrization trick と呼ばれる常套手段である．

次に $\{b_{n'}\}$ が有界な数列であることを示す．

もし，$\{b_{n'}\}$ の部分列 $\{b_{n''}\}$ で ∞ に発散するものがとれるとすると，$a_{n''} \to a$ であるから，固定した任意の x, y に対し，$n'' \geq N$ なら $a_{n''}x + b_{n''} \geq y$ となるある正の整数 N が存在する．このとき，$n'' \geq N$ なら，$G_{n''}(a_{n''}x + b_{n''}) \geq G_{n''}(y)$ となる．

x として H の連続点をとると，$H(x) = \lim_{n''\to\infty} G_{n''}(a_{n''}x + b_{n''})$ さらに，y として，G の連続点をとると，

$$H(x) = \lim_{n''\to\infty} G_{n''}(a_{n''}x + b_{n''}) \geq \lim_{n''\to\infty} G_{n''}(y) = G(y).$$

y は任意に大きくとれるので，$H(x) \geq 1$ がいえる．x は H の任意の連続点なので，これは $H(x)$ が確率分布関数であることと矛盾する．

同様に，$\{b_{n'}\}$ の部分列 $\{b_{n''}\}$ で $-\infty$ に発散するものがとれるとすると，H の任意の連続点 x で $H(x) \leq 0$ となることが示せる．これは $H(x)$ が確率分布関数であることと矛盾する．

したがって，$\{b_{n'}\}$ が有界な数列であることが示せた．よって，$\{n'\}$ のある部分列 $\{n''\}$ をとり $\{b_{n''}\}$ がある有限な値 b に収束するようにできる．

以上により，$\{n\}$ の適当な部分列 $\{n''\}$ をとることにより，$\{a_{n''}\}$ はある有限の正の値 a に収束し，$\{b_{n''}\}$ はある有限の値 b に収束するようにできることが示せた．このことを使い補題が証明できる．

仮定より，$G_n(x) \rightsquigarrow G(x)$ だから，

$$G_{n''}(ax + b) \rightsquigarrow G(ax + b)$$

が明らかに成立する．また，

$$G_{n''}(a_{n''}x + b_{n''}) \rightsquigarrow H(x)$$

が成立する．

$a > 0$ より，

$$\lim_{n''\to\infty} \frac{a_{n''}}{a} \to 1, \quad \lim_{n''\to\infty} \frac{b_{n''} - b}{a} = 0$$

だから，p. 48 下の補題より，

$$H(x) = G(ax + b)$$

となる．したがって，H と G とは同じ型 (type) の分布関数である． □

以上の補題を用いることにより，p. 47 の Lévy による定理が容易に証明できる．

証明 (p. 47 の Lévy による定理)．F が安定分布であることを示すには，安定分布の定義より，任意の正の整数 k に対し，$F^{\star k}$ が F と同じ型であることを示せばよい．

仮定より，$F_n(x) \rightsquigarrow F(x)$ なので，p. 48 上の補題より，

$$F_n^{\star k}(x) \rightsquigarrow F^{\star k}(x) \tag{35}$$

がいえる．

F_n は確率変数 $(S_n - b_n)/a_n$ の分布関数で，$F_n^{\star k}$ は，$(S_n - b_n)/a_n$ と同じ分布をもつ k 個の独立な確率変数の和の分布関数だから，

$$\begin{aligned}
F_n^{\star k}(x) =& P\left(\frac{S_n - b_n}{a_n} + \frac{(S_{2n} - S_n) - b_n}{a_n} + \cdots + \frac{(S_{kn} - S_{(k-1)n}) - b_n}{a_n} \leq x\right) \\
=& P\left(\frac{S_{kn} - kb_n}{a_n} \leq x\right) = P\left(\frac{S_{kn} - b_{kn}}{a_{kn}} \leq \frac{a_n x + kb_n - b_{kn}}{a_{kn}}\right) \\
=& F_{kn}\left(\frac{a_n}{a_{kn}} x + \frac{kb_n - b_{kn}}{a_{kn}}\right).
\end{aligned} \tag{36}$$

定理の仮定より，$F_n(x) \rightsquigarrow F(x)$ であるから，任意の正の整数 k に対し

$$F_{kn}(x) \rightsquigarrow F(x)$$

である．また，(35), (36) より，

$$F_{kn}\left(\frac{a_n}{a_{kn}} x + \frac{kb_n - b_{kn}}{a_{kn}}\right) \rightsquigarrow F^{\star k}(x)$$

となる．

したがって，p. 48 下段の補題より，F と $F^{\star k}$ が同じ型であることが示せた． □

問題 29．F を安定分布の分布関数であるとする．$X_1, X_2, \ldots$ が独立に分布 F にしたがう確率変数であるとき，$S_n := X_1 + \cdots + X_n$ とすると，適当な定数の列 $a_n > 0$, b_n をとることにより，

$$F_n(x) := P\left(\frac{S_n - b_n}{a_n} \leq x\right) \rightsquigarrow F(x)$$

となることを示せ．

参考: レポート課題 (2007 年の駒木先生の時のもの)

問題 13．$X_1, X_2, \ldots$ を独立に $[0,1]$ 上の一様分布にしたがう確率変数列とする．このとき，

$$\lim_{n\to\infty}(X_1 X_2 \cdots X_n)^{1/n} = \mathrm{e}^{-1} \quad \text{a.s.}$$

を示せ.

問題 14．区間 $[0,1]$ 上の確率測度を考える．P_n を $(n+1)$ 個の点 $0, 1/n, 2/n, \ldots, n/n = 1$ にそれぞれ確率 $1/(n+1)$ を与える（離散型) 確率測度とする．$n \to \infty$ で，P_n は $[0,1]$ 上の Lebesgue 測度に弱収束することを示せ.

問題 19．$X_1, X_2, \ldots, X_n$ は独立に同一の分布にしたがう確率変数列，$M_n := \max(X_1, X_2, \ldots, X_n)$ とする．$G(x)$ を X_i の確率分布の分布関数として，

$$\lim_{x\to\infty} \mathrm{e}^x\{1 - G(x)\} = c, \quad c > 0$$

が成立するとき，

$$\lim_{n\to\infty} P(M_n - \log(cn) \leq x) = \mathrm{e}^{-\mathrm{e}^{-x}}$$

となることを示せ.

問題 23．Cauchy 分布は，密度関数が，

$$f(x) := \frac{1}{\pi}\frac{c}{x^2 + c^2} \quad (c > 0)$$

で与えられる確率分布である．Cauchy 分布の特性関数が，

$$\phi(\lambda) = \mathrm{e}^{-c|\lambda|}$$

となることを示せ.

問題 26．$X_1, X_2, \ldots, X_n$ が独立に確率密度 (25) をもつ確率分布にしたがうとき，$(X_1 + X_2 + \cdots + X_n)/n^2$ の分布は X_1 の分布と一致することを特性関数 (24) を用いて示せ.

問題 27.

1) X, Y がそれぞれ平均 μ_X, μ_Y の Poisson 分布に独立にしたがう確率変数のとき，$X + Y$ の分布は平均 $\mu_X + \mu_Y$ の Poisson 分布であることを示せ.

2) Poisson 分布は安定分布ではないことを示せ.

以上

Midterm exam (last year)

1. Explain what you know on the relations among "convergence in probability", "almost sure convergence" and "convergence in the mean". Explain the differences of these notions using the following examples: Let ω be the uniform random variable on $[0,1]$ and let

   $$X_1(\omega)=1, X_2(\omega)=I_{(0,1/2]}(\omega), X_3(\omega)=I_{(1/2,1]}(\omega),$$
   $$X_4(\omega)=I_{(0,1/4]}(\omega), X_5(\omega)=I_{(1/4,1/2]}(\omega), X_6(\omega)=I_{(1/2,3/4]}(\omega), X_7(\omega)=I_{(3/4,1]}(\omega), \dots$$
   $$X_{2^k}=I_{(0,1/2^k]}(\omega), \dots, X_{2^{k+1}-1}(\omega)=I_{(1-1/2^k,1]}(\omega), \dots$$

   $$Y_1(\omega)=1, Y_2(\omega)=2I_{(0,1/2]}(\omega), Y_3(\omega)=2I_{(1/2,1]}(\omega),$$
   $$Y_4(\omega)=4I_{(0,1/4]}(\omega), Y_5(\omega)=4I_{(1/4,1/2]}(\omega), Y_6(\omega)=4I_{(1/2,3/4]}(\omega), Y_7(\omega)=4I_{(3/4,1]}(\omega), \dots$$
   $$Y_{2^k}=2^kI_{(0,1/2^k]}(\omega), \dots, Y_{2^{k+1}-1}(\omega)=2^kI_{(1-1/2^k,1]}(\omega), \dots$$

2. Let $X_1, X_2, \dots$ be independent random variables having the exponential distribution: $P(X \geq c)=e^{-c}$, $c \geq 0$. Let $M_n=\max(X_1,\dots,X_n)$. Obtain the limiting distribution of M_n as $n \to \infty$ after appropriate location shift (depending on n).

3. Let X be a random variable on $\mathbb{R}^1$ and let $\phi(\lambda)=\mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda X})$ be its characteristic function. Explain what you know about the "inversion theorem". Also prove that

   $$\mathrm{E}[|X|^k]<\infty \ \Rightarrow \phi^k(0)=\mathrm{i}^k\mathrm{E}[X^k].$$

Homework No.2 (last year)

Prob.1. Lamperti, Sec.14, Prob.3 (p.81).

Prob.2. Prove that the set of distributions of $X_1, X_2, \ldots$ is tight if $\mathrm{E}[|X_n|]$ is bounded in n. Give an example of a tight family of random variables, such that $\sup_n \mathrm{E}[|X_n|] = \infty$.

Prob.3. The proof of the inversion formula for the characteristic function in Sec.15 of Lamperti uses convolution with the normal distribution $N(0, \sigma^2)$. Discuss that $N(0, \sigma^2)$ can be replaced by $N(\mu\sigma, \sigma^2)$, where μ is fixed and $\sigma \downarrow 0$. How does this modification affect the inversion formula if there exist point masses at α and/or β?

Prob.4. Let $\phi(\lambda) = \mathrm{E}[\mathrm{e}^{\mathrm{i}\lambda X}]$ be the characteristic function of a symmetric random variable on $\mathbb{R}^1$ and suppose that

$$\phi(\lambda) = 1 - |\lambda|^{\alpha} + o(|\lambda|^{\alpha})$$

as $\lambda \to 0$. What can you say about the limiting behavior of the infinite convolution of this distribution (in view of the results on stable distribution)? Can you say something about the behavior of the tail of the distribution of X?

安定分布が中心極限定理における正規分布に似た役割を果たすことを見た (p. 44 の定理)．X_i $(i=1,2,\ldots)$ が独立に同一の分布にしたがう確率変数であるとき，$S_n := X_1 + X_2 + \cdots + X_n$ おく．ある定数列 $a_n > 0$, b_n が存在して，$(S_n - b_n)/a_n$ の分布がある確率分布に弱収束するとき，その収束先の分布が安定分布なのであった．収束先がどの安定分布になるのかは，X_i の分布によって決まるのであるが，それはこの定理からはわからない．

もし，X_i $(i=1,2,\ldots)$ に対し，ある定数列 $a_n > 0$, b_n が存在して，$(S_n - b_n)/a_n$ の分布が安定分布 P に弱収束するとき，X_i の分布は安定分布 P の**吸引域**あるいは**牽引域** (domain of attraction) に属するという．吸引域については詳細な研究が完成しているが，ここでは簡単な例だけを見る．

例．G をある 0 に対し対称な確率分布の分布関数とする．このとき，

$$\lim_{x\to\infty} x\{1 - G(x)\} = c, \quad c > 0 \tag{37}$$

が成立すれば，G は Cauchy 分布に吸引されることが示せる．

問題 30．上の例で，積分

$$\frac{1-\phi(\lambda)}{2\lambda} = \int_0^\infty \frac{1-\cos\lambda x}{\lambda} \mathrm{d}G(x)$$

を適当に評価することにより

$$\phi(\lambda) = 1 - c\pi|\lambda| + \mathrm{o}(\lambda) \quad \text{as} \quad \lambda \to 0$$

を示せ．

問題 31．前問の結果を利用して G が Cauchy 分布に吸引されることを示せ．

8 無限分解可能分布

安定分布は，数学的に自然なクラスであったが，問題 27 で見たように Poisson 分布は安定分布ではない．無限分解可能分布は，安定分布と Poisson 分布を含む自然な確率分布のクラスである．安定分布で分散をもつのは正規分布だけであったが，無限分解可能分布で分散をもつ分布は無数にある．

ひとことで言って，正規分布に対応する確率過程が Brown 運動 (Brown 運動については後でとりあげる) なのに対し，無限分解可能分布に対応する確率過程が Lévy 過程と呼ばれるものである．Lévy 過程は，物理現象や金融工学など現実の現象のモデリングに利用され，最近注目を集めている．

定義．(無限分解可能分布) 確率分布関数 F が，任意の 2 以上の整数 $k = 2, 3, \ldots$ に対し，ある確率分布関数 G_k の k 回のたたみ込み $G_k^{\star k}$ となるとき，F は**無限分解可能** (infinitely divisible) という． □

例．Poisson 分布の特性関数は，

$$\phi(\lambda) = \sum_{n=0}^{\infty} \mathrm{e}^{-\mu} \frac{\mu^n}{n!} \mathrm{e}^{\mathrm{i}\lambda n} = \exp(\mu \mathrm{e}^{\mathrm{i}\lambda} - \mu)$$

である．よって，任意の k に対し，

$$\phi(\lambda) = \exp(\mu \mathrm{e}^{\mathrm{i}\lambda} - \mu) = \left\{\exp\left(\frac{\mu}{k}\mathrm{e}^{\mathrm{i}\lambda} - \frac{\mu}{k}\right)\right\}^k$$

となるから，平均 μ の Poisson 分布は平均 μ/k の Poisson 分布 k 個のたたみ込みであることがわかる．

問題 32．安定分布は無限分解可能分布であることを示せ．

問題 33．ガンマ分布の確率密度関数は，

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)} \dfrac{x^{\alpha-1}}{\beta^\alpha} \exp\left(-\dfrac{x}{\beta}\right) & (x \geq 0), \\ 0 & (x < 0) \end{cases}$$

で与えられる．ただし，$\alpha > 0, \beta > 0$ である．ガンマ分布が無限分解可能であることを示せ．

問題 34．　逆ガウス分布 (p. 44 参照) は無限分解可能であることを示せ．

無限分解可能分布について，三角配列 (triangular array) を用いて特徴付ける以下の定理が知られている．

定理． 任意の正の整数 n に対し，独立に同一の分布にしたがう n 個の確率変数 $X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}$ をとり，その和 $X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)}$ の確率分布関数

$$F_n(x) := P(X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)} \leq x)$$

が，$n \to \infty$ で，ある確率分布関数 $F(x)$ に弱収束するとき，F は無限分解可能である．

□

(証明略)

安定分布について考えたときは，n が変わっても $X_i^{(n)}$ の分布は同じものとして分布の型を扱った．無限分解可能分布は，n により，$X_i^{(n)}$ の分布が変わることを許すので，ずっと多くの分布が含まれることになる．

無限分解可能分布を理解するうえで，以下の複合 Poisson 分布は大変重要である．

例．複合 Poisson 分布

確率変数 $X_1, X_2, \ldots$ は独立にある同一の分布にしたがうものとする．N を期待値 μ の Poisson 分布にしたがう確率変数とする．このとき，N 個の和

$$S_N := \sum_{k=1}^{N} X_k$$

の分布を**複合 Poisson 分布** (compound Poisson distribution) であるという．例えば，Poisson 過程に従って客が来る店の１日の売り上げは，それぞれの客の購買額が独立で同一の分布にしたがうとすると，複合 Poisson 分布になる．

S_N の特性関数 $\psi(\lambda)$ は，X_1 の特性関数を $\phi(\lambda)$ とすると，

$$\begin{aligned}\psi(\lambda) =& \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda S_N}) = \sum_{n=0}^{\infty} \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda S_N} | N = n) P(N = n) \\ =& \sum_{n=0}^{\infty} \mathrm{E}(\mathrm{e}^{\mathrm{i}\lambda (X_1 + \cdots + X_n)}) P(N = n) = \sum_{n=0}^{\infty} \{\phi(\lambda)\}^n \mathrm{e}^{-\mu} \frac{\mu^n}{n!} \\ =& \exp(\mu\phi(\lambda) - \mu)\end{aligned} \tag{38}$$

となる．任意の正の整数 k に対して，

$$\exp(\mu\phi(\lambda) - \mu) = \big\{\exp(\frac{\mu}{k}\phi(\lambda) - \frac{\mu}{k})\big\}^k$$

となる．右辺にあらわれる $\exp\{(\mu/k)\phi(\lambda) - (\mu/k)\}$ は期待値 μ/k の Poisson 分布で混合をとった複合 Poisson 分布の特性関数であるから，S_N の分布は無限分解可能であることがわかる．なお，条件付期待値の厳密な取り扱いは後で行うが，式 (39) の理解に問題はないだろう． □

次の定理から，複合 Poisson 分布はほとんど無限分解可能分布そのものであることがわかる．

定理． 無限分解可能分布は，複合 Poisson 分布そのものか，あるいはある複合 Poisson 分布の列の弱収束する極限の確率分布になっている．逆に，複合 Poisson 分布および複合 Poisson 分布の列が弱収束する極限の確率分布は無限分解可能である． □

(証明略)

この定理より，正規分布やガンマ分布は，ある複合 Poisson 分布の列が弱収束する収束先の分布として表すことができることになる．

例．X_k $(k = 1, 2, \ldots)$ は，確率 1/2 で値 $-\varepsilon$ を確率 1/2 で値 ε をとる独立な確率変数とする．X_k の特性関数は，

$$\phi(\lambda) = \frac{\mathrm{e}^{\mathrm{i}\lambda\varepsilon} + \mathrm{e}^{-\mathrm{i}\lambda\varepsilon}}{2} = \cos(\varepsilon\lambda)$$

である．N を期待値 ε^{-2} の Poisson 分布にしたがう確率変数とする．このとき，N 個の和 $S_N := \sum_{k=1}^N X_k$ の特性関数は，

$$\psi(\lambda) = \exp(\varepsilon^{-2}\phi(\lambda) - \varepsilon^{-2}) = \exp\left\{\frac{\cos(\varepsilon\lambda)-1}{\varepsilon^2}\right\} \to \mathrm{e}^{-\frac{\lambda^2}{2}}, \quad \text{as} \quad \varepsilon \to \infty \tag{39}$$

となり，標準正規分布の特性関数に収束する．このことから正規分布は複合 Poisson 分布列の極限であることがわかる．

無限分解可能分布の特性関数の一般的な形も知られている．無限分解可能分布の分散が存在する場合は，やや簡単な形になる．

定理． (Kolmogorov) 平均 μ，分散 σ^2 の無限分解可能分布の特性関数は，ある $\nu(\mathbb{R}) = \sigma^2$ となる $\mathbb{R}$ 上の有限測度 ν を用いて，

$$\phi(\lambda) = \exp\left(\mathrm{i}\lambda\mu + \int \frac{\mathrm{e}^{\mathrm{i}\lambda x} - 1 - \mathrm{i}\lambda x}{x^2}\nu(\mathrm{d}x)\right) \tag{40}$$

と表せる．逆に，任意の $\mathbb{R}$ 上の有限測度 ν に対し，(40) は分散をもつ無限分解可能分布の特性関数になる．ν は無限分解可能分布の canonical measure と呼ばれる．□

問題 35. ν としてどのような測度をとれば，(40) は Poisson 分布の特性関数になるか．

問題 36. ν としてどのような測度をとれば，(40) は 正規分布の特性関数になるか．

一般の無限分解可能分布については以下の結果が知られている．

定理． (Lévy-Khinchin) 無限分解可能分布の特性関数は，ある $\nu(\{0\}) = 0$, $\int x^2/(1+x^2)\nu(\mathrm{d}x) < \infty$ を満たす $\mathbb{R}$ 上の測度と実数 c, σ を用いて，

$$\phi(\lambda) = \exp\left\{\mathrm{i}c\lambda - \frac{\sigma^2}{2}\lambda^2 + \int\left(\mathrm{e}^{\mathrm{i}\lambda x} - 1 - \frac{\mathrm{i}\lambda x}{1+x^2}\right)\nu(\mathrm{d}x)\right\} \tag{41}$$

と表せる．逆に，$\nu(\{0\}) = 0$, $\int x^2/(1+x^2)\nu(\mathrm{d}x) < \infty$ を満たす任意の $\mathbb{R}$ 上の測度 ν と任意の実数 c, σ に対し，(41) はある無限分解可能分布の特性関数になる．□

以下では，Lévy 測度の解釈について補足する．説明の簡単のために，複合ポアソン分布として，飛びの分布が $a, b \in \mathbb{R}$ の 2 点のみとる分布を考えよう．それぞれの確率を p, q, $p+q=1$ とする．またポアソン過程の intensity を μ とすると，複合ポアソン分布の特性関数は

$$\begin{aligned}\psi(\lambda) &= \exp\left(\mu(p\mathrm{e}^{\mathrm{i}\lambda a} + q\mathrm{e}^{\mathrm{i}\lambda b} - 1)\right)\\ &= \exp\left(\mu(p(\mathrm{e}^{\mathrm{i}\lambda a} - 1) + q(\mathrm{e}^{\mathrm{i}\lambda b} - 1))\right)\end{aligned}$$

となる．

次に飛びのサイズ a, b が共通であるような二つの独立な複合ポアソン分布の畳み込みを考えると，畳み込みの特性関数は

$$\begin{aligned}\psi(\lambda) &= \exp\big(\mu_1(p_1\mathrm{e}^{\mathrm{i}\lambda a} + q_1\mathrm{e}^{\mathrm{i}\lambda b} - 1) + \mu_2(p_2\mathrm{e}^{\mathrm{i}\lambda a} + q_2\mathrm{e}^{\mathrm{i}\lambda b} - 1)\big) \\ &= \exp\big((\mu_1+\mu_2)(\tilde{p}\mathrm{e}^{\mathrm{i}\lambda a} + \tilde{q}\mathrm{e}^{\mathrm{i}\lambda b} - 1)\big), \qquad \tilde{p} = 1 - \tilde{q} = \frac{\mu_1 p_1 + \mu_2 p_2}{\mu_1 + \mu_2}\end{aligned}$$

と書ける．これも一つの複合ポアソン分布である．このように飛びのサイズを共通する二つの複合ポアソン分布を畳み込むと，特定のサイズの飛びについては，どちらの intensity から来たものであるかは区別がつかないことになる．

次に飛びのサイズが a で intensity が μ_1 であるようなポアソン分布と，飛びのサイズが b で intensity が μ_2 であるようなポアソン分布を畳み込むと，その特性関数は

$$\begin{aligned}\psi(\lambda) &= \exp\big(\mu_1(\mathrm{e}^{\mathrm{i}\lambda a} - 1) + \mu_2(\mathrm{e}^{\mathrm{i}\lambda b} - 1)\big) \qquad (42) \\ &= \exp\big((\mu_1+\mu_2)(p'(\mathrm{e}^{\mathrm{i}\lambda a} - 1) + q'(\mathrm{e}^{\mathrm{i}\lambda b} - 1))\big), \quad p' = 1 - q' = \frac{\mu_1}{\mu_1 + \mu_2}\end{aligned}$$

と書ける．これも上の形と区別がつかない．このことより，無限分解可能分布を考える時には，飛びのサイズごとのポアソン分布が畳みこまれていると考え，それぞれの intensity を指定すればよいことがわかる．

以上の観点からは無限分解可能分布を (42) 式のまま表すのがわかりやすい．いま点 a に重み μ_1, 点 b に重み μ_2 を持つ測度を ν と表すと， (42) 式は

$$\psi(\lambda) = \exp\Big(\int_{-\infty}^{\infty}(\mathrm{e}^{\mathrm{i}\lambda x} - 1)\nu(\mathrm{d}x)\Big) \qquad (43)$$

の形に書けるこどわかる．無限分解可能分布の特性関数の一般形は，位置のシフトを除いて

$$\psi(\lambda) = \exp\Big(\int_{-\infty}^{\infty}(\mathrm{e}^{\mathrm{i}\lambda x} - 1 - \frac{\mathrm{i}\lambda x}{1+x^2})\nu(\mathrm{d}x)\Big) \qquad (44)$$

と書けていたが，$-\mathrm{i}\lambda x/(1+x^2)$ の部分は，位置 (平均) を調整することにより，飛びが無限に小さい部分に関して intensity が無限に大きくてもよいことを処理するために必要とされる部分であって，上に示した ν の解釈に影響を与えるものではない．すなわち (44) 式の $\nu(\mathrm{d}x)$ は飛びの大きさが x の部分の intensity を表している．

飛びの大きさが x で intensity が μ のポアソン分布の分散は $x^2\mu$ であるから分散に注目すると，$\nu(\mathrm{d}x)/x^2$ の形の測度を用いればよい．また原点近傍で $x^2\nu(\mathrm{d}x)$ の積分が有限である条件は，位置を調整した後には，無限に細かい飛びの和の分散が有限であれば確率過程が存在することを保証している．さらに，無限分解可能分布はモーメントを持たないことがあるが，それは Lévy 測度の $\pm\infty$ での裾の重さに関連しており，原点回りの挙動とは無関係である．

9 条件付期待値

今まで，独立な確率変数の最大値や和の確率分布について考えてきた．確率過程などのもっと一般の確率システムを扱うためには，独立でない確率変数の組について考える必要がある．そのために必要となるのが，条件付確率や条件付期待値の概念である．(測度論的な) 確率論での条件付期待値の定義は，初めて見る人にはやや取り付きにくいところがあるが，是が非でも理解するべき大変重要なポイントなので考え方に良く慣れておく必要がある．

測度論的な確率論における条件付確率の定義を与える前に，Radon-Nikodym の定理 (の有限測度が絶対連続な場合に限ったもの) の復習をしておく．

定義．P, Q を可測空間 $(\Omega, \mathcal{F})$ 上の有限測度とする．$P(E) = 0$ となる任意の $E \in \mathcal{F}$ について $Q(E) = 0$ が成立するとき，Q は P に対し**絶対連続** (absolutely continuous) であるといい，$Q \ll P$ で表す． □

f が確率空間 $(\Omega, \mathcal{F}, P)$ 上の非負可積分関数であるとき，

$$Q(E) := \int_E f \mathrm{d}P$$

で定義される $(\Omega, \mathcal{F})$ 上の測度は P に対し絶対連続であることは容易にわかる．実はこの逆が成立する．

定理 (Radon-Nikodym)．P, Q を可測空間 $(\Omega, \mathcal{F})$ 上の有限測度とする．Q が P に対し絶対連続のとき，$(\Omega, \mathcal{F}, P)$ 上のある非負可積分関数 f が存在して，任意の $E \in \mathcal{F}$ に対し

$$Q(E) := \int_E f \mathrm{d}P$$

となる． □

定義．確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数 $X_\lambda : \Omega \to \mathbf{R}$ の集合 $\{X_\lambda : \lambda \in \Lambda\}$ (Λ は任意の添字集合，λ は添字) に対し， すべての X_λ が，$\mathcal{X}$ に対して可測となるような最小の Ω の σ-加法族 $\mathcal{X}$ を X_λ $(\lambda \in \Lambda)$ の**生成する** (generate) σ-加法族と呼び，$\sigma(X_\lambda : \lambda \in \Lambda)$ と表す． □

定義から，当然 $\sigma(X_\lambda : \lambda \in \Lambda) \subset \mathcal{F}$ が成立する．このことから，$\sigma(X_\lambda : \lambda \in \Lambda)$ を X_λ $(\lambda \in \Lambda)$ の生成する部分 σ-加法族と呼ぶこともある．

問題 37. 確率変数 X 1 つだけが生成する部分 σ-加法族は，$\mathbb{R}$ の Borel 集合の X による逆像すべてからなる集合，すなわち $\{C : C = X^{-1}(A),\ A \text{ は Borel 集合}\}$ であることを示せ．

確率変数が有限個の値しかとらない場合に，条件付期待値の初等的な定義について復習してから，(測度論的な) 確率論における条件付期待値の定義を与える．

$(\Omega, \mathcal{F}, P)$ が確率空間，X を $x_1, x_2, \ldots, x_n$ のいずれかの値をとる確率変数，Y を $y_1, y_2, \ldots, y_m$ のいずれかの値をとる確率変数とする．初等的な確率論では，条件付確率を

$$P(Y = y_j | X = x_i) := \frac{P(Y = y_j, X = x_i)}{P(X = x_i)},$$

条件付期待値を

$$\mathrm{E}(Y | X = x_i) = \sum_{j=1}^{m} y_j P(Y = y_j | X = x_i)$$

で定義する．また，確率変数 $Z := \mathrm{E}(Y|X)$ を，$X(\omega) = x_i$ であるとき，ω に対し実数値 $\mathrm{E}(Y|X = x_i)$ を返す確率空間上の関数とする．

ここで $A_i := \{\omega : X(\omega) = x_i\}$ とおけば，$\sigma(X)$ の要素は，A_i $(i = 1, 2, \ldots, n)$ の中から任意個を選んでその和集合をとったものなので，2^n 個ある．

任意の i で，$Z(\omega)$ は A_i 上では一定の値をとる．したがって，$Z(\omega)$ は $\sigma(X)$-可測な関数である．

$Z(\omega)$ が各 A_i 上では一定の値 $z_i := E(Y|X = x_i)$ をとることから，

$$\begin{aligned}\int_{A_i} Z \mathrm{d}P =& z_i \int_{A_i} 1 \mathrm{d}P = z_i P(X = x_i) = \sum_j y_j P(Y = y_j | X = x_i) P(X = x_i) \\ =& \sum_j y_j P(Y = y_j, X = x_i) = \int_{A_i} Y \mathrm{d}P\end{aligned}$$

となる．$E \subset \sigma(X)$ は，A_i $(i = 1, 2, \ldots, n)$ のうち幾つかの和集合になっているので，任意の $E \subset \sigma(X)$ に対し，

$$\int_E Z \mathrm{d}P = \int_E Y \mathrm{d}P \tag{45}$$

がいえる．

以上から，確率変数 $Z := \mathrm{E}(Y|X)$ は $\sigma(X)$-可測で，式 (45) を満たすということが解った．(測度論的な) 確率論における条件付期待値は，この 2 つの性質にもとづいて定義される．

定義． Y を $\mathrm{E}(|Y|) < \infty$ を満たす確率変数とする．確率変数 X に対し，条件

1) Z は $\sigma(X)$-可測

2) 任意の $A \subset \sigma(X)$ に対し $\displaystyle\int_A Z \mathrm{d}P = \int_A Y \mathrm{d}P$

を満たす確率変数 Z が存在するとき，Z は X を与えたもとでの Y の**条件付期待値** (conditional expectation) であるという．Radon-Nikodym の定理により，条件付期待値は常に存在する．

とくに，Y が可測集合 C の定義関数 $\mathbf{1}_C(\omega)$ のとき，Z は X を与えたもとでの C の**条件付確率** (conditional probability) であるという． □

この条件付期待値の定義によれば，2 つの異なる確率変数 X_1, X_2 に対し，もし，$\sigma(X_1) = \sigma_(X_2)$ が成立すれば，X_1 を与えたもとでの Y の条件付期待値は，X_2 を与えたもとでの Y の条件付期待値でもあることになる．このことから，条件付期待値について考えるとき，確率変数 X そのものよりも，X の生成する $\mathcal{F}$ の 部分 σ-加法族の方が本質的であることがわかる．

そこで，部分 σ-加法族のもとでの条件付期待値の定義を直接あたえる．確率過程等では，無限個の確率変数を与えたときの条件付確率などを考える必要が生じる．このようなとき，条件付期待値の本質である部分 σ-加法族に基づく条件付期待値の定義を与えておいた方が見通しが良くなる．

定義． Y を $\mathrm{E}(|Y|) < \infty$ を満たす確率変数，$\mathcal{F}'$ を $\mathcal{F}$ の任意の部分 σ-加法族とする．確率変数 Z が条件

1) Z は $\mathcal{F}'$-可測

2) 任意の $A \subset \mathcal{F}'$ に対し $\displaystyle\int_A Z \mathrm{d}P = \int_A Y \mathrm{d}P$

を満たすとき，Z は $\mathcal{F}'$ のもとでの Y の**条件付期待値** (conditional expectation) であるという．

とくに，Y が可測集合 C の定義関数 $\mathbf{1}_C(\omega)$ のとき，Z は $\mathcal{F}'$ のもとでの C の**条件付確率** (conditional probability) であるという． □

定理． Y が $\mathrm{E}(|Y|) < \infty$ を満たす確率変数，$\mathcal{F}'$ が $\mathcal{F}$ の部分 σ-加法族のとき，条件 1), 2) を満たす確率変数 Z が存在する．また，1), 2) の条件を満たす任意の 2 つの確率変数 Z_1, Z_2 に対し，

$$Z_1 = Z_2 \quad \mathrm{a.s.}$$

が成立する． □

証明．仮定より，$\mathrm{E}(|Y|) < \infty$ である．まず，$Y \geq 0$ の場合について考える．集合関数

$$m(A) := \int_A Y(\omega) \mathrm{d}P \tag{46}$$

は，σ-加法族 $\mathcal{F}$ 上の測度を定義する．m を部分 σ-加法族 $\mathcal{F}'$ 上に制限したものは，$\mathcal{F}'$ 上の測度になる．この測度を同じ記号 m で表す．また，P を部分 σ-加法族 $\mathcal{F}'$ 上に制限し

たものは，$\mathcal{F}'$ 上の測度になる．この測度を同じ記号 P で表す．定義式 (46) より，$\mathcal{F}'$ 上の測度 m は P に対し絶対連続であるから，Radon-Nikodym の定理により，

$$m(A) = \int_A Z(\omega)\mathrm{d}P \qquad \text{for} \quad \text{all} \quad A \subset \mathcal{F}'$$

をみたす $\mathcal{F}'$-可測な密度関数 $Z(\omega)$ が存在する．$Z(\omega)$ は条件付期待値の定義の 1), 2) を満たすので，条件付期待値である．

$Y \geq 0$ が成立しない場合について考える．このとき，適当な確率変数 $Y^+ \geq 0, Y^- \geq 0$ をとり，$Y = Y^+ - Y^-$ と表すことができるので，Y^+, Y^- それぞれについて，前と同様に測度 m^+, m^- を定義すれば，Radon-Nikodym の定理により，m^+, m^- それぞれに対応する $\mathcal{F}'$-可測な密度関数 $Z^+(\omega), Z^-(\omega)$ が存在する．$Z := Z^+ - Z^-$ とおけば，$Z(\omega)$ は条件付期待値の定義の 1), 2) を満たすので，条件付期待値である．

次に条件付期待値が測度 0 の集合上での違いを除き一意に決まることを示す．Z_1, Z_2 がともに条件付期待値であるとき，$A := \{\omega : Z_1(\omega) > Z_2(\omega)\}$ とすると，A は $\mathcal{F}'$ に属する．条件付期待値の定義より，

$$\int_A Z_1(\omega)\mathrm{d}P = \int_A Y(\omega)\mathrm{d}P = \int_A Z_2(\omega)\mathrm{d}P$$

だから，

$$\int_A (Z_1 - Z_2)\mathrm{d}P = 0.$$

また，A 上では $Z_1(\omega) - Z_2(\omega) > 0$ だから，$P(A) = 0$. 同様に，$P(\{\omega : Z_1(\omega) < Z_2(\omega)\}) = 0$ だから，$P(\{\omega : Z_1(\omega) \neq Z_2(\omega)\}) = 0$. □

この定理により，条件付期待値は常に存在し，測度 0 の集合での違いを除いて一意に決まることが保証される．$\mathcal{F}'$ のもとでの Y の条件付期待値が，(測度 0 の集合での違いを除いて) 一意に決まることが保証されるので，条件付期待値を $\mathrm{E}(Y|\mathcal{F}')$ と表すことにする．また，$\mathcal{F}' = \sigma(X_\lambda : \lambda \in \Lambda)$ のときは，$\mathrm{E}(Y|\mathcal{F}') = \mathrm{E}(Y|\sigma(X_\lambda : \lambda \in \Lambda))$ を単に $\mathrm{E}(Y|X_\lambda : \lambda \in \Lambda)$ と表す．

次の定理は，条件付確率の性質をまとめたものである．どの性質もよく使われる．とくに 3), 4) に十分慣れておく必要がある．

定理． Y を $\mathrm{E}(|Y|) < \infty$ を満たす確率変数，$\mathcal{F}'$ を $\mathcal{F}$ の部分 σ-加法族とする．このとき，以下が成立する．

1) 条件付期待値 $\mathrm{E}(Y|\mathcal{F}')$ は Y について線形．

2) $Y \geq 0 \quad \text{a.s.}$ のとき，$\mathrm{E}(Y|\mathcal{F}') \geq 0 \quad \text{a.s.}$ であり，

$$|\mathrm{E}(Y|\mathcal{F}')| \leq \mathrm{E}(|Y|\,|\mathcal{F}') \quad \text{a.s.}$$

3) Z が $\mathcal{F}'$-可測で，$\mathrm{E}(|YZ|) < \infty$ のとき，

$$\mathrm{E}(YZ|\mathcal{F}') = Z\mathrm{E}(Y|\mathcal{F}') \quad \mathrm{a.s.}$$

とくに，Y 自体が $\mathcal{F}'$-可測 のとき，

$$\mathrm{E}(Y|\mathcal{F}') = Y \quad \mathrm{a.s.}$$

4) $\mathcal{F}''$ が $\mathcal{F}'$ の部分 σ-加法族のとき，

$$\mathrm{E}\{\mathrm{E}(Y|\mathcal{F}')|\mathcal{F}''\} = \mathrm{E}(Y|\mathcal{F}'') \quad \mathrm{a.s.}$$

□

証明．4) の証明のみあたえる．条件付期待値の定義より，$\mathrm{E}\{\mathrm{E}(Y|\mathcal{F}')|\mathcal{F}''\}$ は $\mathcal{F}''$-可測であるから，任意の $A \in \mathcal{F}''$ に対し，

$$\int_A \mathrm{E}\{\mathrm{E}(Y|\mathcal{F}')|\mathcal{F}''\}\mathrm{d}P = \int_A Y\mathrm{d}P \tag{47}$$

を示せば，$\mathrm{E}\{\mathrm{E}(Y|\mathcal{F}')|\mathcal{F}''\}$ が $\mathrm{E}(Y|\mathcal{F}'')$ と一致することが証明できる．

まず，$A \in \mathcal{F}''$ であるから，

$$\int_A \mathrm{E}\{\mathrm{E}(Y|\mathcal{F}')|\mathcal{F}''\}\mathrm{d}P = \int_A \mathrm{E}(Y|\mathcal{F}')\mathrm{d}P \tag{48}$$

である．さらに，$\mathcal{F}'' \subset \mathcal{F}'$ であることより，$A \in \mathcal{F}'$ だから，

$$\int_A \mathrm{E}(Y|\mathcal{F}')\mathrm{d}P = \int_A Y\mathrm{d}P \tag{49}$$

である．したがって，(48), (49) より，(47) が示せた． □

問題 38．上記の定理の 3) を証明せよ．(ヒント：右辺が $\mathcal{F}'$-可測であることはすぐいえる．あとは $A \subset B'$ 上での積分が YZ の積分と一致することを示せばよい．これは，Z が $\mathcal{F}'$-可測集合の定義関数の場合の証明から始める常套手段を用いることにより示せる．)

初等的な確率のテキストで与えられている，確率密度関数を用いた条件付期待値の定義が問題なく適用できる場合には，初等的な定義による条件付期待値と測度論的な定義による条件付期待値が一致することを確認しよう．

確率変数 X と Y の同時確率分布が密度関数 $f(s,t)$ をもち，f は $\mathbb{R}^2$ 上の正値連続関数とする．このとき，$X = s$ という条件を与えたもとでの Y の条件付期待値の初等的な定義は，

$$\mathrm{E}(Y|X=s) = \frac{\int_{-\infty}^{\infty} tf(s,t)\mathrm{d}t}{\int_{-\infty}^{\infty} f(s,t)\mathrm{d}t} =: g(s)$$

である．

定理． 確率変数 $g(X(\omega))$ は，X を与えたもとでの Y の (測度論的な定義による) 条件付期待値である． □

証明．$g(X)$ は，$\mathcal{F}(X)$ に対し可測であるから，任意の $A \subset \mathcal{F}(X)$ に対し，

$$\int_A g(X)\mathrm{d}P = \int_A Y\mathrm{d}P \tag{50}$$

が成立することを示せばよい．

$(X(\omega), Y(\omega))$ は Ω の点を $\mathbb{R}^2$ の点に移すので，レジュメの p. 5 の定理により，(50) の両辺の積分はもともと Ω 上の積分だが，$\mathbb{R}^2$ 上の積分に移して考えることができる．$A \subset \mathcal{F}(X)$ より，ある Borel 可測集合 C をとり，$A = X^{-1}(C)$ とできる．

(50) の右辺は,

$$\int_A Y\mathrm{d}P = \int_\Omega Y\mathbf{1}_A\mathrm{d}P = \int_{\mathbb{R}^2} t\mathbf{1}_C(s)f(s,t)\mathrm{d}s\mathrm{d}t.$$

左辺は,

$$\begin{aligned}
\int_A g(X)\mathrm{d}P = \int_\Omega g(X)\mathbf{1}_A\mathrm{d}P &= \int_{\mathbb{R}^2} g(s)\mathbf{1}_C(s)f(s,t)\mathrm{d}s\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\int_{-\infty}^{\infty}\tilde{t}f(s,\tilde{t})\mathrm{d}\tilde{t}}{\int_{-\infty}^{\infty}f(s,\tilde{t})\mathrm{d}\tilde{t}}\mathbf{1}_C(s)f(s,t)\mathrm{d}s\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \frac{\int_{-\infty}^{\infty}\tilde{t}f(s,\tilde{t})\mathrm{d}\tilde{t}}{\int_{-\infty}^{\infty}f(s,\tilde{t})\mathrm{d}\tilde{t}}\mathbf{1}_C(s)\left\{\int_{-\infty}^{\infty}f(s,t)\mathrm{d}t\right\}\mathrm{d}s \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{t}f(s,\tilde{t})\mathbf{1}_C(s)\mathrm{d}\tilde{t}\mathrm{d}s
\end{aligned}$$

となり，右辺と一致する． □

10　Brown 運動と確率積分

ここでは，Brown 運動と確率積分の基本的な考え方について簡単に紹介する．普通の微積分の場合と同じように，具体的な問題についての計算に厳密な取り扱いが不可欠かというと必ずしもそうではないので，前節までとは違い，階段過程の近似に基づくできるだけ直観的な理解を目的にする．連続時間の確率過程の理論には微妙な点があるので，厳密なとり扱いについては専門書を参照して下さい．また，Øksendal (2003) は例が豊富な応用家向けのテキストで，工学・経済関係の文献ではよく引用されている．

定義．確率空間 $(\Omega, \mathcal{F}, P)$ 上の，実数値をとるパラメータ t (ここでは $t \geq 0$ を仮定する) をもつ確率変数の族 $X = \{X_t\}_{t \geq 0}$ を**確率過程** (stochastic process) という． □

定義．確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数の族 $\{B_t(\omega) : t \geq 0\}$ が，以下の性質 1, 2, 3, 4 を満たすとき，$\{B_t(\omega) : t \geq 0\}$ を**標準 Brown 運動** (standard Brownian motion) または **Wiener 過程** (Wiener process) と呼ぶ．

1) $B_0 = 0$.

2) $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$ のとき，確率変数 $B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ は独立．

3) 任意の $s, t \geq 0$ に対し，確率変数 $B_{t+s} - B_t$ は平均 0，分散 s の正規分布にしたがう．

4) ほとんどすべての $\omega \in \Omega$ に対し，$B_t = B_t(\omega)$ は t の連続関数．

□

2) の性質をもつ確率過程は独立増分過程と呼ばれる．

Brown 運動は実際に存在する (適当な確率空間上に構成できる) ことが知られている．Brown 運動を構成する際には，上の 4) の条件について若干のテクニカルな注意が必要になる．詳細は確率論の本を参照．

$B_t(\omega)$ は ω を固定すれば（確率１で）t の連続関数となるが，いたるところ t に関し微分不可能であることが知られている．

問題 39．任意の $s, t \geq 0$ に対し $\mathrm{E}(B_s B_t) = \min(s, t)$ となることを示せ．

定義．確率空間 $(\Omega, \mathcal{F}, P)$ の部分 σ-加法族の集合 $\{\mathcal{M}_t\}_{t \geq 0}$ が，任意の $0 \leq s \leq t$ に対し $\mathcal{M}_s \subset \mathcal{M}_t$ を満たすとき，$\{\mathcal{M}_t\}_{t \geq 0}$ を**増大情報系** (フィルトレーレーション, filtration) と呼ぶ． □

[**補足**] Brown 運動の構成についてはさまざまなうまい構成法が知られており，Lamperti にもいくつかの記述があるが，まずは公平なコイン投げにともなうランダムウォークの極限ととらえる理解が基本的である．いま $X_1, X_2, \ldots$ を互いに独立で $P(X_i = 1) = P(X_i = -1) = 1/2$ とする．$S_n = X_1 + \cdots + X_n$ とおく．$S_0, S_1, \ldots, S_n$ を折れ線で結んだ後に，時間軸を $1/n$ 倍，縦軸を $1/\sqrt{n}$ 倍してやると，勾配が $\pm\sqrt{n}$ の線分からなるパスが得られる．これがブラウン運動の近似となる．この折れ線は非常にギザギザしており，特に縦軸方向の変動の絶対値の和は

$$n \times \frac{1}{\sqrt{n}} = \sqrt{n} \ \to \infty \qquad (n \to \infty)$$

となる．このことから，ブラウン運動のパスが微分不可能であり，また有界変動でないことなどが納得される．

定義． Brown 運動 $B_t(\omega)$ に対し，$\mathcal{F}_t := \sigma(B_s : 0 \leq s \leq t)$ を B_t の**自然な増大情報系**と呼ぶ． $\square$

定義． $\{\mathcal{M}_t\}_{t\geq 0}$ を確率空間 $(\Omega, \mathcal{F}, P)$ の増大情報系，$\{g_t(\omega)\}_{t\geq 0}$ を確率過程とする．任意の $t \geq 0$ に対し，確率変数 $g_t(\omega)$ が $\mathcal{M}_t$-可測であるとき，$\{g_t\}_{t\geq 0}$ は M_t-**適合** (M_t-adapted) であるという． $\square$

[**補足**] $\mathcal{F}_t := \sigma(B_s : 0 \leq s \leq t)$ について適合な確率変数とは，時刻 t までのブラウン運動のパス $B_s : 0 \leq s \leq t$ の関数となっているような確率変数と考えればよい．例えば，B_t がある証券の価格過程を表すとすれば，現在までの価格のパスに応じた投資戦略 (つまり，当然であるが未来の価格を見てはいけない) は $\mathcal{F}_t$-適合である．

定義． $\mathcal{L}^2(a, b)$ $(0 \leq a < b)$ を以下の 1, 2, 3 を満たす確率過程 $f = \{f_t\}_{t\geq 0}$ の集合とする．

1) 写像 $(t, \omega) \mapsto f_t(\omega)$ は $\mathcal{F}' \times \mathcal{F}$-可測．ただし，$\mathcal{F}'$ は $\mathbb{R}$ の Borel 集合族，$\mathcal{F}' \times \mathcal{F}$ は $\mathcal{F}'$ と $\mathcal{F}$ の直積 σ-加法族である．(直積 σ-加法族の定義は確率論の本を参照．)

2) 確率過程 f は，$\mathcal{F}_t$-適合．

3) $\mathrm{E}\left[\int_a^b f_t^2 \mathrm{d}t\right] < \infty$.

$\square$

これから，$f \in \mathcal{L}^2(a, b)$ に対し，確率積分 $\int_a^b f_t \mathrm{d}B_t$ を定義する．まず，f が階段過程と呼ばれる簡単な確率過程の場合に積分を定義し，その極限をとることにより一般の確率積分を定義する．ここでは，一般の確率過程を具体的に階段過程で近似して極限をとる部分については省略する．

[**補足**] 確率積分は，数理ファイナンスの文脈で，特定の投資戦略のもとでの資金額を表すと見ると，理解しやすい．いまブラウン運動が 1 単位 (1 枚) の「証券」の価格を表すとしよう．証券は株主が有限責任のみを負うという仕組みのために，その価格は常に正であるが，数学的には負も許して「加法的」に考えたほうがわかりやすいので，負の値も許して考えることとする．投資戦略は，それまでの価格の動きに応じて ($\mathcal{F}_t$-適合) 何枚の証券を保有するか，を定める戦略とする．パスに依存する関数であるから，戦略自体も確率変数となる．

ω を固定すれば $B_t(\omega)$ は t の連続関数となるので，$\int f_t(\omega) \mathrm{d}B_t(\omega)$ は ω を固定して考えて普通の Stieltjes 積分として定義すれば良さそうに思えるが，$B_t(\omega)$ は有界変動でなく，Stieltjes 積分の前提条件を満たさないので，別の扱いが必要になる．

確率過程 $f = \{f_t\}_{t\geq 0}$ が

$$f_t(\omega) = \sum_{j=1}^{n} e_j(\omega) \mathbf{1}_{[t_{j-1}, t_j)}(t)$$

の形に表されるとき f を階段過程とよぶ．ここで，$e_j(\omega)$ は $\mathcal{F}_{t_{j-1}}$-可測で有界な確率変数である．また，$a=t_0\le t_1\le\cdots\le t_n=b$ は区間 $[a,b]$ の t によらない固定された分点である．このとき，確率積分を

$$\int_a^b f_t \mathrm{d}B_t = \int_a^b f_t(\omega)\mathrm{d}B_t(\omega) := \sum_{j=1}^n e_j(\omega)\{B_{t_j}(\omega)-B_{t_{j-1}}(\omega)\}$$

と定義する．確率積分は確率変数である．

[**補足**] 階段過程は，あらかじめ定められた時刻 t_j (例えば毎朝 8 時) のみに取引の許されるような投資戦略に対応する．つまり次の時点までは保有枚数を変えられないこととなる．時刻 t_{j-1} で e_j 枚の証券を保有すれば，$e_j(\omega)\{B_{t_j}(\omega)-B_{t_{j-1}}(\omega)\}$ は $[t_{j-1},t_j]$ での儲け (あるいは損失) を表す．そうすると確率積分は時刻 0 からの累積の儲け，すなわち資金，を表すこととなる．ここでも，投資戦略が未来に依存してはいけないこと点が本質的である．さて，ここでは t_j を非確率的な時刻としているが，投資の文脈で考えると，いわゆる stopping time を許すほうが自然である．

定理． $\{f_t\}_{t\ge 0}$ が階段過程のとき，

$$\mathrm{E}\left[\left(\int_a^b f_t\mathrm{d}B_t\right)^2\right] = \mathrm{E}\left[\int_a^b \{f_t(\omega)\}^2\mathrm{d}t\right]$$

□

証明.

$$\begin{aligned}
\mathrm{E}\left[\left(\int_a^b f_t\mathrm{d}B_t\right)^2\right] =& \mathrm{E}\left[\left(\sum_{j=1}^n e_j(\omega)\{B_{t_j}(\omega)-B_{t_{j-1}}(\omega)\}\right)^2\right]\\
=&\mathrm{E}\left[\sum_{j=1}^n \{e_j(\omega)\}^2\{B_{t_j}(\omega)-B_{t_{j-1}}(\omega)\}^2\right]\\
=&\mathrm{E}\left[\sum_{j=1}^n \{e_j(\omega)\}^2(t_j-t_{j-1})\right]\\
=&\mathrm{E}\left[\int_a^b \{f_t(\omega)\}^2\mathrm{d}t\right]
\end{aligned}$$

□

$\mathcal{L}^2(a,b)$ の任意の要素 f に対し，適当な階段過程の列 $f^{(n)}$ $n=1,2,\dots$ をとり，

$$\lim_{n\to\infty}\mathrm{E}\left[\int_a^b \{f_t(\omega)-f_t^{(n)}(\omega)\}^2\mathrm{d}t\right]=0$$

とできることが知られている．f の確率積分 (伊藤積分) は,

$$\int_a^b f_t(\omega)\mathrm{d}B_t(\omega) := \lim_{n\to\infty}\int_a^b f_t^{(n)}(\omega)\mathrm{d}B_t(\omega) \quad (2\,\text{次平均収束})$$

で定義される．この極限は存在して階段過程列のとり方によらないことが知られている．

定理． $f \in \mathcal{L}^2(a,b)$ のとき,

$$\mathrm{E}\left[\left(\int_a^b f_t \mathrm{d}B_t\right)^2\right] = \mathrm{E}\left[\int_a^b \{f_t(\omega)\}^2 \mathrm{d}t\right].$$

□

(証明略)

定理． $f, f^{(n)} \in \mathcal{L}^2(a,b)$ で,

$$\lim_{n\to\infty} \mathrm{E}\left[\int_a^b \{f_t(\omega) - f_t^{(n)}(\omega)\}^2 \mathrm{d}t\right] = 0$$

のとき,

$$\int_a^b f_t(\omega)\mathrm{d}B_t(\omega) := \lim_{n\to\infty}\int_a^b f_t^{(n)}(\omega)\mathrm{d}B_t(\omega) \quad (2\,\text{次平均収束})$$

□

証明．前定理よりいえる． □

[**補足**] 階段過程から極限操作をすると，連続的に保有枚数 f_t を変化させられるような投資戦略を許すこととなる．このような操作はなかなか考えにくい．連続的に保有枚数を変化させることができると「無限小の未来を見てはいけない」ということとの対応がわかりにくくなる．数学的には平均二乗収束という形でうまく問題を回避していると見ることもできよう．例えば Black-Scholes 公式にともなう複製戦略は価格のなめらかな関数であり，価格自体は有界変動ではないから，複製戦略自体も有界変動でなくなり，その現実的な意味は必ずしも明らかではない．

例．確率積分

$$\int_0^T B_t \mathrm{d}B_t$$

を考える．B_t が $\mathcal{F}_t$-適合であることは自明である．区間 $[0,T]$ を n 等分して,

$$t_j^{(n)} := \frac{T}{n}(j-1)$$

とし，

$$f_t^{(n)}(\omega) := \sum_{j=1}^{n} B_{t_j}(\omega)\mathbf{1}_{[t_j,t_{j+1})}(\omega)$$

とおく．

$$\begin{aligned}\mathrm{E}\left[\int_0^T (B_t - f_t^{(n)})^2 \mathrm{d}t\right] =& \mathrm{E}\left[\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(B_t - B_{t_{j-1}})^2\mathrm{d}t\right] = \sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}(t-t_{j-1})\mathrm{d}t \\ =& \sum_{j=1}^{n}\frac{1}{2}(t_j - t_{j-1})^2 = \frac{1}{2}\frac{T^2}{n}\end{aligned}$$

より，$\lim_{n\to\infty}\mathrm{E}\left[\int_0^T (B_t - f_t^{(n)})^2\mathrm{d}t\right] = 0$.

したがって，

$$\begin{aligned}\int_0^T B_t \mathrm{d}B_t &= \lim_{n\to\infty}\int_0^T f^{(n)}\mathrm{d}B_t = \lim_{n\to\infty}\sum_{j=1}^{n} B_{\frac{T}{n}(j-1)}(B_{\frac{T}{n}j} - B_{\frac{T}{n}(j-1)}) \\ &= \lim_{n\to\infty}\sum_{j=1}^{n}\left\{\frac{1}{2}(B^2_{\frac{T}{n}j} - B^2_{\frac{T}{n}(j-1)}) - \frac{1}{2}(B^2_{\frac{T}{n}(j-1)} - 2B_{\frac{T}{n}(j-1)}B_{\frac{T}{n}j} + B^2_{\frac{T}{n}j})\right\} \\ &= \lim_{n\to\infty}\sum_{j=1}^{n}\left\{\frac{1}{2}(B^2_{\frac{T}{n}j} - B^2_{\frac{T}{n}(j-1)}) - \frac{1}{2}(B_{\frac{T}{n}j} - B_{\frac{T}{n}(j-1)})^2\right\}\end{aligned}$$

ここで，

$$\sum_{j=1}^{n}(B^2_{\frac{T}{n}j} - B^2_{\frac{T}{n}(j-1)}) = B_T^2$$

であり，

$$\lim_{n\to\infty}\sum_{j=1}^{n}(B_{\frac{T}{n}j} - B_{\frac{T}{n}(j-1)})^2 = T \qquad (2\text{ 次平均収束}) \tag{51}$$

だから，

$$\int_0^T B_t\mathrm{d}B_t = \frac{1}{2}B_T^2 - \frac{1}{2}T. \tag{52}$$

となる．なお，式 (52) は後で見る伊藤の公式を使えばすぐいえるので，実際の計算ではいちいち階段過程による近似にもどる必要は無い． □

問題 40．式 (51) を示せ．

問題 41．$f, g \in \mathcal{L}^2(0,T)$, $0 \le a < b < T$，$\alpha, \beta \in \mathbb{R}$ のとき，

1) $\mathrm{E}\left[\int_a^b f_t \mathrm{d}B_t\right] = 0.$

2) $\int_a^b (\alpha f_t + \beta g_t)\mathrm{d}B_t = \alpha \int_a^b f_t \mathrm{d}B_t + \beta \int_a^b g_t \mathrm{d}B_t.$

が成立することが知られている．f, g が階段過程の場合に 1, 2 を示せ．

[**補足**] $\int_0^T B_s \mathrm{d}B_s$ はその時点での証券の価格に比例した枚数の証券を保有する戦略の資金過程を表している．ここで，その意味するところを理解するために，離散時間版を考えてみよう．X_i が第 i 日の証券一枚の値動きを表すとする．$S_n = X_1 + \cdots + X_n$ は第 n 日の夜の価格に対応している．いま第 n 日の朝の保有枚数を S_{n-1} とするような投資戦略を考えると，第 n 日の儲けは $S_{n-1}x_n$ と表される．第 n 日までの儲けの和，すなわち資金は

$$\sum_{i=1}^n S_{i-1}x_i, \qquad (S_0 \equiv 0)$$

と表わされる．

$$S_i^2 - S_{i-1}^2 = 2S_{i-1}x_i + x_i^2$$

を用いると，容易に

$$\sum_{i=1}^n S_{i-1}x_i = \frac{1}{2}S_n^2 - \frac{1}{2}\sum_{i=1}^n x_i^2$$

となることがわかる．特にコイン投げの場合 $(x_i = \pm 1)$ には

$$\sum_{i=1}^n S_{i-1}x_i = \frac{1}{2}S_n^2 - \frac{n}{2}$$

となる．これが (52) 式に対応する．

定義． $\{\mathcal{M}_t\}_{t\geq 0}$ を確率空間 $(\Omega, \mathcal{F}, P)$ の増大情報系とする．このとき，確率過程 $\{X_t\}_{t\geq 0}$ が 1), 2), 3) を満たすとき，$\{X_t\}_{t\geq 0}$ は増大情報系 $\{\mathcal{M}_t\}_{t\geq 0}$ について**マルチンゲール** (martingale) であるという．

1) 任意の $t \geq 0$ に対し $\mathrm{E}[|X_t|] < \infty$.

2) 確率過程 $\{X_t\}_{t\geq 0}$ は $\mathcal{M}_t$-適合.

3) 任意の $s \geq t$ に対し，$\mathrm{E}[X_t|\mathcal{M}_s] = X_s \quad \text{a.s.}$

□

例．Brown 運動 B_t は $\mathcal{F}_t := \sigma(B_s : 0 \leq s \leq t)$ についてマルチンゲールになる．マルチンゲールの定義のうち 1) を満たすことは容易にいえる．2) は $\mathcal{F}_t$ の定義より自明．3) は，$\mathrm{E}[B_t|\mathcal{F}_s] = \mathrm{E}[B_t - B_s|\mathcal{F}_s] + \mathrm{E}[B_s|\mathcal{F}_s] = B_s$ からわかる．なお，Brown 運動が独立増

分過程であることより $\mathrm{E}[B_t - B_s|\mathcal{F}_s] = 0$ が，定義より B_s が $\mathcal{F}_s$-可測であることから，$\mathrm{E}[B_s|\mathcal{F}_s] = B_s$ がいえる． □

定理． 任意の $T > 0$ に対し，$\{f_t(\omega)\}_{t\geq 0} \in \mathcal{L}^2(0,T)$ のとき，

$$X_t := \int_0^t f_s \mathrm{d}B_s$$

と定義すると，確率過程 $\{X_t\}_{t\geq 0}$ は $\{\mathcal{F}_t\}_{t\geq 0}$ についてマルチンゲールになる． □

（伊藤の）確率積分がマルチンゲールになることは重要である．

参考文献

Øksendal, B. (2003). Stochastic Differential Equations: An Introduction with Applications, 6th ed., Springer, Berlin.

参考: レポート課題 (2007年の駒木先生の時のもの)

問 1. $X_1, X_2, X_3, \ldots$ は，独立に同一の平均 μ 分散 σ^2 の分布にしたがう確率変数列である．N は X_i $(i = 1, 2, 3, \ldots)$ とは独立な非負の整数値をとる確率変数で，$\mathrm{E}(N^2) < \infty$ である．このとき，

$$S = X_1 + \cdots + X_N$$

と定義すると，S の分散について，

$$\mathrm{var}(S) = \sigma^2 \mathrm{E}(N) + \mu^2 \mathrm{var}(N)$$

が成立することを示せ.

問 2. X は確率空間 $(\Omega, \mathcal{F}, P)$ 上の $\mathrm{E}(X^2) < \infty$ を満たす確率変数である．$\mathcal{F}_1$, $\mathcal{F}_2$ はそれぞれ $\mathcal{F}$ の部分 σ-加法族で，$\mathcal{F}_1 \subset \mathcal{F}_2$ であるとする．このとき，条件付期待値の性質を用いて，

$$\mathrm{E}[\{X - \mathrm{E}(X|\mathcal{F}_2)\}^2] \leq \mathrm{E}[\{X - \mathrm{E}(X|\mathcal{F}_1)\}^2]$$

を示せ.

問 3. X, Y は確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数で，$\mathcal{F}'$ は $\mathcal{F}$ の部分 σ-加法族である．$\mathrm{E}(Y|\mathcal{F}') = X$ と $\mathrm{E}(X^2) = \mathrm{E}(Y^2) < \infty$ が成立するとき，

$$X = Y \quad \mathrm{a.s.}$$

となることを示せ.

11　伊藤の公式と確率微分方程式

確率過程 X_t が

$$X_t = X_0 + \int_0^t u_s(\omega)\mathrm{d}s + \int_0^t v_s(\omega)\mathrm{d}B_s \tag{53}$$

で与えられているとする．ここで，X_0 は Brown 運動とは独立な確率変数，$u_s(\omega)$, $v_s(\omega)$ は $\mathcal{F}_s$-適合な確率過程である．$u_s(\omega)$, $v_s(\omega)$ の満たすべき条件の詳細にはここでは立ち入らない．(53) を，積分形を使わずに形式的に

$$\mathrm{d}X_t = u_t(\omega)\mathrm{d}t + v_t(\omega)\mathrm{d}B_t \tag{54}$$

と表記することが多い．

[**補足**] 投資戦略の例で考えると，$\mathrm{d}t$ 項は単位時間あたりに価格が必ず 1 上がっていくような「安全資産」と考えるとよい．u_t は時刻 t までのブラウン運動のパスを見て，安全資産を何単位保有するかを表している．価格モデルは通常は「積型」で考えるから，安全資産が一定の利子率 r で増加するとするのが簡単であるが,「和型」で考えれば単位時間あたりの増加が一定とすればいいわけである．積型の場合でも，時間スケールを変換してやって，安全資産が e 倍となる時間の長さを単位時間としてやれば，$r = 1$ としても一般性を失わない．

定理． (伊藤の公式)

X_t を (54) で与えられる確率過程，$\varphi(t, x)$ を $t \in [0, \infty)$, $x \in \mathbb{R}$ の滑らかな関数とする．このとき，$Y_t := \varphi(t, X_t)$ に対し，

$$\mathrm{d}Y_t = \frac{\partial \varphi}{\partial t}(t, X_t)\mathrm{d}t + \frac{\partial \varphi}{\partial x}(t, X_t)\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 \varphi}{\partial x^2}v^2\mathrm{d}t$$

が成り立つ．□

伊藤の公式は，変数変換の全微分の関係を表し，

1)　$\mathrm{d}\varphi(t, X_t)$ の 2 次までの形式的な Taylor 展開

$$\begin{aligned}\mathrm{d}Y_t =& \mathrm{d}\varphi(t, X_t)\\ =& \frac{\partial \varphi}{\partial t}(t, X_t)\mathrm{d}t + \frac{\partial \varphi}{\partial x}(t, X_t)\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 \varphi}{\partial t^2}(t, X_t)(\mathrm{d}t)^2 + \frac{\partial^2 \varphi}{\partial t \partial x}(t, X_t)\mathrm{d}t\mathrm{d}X_t\\ &+ \frac{1}{2}\frac{\partial^2 \varphi}{\partial x^2}(t, X_t)(\mathrm{d}X_t)^2\end{aligned} \tag{55}$$

を求め，３次以上の項は無視する．

2)　$\mathrm{d}X_t$ に $u(s, \omega)\mathrm{d}t + v(t, \omega)\mathrm{d}B_t$ を代入する．

3)　$(\mathrm{d}t)^2$, $\mathrm{d}t\,\mathrm{d}B_t$ は 0 で置き換え，$(\mathrm{d}B_t)^2$ は $\mathrm{d}t$ で置き換える．

という簡単な手続きに対応している．

その際に，$\mathrm{d}B_t = O(\sqrt{\mathrm{d}t})$, $(\mathrm{d}B_t)^2 = \mathrm{d}t$ と見て，$O(\mathrm{d}t)$ までの項を考え，$(\mathrm{d}X_t)^2 = v_t^2(\mathrm{d}B_t)^2$ となることを使っている．

例．p. 62 で見た確率積分

$$\int_0^t B_s \mathrm{d}B_s$$

を考える．これを，確率積分ではない普通の積分とみると，$\frac{1}{2}B_t^2$ になる．$\frac{1}{2}B_t^2$ を展開して伊藤の公式を適用することにより，$\frac{1}{2}B_t^2$ に加える必要のある修正項を求める．$\varphi(x) := (1/2)x^2$ とおくと，$\frac{1}{2}B_t^2 = \varphi(B_t)$ である．$\varphi(x)$ は t を含まないので，形式的な 2 次までの Taylor 展開は,

$$\mathrm{d}\left(\frac{1}{2}B_t^2\right) = B_t \mathrm{d}B_t + \frac{1}{2}(\mathrm{d}B_t)^2$$

となる．伊藤の公式より $(\mathrm{d}B_t)^2 = \mathrm{d}t$ だから

$$\mathrm{d}\left(\frac{1}{2}B_t^2\right) = B_t \mathrm{d}B_t + \frac{1}{2}\mathrm{d}t.$$

これを積分形に直せば,

$$\int_0^t B_s \mathrm{d}B_s = \frac{1}{2}B_t^2 - \frac{1}{2}t$$

となる．

問題 42．伊藤の公式を用いて

$$\int_0^t s \mathrm{d}B_s = tB_t - \int_0^t B_s \mathrm{d}s$$

を示せ．

ランダムな要素を含まないダイナミックなシステムは微分方程式

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = b_t(x_t)$$

で記述できることが多い．これは,

$$\mathrm{d}x_t = b_t(x_t)\mathrm{d}t$$

とも表せる．これにランダムな項が加わった

$$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}B_t$$

が**確率微分方程式** (stochastic differential equation) である．$b_t(X_t)$ を**ドリフト係数** (drift coefficient)，$\sigma_t(X_t)$ を**拡散係数** (diffusion coefficient) とよぶ．X_t 自身を diffusion とよぶことが多い．X_t の「将来の」変動 $\mathrm{d}X_t$ が現時点の値 X_t のみに依存するので，diffusion はマルコフ過程となる．さらに，b,σ を X_t のみの関数として，$b(X_t), \sigma(X_1)$ とすれば時間的に homogeneous なマルコフ過程となる．

直観的には，Brown 運動の形式的な時間微分 $\mathrm{d}B_t/\mathrm{d}t$ を，工学でよく用いる（連続時間版の）正規白色雑音（正規分布のホワイトノイズ）とみなすと都合が良い．

ここでは，確率微分方程式の解の存在と一意性に関する問題については扱わず，いくつかの具体例について見る．

例．Ornstein-Uhlenbeck 過程

確率微分方程式

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma \mathrm{d}B_t$$

は **Ornstein-Uhlenbeck 方程式** と呼ばれる．$X_t = \mathrm{e}^{\mu t} Y_t$ とおくと，

$$\mathrm{d}X_t = \mu \mathrm{e}^{\mu t} Y_t \mathrm{d}t + \mathrm{e}^{\mu t} \mathrm{d}Y_t$$

より，

$$\mathrm{e}^{\mu t}\mathrm{d}Y_t = \mathrm{d}X_t - \mu X_t \mathrm{d}t = \sigma \mathrm{d}B_t$$

だから，

$$Y_t = Y_0 + \sigma \int_0^t \mathrm{e}^{-\mu s} \mathrm{d}B_s$$

となる．Y_0 は Brown 運動とは独立な確率変数である．$Y_0 = X_0$ だから，

$$X_t = \mathrm{e}^{\mu t} Y_t = \mathrm{e}^{\mu t} X_0 + \sigma \int_0^t \mathrm{e}^{\mu(t-s)} \mathrm{d}B_s \tag{56}$$

となる．

(56) の確率積分の被積分関数は B_s を含んでいないので，確率過程 X_t の共分散関数は容易に求めることができる．初期値が $X_0 = 0$ のとき X_t と $X_{t+\tau}$ $(\tau \geq 0)$ との共分散は，

$$\begin{aligned}
&\mathrm{E}\big[\{X_t - \mathrm{E}(X_t)\}\{X_{t+\tau} - \mathrm{E}(X_{t+\tau})\}\big] = \mathrm{E}\big[X_t X_{t+\tau}\big] \\
=&\mathrm{E}\left[\left\{\sigma \int_0^t \mathrm{e}^{\mu(t-s)} \mathrm{d}B_s\right\}\left\{\sigma \int_0^{t+\tau} \mathrm{e}^{\mu(t+\tau-s)} \mathrm{d}B_s\right\}\right] \\
=&\sigma^2 \mathrm{E}\left[\left\{\int_0^t \mathrm{e}^{\mu(t-s)} \mathrm{d}B_s\right\}\left\{\int_0^t \mathrm{e}^{\mu(t+\tau-s)} \mathrm{d}B_s + \int_t^{t+\tau} \mathrm{e}^{\mu(t+\tau-s)} \mathrm{d}B_s\right\}\right] \\
=&\sigma^2 \mathrm{E}\left[\left\{\int_0^t \mathrm{e}^{\mu(t-s)} \mathrm{d}B_s\right\}\left\{\int_0^t \mathrm{e}^{\mu(t+\tau-s)} \mathrm{d}B_s\right\}\right] \\
=&\sigma^2 \int_0^t \mathrm{e}^{2\mu(t-s)+\mu\tau} \mathrm{d}s = \frac{\sigma^2}{2\mu}(\mathrm{e}^{2\mu t} - 1)\mathrm{e}^{\mu\tau}
\end{aligned}$$

となる． □

例．力学や回路学で，微分方程式

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t)+c\frac{\mathrm{d}}{\mathrm{d}t}x(t)+kx(t)=0$$

は基本的である．c, d は実定数である．$x(t)$ が質量 1 の質点の時刻 t における位置を表すものと解釈して，システムに新たに正規白色雑音の外力 $\mathrm{d}B_t/\mathrm{d}t$ が加わった状況を考えると微分方程式は形式的に

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}X(t)+c\frac{\mathrm{d}}{\mathrm{d}t}X(t)+kX(t)=\sigma^2\frac{\mathrm{d}B_t}{\mathrm{d}t} \tag{57}$$

となる．x を X で書き換えてある．これを確率微分方程式として扱うことを考える．

$$Y(t)=\begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}:=\begin{pmatrix} X(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}X(t) \end{pmatrix}$$

とおくと，(57) は，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}Y(t)&=\begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}Y_1(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}Y_2(t) \end{pmatrix}=\begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}X(t) \\ \frac{\mathrm{d}^2}{\mathrm{d}t^2}X(t) \end{pmatrix}=\begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}X(t) \\ -c\frac{\mathrm{d}}{\mathrm{d}t}X(t)-kX(t)+\sigma^2\frac{\mathrm{d}B_t}{\mathrm{d}t} \end{pmatrix}\\ &=\begin{pmatrix} Y_2(t) \\ -cY_2(t)-kY_1(t)+\sigma^2\frac{\mathrm{d}B_t}{\mathrm{d}t} \end{pmatrix}\end{aligned}$$

より，

$$\begin{pmatrix} \mathrm{d}Y_1(t) \\ \mathrm{d}Y_2(t) \end{pmatrix}=\begin{pmatrix} 0 & 1 \\ -k & -c \end{pmatrix}\begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}\mathrm{d}t+\begin{pmatrix} 0 \\ \sigma^2 \end{pmatrix}\mathrm{d}B_t$$

という 2 次元の確率微分方程式として定式化できる．ここで，

$$F:=\begin{pmatrix} 0 & 1 \\ -k & -c \end{pmatrix},\qquad G:=\begin{pmatrix} 0 \\ \sigma^2 \end{pmatrix}$$

とおけば，

$$\mathrm{d}Y(t)=FY(t)+G\mathrm{d}B_t$$

となる．$Y(t)=\exp(Ft)Z(t)$ とおくと，

$$\mathrm{d}Y(t)=F\exp(Ft)Z(t)\mathrm{d}t+\exp(Ft)\mathrm{d}Z(t)$$

より，

$$Z(t)=Z(0)+\int_0^t\exp(-Fs)G\mathrm{d}B_s$$

となる．したがって，

$$Y(t)=\exp(Ft)X(0)+\exp(Ft)\int_0^t\exp(-Fs)G\mathrm{d}B_s$$

が得られる．□

例．幾何 Brown 運動

　微分方程式

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = rx(t)$$

の解は $x(t) = x(0)\mathrm{e}^{rt}$ となる．成長率 r にノイズの加わった方程式

$$\mathrm{d}X_t = rX_t\mathrm{d}t + \sigma X_t\mathrm{d}B_t$$

により決まる確率過程は**幾何ブラウン運動** (geometric Brownian motion) と呼ばれ，金融工学等でよく利用される．

　両辺を $1/X_t$ 倍して

$$\frac{\mathrm{d}X_t}{X_t} = r\mathrm{d}t + \sigma\mathrm{d}B_t$$

より，

$$\int_0^t \frac{\mathrm{d}X_s}{X_s} = rt + \sigma B_t$$

となる．伊藤の公式より，

$$\mathrm{d}(\log X_t) = \frac{\mathrm{d}X_t}{X_t} - \frac{1}{2}(\mathrm{d}X_t)^2 = \frac{\mathrm{d}X_t}{X_t} - \frac{\sigma^2}{2}\mathrm{d}t$$

だから，

$$\int_0^t \frac{\mathrm{d}X_s}{X_s} = \log\frac{X_t}{X_0} + \frac{\sigma^2}{2}t$$

となり，

$$\log\frac{X_t}{X_0} = rt - \frac{\sigma^2}{2}t + \sigma B_t$$

を得る．したがって，

$$X_t = X_0\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma B_t\right\}$$

となり，X_t が得られた．

　初期値を $X_0 = x_0$ に固定すると，$\mathrm{E}(X_t) = x_0\mathrm{e}^{rt}$ となり，期待値はノイズの無い場合と一致することが以下のように確認できる．

　伊藤の公式より，

$$\mathrm{d}\exp(\sigma B_t) = \sigma\exp(\sigma B_t)\mathrm{d}B_t + \frac{1}{2}\sigma^2\exp(\sigma B_t)\mathrm{d}t$$

だから，

$$\exp(\sigma B_t) = \int_0^t \sigma\exp(\sigma B_s)\mathrm{d}B_s + \frac{1}{2}\sigma^2\int_0^t \exp(\sigma B_s)\mathrm{d}s$$

両辺の期待値をとって，$y(t) := \mathrm{E}(\exp(\sigma B_t))$ とおくと，伊藤積分の期待値は 0 なので，

$$y(t) = \frac{1}{2}\sigma^2\int_0^t y(s)\mathrm{d}s$$

微分方程式に直すと，

$$\frac{\mathrm{d}}{\mathrm{d}t}y(t) = \frac{1}{2}\sigma^2 y(t)$$

だから，

$$y(t) = \exp\left(\frac{1}{2}\sigma^2 t\right)$$

となる．($B_0 = 0$ とした．)

したがって，

$$\mathrm{E}(X_t) = x_0 \mathrm{E}\left[\exp\left\{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma B_t\right\}\right] = x_0 \exp(rt)$$

となる．B_t の分布が $\mathrm{N}(0,t)$ であることを利用して直接確認をすることもできる． □

例．Brownian bridge

確率微分方程式

$$\mathrm{d}X_t = \frac{b - X_t}{1-t}\mathrm{d}t + \mathrm{d}B_t \quad (X_0 = a,\ 0 \leq t < 1)$$

にしたがう確率過程 X_t は**ブラウニアン・ブリッジ** (Brownian bridge) と呼ばれ，いろいろなところに現れる．

伊藤の公式より，

$$\mathrm{d}\frac{X_t - b}{1-t} = \frac{\mathrm{d}X_t}{1-t} + \frac{X_t - b}{(1-t)^2}\mathrm{d}t$$

だから

$$\mathrm{d}X_t = (1-t)\mathrm{d}\frac{X_t - b}{1-t} + \frac{b - X_t}{1-t}\mathrm{d}t$$

となり，これともとの確率微分方程式から，

$$(1-t)\mathrm{d}\frac{X_t - b}{1-t} = \mathrm{d}B_t$$

を得る．よって，

$$\frac{X_t - b}{1-t} = a - b + \int_0^s \frac{1}{1-s}\mathrm{d}B_s$$

より

$$X_t = (1-t)a + tb + (1-t)\int_0^t \frac{1}{1-s}\mathrm{d}B_s$$

となる．この式より，任意の $0 < t_1 < t_2 < \cdots < t_k < 1$ に対し，$X_{t_1}, X_{t_2}, \ldots, X_{t_k}$ の同時分布は k 次元正規分布となることがわかる．

伊藤積分の期待値は 0 なので，

$$\mathrm{E}(X_t) = (1-t)a + tb.$$

X_s と X_t $(t \geq s)$ との共分散は

$$
\begin{aligned}
\mathrm{Cov}(X_s, X_t) =& \mathrm{E}\{(X_s - \mathrm{E}(X_s))(X_t - \mathrm{E}(X_t))\} \\
=& \mathrm{E}\left[\left\{(1-s)\int_0^s \frac{1}{1-u}\mathrm{d}B_u\right\}\left\{(1-t)\int_0^t \frac{1}{1-v}\mathrm{d}B_v\right\}\right] \\
=& (1-s)(1-t)\mathrm{E}\left\{\left(\int_0^s \frac{1}{1-u}\mathrm{d}B_u\right)\left(\int_0^s \frac{1}{1-v}\mathrm{d}B_v + \int_s^t \frac{1}{1-v}\mathrm{d}B_v\right)\right\} \\
=& (1-s)(1-t)\int_0^s \frac{1}{(1-u)^2}\mathrm{d}u \\
=& (1-s)(1-t)\left[\frac{1}{1-u}\right]_{u=0}^s \\
=& s(1-t)
\end{aligned}
$$

となる．

なお，Browinan bridge は時刻 0 に $X_0 = a$ から出発する Brown 運動に対し，時刻 1 に $X_1 = b$ となるという条件を付けて得られる確率過程であることが知られている． □

問題 43. (円周上の Brown 運動) 複素数値をとる確率過程 $X(t)$ が

$$
X(t) = X_1(t) + \mathrm{i}X_2(t) = \exp(\mathrm{i}B_t)
$$

で定義されるとき，$X_1(t)$, $X_2(t)$ は確率微分方程式

$$
\begin{aligned}
\mathrm{d}X_1(t) =& -\frac{1}{2}X_1(t)\mathrm{d}t - X_2(t)\mathrm{d}B_t, \\
\mathrm{d}X_2(t) =& -\frac{1}{2}X_2(t)\mathrm{d}t + X_1(t)\mathrm{d}B_t
\end{aligned}
$$

を満たすことを示せ.

問題 44. 確率積分の部分積分

X_t, Y_t がそれぞれ確率微分方程式

$$
\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}B_t, \quad \mathrm{d}Y_t = c_t(Y_t)\mathrm{d}t + \tau_t(Y_t)\mathrm{d}B_t
$$

にしたがう確率過程であるとき，

$$
\mathrm{d}(X_tY_t) = X_t\mathrm{d}Y_t + Y_t\mathrm{d}X_t + \mathrm{d}X_t \cdot \mathrm{d}Y_t
$$

が成り立つことを確認せよ．これより，確率積分の部分積分の公式

$$
\begin{aligned}
\int_0^t X_s\mathrm{d}Y_s =& X_tY_t - X_0Y_0 - \int_0^t Y_s\mathrm{d}X_s - \int_0^t \mathrm{d}X_s \cdot \mathrm{d}Y_s \\
=& X_tY_t - X_0Y_0 - \int_0^t Y_s\mathrm{d}X_s - \int_0^t \sigma_s(X_s)\tau_s(Y_s)\mathrm{d}s
\end{aligned}
$$

が得られる．

12　生成作用素

今まで扱ってきた

$$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}B_t$$

の型の確率微分方程式にしたがう確率過程は,（ドリフト係数 $b_t(X_t)$ と拡散係数 $\sigma_t(X_t)$ に対する一定の条件のもとで）時間 t に対して連続なパスをもつ一意的な解をもつことが知られている．このような確率過程 X_t は **拡散過程** (diffusion process) と呼ばれる．拡散過程の性質を調べる方法には，確率微分方程式に基づく直接的な方法の他に，生成作用素を利用する方法がある．ここでは生成作用素に基づく方法について直観的な理解をめざす.

X_t を (簡単のため時間に依存しない) 確率微分方程式

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t \tag{58}$$

にしたがう確率過程とする．このとき，X_t は時間的に一様な (time homogeneous) **マルコフ過程** (Markov process) になる.

X_t の**生成作用素** (generator) A を

$$Af(x) = \lim_{t\to 0} \frac{\mathrm{E}^x\{f(X_t)\} - f(x)}{t}$$

で定義する．ただし，$f : \mathbb{R} \to \mathbb{R}$ は 適当な正則条件を満たす任意の関数であり，E^x は初期条件 $X_0 = x$ のもとでの期待値を表わす.

分布収束の議論を思い出すと，分布を累積分布関数の形で与えることと，任意の有界連続関数の期待値を与えることは同値であった．特性関数の場合のように，期待値をとる関数の族は必要に応じてさらに狭くとることもできる．そこで，diffusion process の周辺分布が微小時間でどのような変化するかを調べるには，適当な関数の族について $\mathrm{E}^x\{f(X_t)\}$ を調べればよいことが納得できる．時間的に一様なマルコフ過程の周辺分布の時間発展は，各時刻で同様であり，生成作用素 A は時刻には依存しない．A は微小時間の時間発展を微分の形で記述していると理解すればよい.

A は，b, σ を用いて表すことができる．X_t^x を (58) にしたがう初期値 $X_0 = x$ の確率過程とすると,

$$X_t = x_0 + \int_0^t b_s(X_s)\mathrm{d}s + \int_0^t \sigma_s(X_s)\mathrm{d}B_s$$

となる．伊藤の公式より,

$$\begin{aligned}\mathrm{d}f(X_t) =& \frac{\mathrm{d}f}{\mathrm{d}x}(X_t)\mathrm{d}X_t + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_t)(\mathrm{d}X_t)^2 \\ =& \frac{\mathrm{d}f}{\mathrm{d}x}(X_t)\{b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t\} + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_t)\sigma^2(X_t)\mathrm{d}t\end{aligned}$$

だから,

$$f(X_t) = f(x) + \int_0^t \left\{\frac{\mathrm{d}f}{\mathrm{d}x}(X_s)b(X_s) + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_s)\sigma^2(X_s)\right\}\mathrm{d}s + \int_0^t \frac{\mathrm{d}f}{\mathrm{d}x}(X_s)\sigma(X_s)\mathrm{d}B_s.$$

両辺の期待値をとると,

$$\mathrm{E}^x\{f(X_t)\} = f(x) + \int_0^t \left\{ \frac{\mathrm{d}f}{\mathrm{d}x}(X_s)b(X_s) + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(X_s)\sigma^2(X_s) \right\} \mathrm{d}s$$

となる．したがって，生成作用素の定義より,

$$Af(x) = \lim_{t\to 0} \frac{\mathrm{E}^x\{f(X_t)\} - f(x)}{t} = b(x)\frac{\mathrm{d}f}{\mathrm{d}x}(x) + \frac{1}{2}\sigma^2(x)\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x)$$

となる．

ドリフト係数，拡散係数が時間に依存する確率微分方程式

$$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}B_t$$

の生成作用素は,

$$A_t f(x) = b_t(x)\frac{\mathrm{d}f}{\mathrm{d}x}(x) + \frac{1}{2}\sigma_t^2(x)\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x)$$

であることが同様に示せる．

拡散過程の場合，時刻 s に $X_s = x$ であるという条件のもとでの X_{s+t} の確率分布は x, s, t で決まる．この条件付確率の密度関数を，推移確率密度関数と呼び $p(t, y|s, x)$ で表す．このとき,

$$p(z, u|x, s) = \int_{-\infty}^{\infty} p(z, u|y, t)p(y, t|x, s)\mathrm{d}y \qquad (x, y, z \in \mathbb{R}, \quad u > t > s \geq 0)$$

が成立する．この式を **チャップマン - コルモゴロフの方程式** (Chapman-Kolmogorov equation) という．

拡散過程 X_t が生成作用素

$$A_t = b(t, x)\frac{\partial}{\partial x} + a(t, x)\frac{\partial^2}{\partial x^2}$$

をもつとき,

$$\frac{\partial p}{\partial s}(y, t|x, s) = -b(s, x)\frac{\partial p}{\partial x}(y, t|x, s) - a(s, x)\frac{\partial^2 p}{\partial x^2}(y, t|x, s) \tag{59}$$

$$\frac{\partial p}{\partial t}(y, t|x, s) = -\frac{\partial}{\partial y}\{b(t, y)p(y, t|x, s)\} + \frac{\partial^2}{\partial y^2}\{a(t, x)p(y, t|x, s)\} \tag{60}$$

が成立する．

(59) を**コルモゴロフの後向き方程式** (Kolmogorov's backward equation) という．後向き方程式の右辺には，生成作用素 (に負号をつけたもの) $-A_t$ が現れていることに注意する．

また，(60) を**コルモゴロフの前向き方程式** (Kolmogorov's forward equation) という．コルモゴロフの前向き方程式を (特に物理学では) **フォッカー・プランク方程式** (Fokker-Planck equation) と呼ぶことも多い．物理学・工学などの応用では前向き方程式を適当な初期条件・境界条件のもとで解いて分布を求めることがよく行われる．応用上は，確率微分方程式と前向き方程式は補いあう関係にある．前向き方程式を解析的に扱うためのいろいろな方法があるが，ここでは深入りしない．なお，解析的に密度の形が完全に求まるのは限られた場合である．

(60) の右辺に現れている

$$A_t^* f(y) = -\frac{\partial}{\partial y}(bf) + \frac{\partial^2}{\partial y^2}(af)$$

で定義される作用素 A_t^* は A_t の**共役作用素** (adjoint operator) と呼ばれる．ここでいう A^* が A の共役作用素であるということのもともとの定義は，$x \to \infty$, $x \to -\infty$ で十分速く 0 に近づく十分滑らかな関数の組 $f(x)$, $g(x)$ に対し，

$$\int (Af)(x)g(x)\mathrm{d}x = \int f(x)(A^*g)(x)\mathrm{d}x$$

が成立するということである．

実際，A_t, A_t^* に関してこの関係が成立する事が部分積分により次のように確認できる．

$$\begin{aligned}
\int (A_t f)(x)g(x)\mathrm{d}x &= \int \left\{ b(t,x)\frac{\partial f}{\partial x}(x) + a(t,x)\frac{\partial^2 f}{\partial x^2}(x) \right\} g(x)\mathrm{d}x \\
&= \int \left\{ b(t,x)\frac{\partial f}{\partial x}(x) \right\} g(x)\mathrm{d}x + \int \left\{ a(t,x)\frac{\partial^2 f}{\partial x^2}(x) \right\} g(x)\mathrm{d}x \\
&= -\int f(x)\frac{\partial}{\partial x}\{b(t,x)g(x)\}\,\mathrm{d}x - \int \frac{\partial f}{\partial x}(x)\frac{\partial}{\partial x}\{a(t,x)g(x)\}\,\mathrm{d}x \\
&= -\int f(x)\frac{\partial}{\partial x}\{b(t,x)g(x)\}\,\mathrm{d}x + \int f(x)\frac{\partial^2}{\partial x^2}\{a(t,x)g(x)\}\,\mathrm{d}x \\
&= \int f(x)(A_t^* g)(x)\mathrm{d}x.
\end{aligned}$$

次に，後向き方程式・前向き方程式を（形式的な議論により）導出する．

X_t を $\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}B_t \quad (t \geq s)$ にしたがう拡散過程とする．初期条件として $X_s = x$ を仮定する．チャップマン - コルモゴロフの方程式

$$p(y,t|x,s) = \int_{-\infty}^{\infty} p(y,t|z,u)p(z,u|x,s)\mathrm{d}z \qquad (x,y,z \in \mathbb{R}, \quad t > u > s \geq 0)$$

より，y, t を固定して $f(X_u, u) = p(y,t|X_u,u)$ とおくと，

$$\mathrm{E}^{X_s = x}[f(X_u, u)] = f(x,s) \quad (t > u \geq s)$$

となり左辺は u に依らないことがわかる．伊藤の公式より

$$\begin{aligned}
\mathrm{d}f(X_u,u) &= \frac{\partial f}{\partial u}(X_u,u)\mathrm{d}u + \frac{\partial f}{\partial x}(X_u,u)\mathrm{d}X_u + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_u,u)(\mathrm{d}X_u)^2 \\
&= \frac{\partial f}{\partial u}(X_u,u)\mathrm{d}u + \frac{\partial f}{\partial x}(X_u,u)b_u(X_u)\mathrm{d}u + \frac{\partial f}{\partial x}(X_u,u)\sigma_u(X_u)\mathrm{d}B_u + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_u,u)\sigma_u^2(X_u)\mathrm{d}u.
\end{aligned}$$

両辺の期待値をとり $u=s$ とおくと，

$$0 = \frac{\partial f}{\partial s}(x,s)\mathrm{d}s + \frac{\partial f}{\partial x}(x,s)b_s(x)\mathrm{d}s + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_s,s)\sigma_s^2(x)\mathrm{d}s$$

となり，後向き方程式 (59) が得られる．ここでの (x の微分に関する) 計算は，拡散過程の生成作用素の導出の計算と同じである．

次に前向き方程式を導出する．f をコンパクトな区間の外では 0 になるような滑らかな関数とすると，$s<t<t+\tau$ として，チャップマン - コルモゴロフの方程式より

$$\begin{aligned}
&\int \lim_{\tau\to 0}\frac{1}{\tau}\{p(y,t+\tau|x,s) - p(y,t|x,s)\}f(y)\mathrm{d}y \\
&= \lim_{\tau\to 0}\frac{1}{\tau}\left\{\iint p(y,t+\tau|z,t)p(z,t|x,s)f(y)\mathrm{d}z\mathrm{d}y - \int p(y,t|x,s)f(y)\mathrm{d}y\right\} \\
&= \lim_{\tau\to 0}\frac{1}{\tau}\left[\int p(y,t|x,s)\left\{\int p(z,t+\tau|y,t)f(z)\mathrm{d}z\right\}\mathrm{d}y - \int p(y,t|x,s)f(y)\mathrm{d}y\right] \\
&= \int p(y,t|x,s)\lim_{\tau\to 0}\frac{1}{\tau}\left\{\int p(z,t+\tau|y,t)f(z)\mathrm{d}z - f(y)\right\}\mathrm{d}y \\
&= \int p(y,t|x,s)(A_t f)(y)\mathrm{d}y \\
&= \int (A_t^* p)(y,t|x,s)f(y)\mathrm{d}y
\end{aligned}$$

が得られる．f は任意なので，前向き方程式 (60) が示せた．

解析的に解ける扱いやすい例について，前向き方程式の具体的な形を挙げる．

例．Brown 運動

$$\mathrm{d}X_t = \mathrm{d}B_t$$

より，

$$A = \frac{1}{2}\frac{\partial^2}{\partial x^2}.$$

推移確率密度関数は，

$$p(y,t|x,s) = \frac{1}{\sqrt{2\pi(t-s)}}\exp\left\{-\frac{(y-x)^2}{2(t-s)}\right\}.$$

前向き方程式は，

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}$$

後向き方程式は,
$$\frac{\partial p}{\partial s} = -\frac{1}{2}\frac{\partial^2 p}{\partial y^2}$$
である．推移確率密度関数が前向き・後向き方程式を満たすことが確認できる． □

例．ドリフト付き Brown 運動
　確率微分方程式
$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}B_t$$
の初期条件 $X_0 = 0$ のもとでの解は,
$$X_t = X_0 + \mu t + \sigma B_t$$
である．推移確率密度関数も容易に求められる．前向き方程式は,
$$\frac{\partial p}{\partial t} = -\mu\frac{\partial p}{\partial y} + \frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial y^2}$$
である． □

例．Ornstein-Uhlenbeck 過程
　確率微分方程式
$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma \mathrm{d}B_t$$
の解は
$$X_t = \mathrm{e}^{\mu t}X_0 + \sigma\int_0^t \mathrm{e}^{\mu(t-s)}\mathrm{d}B_s \tag{61}$$
となることを前に見た．
　推移確率密度関数は
$$p(y,t|x,s) = \sqrt{\frac{\mu}{\pi\sigma^2\{\mathrm{e}^{2\mu(t-s)}-1\}}}\exp\left[-\frac{\mu}{\sigma^2\{\mathrm{e}^{2\mu(t-s)}-1\}}\left\{y - x\mathrm{e}^{\mu(t-s)}\right\}^2\right]$$
となることが容易にわかる．
　前向き方程式は,
$$\frac{\partial p}{\partial t} = -\mu\frac{\partial}{\partial y}(yp) + \frac{1}{2}\sigma^2\frac{\partial^2 p}{\partial y^2}$$
となる． □